WDSA - CCWI
Proceedings Book
Valencia 18th- 22nd July

2022

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

edUPV
Universitat Politècnica de València

2nd International Join Conference on Water Distribution System Analysis (WDSA) & Computing and Control in the Water Industry (CCWI)

WDSA CCWI 2022

Department of Hydraulic Engineeering and Environment
Universitat Politècnica de València (Valencia Tech)
VALENCIA, SPAIN
18-22 JULY, 2022

edUPV
Universitat Politècnica de València

Congresos UPV
Proceedings of the 2nd International Join Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI)

The contents of this publication have been evaluated by the Scientific Committee which it relates and the procedure set out http://ocs.editorial.upv.es/index.php/WDSA-CCWI/WDSA-CCWI2022/about/editorialPolicies

© **Scientific Editors**
Pedro L. Iglesias Rey
Fernando Martínez Alzamora
F. Javier Martínez Solano

© texts: The authors

© **Publisher**
Editorial Universitat Politècnica de València, 2022
http://www.lalibreria.upv.es / Ref.: 6705_01_01_01

ISBN: 978-84-9048-982-6

DOI: http://dx.doi.org/10.4995/WDSA-CCWI2022.2022.17011

Proceedings of the 2nd International Join Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI)
This book is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Editorial Universitat Politècnica de Valencia
Based on a work in http://ocs.editorial.upv.es/index.php/WDSA-CCWI/WDSA-CCWI2022

PROLOGUE

Welcome to the Conference Book of the 2nd International Conference on Water Distribution System Analysis (WDSA) and Computer and Control for the Water Industry (CCWI). After several years of pandemic, we were able to meet again to exchange research, developments and experiences related to the water industry and water distribution systems. The Universidad Politècnica de València (UPV) was in charge of welcoming the researchers from all over the world and hosted them during the week of the conference.

The most recent events have made it clear that infrastructures related to the water industry, especially supply, drainage and sanitation, are essential to ensure not only the survival of cities but also their sustainability and progress. In such difficult times, water supply has become a strategic element; water drainage has become possibly one of the greatest challenges related to climate change; and sewerage systems have become a way to control the sustainability of the system and at the same time serve as a control and warning of diseases and pandemics.

The importance of the water industry is palpable throughout the world. Beyond the existence of the Sustainable Development Goals, specifically Goal 6 (Clean Water and Sanitation), there is a growing interest in investing in water infrastructure. In the United States, the Bipartisan Infrastructure Law will deliver more than $50 billion to the Environmental Protection Agency to improve drinking water, wastewater, and stormwater infrastructure. The European Union, through its Cohesion Funds, will allocate 15 billion euros to water management. The Government of Spain has approved a Strategic Project for the Recovery and Economic Transformation of the country focused on the Digitalization of the Water Cycle. The project aims to modernize the sector and move towards more efficient and sustainable management. This project has an investment of more than 3,000 million euros and is estimated to create around 3500 jobs.

The main theme of this Conference is ‘Smart Water and Circular Economy: the next challenges’. After decades of evolution from the first mathematical models to the emergence of hydro-informatics, an evolution of network management using a digital environment and the tools associated with artificial intelligence has been necessary. The perception of water has also changed. It is no longer a service but a product. Integral use within a circular economy will become critical in the coming decades.

In the face of a rapidly changing world, the challenges posed to water distribution systems have evolved, demanding innovative and integrated approaches to tackle emerging complexities. This conference served as a crucible of ideas and solutions, a testament to our commitment to addressing these new challenges and finding sustainable pathways for the water industry. The field of hydroinformatics has emerged as a guiding light in this endeavor, harnessing the power of advanced technologies, data analytics, and modeling techniques to optimize water distribution systems.

With great enthusiasm and an overwhelming response, we received 300 proposals from experts, researchers, and practitioners across the globe. The sheer number of submissions underscores the significance of the conference's themes and the growing interest in this critical area of study. The collective passion and dedication demonstrated through these proposals promise to enrich our understanding of water distribution systems and propel us forward.

After a thorough and meticulous evaluation process, 270 abstracts were selected, reflecting the outstanding quality and relevance of the research conducted in the water industry and distribution systems. We extend our heartfelt gratitude to the authors whose exceptional work was approved for presentation at this esteemed conference.

In a testament to the depth and breadth of the research shared, 225 final papers were presented during the conference. These works encapsulate groundbreaking research, practical case studies, and novel approaches that promise to shape the future of water industry. As you delve into these pages, you will encounter a wealth of knowledge, ideas, and innovations that inspire new perspectives and avenues for progress.

While all the contributions to this conference are valuable and instrumental in advancing the field, it is important to note that only 145 papers are featured in this conference book. This selection was made to ensure a diverse representation of topics and perspectives, fostering a comprehensive and impactful dialogue among our esteemed participants. Among these carefully chosen papers are those that belong to the esteemed Industry Track, which, while essential, did not require formal paper submissions. Their invaluable insights and experiences further enrich the fabric of this conference, bridging the gap between theory and real-world applications. In addition, several outstanding papers are part of the Track of "The Battle of Intermittent Water Supply," a special collection that in process of been published in a dedicated Special Issue of the prestigious Journal of Water Resources Planning and Management. The inclusion of these papers in our conference underscores the importance of disseminating valuable knowledge and promoting rigorous research within the scientific community and beyond.

In the spirit of recognition and collaboration, we are pleased to announce that 14 papers were selected for publication in a special issue of the Journal of Hydroinformatics. These contributions exemplify excellence in research and exemplify the conference's aim to promote scholarly work and its broader impact on the water industry. The authors of these papers have demonstrated a commitment to advancing the science of Hydroinformatics and are at the forefront of driving positive change in the water sector.

This conference book stands as a testament to the collective pursuit of excellence, embracing innovation, and fostering collaboration to create a sustainable water future. We extend our heartfelt appreciation to all the authors, reviewers, and participants whose unwavering dedication has made this event a resounding success. The depth and breadth of knowledge shared here will undoubtedly contribute to the advancement of water distribution system analysis and the use of computer and control systems in the water industry.

Finally, we would like to thank all those who have made this event possible:

- To the WDSA and the CCWI for giving us the opportunity to organize this event.
- To the UPV that has offered all its infrastructure at our disposal.

- To the Department of Hydraulic Engineering and Environment, for their constant help.
- To the sponsors and collaborators for their invaluable support in the organization.
- To the members of the Steering Committee for their vigilance, suggestions and indications that have guided us.
- To the members of the Scientific Committee for their rigorous and hard work in reviewing the papers presented.
- To the members of the Local Organizing Committee, all their professors, students and collaborators. We would not have been able to organize this event without them.

Thank you all for making this conference a source of inspiration and enlightenment for all.

Conference Co-Chairs
Pedro L. Iglesias-Rey
Fernando Martínez Alzamora

CONFERENCE COMMITTEES

Steering Committee

Dominic Boccelli	University of Arizona (USA)
Joby Boxal	The University of Sheffield (UK)
Steven Buchberger	University of Cincinnati (USA)
David Butler	Univ. of Exeter (UK)
Kegong Diao	De Montfort University (UK)
Ives Filion	Queen's University (Canada)
Orazio Giustolisi	Technical University of Bari (Italy)
Walter Grayman	Walter M. Grayman Consulting Engineer (USA)
Pedro Iglesias	Universitat Politècnica de València (Valencia Tech) (Spain)
Zoran Kapelan	TU Delft (Netherlands)
Javier Martínez	Universitat Politècnica de València (Valencia Tech) (Spain)
Fernando Martínez	Universitat Politècnica de València (Valencia Tech) (Spain)
Regan Murray	US EPA (USA)
Avi Ostfeld	Technion - Israel Institute of Technology (Israel)
Juan Saldarriaga	Universidad de Los Andes (Colombia)
Dragan Savic	KWR-Water Research Institute (Netherlands)
Angus Simpson	The University of Adelaide (Australia)
Ivan Stoianov	Imperial College London (UK)
Kobus van Zyl	The University of Auckland (New Zeland)

Local Organizing Committee

The Local Organizing Committee is formed by staff and students of the Department of Hydraulic Engineering and Environment of the UPV.

Pedro L. Iglesias-Rey
F. Javier Martínez-Solano
Vicente S. Fuertes Miquel
Francisco J. García Mares
Elena Gómez Sellés
Gonzalo López Patiño
Christian Xavier Briceño Leon
Gustavo A. Delgado Sandoval
Melvin A. García Espinal
Angélica G. Simbaña Pumisacho
Elisa Martínez Martínez
Fernando Martínez-Alzamora
Pilar Conejos Fuertes
Jorge García-Serra
Salvador García Todolí
Miguel Angel Jiménez Bello
Javier Soriano-Olivares
Leonardo Bayas Jimenez
Arianne I. Estrada Estrada
Alejandra Guarachí Quiñones
F. Sarahi Pecina Nuñez
Ester Martínez Martínez

Scientific Committee

Mohsen Aghashahi	Texas A&M University (USA)
Francisco Arregui	Univ. Politècnica de València (Spain)
Luigi Berardi	Università degi Studi "G. d'Annunzio" (Italy)
Mirjam Blokker	KWR-Water Research Institute (Netherlands)
Dominic Boccelli	University of Arizona (USA)
Joby Boxal	The University of Sheffield (UK)
Bruno Brunone	Università degli Studi di Perugia (Italy)
Steven Buchberger	University of Cincinnati (USA)
David Butler	Univ. of Exeter (UK)
Ricardo Cobacho	Univ. Politècnica de València (Spain)
Didia Covas	Instituto Superior Técnico (Portugal)
Enrico Creaco	Univesity of Pavia (Italy)
Maria C. Cunha	Universidade de Coimbra (Portugal)
Graeme Dandy	The University of Adelaide (Australia)
Jochen Deuerlein	3S Consult (Germany)
Kegong Diao	De Montfort University (UK)
Demetrios G. Eliades	University of Cyprus (Cyprus)
Raziyeh Farmani	University of Exeter (UK)
Ives Filion	Queen's University (Canada)
Marco Franchini	Università degli Studi di Ferrara (Italy)
Orazio Giustolisi	Technical University of Bari (Italy)
Walter Grayman	Walter M. Grayman Consulting Engineer (USA)
Pedro L. Iglesias	Univ. Politècnica de València (Spain)
Adesola Ilemobade	University of Witwatersrandt (South Africa)
Donghwi Jung	Korea University (South Korea)
Zoran Kapelan	TU Delft (Netherlands)
Bryan Karney	University of Toronto (Canada)
Joong Hoon Kim	Korea University (South Korea)
Martin Lambert	The University of Adelaide (Australia)
Kevin Lansey	The Univesity of Arizona (USA)
Juneseok Lee	Manhattan College (USA)
P. Amparo Lopez	Univ. Politècnica de València (Spain)
Christos Makropoulos	National Technical University of Athens, (Greece)
Javier Martínez	Univ. Politècnica de València (Spain)
Fernando Martínez	Univ. Politècnica de València (Spain)
Daniel Mora	Univesity of Talca (Chile)
Regan Murray	US EPA (USA)
Lindell Ormsbee	University of Kentucky. College of Engineering (USA)
Avi Ostfeld	Technion - Israel Institute of Technology (Israel)
Olivier Piller	INRAE, ETBX Research Unit (France)
Heber Pimentel	Universidade Federal da Paraiba (Brazil)
Dusan Prodanovic	University of Belgrade (Serbia)
Raido Puust	Tallinn University of Technology (Estonia)

cc BY NC SA

Helena Ramos	Instituto Superior Técnico (Portugal)
Juan Saldarriaga	Universidad de Los Andes (Colombia)
Dragan Savic	KWR-Water Research Institute (Netherlands)
Hwee See	Mott MacDonald (UK)
Angus Simpson	The University of Adelaide (Australia)
Robert Sitzenfrei	University of Innsbruck (Austria)
Javier Soriano	Univ. Politècnica de València (Spain)
Vanessa Speight	The University of Sheffield (UK)
David Steffelbauer	Norwegian University of Science and Technology (Norway)
Ivan Stoianov	Imperial College London (UK)
Ricardo Taormina	TUDelft (Netherlands)
Ezio Todini	Freelance Consultant (Italy)
James G. Uber	CityLogics (USA)
Kobus van Zyl	The University of Auckland (New Zeland)

Battle of Water Networks Committee

The members of the Committee in charge of the proposal and evaluation of the Intermittent Water Supply Battle are:

Pedro L. Iglesias	Univ. Politècnica de València (Spain)
Javier Martínez	Univ. Politècnica de València (Spain)
Fernando Martínez	Univ. Politècnica de València (Spain)
Avi Ostfeld	Technion - Israel Institute of Technology (Israel)
Gustavo Delgado	Univ. Politècnica de València (Spain)

SPONSORS AND COLLABORATORS

The development of the conference would not have been possible without the personal, emotional, financial and material support of all the entities, organizations and institutions that supported the conference from the beginning. We would like to express the gratitude of the Organizing Committee to Sponsors and Collaborators

Sponsors

IDRICA | MOLECOR

TALIS | BERMAD

cátedra aguas de valencia | iiama | EWRI ASCE

Collaborators

VLC VALÈNCIA | global omnium

GENERALITAT VALENCIANA parc natural de l'albufera | ESCUELA TÉCNICA SUPERIOR INGENIERÍA INDUSTRIAL VALENCIA | Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

WDSA CCWI 2022

INDEX

Investigating an alternative to exhumed grey cast iron water pipes for small-scale fatigue tests
Edward D'Arcy Aston John, Joby Boxall, Richard Collins, Elisabeth Bowman & Luca Susmel 1

Water Quality in Drinking Water Distribution Systems: A Whole-Systems Approach to Decision Making
Sally L Weston, Manuel Rodriguez & Sonja Behmel 13

A Model of Intermittent Water Supply Simulating the Inequitable Distribution of Water
Matthew Henry MacRorie, Sally Weston, Vanessa Speight, Robin Price & Richard Collins 20

Characterizing the effects of water distribution system topology modifications on its dynamic behaviour through connectivity metrics
Valentina Marsili, Stefano Alvisi, Filomena Maietta, Caterina Capponi, Silvia Meniconi, Bruno Brunone & Marco Franchini 28

Local control schemes for real-time optimization of variable speed pumps
Elad Salomons, Mashor Housh & Uri Shamir 38

Experimental study of mixing phenomenon in water distribution networks under real-world conditions
Reza Yousefian & Sophie Duchesne 47

Optimal Operation in Sectorized Networks with Intermittent Water Distribution
Rui Gabriel Souza, Gustavo Meirelles & Bruno Brentan 58

Is water quality based stormwater management actually feasible? A SWMM-based study on the trade-offs of various stormwater management approaches
Kristjan Suits, Anatoli Vassiljev & Ivar Annus, Nils Kändler 69

Comparison of second and third order algorithms for steady state resolution of water distribution networks
Enrico Creaco, Armando Di Nardo, Michele Iervolino & Giovanni Francesco Santonastaso 84

Sensitivity analysis of water distribution networks deterioration in hydraulic and economic parameters
Leandro Alves Evangelista, Gustavo Meirelles Lima & Bruno Melo Brentan 93

Multi-leak detection and isolation in water distribution network
Debora Alves, Joaquim Blesa, Eric Duviella & Lala Rajaoarisoa 108

Smart water applications vs information and communication technologies – an integrative selection framework
Martin Oberascher, Wolfgang Rauch & Robert Sitzenfrei 120

Comparison between the Top-down and Bottom-up approach for the diffuse-dispersive phenomenon analysis
Stefania Piazza, E. J. Mirjam Blokker, Mariacrocetta Sambito & Gabriele Freni 131

COSMOS – A framework for containerised, distributed creation, execution and analysis of hydraulic water distribution system models
Georg Arbesser-Rastburg, David Camhy & Daniela Fuchs-Hanusch 140

Water Consumption variation in Latin America due to COVID-19 Pandemic
Catalina Ortiz, William Clavijo, Jürgen Mahlknecht & Juan Saldarriaga 154

Analysis of online pressure for resilience phase characterisation of leakages/burst events
Sotudeh Hoseini Ghafari, Jorge Francés-Chust, Olivier Piller & David Ayala-Cabrera 165

The impact of drinking water network model spatial and temporal scale on hydraulic metrics indicating discolouration risk
Reinar Lokk, Mirjam Blokker, Joby Boxall, Michele Romano, Anna Provost & Stewart Husband 179

Leak detection and location in a real water distribution network using a model-based technique
Bruno Ferreira, Nelson Carriço & Dídia Covas 193

Dual estimation of iron oxide deposition on drinking water PVC pipes using calibrated turbidity data and brightfield microscopy in a full-scale laboratory system
Artur Sass Braga & Yves Filion 205

Model fast deposition of fine iron oxide particles on PVC pipe mains during the passage of a suspended particle plume in a full-scale laboratory system
Artur Sass Braga & Yves Filion 220

The identification of variable shear strength of particles deposited on drinking water PVC pipes after the passage of a suspended particle plume in a full-scale laboratory system
Artur Sass Braga & Yves Filion 226

W-NET4.0 – Integrated Platform and Digital Twin for Small and Medium Sized Water Supply Utilities
Jochen Werner Deuerlein, Thomas Bernard, Naga Mamatha Gonuguntla, Jorge Thomas, Armin Canzler, Heiko Keifenheim, Susanne Wiese, Ruediger Hoeche, Joachim Rapp & Salomé Parra 236

Pumping Station Design with an Analysis of Variability of Demand and Considering Techno-Economic and Environmental Criteria through the AHP Method
Christian Xavier Briceno León, Pedro Luis Iglesias Rey, Francisco Xavier Martinez Solano & Enrico Creaco 248

An improved control strategy for high-pressure pumping irrigation systems
Ye Wang, Qi Zhao, Erik Weyer, Wenyan Wu, Angus Simpson & Ailsa Willis 265

Water distribution system design with behind-the-meter solar energy under various discount rates
Qi Zhao, Angus Simpson, Wenyan Wu & Ailsa Willis 274

Graph based method for critical pipe analysis in urban drainage networks and the effect of loop degree
Aun Dastgir, Martin Oberascher, Sina Hesarkazazzi & Robert Sitzenfrei 287

NRW Estimation and Localization in Water Distribution Networks via Hydraulic Model Calibration using 24/7 Continuously Monitoring Data
Alvin Wei Ze Chew, Zheng Yi Wu, Rony Kalfarisi, Meng Xue, Jocelyn Pok, Jianping Cai, Kah Cheong Lai, Sock Fang Hew & Jia Jie Wong 297

Acoustic Data Analysis Framework for Near Real-Time Leakage Detection and Localization for Smart Water Grid
Alvin Wei Ze Chew, Zheng Yi Wu, Rony Kalfarisi, Meng Xue, Jocelyn Pok, Jianping Cai, Kah Cheong Lai, Sock Fang Hew & Jia Jie Wong 311

Modelling of Air Pocket Entrapment during Pipe Filling Events in Intermittent Water Supply Systems
Joao Ferreira, David Ferras, Didia Covas & Zoran Kapelan 324

A decision-making approach to assess and prioritise intervention solutions in water distribution systems
Marta Cabral, Dália Loureiro & Dídia Covas 334

Measuring drinking water temperature changes in a distribution network
Mirjam Blokker, Joost van Summeren & Karel van Laarhoven 350

Small community water supplies in the Isiolo County, Kenya
Paolo Vezza, Roberto Arnesano, Fulvio Boano & Andrea Bessone 364

Advancing towards semi-automatic labeling of GPR images to improve visualizations of pipes and leakage in water distribution networks using multi-agent systems and machine learning techniques
Gemma Stanton & David Ayala-Cabrera 370

The Use of Resilience Metrics to Support Decision Making in Drinking Water Systems
Joana Carneiro, Dália Loureiro & Dídia Covas 382

AQUARELLUS: a numerical tool to calculate accumulation of particulate matter in distribution systems
Joost van Summeren, Amitosh Dash, Mark Morley, Luuk de Waal & Jip van Steen 397

Advanced fire flow risk analysis using EPANET
Alexander Nicholas Sinske, Altus Hugo de Klerk & George Adrian van Heerden 412

Data Driven Approach for Equitable Supply in Water Networks
Adhityan R, Varghese Kurian & Sridharakumar Narasimhan 426

UrbanLemma: A serious game to support adoption of sustainable urban drainage solutions
Aashna Mittal, Jessica Nguyen, Lisa Scholten & Zoran Kapelan 438

Modelling water mixing and renewal in storage tanks
Didia Isabel Cameira Covas, Alexandre Pinheiro, Sofia Vaz, Laura Monteiro, Nuno Martins & Ana Margarida Ricardo 447

Discovering differences in iron and manganese behaviour in service reservoirs
Anastasia Doronina, Stewart Husband, Joby Boxall & Vanessa Speight 462

Flow dynamics in a pipe containing uneven roughness elements
Katrin Kaur, Anatoli Vassiljev, Ivar Annus, Murel Truu & Nils Kändler 475

GIS-integrated pluvial flood risk assessment methodology for urban areas
Murel Truu, Ivar Annus & Nils Kändler 486

TEmporal Scale SIgmoid Curve (TESIC): A tool to characterize short-term demand variability at water supply systems
Emilio Ruiz Gómez, Sarai Díaz García & Javier González Pérez 501

Multidimensional data generation in water distribution systems using the Cycle-GAN
Sehyeong Kim & Donghwi Jung 510

An experimental study on early leak localization in drinking water networks using pressure measurements
Yannick Deleuze, Arley Nova Rincón, Yves-Marie Batany, Teodulo Abril, Damien Chenu & Nicolas Roux 516

Water Distribution Network operation optimization: an industrial perspective
Marco Lauricella, Felix Lenders & Matthias Biskoping 528

Biokinetic and artificial intelligence models for the simulation of nitrous oxide emissions from wastewater treatment plants
Siddharth Seshan, Johann Poinapen, Jules B. van Lier & Zoran Kapelan 540

Flow ratio and pressure control in a low losses WDS – Analysis of dynamics
Bogumil Ulanicki, Philippe Beaujean, Juliaan Plancke & Bastien Leruth 547

PRVs working in parallel configurations – Analysis of dynamics
Bogumil Ulanicki, Philippe Beaujean, Juliaan Plancke & Bastien Leruth 559

Hydraulic State Estimation: pilot implementation in a water distribution system
Emilio Ruiz Gómez, Sarai Díaz García & Javier González Pérez 574

Modelling a comprehensive relation between water quality and cement degradation in the drinking water distribution
Karel Antonie van Laarhoven, Alex Hocking & Martin Korevaar 579

On the use of SINDy for WDN
Mario Castro-Gama 591

Household behaviour and energy loss in intermittent water supply networks
Hamidreza Mohabbat & David D.J. Meyer 598

Advances in premise plumbing modeling
William Platten, Regan Murray, Juneseok Lee, Robert Janke, Terra Haxton, Walter Grayman, Jonathan Burkhardt & Steven Buchberger 611

Disinfection residual behaviour within drinking water distribution systems
Jade Rogers, Katherine Fish, Vanessa Speight, Hunter Fairley, Graham Knowles & Joby Boxall 619

An approach to improve drainage networks based on the study of flood risk
Leonardo Bayas-Jiménez, F. Javier Martínez-Solano, Pedro L. Iglesias-Rey & Fulvio Boano 628

Survey Exploring Customer Complaint Management and Smart Water Technology Adoption Across U.S. Water Systems
Morgan DiCarlo, Emily Zechman Berglund, Nikhil Kaza, Andrew Grieshop, Luke Shealy & Adam Behr 640

Estimation of the spatial distribution of substances in an anaerobic digestion tanks with CFD
Soroush Dabiri, Prashant Kumar & Wolfgang Rauch 649

Optimal design of sewer networks including topographic criteria and drop manholes
Juan Saldarriaga, Juana Herrán & Pedro L. Iglesias-Rey 656

Roadmap towards smart wastewater treatment facilities
Daniel Aguado García, Henri Haimi, Michela Mulas & Francesco Corona 666

Metadata: a must for the digital transition of wastewater treatment plants
Daniel Aguado García, Frank Blumensaat, Juan Antonio Baeza, Kris Villez, Mª Victoria Ruano, Oscar Samuelsson, Queralt Plana & Janelcy Alferez 680

Optimal design-for-control of water distribution networks via convex relaxation
Filippo Pecci, Ivan Stoianov & Avi Ostfeld 688

An optimization framework for large water distribution systems based on complex network analysis
Robert Sitzenfrei 696

Long-term transitioning of water distribution systems: ERC water-futures project
Dragan Savic, Barbara Hammer, Phoebe Koundouri & Marios Polycarpou 706

A review on staged design of water distribution networks
Lydia Tsiami, Christos Makropoulos & Dragan Savic 718

Performance evaluation of buried uPVC pipes by numerical simulation of soil-pipe interaction
Milad Latifi, Ramiz Beig Zali, Raziyeh Farmani, Akbar Javadi & Suhayl Zulfiquar 729

Leak localisation method using a detailed hydraulic model combined with high resolution pressure sensors applied to a real network
Sergi Grau Torrent, Ramon Pérez Magrané, Xavier Torret Requena & David Casado Ruiz 741

Development of a graph dynamic sectorization tool using QGIS/PostgreSQL to plan and operate water distribution networks
Xavier Torret Requena, Sergi Grau Torrent, Ramon Pérez Magrané & Enric Amat Miralles 750

Effect of swabbing cleaning method on biofilm communities of a drinking water distribution system in Madrid (Spain)
Carolina Calero Preciado, Emelina Ruth Rodríguez Ruiz & Manuel José Arias Guedón 757

Dynamic edge betweenness centrality and optimal design of water distribution networks
Mohsen Hajibabaei, Sina Hesarkazzazi, Amin Minaei, Dragan Savić & Robert Sitzenfrei 767

Hydraulic and Co-located Pipe Criticalities in the Rehabilitation of Water Distribution Mains
Amin Minaei, Mohsen Hajibabaei, Enrico Creaco & Robert Sitzenfrei 778

A Bayesian Generative Adversarial Networks (GAN) to generate synthetic time-series data, application in combined sewer flow prediction
Amin Ebrahim Bakhshipour, Alireza Koochali, Ulrich Dittmer, Ali Haghighi, Ahmed Sheraz & Andreas Dengel 791

Sustainable wastewater treatment solutions for water-smart circular economy
Daniel Aguado García, Henri Haimi, Anna Mikola, Ana Soares & Ulf Jeppsson 801

Machine learning methodologies to predict possible water quality anomalies as a support tool for online monitoring of organic parameters
Leonid Kadinski, Jonas Schuster, Gopinathan R. Abhijith, Cao Hao, Anissa Grieb, Thomas Meier, Pu Li, Mathias Ernst & Avi Ostfeld F. ASCE 812

Multi-objective optimization of water distribution system under uncertainty using robust optimization
Sriman Pankaj Boindala & Avi Ostfeld F. ASCE 818

Evaluation of physical and water quality indicators in aductors in hydraulic transient events
Mariele de Souza Parra Agostinho Agostinho, Danieli Mara Ferreira & Cristovão Vicente Scapulatempo Fernandes 829

Pressure sensor placement in water networks based on sensitivity matrix: a real instrumentation project plan
Joan van Eeckhout, Josep Cugueró & Ramon Pérez 844

Disinfection by products estimation in a water distribution network
Laura Vinardell Magre, Irene Jubany Güell & Ramon Pérez Magrané 859

Infrastructure benchmarking for semi-real urban stormwater networks
Sina Hesarkazzazi, Mohsen Hajibabaei, Aun Dastgir & Robert Sitzenfrei 869

Methodology for analysis of the SDGs compliance in urban water systems. Implementation in case studies in the Valencian Community
Camila Andrea Garcia Rodriguez, Pilar Conejos Fuertes, Jaime Castillo Soria, Petra Amparo Lopéz Jiménez & Modesto Perez Sanchez 878

Modelling consumers in intermittent water supplies: a comparative review of EPANET - based methods
Omar Abdelazeem & David Donald James Meyer 892

Robustness of profile sampling in detecting dissolved lead in houshold drinking water
Amitosh Dash, Jip van Steen & Mirjam Blokker 907

Roadmap Drinking Water Distribution: an inventory of the challenges of the Dutch Water companies and research needs towards a future proof water supply
Mirjam Blokker, Karel van Laarhoven, Petra Holzhaus & Jan Vreeburg 922

Data Driven Monitoring of Water Distribution Networks
Rohit Raphael & Sridharakumar Narasimhan 928

Coupling Agent-based Modeling with Water Distribution System Models to Simulate Social Distancing and Water Infrastructure Performance during COVID-19
Brent Vizanko, Leonid Kadinski, Avi Ostfeld, Emily Berglund & Christopher L Cummings 938

Iteratively tuning the regularisation parameter in an inverse method for localising leaks in water distribution networks
Yuanyang Liu, Ivan Stoianov, Filippo Pecci 947

Determination of the cost of extraction and sale price of water for drinking use: a case study
Mª Magdalena Sánchez Astello, L.A. Bravo Anaya 957

Which flow and pressure constraints for sustainable operation of Water Distribution Systems?
Olivier Piller, Jochen Deuerlein, Sylvan Elhay & Angus Simpson 966

Important Factors For Water Main Break Prediction Across 13 Canadian Systems
Rebecca Dziedzic, Sadaf Gharaati 973

Valve Loss Curve Estimation Using Real-Time SCADA Data
SM Masud Rana, Gal Perelman, Elad Salomons & James G Uber 987

Multi-objective insights and analysis on data driven classifiers for anomaly detection in water distribution systems
Elizabeth Pauline Carreño-Alvarado, Mayra Hernández Alba, Gilberto Reynoso Meza 996

Analysis of PDA-based Water Distribution System Suspension Risk using statistical and machine learning method
Yoojin Oh, Haekeum Park, Jinseok Hyung, Taehyeon Kim, Kibum Kim & Jayong Koo 1007

Efficiency analysis and evaluation model development of water distribution system rebuilding project using DEA method
Haekeum Park, Kibum Kim, Jinseok Hyung, Taehyeon Kim & Jayong Koo 1013

Mass Balance Calibration of Water Distribution Networks: application to a real case study
Daniele Biagio Laucelli, Orazio Giustolisi, Francesco Gino Ciliberti & Luigi Berardi 1020

Energy diagnosis of pressurized water systems with the ENERGOS tool
Elena Gómez, Roberto del Teso, Enrique Cabrera, Elvira Estruch-Juan, Pascual Jose Maximino, Miguel Ortiz, Guillermo del Pozo & Carlos Marco 1035

pySIMDEUM - An open-source stochastic water demand end-use model in Python
David Bernhard Steffelbauer, Bram Hillebrand, Mirjam Blokker 1049

Designing advanced asset management for Water Distribution Networks: application to a real case study
Daniele Biagio Laucelli, Serena Spagnuolo, Antonio Rinaldi, Gianluca Perrone, Luigi Berardi & Orazio Giustolisi 1057

Domain Analysis of Water Distribution Networks
Luigi Berardi, Francesco Gino Ciliberti, Daniele Biagio Laucelli & Orazio Giustolisi 1073

Development of prediction model of ozone dosage and residual ozone concentration using machine learning methods in ozone process of drinking water treatment process
Jaeyoung Kwon, Jinseok Hyung, Taehyeon Kim, Haekuem Park, Jayong Koo 1085

Why aren't surrogate reliability indices so reliable? Can they be improved?
Joaquim José de Oliveira Sousa, João Muranho, Marco Bonora & Mario Maiolo 1092

Energy equations to analyse pressurized water transport systems
Roberto del Teso March, Elena Gómez Sellés, Elvira Estruch-Juan & Enrique Cabrera Marcet 1106

Unsteady Friction Modeling Technique for Lagrangian Approaches in Transient Simulations
Mohamad Zeidan, Avi Ostfeld 1114

The Shape of Water Distribution Networks - Describing local structures of water networks via graphlet analysis
Bulat Kerimov, Franz Tscheikner-Gratl, Riccardo Taormina & David Steffelbauer 1122

A stochastic sewer model to predict pipe flows and pollutant loads in an urban drainage system
William Addison Atkinson, Albert Chen, Fayyaz Memon & Jan Hofman 1136

IoT.H2O - Supervision and Control for Water Systems based on Low-Cost Hardware and Open-Source Software
Harald Roclawski, Laura Sterle, Martin Böhle, Thomas Krätzig 1141

Grey-box dynamic model of a drinking water treatment plant
Blaž Korotaj, Mario Vašak, Hrvoje Novak 1156

A Thorough Cybersecurity Dataset for Intrusion Detection in Smart Water Networks
Riccardo Taormina, Andrés F. Murillo, Nils Ole Tippenhauer & Stefano Galelli 1171

PathoInvest: Pathogen Contamination Investigations during Emergencies
Sotirios Paraskevopoulos, Stelios Vrachimis, Marios Kyriakou, Pavlos Pavlou, Dimitris Kouzapas, George Milis, Patrick Smeets, Demetrios G. Eliades, Gertjan Medema, Marios Polycarpou, Christos Panayiotou 1182

Optimization of reservoir treatment levels considering uncertainty in mixing at cross junctions in water distribution systems using Info-gap decision theory
Sriman Pankaj Boindala, G Jaykrishnan, Avi Ostfeld 1196

Numerical and experimental investigation on leak discharge-pressure relationship for linear elastic deforming pipes
Yuanzhe Li, Jinliang Gao, Jianxun Chen, Hangwei Zheng, Guosheng Sun & Chengzhi Zheng 1211

High resolution water pressure monitoring for the assessment of fatigue damage in water distribution pipes
Carlos Jara-Arriagada, Ivan Stoianov 1220

A Nonlinear Model Predictive Control Framework for Dynamic Water Network Optimization
Ernesto Arandia, James Uber 1236

Leakage control of water distribution system by drop-restore pressure based on viscoelastic mechanism
Jinliang Gao, Jianxun Chen, Wenyan Wu, Liqun Deng, Kunyi Li & Yi Yuan 1251

Study on iron dispersion law and control measures of dead-end branch pipe
Jianxun Chen, Jinliang Gao, Wenyan Wu, Guanghui Wang, Qiang Ding, Shihua Qi & Bo Li 1264

Analysis of friction models during simulations of filling processes in single pipelines
Vicente S. Fuertes-Miquel, Óscar E. Coronado-Hernández, Mohsen Besharat & Pedro L. Iglesias-Rey 1279

Investigation of groundwater consumption to cope with the inadequate piped water supply in continuous and intermittent supply systems: A case study in Bangalore, India.
B.N Priyanka, Bharanidharan B, Sheetal Kumar KR, MS Mohan Kumar, VV Srinivas, Sanjana R Nibgoor & Kishore Y 1290

Reduction of the search space for the optimization problem of the design of the pumping station through the automatic identification of infeasible flow distributions
Jimmy H. Gutierrez-Bahamondes, Daniel Mora-Melia, Bastian Valdivia-Muñoz & Pedro L. Iglesias-Rey 1298

Non-intrusive water usage classification considering limited training data
Pavlos Pavlou, Stelios Vrachimis, Demetrios G. Eliades & Marios M. Polycarpou 1309

Water for firefighting: A comparative study across several cities
Adeshola Ayodeji Ilemobade 1324

Supervised machine learning models for leak detection in water distribution systems
Lochan Basnet, James Levis, Downey Brill, Ranji Ranjithan & Kumar Mahinthakumar 1331

Water distribution network disruptive events. Generation and exploration of an incident hub to increase the network preparedness
David Ayala-Cabrera, Jorge Francés-Chust, Sotudeh Hoseini-Ghafari, Gemma Stanton & Joaquín Izquierdo 1346

Experimental analysis for the losses assessment in water distribution systems
Giovanna Darvini, Luciano Soldini 1356

A critical review of leakage detection strategies including pressure and water quality sensor placement in water distribution systems – sole and integrated approaches for leakage and contamination intrusion
Shweta Rathi 1364

Biofungus: fungus MBBR pilot plant on Murcia Este WWTP
E. Mena, A. Gadea, A. Monreal-Bernal, S. López-García, V. Garre, A. J. Lara-Guillén 1377

Water consumption analysis during night hours of residential customers
Arianne Isabella Estrada Estrada, Francisco Arregui de la Cruz, Javier Soriano Olivares, Román Ponz Carcelén & David Torres Toro 1388

Development of a tool for the optimization and regulation of hydraulic microgeneration systems adapted to the demand and flow variations aimed at the clean energy recovery in water supply networks
Melvin Alfonso Garcia Espinal, Modesto Pérez-Sánchez , Francisco Javier Sánchez-Romero , P. Amparo López Jiménez & Pilar Conejos Fuertes 1403

Potential reduction of leakage volume by combining dynamic pressure management and energy recovery in Valencia
Alejandra Guarachi Quiñones, Francisco Javier Martínez Solano, Carmen Sánchez Briones & Andross Pérez Lleó 1415

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14005

INVESTIGATING AN ALTERNATIVE TO EXHUMED GREY CAST IRON WATER PIPES FOR SMALL-SCALE FATIGUE TESTS

Edward John[1], Joby Boxall[2], Richard Collins[3], Elisabeth Bowman[4], and Luca Susmel[5]

[1,2,3,4,5]Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, UK

[1] *edajohn1@sheffield.ac.uk*, [2] *j.b.boxall@sheffield.ac.uk*, [3] *r.p.collins@sheffield.ac.uk*,
[4] *e.bowman@sheffield.ac.uk*, [5] *l.susmel@sheffield.ac.uk*

Abstract

Reducing and preventing leakage is a priority for water distribution network managers in many countries, including the UK. Understanding the mechanisms that cause leaks to form, and developing the ability to model these processes, will enable proactive replacement of water pipes before they start to leak. Smaller diameter Grey Cast Iron (GCI) water pipes are understood to experience biaxial, repeating loads, so fatigue cracking may be a cause of leakage for these pipes. To investigate this fatigue cracking mechanism a small-scale biaxial fatigue experiment is under development at The University of Sheffield. A large number of small diameter, un-corroded GCI pipes are needed to serve as specimens for this experiment. Therefore, in this work off-the-shelf BS416-2 DN 50 mm soil pipes are explored as an alternative to using exhumed pipes, which are often corroded and ≥76.2 mm diameter. The graphite microstructure and tensile stress-strain behaviour of a BS416-2 pipe were characterised and compared with literature data for exhumed spun-cast GCI water pipes, and a good agreement was found. This work concluded that BS 416-2 soil pipes can be used to represent spun-cast GCI water pipes in small-scale destructive experiments.

Keywords
Water distribution pipe, grey cast iron, leakage, material properties, graphite microstructure.

1 INTRODUCTION

Reducing and preventing leakage is a priority for water distribution network managers in many countries, including the UK. Understanding the mechanisms that cause leaks to form, and developing the ability to model these processes, will enable proactive replacement of water pipes before they start to leak. Grey Cast Iron (GCI) pipes are amongst the oldest pipes still in service, most having been installed pre-1960, and are still common in many UK networks [1]. One way that GCI pipes can develop leaks is by forming through-wall cracks [2].

Smaller diameter GCI pipes experience biaxial stress states; internal water pressure causes a pipe to experience stress acting around its circumference [3] and bending loads, such as vehicles and soil response to moisture change, cause stress acting in the pipe's axial direction [4], [5]. These loads are time variable with the potential to cycle tens or even hundreds of times per day [6], [7].

Developing a fundamental understanding of how GCI pipes develop through-wall cracks in response to these loads must be done experimentally so that the loading can be controlled, tests can be repeated, and the failure mode can be observed. For example, the formation of fatigue cracks in GCI pipes has been proposed as a potential leak initiation mechanism [8], [9], but physical tests are needed to confirm this and validate a suitable fatigue failure criterion. These tests must include cyclic, biaxial stress conditions to investigate the effect of this loading type on the time it takes for a leaking crack to form. However, no process or equipment has previously been developed that can apply cyclic, biaxial stresses to a GCI water pipe.

A small-scale experiment that is capable of fatigue testing GCI pipes using biaxial stress histories is under development at the University of Sheffield to investigate the fatigue cracking leakage mechanism. For the specimens in this experiment, small diameter, un-corroded GCI pipes are needed. Smaller diameter pipes require lower bending loads to generate a given stress magnitude, enabling higher loading frequencies and shorter test durations. Using un-corroded pipes means artificial corrosion pit geometries can be added in a controlled way. Exhumed pipes are typically ≥76.2 mm nominal diameter and in variable states of deterioration, sometimes featuring deep corrosion pits or through-wall cracks [10], [11]. Newly manufactured GCI water pipes without corrosion would be more suitable but unfortunately new GCI water pipes are no longer widely manufactured because ductile iron, steel, and plastic pipes are used in new installations [1]. A potential alternative to new GCI water pipes is off-the-shelf soil pipes made in accordance with BS 416-2, which are easy to obtain, manufactured from GCI, and available with nominal diameters from 50 mm to 150 mm [12].

The work presented in this paper investigates whether BS 416-2 soil pipes can be used to represent GCI water pipes in small-scale destructive experiments where small diameter, un-corroded pipes are needed. Two aspects were considered in this investigation: the pipe's graphite microstructure, and the pipe's tensile stress-strain response. The graphite microstructure of GCI is widely understood to have a strong influence on the macro properties of the material [13]–[15]. GCI pipes are more vulnerable to tensile failure, due to their compressive strength being 1.3 to 4 times larger than their tensile strength [16], [17], so matching the tensile stress-strain and failure behaviour is essential. For the BS 416-2 pipe's microstructure and stress-strain behaviour to be suitable they must be similar to literature data for exhumed GCI water pipes.

The structure of this paper is as follows: firstly, available information from the literature regarding the observed microstructure and tensile stress-strain behaviour is presented; then, the methods used to assess the BS 416-2 pipe are explained; and finally, the results obtained are compared with the literature data to assess the similarity of BS416-2 pipes to exhumed GCI water pipes.

2 GCI WATER PIPE PROPERTIES FROM LITERATURE

2.1 Background

A defining feature of GCI is the microscopic graphite flakes contained within the iron's microstructure. These graphite flakes act as cracks within the material, causing GCI to demonstrate brittle tensile behaviour [14], [17]. Graphite flakes also enable graphitic corrosion, which can result in deep corrosion pits forming [18]. The size and distribution of the graphite flakes is known to strongly affect the material's mechanical properties, with larger flakes generally resulting in a lower bulk tensile strength [16]. The size and distribution of the graphite flakes is determined by the chemical composition of the GCI and the rate at which the iron cooled, with slower cooling generally allowing larger flakes to form [14]. The cooling rate can be influenced by the pipe's wall thickness, the ambient temperature, and the manufacturing process used.

GCI pipes broadly fall into two categories which are determined by the manufacturing technique used. In the UK, pit-cast iron pipes were manufactured from the mid-1800s until the 1920s using vertical sand moulds, which generally resulted in slow cooling times and larger graphite flakes. Spun-cast pipes were manufactured from the 1920s to 1960s and were cast in horizontal, spinning, water-cooled metal moulds, resulting in fast cooling times and smaller graphite flakes [1], [19]–[21]. The BS 416-2 pipe in this investigation has been produced using a spin-casting technique.

2.2 Graphite Flake Microstructure

Most previous studies report the graphite flake morphology and sizes of the GCI pipes examined according to ASTM A247-19 [22], or a preceding version of the standard. A247-19 describes the graphite microstructure according to its size, form, and distribution.

Literature data for pit-cast iron pipes give flake sizes of between Class 5 and Class 1, or 40 μm to ≥640 μm, with the large flakes tending to occur in the central region of the pipe wall [11], [14], [16]. Similar data for spun-cast pipes gives sizes between Class 8 and Class 4, or <10 μm to 160 μm, with the larger flakes often occurring towards the inside wall of the pipe [11], [14], [16].

These size ranges therefore broadly match the theory that pit-cast pipes should have larger flakes than spun-cast pipes. The trends across the pipe walls also suggest that the centre of the wall cools most slowly in pit-cast pipes, whereas the inside wall cools most slowly in spun-cast pipes. The images in Figure 1 illustrate the typical difference observed in graphite flake size and distribution between pit and spun-cast pipes.

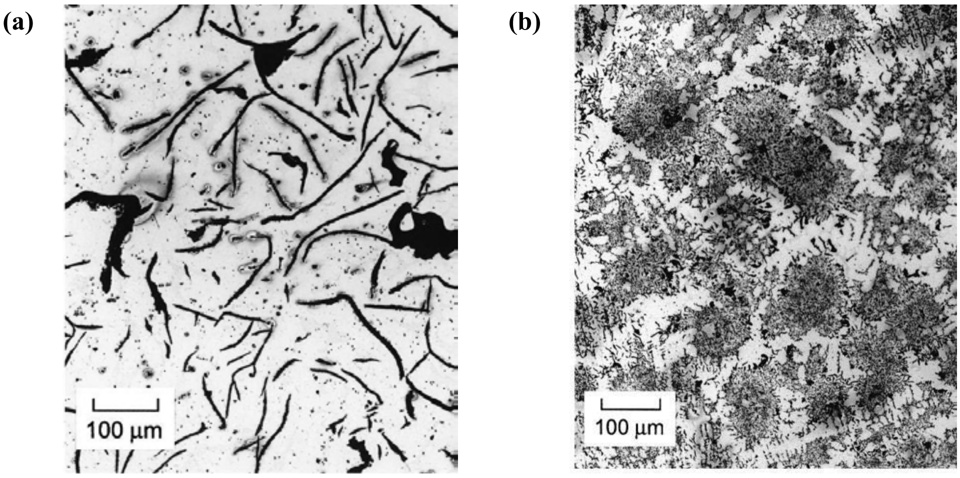


*Figure 1: Examples of **(a)** pit-cast and **(b)** spun-cast graphite microstructures from Makar and Rajani [14].*

For both pit and spun-cast pipes the graphite form reported is exclusively Type VII flake graphite, apart from one spun cast pipe inspected by Makar and Rajani [14], which featured some Type III compacted graphite at its outer edge. The authors attributed this irregularity to potential extended heat treatment.

In terms of graphite distribution, pit cast-pipes are reported to display Distributions A, B, and C (see Table 1) [11], [14], [16]. Spun-cast pipes are generally dominated by Distribution D, although Distributions A and C are often present near the inside wall of the pipe [14], [16].

Table 1: Description of ASTM graphite distributions, according to Makar and Rajani [14].

Flake Distribution	Description
A	Uniformly distributed, apparently randomly oriented flakes (see Figure 1a)
B	Rosette pattern of graphite flakes
C	Randomly oriented flakes of widely varying sized
D	Very fine pattern of flakes surrounding areas without graphite (see Figure 1b)
E	Graphite flakes have preferred orientation and appear in quasi-regular pattern

interest in this critical area of study. The collective passion and dedication demonstrated through these proposals promise to enrich our understanding of water distribution systems and propel us forward.

After a thorough and meticulous evaluation process, 270 abstracts were selected, reflecting the outstanding quality and relevance of the research conducted in the water industry and distribution systems. We extend our heartfelt gratitude to the authors whose exceptional work was approved for presentation at this esteemed conference.

In a testament to the depth and breadth of the research shared, 225 final papers were presented during the conference. These works encapsulate groundbreaking research, practical case studies, and novel approaches that promise to shape the future of water industry. As you delve into these pages, you will encounter a wealth of knowledge, ideas, and innovations that inspire new perspectives and avenues for progress.

While all the contributions to this conference are valuable and instrumental in advancing the field, it is important to note that only 145 papers are featured in this conference book. This selection was made to ensure a diverse representation of topics and perspectives, fostering a comprehensive and impactful dialogue among our esteemed participants. Among these carefully chosen papers are those that belong to the esteemed Industry Track, which, while essential, did not require formal paper submissions. Their invaluable insights and experiences further enrich the fabric of this conference, bridging the gap between theory and real-world applications. In addition, several outstanding papers are part of the Track of "The Battle of Intermittent Water Supply," a special collection that in process of been published in a dedicated Special Issue of the prestigious Journal of Water Resources Planning and Management. The inclusion of these papers in our conference underscores the importance of disseminating valuable knowledge and promoting rigorous research within the scientific community and beyond.

In the spirit of recognition and collaboration, we are pleased to announce that 14 papers were selected for publication in a special issue of the Journal of Hydroinformatics. These contributions exemplify excellence in research and exemplify the conference's aim to promote scholarly work and its broader impact on the water industry. The authors of these papers have demonstrated a commitment to advancing the science of Hydroinformatics and are at the forefront of driving positive change in the water sector.

This conference book stands as a testament to the collective pursuit of excellence, embracing innovation, and fostering collaboration to create a sustainable water future. We extend our heartfelt appreciation to all the authors, reviewers, and participants whose unwavering dedication has made this event a resounding success. The depth and breadth of knowledge shared here will undoubtedly contribute to the advancement of water distribution system analysis and the use of computer and control systems in the water industry.

Finally, we would like to thank all those who have made this event possible:

- To the WDSA and the CCWI for giving us the opportunity to organize this event.
- To the UPV that has offered all its infrastructure at our disposal.

3 METHODOLOGY

3.1 Materials

A 3000 mm long, 50 mm nominal diameter BS 416-2 soil pipe (product code MS2001 [24]) manufactured by Hargreaves Foundry (Halifax, UK) was obtained so that the properties of the pipe could be characterised and compared with the literature data for exhumed GCI water pipes. DN50 was selected because these pipes are intended for use in scaled-down tests where a small diameter is beneficial.

The measured outside diameter of the pipe averaged 58.11 mm, and the wall thickness averaged 3.54 mm. The manufacturer confirmed via private communication that the pipe supplied was manufactured using a centrifugal casting process (spin-casting) in a water-cooled metal mould. The pipe was supplied with internal and external coatings of epoxy resin, visible in Figure 3, which were removed by grit-blasting.

3.2 Microscope Inspection

The microscope inspection aimed to characterise the graphite microstructure of the material across the full thickness of the pipe wall. Three small material samples were cut from the BS416-2 pipe and mounted in Bakelite so that the circumferential cross-section was visible, as shown by Figure 3. The mounted samples were then ground using P280, P800, and P1200 grit paper. Following this the samples were polished using 9 μm then 3 μm diamond paste, and lastly 0.05 μm aluminium oxide.

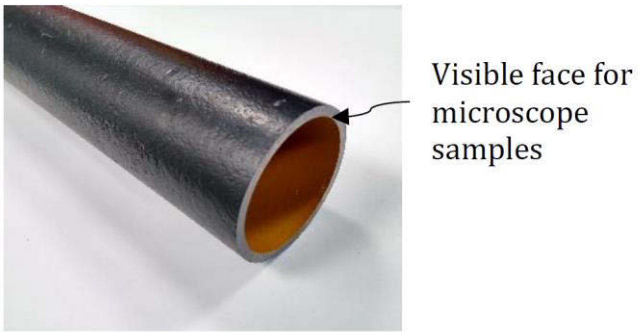


Figure 3: Image of a BS416-2 soil pipe showing the face selected for microscope inspection.

ASTM standard A247-19 [22] was adhered to for evaluation of the graphite microstructure. As a result, microscope images were taken at x100 magnification. To capture the expected variation in flake size across the pipe wall, a series of images were taken, spanning in a straight line from the outside wall to the inside wall. This was repeated for three separate samples, with a total image count of 21 for the first sample and 33 each for the second two samples. The intervals between the images for a given specimen were approximately equal.

Each image was processed using the MatLab Image Processing Toolkit to determine the length and area of the graphite present. A247-19 [22] does not specify the process for measuring flake length, so the largest straight line distance between any pair of points around the flake edge was used. Objects smaller than 1 μm were excluded as it was not possible to confirm whether these were graphite or other features due the resolution of the images. The results from the three samples were combined to provide a more representative view of the pipe's graphite microstructure.

3.3 Tensile Testing

As recommended by ASTM E8/E8M [25] for characterising the UTS of large diameter tubular products, a specimen was cut from the wall of the BS 416-2 pipe with the geometry shown by

Figure 4. Five specimens were cut from the same length of pipe as the microscope samples using wire-EDM cutting.

The specimens were tested to failure using a Shimadzu 300 kN electronic test machine. Grips with curved inserts were made so the specimens could be held securely without crushing the curved end sections. A 50 mm extensometer was fitted to the gauge section of the specimens during testing to record extension. Upon failure, the thickness and width of the fracture surface was measured manually using digital callipers. These measurements were used to calculate the cross-sectional area at the point of failure, which was in turn used to calculate the applied stress.

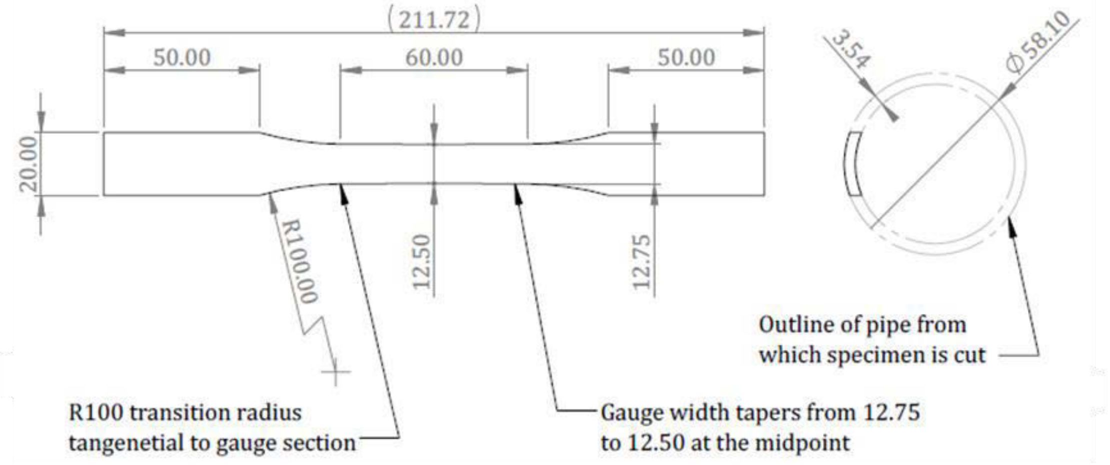

(211.72)
50.00
60.00
50.00
20.00
R100.00
12.50
12.75
3.54
ϕ58.10
Outline of pipe from which specimen is cut
R100 transition radius tangenetial to gauge section
Gauge width tapers from 12.75 to 12.50 at the midpoint

Figure 4: Tensile specimen design used for characterising the UTS of the BS 416-2 DN50 pipe. 3rd angle projection, not to scale.

4 RESULTS AND DISCUSSION

4.1 Graphite Microstructure

The majority of graphite observed in the BS416-2 pipe fell into ASTM size Class 8 (<10 μm), with some larger Class 7, 6, and 5 flakes also present. As shown by Figure 5, the breakdown of graphite size was found to be very similar for the two quarters of the wall closest to the outside surface, with size Class 8 graphite making up around 85% of the total graphite area, and slightly more than 1% of the area being Class 6. The largest graphite was found to occur closer to the inside wall, with about 3.5% of the graphite area in the wall quarter closest to the inside edge being Class 5. The largest individual piece of graphite observed in this region was 76.6 μm long. These results are in agreement with the literature values for spun cast pipes reported above where size Class 8 to 4 were observed, with the larger graphite tending to occur towards the pipe inside wall.

The graphite observed in the BS416-2 pipe was mainly Type VII flake graphite (see Figure 6), with a small amount of Type III compacted graphite present near the inside wall of the pipe. Where Type III graphite occurred it was mainly found near regions of iron devoid of graphite, such as the example shown in Figure 6d. Note that in the top region of Figure 6d the faint white lines may indicate the presence of pearlite. The dominance of Type VII flake graphite is expected for a spun-cast GCI water pipe, based on the literature review findings, however the occurrence of Type III compacted graphite near the pipe inside wall is more unusual and is perhaps linked to the presence of the suspected pearlite grains.

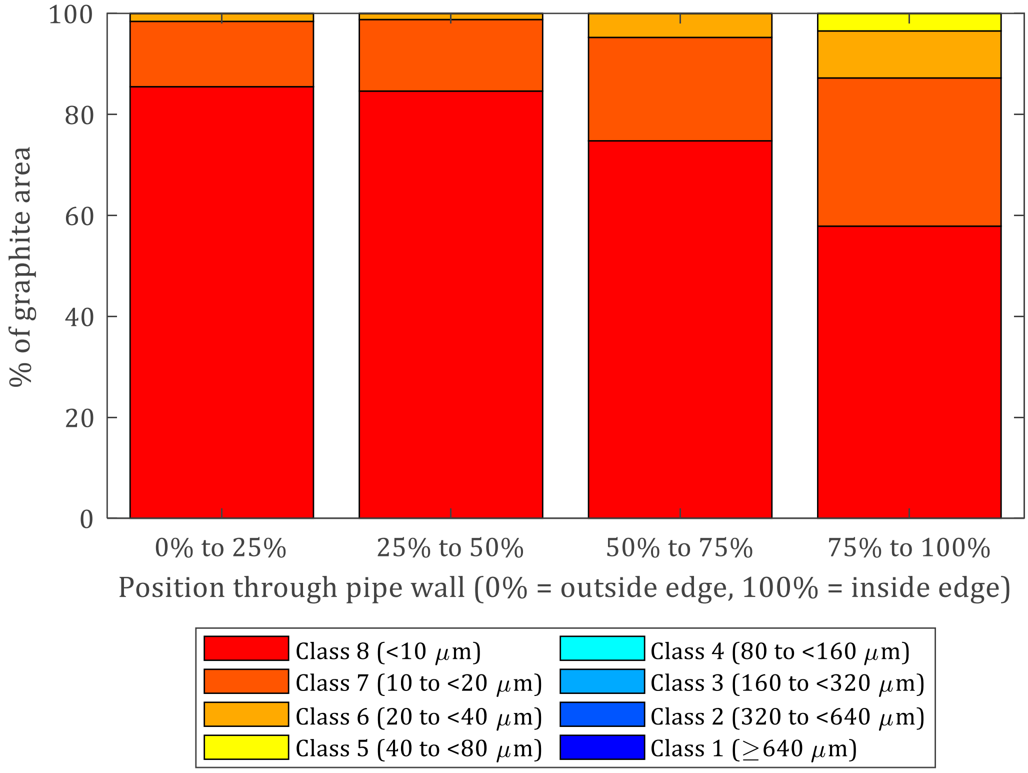


Figure 5: Proportion of graphite, by area, in each ASTM size class across the wall of the BS 416-2 DN50 pipe.

Nearly all observations made in the mid-region of the pipe wall showed Distribution D graphite (see Figure 6b). Around 10% of observations near the outside wall and more than 60% of the observations near the inside wall were Distribution A or C (Figure 6a and 6c). These observed distributions match those expected for a spun-cast GCI, based on the literature review above. This can be seen visually by comparing Figure 1b with Figure 6b, although the difference in scale must be noted.

In summary, the graphite sizes, types, and distributions observed in the BS416-2 pipe match the literature data for spun-cast water pipes well. The outer and middle regions of the pipe wall are dominated by fine (Class 8 and Class 7), Distribution D, Type VII graphite flakes (Figure 6b) which are typical of spun-cast pipes examined by previous authors [14], [17]. The region of cast iron closest to the inside wall of the pipe varied from this norm, with ~500 μm diameter patches of Distribution A or C, Type VII flakes with sizes up to Class 5 (Figure 6c). These areas with large flake sizes are likely due to the longer cooling time experienced by the iron closest to the inside wall, giving more time for larger flakes to form, and are consistent with observations in the literature [14], [17]. Also found close to the inside wall were possible areas of pearlite, surrounded by Type III compacted graphite (Figure 6d). Pearlite grains are associated with cast iron that has cooled more slowly [14] and Type III graphite is reportedly common in annealed malleable iron castings [22]. Therefore, the occurrence of the regions containing pearlite and Type III graphite is also probably associated with the slower cooling rate of the pipe inside wall.

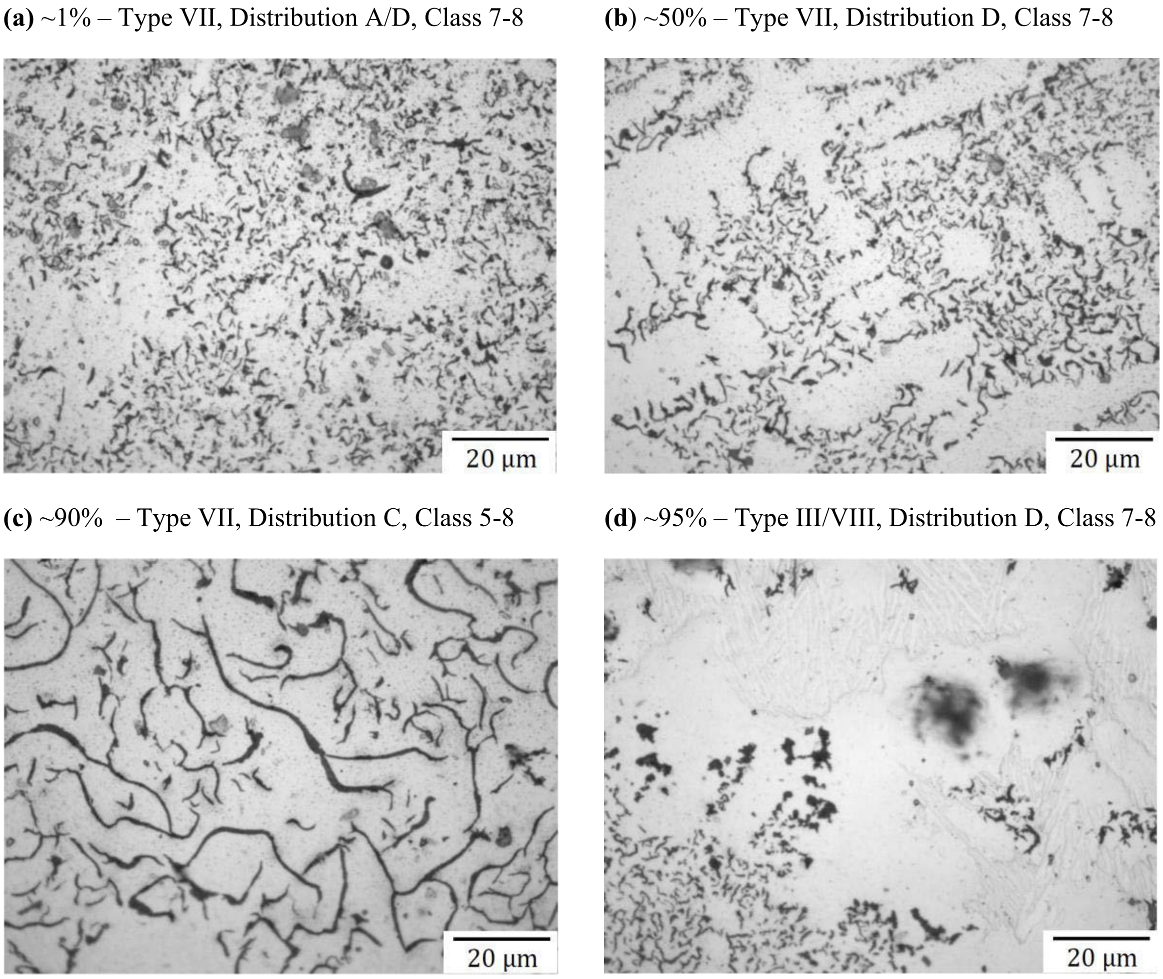

(a) ~1% – Type VII, Distribution A/D, Class 7-8
(b) ~50% – Type VII, Distribution D, Class 7-8
(c) ~90% – Type VII, Distribution C, Class 5-8
(d) ~95% – Type III/VIII, Distribution D, Class 7-8
20 µm
20 µm
20 µm
20 µm

Figure 6: Selection of x100 magnification images showing the graphite microstructure of the BS416-2 pipe. ~X% refers to the approximate position of the image, where ~0% would be the outside wall and ~100% would be the inside wall.

4.2 Stress-Strain Behaviour

The stress-strain results of the five tensile specimens tested are shown in Figure 7. From these stress-strain curves the initial elastic modulus, failure stress and failure strain were determined; these values are provided in Table 2. The 0.2% offset yield stress could not be calculated for any of the BS416-2 specimens as failure occurred before the offset slope was met.

The average initial elastic modulus of the five specimens was found to be 126 GPa, with a range of +13% -9%. This spread is quite large considering that the five specimens were cut from the same section of the same pipe, however, this is expected for GCI. Repeated tensile tests of the same type of specimen cut from the same spun-cast pipes by Makar and McDonald [16] returned elastic moduli ranging between 114 GPa and 184 GPa for one pipe (approximately ±23%), and between 78 GPa and 167 GPa for the other (approximately ±36%). The cause of this variation in elastic modulus within the same pipe is likely due to local differences in microstructure, as observed during the microscope investigation. The BS416-2 pipe elastic modulus measurements all fall well within the range of values reported in the literature for spun-cast water pipes.

The average tensile strength of the five BS416-2 specimens was 292 MPa, with a range of +4.8% -4.5%. This spread is also considerably less than the range of +20% -22% observed by Makar and McDonald [16] for seven identical specimens cut from one ex-service spun cast pipe with an average tensile strength of 213 MPa (Figure 7, 'SC2 9mm thick specimen failure points'). The tensile strengths measured for the BS416-2 pipe are close to the highest tensile strengths reported for spun-cast pipes in literature, falling within or just above the 95th percentile (see Figure 2). Thinner GCI castings are known to demonstrate higher tensile strengths, even when the same molten iron mixture is used [26]. The BS416-2 pipes have a wall thickness of around 3.54 mm whereas the pipes tested in the literature generally had wall thicknesses between 8 mm and 12.7 mm. Therefore, the thinner walls of the BS416-2 pipe is likely the cause of the relatively high tensile strength measured.

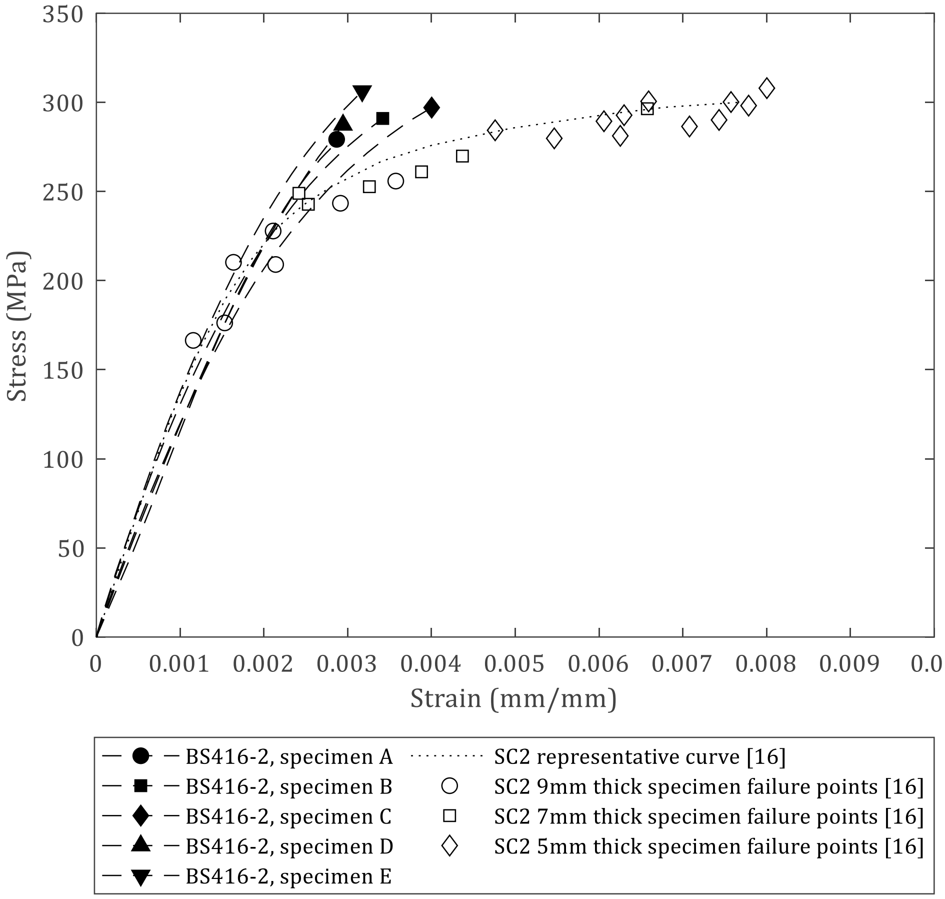


Figure 7: Tensile stress-strain curves for the five BS416-2 pipe tensile specimens tested as part of this work. The representative stress-strain curve and individual test failure points for spun-cast pipe "SC2" from [16] are provided for comparison.

Table 2: BS416-2 tensile stress-strain parameters for each of the five specimens tested as part of this work.

Specimen	Initial E (GPa)	UTS (MPa)	Failure Strain (mm/mm)
A	115	279	0.0029
B	134	291	0.0034
C	121	297	0.0040
D	120	287	0.0029
E	142	306	0.0032

The average failure strain of the BS416-2 pipe specimens was 0.0033, with a range of +21% -12%. As with the other results, this spread is much smaller than that observed by Makar and McDonald [16] for similar tests. The magnitude of the failure strain is lower than the failure strains reported in the literature for spun-cast pipes, the smallest literature value being around 0.005 (see Figure 2). However, these literature values are usually from the strongest specimens tested from each pipe which the references highlight as representative [15], [16]. Fortunately, Makar and McDonald [16] provide the data from repeats for one of the spun-cast pipes they investigated, "SC2", and these values are plotted on Figure 7. These repeats show a huge variation in failure strain with some specimens displaying ductile behaviour, while others are very brittle. Generally, Makar and McDonald [16] found that specimens which included nearly the full wall thickness (9mm thick) were more brittle, and failed at lower strains, whereas specimens that had the inside wall material removed (5mm thick) were more ductile and demonstrated higher failure strains. The BS416-2 tensile specimens tested in the present work included the full wall thickness, so their brittle behaviour matches Makar and McDonald's [16] observations for the 9mm thick specimens, which also included most of the wall thickness. A possible cause of Makar and McDonald's thinner specimens demonstrating more ductile behaviour than the full wall specimens is that the material removed during thinning contained the largest graphite flakes; in other words, thinning the specimens probably removed the largest internal defects.

In summary, for applied stresses below 200 MPa the BS416-2 pipe's tensile stress-strain matches literature data for spun-cast pipes well. The failure behaviour is much more brittle than the "representative" stress-strain curves for spun-cast pipes provided in the literature which show more ductile behaviour [15]–[17]. However, closer examination of these literature results shows that more brittle behaviour can occur when the whole pipe wall is sampled, as is the case for the tests conducted for the present work. The magnitude of the pipe's tensile strength is high compared to the literature data, but this is likely to be a result of the BS416-2 pipe's thinner walls. It is also clear that further work is required to fully characterise and understand the complex and highly variable material that is spun-cast grey iron.

5 CONCLUSIONS

The graphite microstructure and tensile stress-strain behaviour of an off-the-shelf BS416-2 soil pipe were characterised and compared with literature data for spun-cast GCI water pipes. The motivation behind this work was that a large number of small-diameter, un-corroded GCI pipes are needed for a planned fatigue experiment. Through the comparisons performed it was concluded that BS 416-2 soil pipes can be used to represent spun-cast GCI water pipes in small-scale destructive experiments. Future work will aim to characterise the compressive stress-strain and fatigue behaviour of these BS416-2 pipes.

6 ACKNOWLEDGEMENTS

This work was funded by UK Water Industry Research (UKWIR) and the ESPRC through the Water Infrastructure and Resilience (WIRe) Centre for Doctoral Training. The authors would like to thank Dennis Dellow from UKWIR for providing an experienced industry perspective on the project. Also a special thanks to the technicians Paul Blackbourne, Sam Gibson, Geoffrey Hibberd, and Tesoro Monaghan for their work which enabled the microscopy and tensile testing to take place. For the purpose of open access, the author has applied a creative commons attribution (CC BY) license to any author accepted manuscript versions arising.

7 REFERENCES

[1] N. A. Barton, T. S. Farewell, S. H. Hallett, and T. F. Acland, “Improving pipe failure predictions: Factors effecting pipe failure in drinking water networks,” Water Research, vol. 164, p. 114926, 2019, doi: 10.1016/j.watres.2019.114926.

[2] S. Rathnayaka, B. Shannon, C. Zhang, and J. Kodikara, “Introduction of the leak-before-break (LBB) concept for cast iron water pipes on the basis of laboratory experiments,” Urban Water Journal, 2017, doi: 10.1080/1573062X.2016.1274768.

[3] M. Larson and L. Jönsson, “Elastic Properties of Pipe Materials during Hydraulic Transients,” Journal of Hydraulic Engineering, vol. 117, no. 10, pp. 1317–1331, 1991, doi: 10.1061/(asce)0733-9429(1991)117:10(1317).

[4] D. Chan, C. P. K. Gallage, P. Rajeev, and J. Kodikara, “Field performance of in-service cast iron water reticulation pipe buried in reactive clay,” Canadian Geotechnical Journal, 2015, doi: 10.1139/cgj-2014-0531.

[5] C. Randeniya, D. J. Robert, C. Q. Li, and J. Kodikara, “Large-scale experimental evaluation of soil saturation effect on behaviour of buried pipes under operational loads,” Canadian Geotechnical Journal, vol. 57, no. 2, pp. 205–220, 2020, doi: 10.1139/cgj-2018-0544.

[6] H. Rezaei, B. Ryan, and I. Stoianov, “Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks,” Procedia Engineering, vol. 119, no. 1, pp. 253–262, 2015, doi: 10.1016/j.proeng.2015.08.883.

[7] Department for Transport, “AADF Data - major and minor roads,” Gov.uk, 2020. https://roadtraffic.dft.gov.uk/downloads (accessed Apr. 19, 2021).

[8] W. Brevis, L. Susmel, and J. Boxall, “Investigating in-service failures of water pipes from a multiaxial notch fatigue point of view: A conceptual study,” Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 229, no. 7, 2015, doi: 10.1177/0954406214553020.

[9] R. Jiang, S. Rathnayaka, B. Shannon, X.-L. Zhao, J. Ji, and J. Kodikara, “Analysis of failure initiation in corroded cast iron pipes under cyclic loading due to formation of through-wall cracks,” Engineering Failure Analysis, vol. 103, pp. 238–248, 2019, doi: 10.1016/j.engfailanal.2019.04.031.

[10] D. A. Jesson et al., “Thermally induced strains and stresses in cast iron water distribution pipes: An experimental investigation,” Journal of Water Supply: Research and Technology - AQUA, 2010, doi: 10.2166/aqua.2010.078.

[11] H. Mohebbi, D. A. Jesson, M. J. Mulheron, and P. A. Smith, “The fracture and fatigue properties of cast irons used for trunk mains in the water industry,” Materials Science and Engineering A, 2010, doi: 10.1016/j.msea.2010.05.071.

[12] British Standards Institution, “Discharge and ventilating pipes and fittings , sand-cast or spun in cast iron - Part 2: Specifications for socketless systems. BS 416-2:1990.” British Standards Publications, 1990. [Online]. Available: https://bsol-bsigroup-com.sheffield.idm.oclc.org/Bibliographic/BibliographicInfoData/000000000000225830

[13] P. A. Blackmore and K. Morton, “Structure-property relationships in graphitic cast irons,” International Journal of Fatigue, vol. 4, no. 3, pp. 149–155, 1982, doi: 10.1016/0142-1123(82)90042-1.

[14] J. M. Makar and B. Rajani, “Gray Cast-Iron Water Pipe Metallurgy,” Journal of Materials in Civil Engineering, 2000, doi: 10.1061/(asce)0899-1561(2000)12:3(245).

[15] H. M. S. Belmonte, M. Mulheron, and P. A. Smith, “Weibull analysis, extrapolations and implications for condition assessment of cast iron water mains,” Fatigue and Fracture of Engineering Materials and Structures, 2007, doi: 10.1111/j.1460-2695.2007.01167.x.

[16] J. M. Makar and S. E. McDonald, “Mechanical Behavior of Spun-Cast Gray Iron Pipe,” Journal of Materials in Civil Engineering, vol. 19, no. 10, pp. 826–833, 2007, doi: 10.1061/(ASCE)0899-1561(2007)19.

[17] M. v. Seica and J. A. Packer, “Mechanical Properties and Strength of Aged Cast Iron Water Pipes,” Journal of Materials in Civil Engineering, 2004, doi: 10.1061/(asce)0899-1561(2004)16:1(69).

[18] R. Logan et al., “Graphitic corrosion of a cast iron trunk main: Implications for asset management,” 2014. doi: 10.2495/UW140351.

[19] UKWIR, “Water Industry Information & Guidance Note: Ductile Iron Pipes and Fittings,” 2006. [Online]. Available: https://www.water.org.uk/wp-content/uploads/2018/11/ign-4-21-01.pdf

[20] British Standards Institution, “Specification for cast iron pipes and special castings for water, gas and sewage. BS 78:1917.” British Standards Publications, London, 1917. [Online]. Available: https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030338181

[21] British Standards Institution, “Centrifugally cast (spun) iron pressure pipes for water, gas & sewage. BS 1211:1958.” British Standards Publications, London, 1958. [Online]. Available: https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000000034043

[22] ASTM International, “Standard Test Method for Evaluating the Microstructure of Graphite in Iron Castings A247-19.” ASTM Compass, West Conshohocken, PA, USA, 2019. doi: 10.1520/A0247-19.

[23] B. Rajani and J. Makar, “A methodology to estimate remaining service life of grey cast iron water mains,” Canadian Journal of Civil Engineering, 2000, doi: 10.1139/l00-073.

[24] Hargreaves Foundry, “Mech 416 Cast Iron Mechanically Jointed Soil System,” www.hargreavesfoundry.co.uk, 2021. https://www.hargreavesfoundry.co.uk/cast_iron_drainage/mech416/ (accessed Sep. 06, 2021).

[25] ASTM International, “Standard test methods for tension testing of metallic materials ASTM E8/E8M - 16a.” ASTM Compass, West Conshohocken, PA, USA, 2016. doi: 10.1520/E0008.

[26] British Standards Institution, “Founding - Grey cast irons. BS EN 1561:2011.” British Standards Publications, 2011. [Online]. Available: https://bsol-bsigroup-com.sheffield.idm.oclc.org/Bibliographic/BibliographicInfoData/000000000030214938

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14007

WATER QUALITY IN DRINKING WATER DISTRIBUTION SYSTEMS: A WHOLE-SYSTEM APPROACH TO DECISION MAKING

S.L. Weston[1], A. Scheli[2], S. Behmel[3], and M. Rodriguez[4]

[1,4] ÉSAD, Université Laval, Quebec City, Canada
[2,3] WaterShed Monitoring, Quebec City, Canada

[1] 0000-0003-1443-2253, slwes@ulaval.ca, [4] 0000-0003-2010-6438

Abstract

Water utilities are responsible for continuously providing safe water to consumers through their drinking water distribution systems. Avoiding siloed approaches that may lengthen critical response times in the case of a water quality hazard can be accomplished through a whole-system approach. To achieve this objective, providers need to pursue the state-of-the-art knowledge and techniques across all facets of water quality management as our understanding of these complex infrastructures continually evolves.

This study provides a comprehensive and up-to-date bibliometric study of water quality in drinking water distribution systems over the first twenty years of the 21st century. Analysis of the relevant literature reveals how the research landscape has expanded in terms of number of publications made, variety of topics, and geographic diversity. Each region has a unique 'research identity' in the different topics focused upon, yet the presented inter-dependency of factors impacting water quality emphasises the opportunities for sharing of best practices.

Keywords
Water quality, distribution systems, decision support, water utilities, bibliometric study, drinking water management

1 INTRODUCTION

Drinking Water Distribution Systems (DWDS) can interact with the water they transmit, affecting the quality of water received by consumers. Some of the spatial-temporal variability measured in typical water quality parameters can be predicted and explained by known factors, such as those related to increased temperatures [1,2], planned operations and maintenance [3,4], and distribution materials and configuration [5,6]. However, random unforeseen events can drive potentially rapid deterioration of water quality, including microbiological and chemical concerns. Utilities have an obligation to intervene and manage such risks, but siloed and reactive approaches can lengthen response times. This could lead to events that may expose the population to regulated and non-regulated contaminants in drinking water, and ultimately public dissatisfaction and increased repercussive costs.

Managing the vast infrastructures of DWDS is a complex challenge, and providers continually pursue the state-of-the-art knowledge and techniques to maintain and elevate water quality beyond regulation compliance. Extensive research and the accrued experience of system operators built over many decades tackle how different facets of water quality behave individually and interact together within the system [7]. Diverse approaches have contributed new pieces of knowledge and practice to the wide field of DWDS water quality; however, our understanding is not complete and necessitates constant evolution. To effectively manage water quality events now and into the future, drinking water managers require a whole-system approach to water quality that can instigate the adoption of preventative and curative actions.

BY NC SA

Bibliometric studies are the quantitative analysis of literature within a subject to understand the major concepts and temporal and geographical patterns within sub-disciplines. From these studies, researchers and practioneers can identify the present status, topic hotspots and gaps, and form a consensus across the relevant literature, which may change over time. Few papers have previously performed bibliometric analysis within the field of water quality in DWDS; yet they have focused exclusively on a particular topic, such as disinfection by-products [8] and disinfectant residual stability [9], or location, such as drinking water research in Africa [10]. A bibliometric study conducted from 1992 to 2011 by Fu, Wang and Ho [11] outlined a more holistic overview across multiple topics relevant to water quality in DWDS. Yet, the rapidly changing pace and diversity of drinking water research means that this study needs to be updated as it is now out of date and potentially limited when extrapolating to future challenges.

This study aims to gain an integrated understanding of the events and factors that impact water quality in DWDS. A comprehensive and up to date bibliometric study at the macro-level is presented, detailing the global research landscape and how it has evolved over the first twenty years of the 21st century. We will highlight the main trends in water quality research and explore the most consistently prioritised areas of research.

2 APPROACH

2.1 Sourcing Literature

Potential literature was sourced from the well-known and established SCOPUS database; chosen as SCOPUS has a wider journal base than its other citation databases [12]. Records published in English between 2000 and 2020 were searched using the keywords "water quality distribution system". The vast range of topics relating to this field and the variety of terminology deployed (e.g., "water distribution network", "drinking water distribution system", "piped water network"), meant that a narrower scope could not be used. Duplicates, retracted articles, and anonymous records were dismissed through a preliminary screening. Therefore, a total of 13,019 records were exported.

Three eligibility criteria were applied to filter records relevant to the aim of the study. Firstly, it was required that the literature primarily focus on knowledge and/or management of water quality in DWDS between the treatment works and the customer boundary. Secondly, only continuously pressurised systems with surface or groundwater sources were considered. Studies relating to intermittent water supply, reclaimed water, or ultrapure water were not included in this review. Thirdly, epidemiological studies were deemed to be outside the scope of this study.

Potential records were carefully manually evaluated against the eligibility criteria using the title, keywords and abstract. Where there was ambiguity, the full publication was retrieved to enable further examination. Any records that did not meet the requirements were removed, leaving 1868 eligible literature sources. Of these, 91% were available to be downloaded as full publications. Identifying information for the 1868 individual sources were compiled into an Excel spreadsheet for bibliometric analysis, such as title, author names, year of publication, document type (e.g., journal paper, conference paper, report, book chapter), source, and digital object identifier (DOI).

2.2 Classifying Literature

To determine the research landscape, each publication was classified into a theme that described the study's main area and 'tagged' with up to three different topics within that overarching theme. As an example, a paper titled "Do Transients Contribute to Turbidity Failures of Water Distribution Systems" [13] was classified within the *Hydraulic Behaviour* theme and tagged with the *Dynamic Hydraulics*, *Particle Mobilisation*, and *Discolouration* topics based on the content of

BY NC SA

the paper. Nine themes and 47 topics were conceptualised. These themes and example topics are: 1) infrastructure – corrosion, pipe material, leaching; 2) chemicals & treatment decay – chlorine, disinfection by-products, metals; 3) contamination – intrusion, backflow; 4) hydraulic behaviour – stagnation, dynamic hydraulics, pressure; 5) microbiology – biofilm, coliforms, fungi; 6) water aesthetic – discolouration, particles, taste and odour; 7) utility management – flushing, utility management, regulation compliance; 8) technology – sensors, monitoring, simulations; and 9) external factors – temperature, treatment implications, social aspects.

3 RESULTS

3.1 Temporal Changes in Publications

A database of 1868 publications relating to water quality in DWDS were found to be published between 2000 and 2020. Two-thirds of the literature were journal papers (1231), 31% were conference papers (579), and the remaining sources (58) consisted of books, magazine articles, and technical reports. Figure 1 shows how the publication rate varied considerably from average of 89 publications per year. The lowest number of publications (40) occurred in 2002, where the highest number of publications (129) occurred in 2018. Overall, there was a significant positive increase in relevant publications over the two decades (Pearson coefficient 0.75).

In the majority of the years investigated, more journal papers relevant to this study were found than conference papers. The only exception was in 2006 when 66 conference papers were determined to be eligible, compared to 41 journal papers. This finding aligns with a burst of conference papers published between 2005 and 2010, most likely due to an alignment of several major conference series: World Environmental and Water Congress, Water Distribution Systems Analysis Conference, Computing and Control in the Water Industry Conference, Water Quality and Technology Conference and AWWA Annual Conference and Exposition.

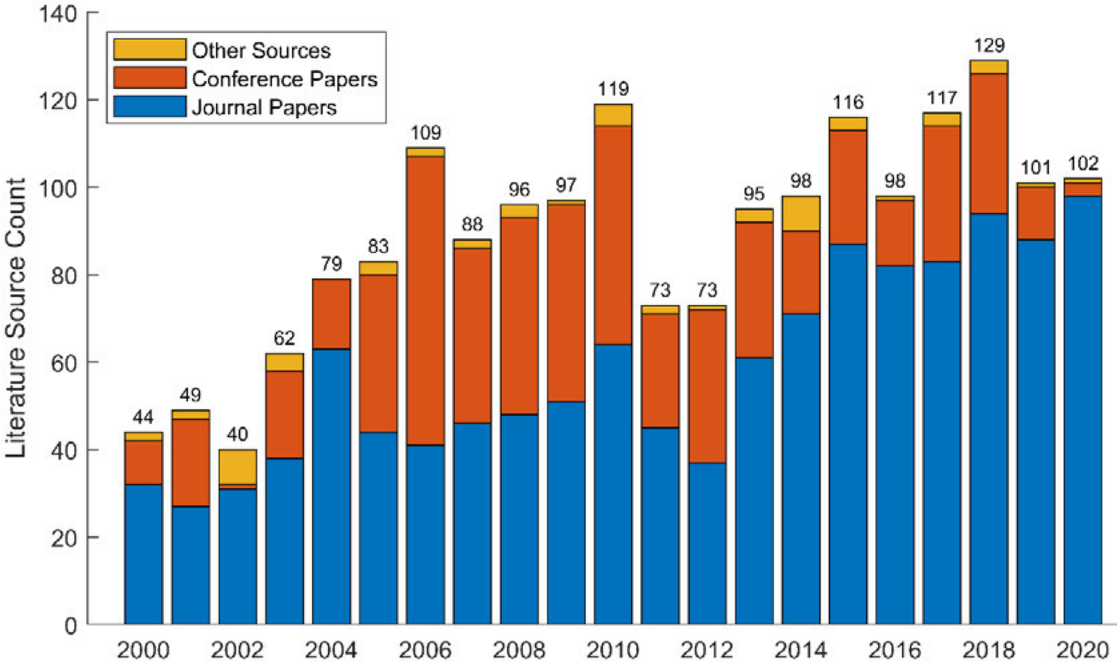


Figure 1. Eligible literature published per year between 2000 and 2020.

BY NC SA

3.2 Geographical Changes in Publications

Of the 1868 publications investigated in this study, three quarters (1391) referenced a study location at the community, region or national scale. It is important to note this is not where the authors' associated institution is based, but rather where the research was performed or situated. A total of 58 countries were represented across North American, South America, Europe, Africa, Middle East, Asia, and Oceania. Figure 2 presents the number of studies published per year for each of these regions.

Half of the total literature (970) was based in North America and Europe, predominantly the United States of America, Canada, the UK, the Netherlands, and France. The research dominance of these countries in terms of productivity is well recognised across most scientific fields [14]. Studies centring in North America peaked in 2005-2006. It appears that this research peak corresponds to a surge of literature focusing on contamination and monitoring technologies that is most likely in reaction to the "9/11" terrorist attack in New York City. This catastrophic event brought to the research forefront DWDS' vulnerability against deliberate threats. Over the two decades, literature based in Europe was published at an increasing rate (Pearson coefficient 0.84) with a growing number of studies in the later years based in Central and Eastern European countries, such as Romania and Poland.

All seven regions were mentioned in the literature for the first time in 2009 and for seven of the following years, which suggests a turning point in the diversity of water quality research in DWDS. Studies based in Africa, the Middle East, and South America have grown in number, but these regions still remain a small share of the total field (109, 5.8%). In contrast, literature representing Asian DWDS has rapidly increased (Pearson coefficient 0.91). Accelerated rates of publishing from China, India, and South Korea in particular mean that, in the last five years investigated (2016-2020), there are more studies investigating water quality in DWDS in Asia than in North America.

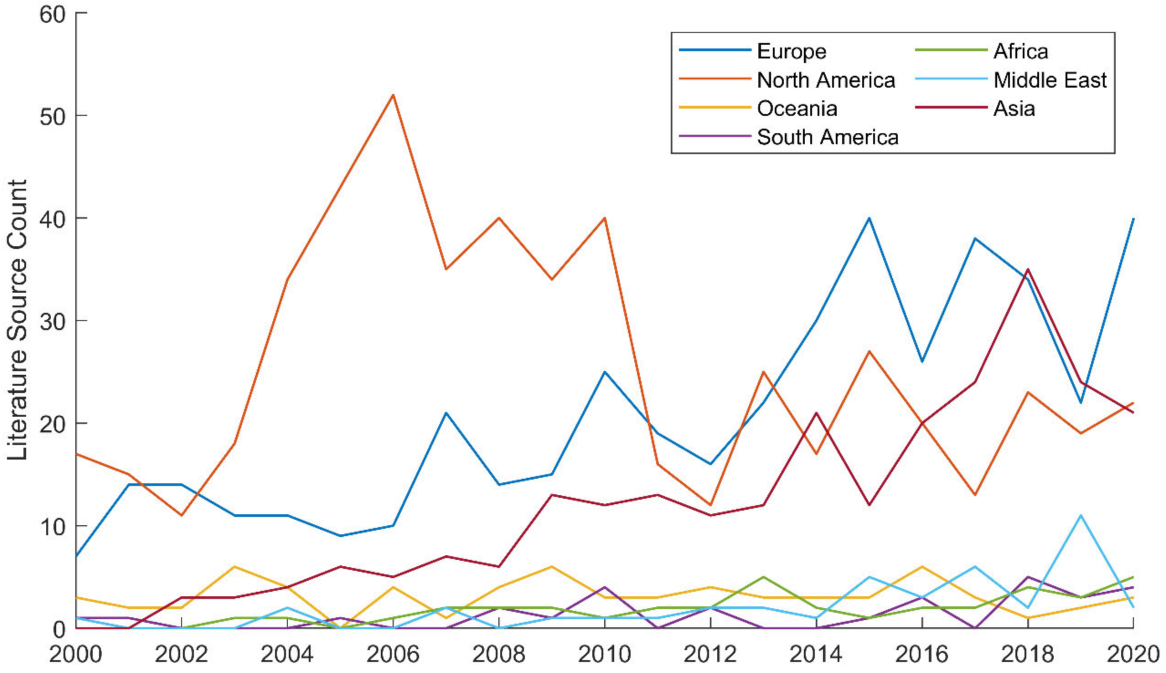


Figure 2. Study locations represented by the literature categorised into regions.

cc BY NC SA

3.3 Research Themes

The most popular themes and the corresponding number of literature sources were technology (528), chemicals and treatment decay (362), and microbiology (337). Given the exponential advances made since 2000, it was unsurprising that technology was the most common theme. Significantly greater data storage and processing power mean complex simulations have become normalised, aided by cheaper and more accessible devices to measure water quality. The chemicals and treatment decay theme (referring to research involving chemical reactions or decay of chemicals added during water treatment) has experienced the greatest expansion in terms of publications per year (Pearson coefficient 0.76). Less prevalent themes of infrastructure, contamination, hydraulic behaviour, water aesthetic, utility management, and external factors remained stable over the two decades investigated. Overall, these less prevalent themes averaged five publications per year, where technology, chemicals and treatment decay, and microbiology averaged 25, 17, and 16 publications per year, respectively.

Each geographical region has a unique profile of themes studied relative to the total number of publications associated in those countries, which is represented in Figure 3. These 'research identities' exist as a product of multiple factors, such as national/regional priorities, personal interests of researchers, funding opportunities, and environmental contexts.

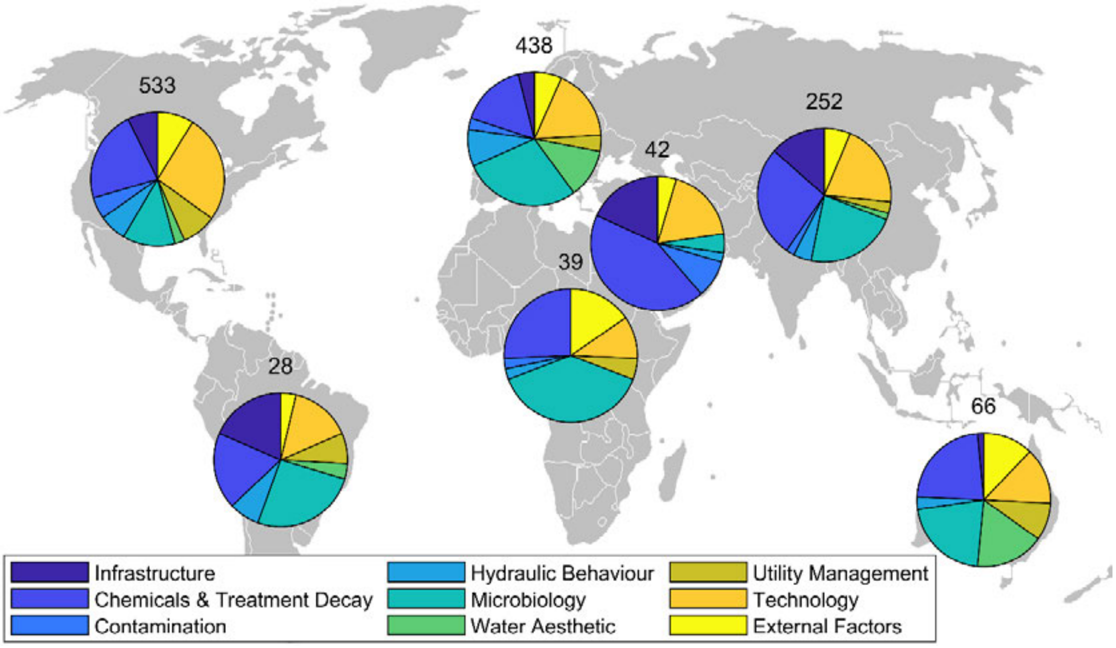


Figure 3. Geographical representation of the different themes. The total number of publications over the study period is presented for each region.

3.4 Research Topics

Each literature source was 'tagged' with up to three different topics that aimed to capture the work's focus. Three-quarters of publications (1412) were assigned three topics, 425 publications were assigned two topics, and 31 publications were assigned the minimum of a single topic. The 'top five' topics and their respective frequencies of use were simulations (647), chlorine (369), biofilm (368), intrusion (327), and monitoring (310). In contrast, the 'bottom five' topics were backflow (4), pipe maintenance (4), pumps (12), parasites (12), and fungi (13).

CC BY NC SA

To determine if topics occurred in frequent combination, the topics were coded into pairs, e.g., transient-intrusion, chlorine-water age, metals-discolouration. When the order of the topics assigned to the publication was ignored, a total of 617 unique pairs were found, out of a possible 1081 pairs using 47 different topics. The most popular topic pairs and their respective frequencies were simulations-chlorine (177), simulations-intrusion (169), monitoring-intrusion (144), simulations-monitoring (131), and simulations-velocity (86).

Figure 4 visually depicts the pairings found in this study where line thickness indicates how frequently the respective pair was found in the literature. Only pairs that occurred at least five times are included, and the colours are simply for visual clarity. This figure visually accentuates the breath and interconnectivity of topics relating to water quality in DWDS.

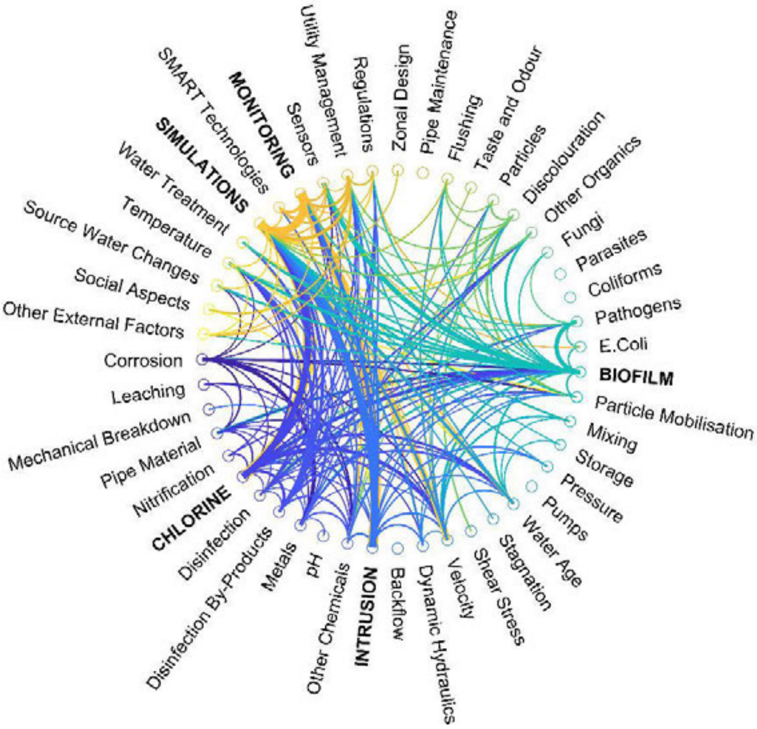

Utility Management
Regulations
Zonal Design
Pipe Maintenance
Flushing
Taste and Odour
Particles
Discolouration
Other Organics
Fungi
Parasites
Coliforms
Pathogens
E.Coli
BIOFILM
Particle Mobilisation
Mixing
Storage
Pressure
Pumps
Water Age
Stagnation
Shear Stress
Velocity
Dynamic Hydraulics
Backflow
INTRUSION
Other Chemicals
pH
Metals
Disinfection By-Products
Disinfection
CHLORINE
Nitrification
Pipe Material
Mechanical Breakdown
Leaching
Corrosion
Other External Factors
Social Aspects
Source Water Changes
Temperature
Water Treatment
SIMULATIONS
SMART Technologies
MONITORING
Sensors

Figure 4. Illustration of the topic pairs observed in the literature. The five most prevalent topics are highlighted in bold font.

4 CONCLUDING REMARKS

This study presents an up-to-date overview of research published between 2000 and 2020 that examines water quality in drinking water distribution systems. The findings show how this field has expanded over the two decades in terms of number of publications made, variety of topics, and geographic diversity that represents the increased research capacity of countries in the Global South. Relevant research is now being performed around the globe that reflects the universal need for safe drinking water. Topic analysis has indicated that intrusion, chlorine, biofilm and monitoring are currently prevalent disciplines of research, enabled by increasingly advanced simulations and technologies. This study has exhibited the inter-dependency of factors impacting water quality, yet also highlighted how different focuses are unique to each region. It is important to acknowledge these changing priorities and perspectives, and the opportunities they provide for cross-border thinking and sharing of best practice.

This work forms the first stage of a larger project to develop an integrated decision support system that consolidates scientific knowledge on the spatial and temporal variability of water quality with recommendations from government entities and best practices for municipal management. The literature database developed will serve as an expert knowledge base for the decision support system, which will be incorporated into WaterShed Monitoring's data storage and management software Enki©.

5 REFERENCES

[1] C. Agudelo-Vera *et al.*, "Drinking Water Temperature Around the Globe: Understanding, Policies, Challenges and Opportunities," *Water (Switzerland)*, vol. 12, no. 4, 2020.

[2] J. Novakova and J. Rucka, "Undesirable Consequences of Increased Water Temperature in Drinking Water Distribution System," *MM Sci. J.*, vol. 2019, no. December, pp. 3695–3701, 2019.

[3] D. Vitanage, C. Copelin, N. Karunatilake, and T. Vourtsanis, "Assessment of the Impact of Mains Cleaning to Improve Distribution Water Quality," *Water Sci. Technol. Water Supply*, vol. 3, no. 1–2, pp. 71–78, 2003.

[4] H. Mala-Jetmarova, N. Sultanova, and D. A. Savic, "Lost in Optimisation of Water Distribution Systems? A Literature Review of System Operation," *Environ. Model. Softw.*, vol. 93, pp. 209–254, 2017.

[5] M. Li, Y. Wang, Z. Liu, Y. Sha, G. V. Korshin, and Y. Chen, "Metal-release Potential from Iron Corrosion Scales Under Stagnant and Active Flow, and Varying Water Quality Conditions," *Water Res.*, vol. 175, p. 115675, 2020.

[6] M. F. K. Pasha, M. K. Shiva, and J. W. Rentch, "Effect of Water Distribution System Configuration on Nodal Pressure, Water Quality, and Pumping Energy Uncertainties," in *World Environmental and Water Resources Congress*, 2016.

[7] S. Curnin and B. Brooks, "Making Waves: How Do We Prepare for the Next Drinking Water Disaster?," *Water Res.*, vol. 185, p. 116277, 2020.

[8] Y. Tang *et al.*, "Bibliometric Review of Research Trends on Disinfection By-products in Drinking Water During 1975–2018," *Sep. Purif. Technol.*, vol. 241, no. April 2019, p. 116741, 2020.

[9] R. A. Li, J. A. McDonald, A. Sathasivan, and S. J. Khan, "Disinfectant Residual Stability Leading to Disinfectant Decay and By-product Formation in Drinking Water Distribution Systems: A Systematic Review," *Water Res.*, vol. 153, pp. 335–348, 2019.

[10] E. W. Wambu and Y. S. Ho, "A Bibliometric Analysis of Drinking Water Research in Africa," *Water SA*, vol. 42, no. 4, pp. 612–620, 2016.

[11] H. Z. Fu, M. H. Wang, and Y. S. Ho, "Mapping of Drinking Water Research: A Bibliometric Analysis of Research Output During 1992-2011," *Sci. Total Environ.*, vol. 443, pp. 757–765, 2013.

[12] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas, "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses," *FASEB J.*, vol. 22, no. 2, pp. 338–342, 2008.

[13] S. L. Jones, R. P. Collins, and J. B. Boxall, "Do Transients Contribute to Turbidity Failures of Water Distribution Systems?," in *Pressure Surges*, 2015, no. November, pp. 297–308.

[14] Nature, "The Ten Leading Countries in Natural-Sciences Research," *Nature*, 29-Apr-2020.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14013

A MODEL OF INTERMITTENT WATER SUPPLY SIMULATING THE INEQUITABLE DISTRIBUTION OF WATER

Matthew H MacRorie[1], Vanessa L Speight[2], Sally L Weston[3], Robin Price[4] and Richard P Collins[5]

[1,2,5]Department of Civil and Structural Engineering, University of Sheffield
Sheffield, Yorkshire, UK
[3]ÉSAD, Laval Université, Québec, Canada
[4]Anglian Water, Lancaster House, Huntingdon, Cambridgeshire, UK

[1] *mhmacrorie1@sheffield.ac.uk, [2] v.speight@sheffield.ac.uk, [3] sally.weston@sheffield.ac.uk,*
[4] robinprice@wre.org.uk, [5] r.p.collins@sheffield.ac.uk.

Abstract

Over one billion people worldwide with access to piped water experience Intermittent Water Supply (IWS), where consumers receive water for only a fraction of the day or week. A widely observed problem associated with IWS is the inequitable distribution of water across the network. This results in different consumers in the network receiving different volumes of water. Modelling the inequity within IWS systems remains an open research field. To date, simulations have often adapted hydraulic modelling software to understand the distribution of water with little attention to the consumer interaction with the network. This paper proposes a conceptual model based on a more holistic understanding of water distribution in IWS systems. Understanding the spatial and temporal variation in received volumes, as well as the variation in consumer access within the network, enables a more representative simulation of the inequitable quantity of water received by consumers.

Keywords
Intermittent water supply, Inequity, Modelling distribution systems.

1 BACKGROUND

Intermittent operation of piped water supply networks is prevalent throughout the world; estimates suggest IWS is operated in 41% of networks in lower and middle-income countries [1]. This is despite the fact there are few, if any, systems that have been intentionally designed to operate intermittently [2]. IWS can affect water quality; frequent de-pressurisation of the network during 'dry-tap' periods creates contamination pathways, causing a substantially higher risk of intrusion of environmental contaminants than in continuous water supply systems [3]. The induced wet-dry cycles and high flow rates influence biofilm detachment, whilst water stagnation and household storage of water increases retention time and bacteriological growth [2]. Moreover, dynamic hydraulic forces in IWS may accelerate infrastructure deterioration, leading to increased leakage rates [4]. Different studies have found that the impact of such mechanisms vary significantly between systems in different locations [5], [6]. Despite this, it is estimated IWS causes significant impacts to public health resulting in an estimated 4.5 million diarrhoeal disease cases and 1560 deaths per year [7].

1.1 Inequitable Distribution of Water

Attempts to quantify volumes of water received by different users in an IWS network have often shown widespread inequity. A study in Hubli-Dharwad, India, used household surveys to calculate the consumed volumes of water of different consumer groups. The authors found the water usage ranged from 21.0 to 97.1 Litres per capita per day (LPCD) [8]. In Kathmandu, Nepal, a household survey (n=369) measured the inequality in distributed water [9]. It found dependence on tankered water to supplement water demand varied from 8 - 51% between different household groups, indicating large differences in the adequacy of the piped supply. The variation in Kathmandu produced a Hoover Index for the city of 0.51 indicating that 51% of the supply hours need to be redistributed in order to achieve equality of supply hours.

This paper proposes grouping the factors influencing the inequitable quantity of water received by consumers into three broad categories:

(a) Supply characteristics

(b) Network hydraulics

(c) Consumer access (to the network connection and storage volume)

1.2 Supply Characteristics

Observations from systems across the globe have recorded the wide range of intermittency regimes. These have been categorised into unreliable, irregular and predictable supply [10]. Unreliable supply means water is supplied at random times with gaps between supply periods ranging from days to weeks. This induces the greatest hardship for consumers and can lead to drastic action such as local protests [6]. Irregular supply means a reliable total volume supplied but with unknown delivery timings, while predictable supply means a consistent supply schedule with a guaranteed volume received each week. With enough storage, predictable supply can enable households to mitigate the majority of the effects of intermittency [10]. In Kathmandu, consumers placed equal expectations on improvement in regularity of supply as they did for the total volume supplied, highlighting the value placed on predictability of supply [9]. In summary, the mode of operation of an IWS network has a profound impact on the adequacy of the supplied water.

1.3 Network Hydraulics

Erickson et al. [6] found large variation in received supply between differently operated IWS networks and within networks in Panama: “walking 50 yards up a hill in Zone 1 could take you from a house where supply rarely went out to a house where supply went out most afternoons” [6]. Spatial variation in hydraulic conditions resulted in service quality being unequal between neighbours as well as between neighbourhoods.

An unequal distribution of pressure across IWS networks has also been widely observed, for example, Ghorpade et al [11] observe inequitable pressure in IWS systems in India. This is supported by Andey & Kelkar [12] who found significant variation in the measured pressure across four IWS networks also in India. Sánchez-Navarro et al [13] recorded the pressure at 347 points within an IWS network in Chihuahua, Mexico, over 3 years. The results showed significant variation in the local pressure at different points in the network. Chandapillai et al. [14] conclude that the pressure dependency in IWS networks make them innately inequitable.

1.4 Consumer Access

Having sufficient storage availability can mitigate the impact of intermittency and ensure an adequate volume of water for the household. In Jaipur, 77% of respondents found a three-hour supply period to be adequate, where 76% of respondents in Panaji said that their five-hour supply period was inadequate [12]. The authors suggest this discrepancy is a result of greater household storage volume in Jaipur compared to Panaji.

In Hubli-Dharwad, a survey of the population was used to define their access to the piped network [8]. The study found a range of consumer access that they classified by (i) connection access (i.e. whether the connection was shared with neighbours) and (ii) storage volume (i.e. whether they had an overhead storage container or not). The volumes received varied accordingly, with a difference of 76.1 LPCD between the lowest and highest access groups. In Greater Amman, Jordan, household surveys revealed a stark contrast in the available household storage volume with average volumes of 3.12 and 16.24m^3 for low and high-income households, respectively [15]. This appeared to correlate with the average consumption values of the two groups of 32.68 and 70.24 m^3 per quarter.

1.5 Modelling IWS

Modelling the distribution of water in IWS systems is an ongoing challenge. The majority of literature aiming to simulate IWS has thus far focussed on using hydraulic modelling software [16], [17], such as EPANET, which are designed for continuously operated systems, i.e., water is distributed according to demand. Various attempts have been made to adapt existing hydraulic models to better represent intermittent systems. Batish [18] modelled demand nodes as reservoirs in EPANET to imitate pressure dependent demand. The effect of consumer storage has been modelled by replacing the reservoirs with tanks, which fill during supply hours resulting in volume dependent demand [19]. These models are more realistic as they are able to represent how the storage volume available to a consumer limits the quantity of water they can receive during a supply period.

De Marchis et al [16] consider the effects of the filling process when water supply is turned on in the network. The unique hydraulic conditions during the filling and emptying stages of IWS networks may be significant in determining the received quantities of water, however, there is insufficient data to validate their model. These models are limited to only analysing the network hydraulics of the system and partial effects of consumer storage volume. The wider structure of the system such as the supply characteristics and other consumer access issues are not captured. Additionally existing models have a fundamental problem: they are difficult to implement in IWS settings due to the requirement for detailed network data that is often highly uncertain.

In response to this problem, an alternative approach has taken the principle of parsimony to model an IWS network in order to maximise learning with minimal input requirements [20]. The author simplifies a hydraulic network into a single consumer and a single leak. The model reveals the relationship between the bulk values of consumer usage and leakage against changes in supply duration. Taylor et al. [20] propose a new parameter termed consumer demand satisfaction (CDS), which is the degree to which consumers are satisfied with the volume of water they receive. Crucially, CDS defines the point at which consumers will turn off their taps, re-directing water to other areas in the network. In this way, the model begins to explore the interplay between consumers and the network; however, the variability in the distribution of water within the network cannot be reflected.

The models to date cannot incorporate all the factors highlighted. The modelling techniques revolve around the network with much less attention given to how consumers use and store the delivered water. The temporal effects of these factors require particular attention. The first step to develop more representative models is to interrogate all the factors influencing the quantity of water received by a consumer over a period of time.

2 METHODS: AN ANALYSIS OF THE QUANTITY OF WATER RECEIVED BY A CONSUMER IN AN IWS SYSTEM

This investigation is split into three sections following the three categories defined previously. Firstly, the influence of supply characteristics is considered. Secondly, the network hydraulics are assessed through analysis of a single supply period. Finally, the impacts of consumer access are discussed. Figure 1 summarises the key determining factors and their grouping.

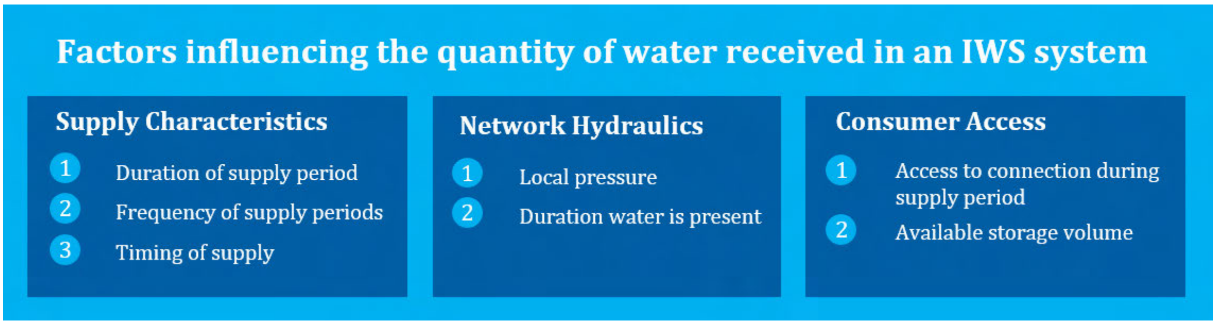

Factors influencing the quantity of water received in an IWS system
Supply Characteristics
1 Duration of supply period
2 Frequency of supply periods
3 Timing of supply
Network Hydraulics
1 Local pressure
2 Duration water is present
Consumer Access
1 Access to connection during supply period
2 Available storage volume

Figure 1. Categorisation of the factors influencing the volume of water received by consumers in IWS systems

2.1 Supply Characteristics

The supply characteristics govern all aspects of the quantity of water received by a consumer in an IWS system. A longer supply period allows water to reach further in the network for longer, thus, supplying water to more consumers. More frequent supply periods reduce the duration of the non-supply period reducing the reliance on storage volume.

Supply Timing: When the supply is on, households use the water to fill storage and complete domestic tasks. The timing of supply impacts how usable water is during these supply hours. Inconvenient timings, such as a supply period during the night, make it far less convenient to use water during supply, therefore limiting a consumer's access to water. The impact of supply timing is mitigated by having sufficient storage, since household tasks do not need to be concentrated around the delivery of water. Consequently, the timing of supply compounds the disadvantage of inadequate storage and disproportionately reduces access to water for those with limited storage.

2.2 Network Hydraulics

This analysis aims to identify the principles governing the distribution of water in an IWS network, focusing on the factors that affect the quantity of water received in different locations. The volume of water which leaves the network through a leak or consumer connection, is a function of time t, pressure H and orifice area A as per equation (1):

$$V_L = C_d . t . A . (2gH)^{\alpha} \tag{1}$$

$C_d = Factor\ to\ account\ for\ orifice\ shape$
$g\ = Gravitational\ acceleration$
$\alpha = Factor\ to\ account\ for\ the\ flow\ rate's\ pressure\ dependency$

The volume of water that leaves a given orifice will therefore depend on the duration that water is present and the local water pressure. An IWS cycle consists of several stages that impact the magnitude of these two parameters across the network.

Filling: When the water is turned on, it fills the network from the inlet sources. The velocity at which water travels through the network will be dependent on the following factors: inlet pressure, relative elevation, frictional losses from the pipes, water demand and the ease at which air is released from the pipes. Therefore, the time it takes to arrive at a particular location in the network, will depend on these factors in combination with the distance from the inlet.

Water Delivery: Once the network has filled, the Bernoulli equation shows that the pressure at any point in the network is a function of the relative elevation and the frictional losses. It follows that locations in the network which are closest to the inlet location and at a low elevation will have the highest local pressures.

CC BY NC SA

Emptying: When the input supply is turned off, the network will drain out of the orifices (leaks and consumer taps). The network will drain under gravity meaning it will empty from the higher elevations down to the lower elevations. As in the case of the filling stage, this results in different durations of water being present in different locations.

These stages of the supply period result in differential local pressures and supply durations across the network. The orifice equation shows that this will result in different total volumes of water leaving orifices in different locations. As a result, leakage volumes and the volume of water received by different consumers will vary across the network.

2.3 Consumer Access

Consumer Access during the supply period: The delivery of water from the inlet sources to consumer connections is only part of what determines how much water individual consumers receive. During supply, if a connection serves more than one household, the received supply time is divided by the number of groups sharing the connection. An additional consideration is the rate at which water can realistically be used by a household. Once storage vessels are filled, households must use water as effectively as possible for household tasks. There is a limitation on what is physically possible which may be less than the flow rate of water out of the tap (depending on the local pressure).

Storage Volume: The storage volume and the length of time between supply periods dictate the access to water during the non-supply period. For instance, we can compare two systems that both supply seven hours of water per week. In the first system, a household receives water for seven hours once a week; while in the second system, a household receives water for one hour seven times a week. The individual household storage volume will drastically alter the quantity of usable water received over the course of the week in the two scenarios. Without a very large storage tank, the household water supply in the first scenario will be severely restricted while in the second scenario a small storage volume will not have such an effect.

3 RESULTS: A CONCEPTUAL MODEL OF IWS SYSTEMS

The analysis of how water is distributed from source to point-of-use results in the conceptual model shown in Figure 2. It follows the flow of water through the system and therefore how different consumers will receive different volumes of water. The model operates over a fixed period of time, be that a week or month, to establish the operation of the system and the volumes of water received by consumers.

The output of the model is the received volumes and associated consumer demand satisfaction of the different consumer groups. Utilising demand satisfaction ensures water is re-distributed to other groups simulating taps being turned off. The range in consumer demand satisfaction in turn describes the inequity in the system. The number of network ‘zones’ and ‘access groups’ required to represent the system is dictated by the local context.

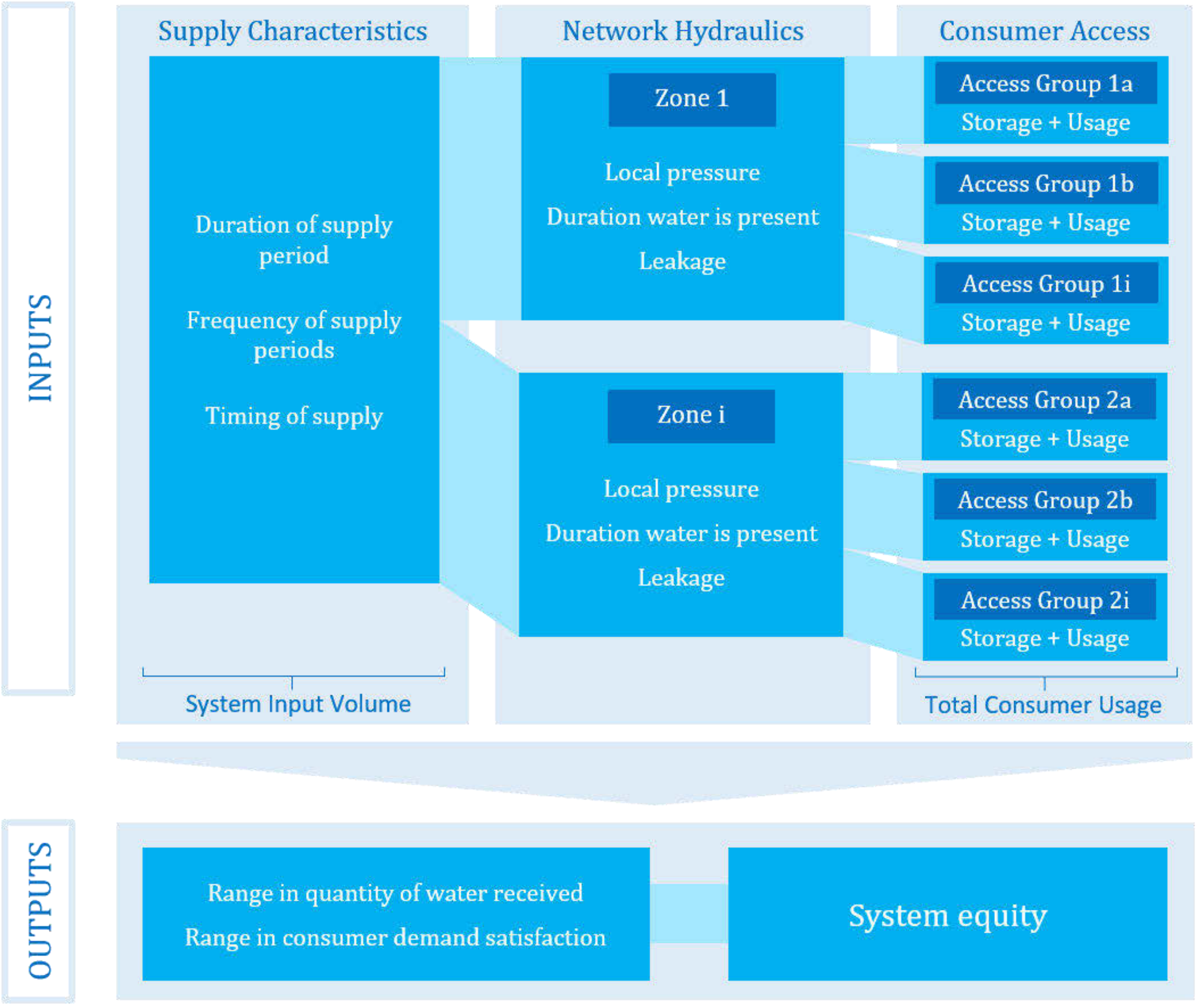

INPUTS
Supply Characteristics
Duration of supply period
Frequency of supply periods
Timing of supply
System Input Volume
Network Hydraulics
Zone 1
Local pressure
Duration water is present
Leakage
Zone i
Local pressure
Duration water is present
Leakage
Consumer Access
Access Group 1a
Storage + Usage
Access Group 1b
Storage + Usage
Access Group 1i
Storage + Usage
Access Group 2a
Storage + Usage
Access Group 2b
Storage + Usage
Access Group 2i
Storage + Usage
Total Consumer Usage
OUTPUTS
Range in quantity of water received
Range in consumer demand satisfaction
System equity

Figure 2. A conceptual model of the distribution of water through IWS systems

4 DISCUSSION

The model illustrated by Figure 2 enables an understanding of how an IWS system results in different quantities of water received by consumers. The hydraulic behaviour of IWS networks results in different pressures and duration that water is present for different zones in the network. The inequitable quantity of water received across the network is compounded by differences in access to the connection. Consumers sharing a connection must divide the volume supplied between themselves. Access to water during the non-supply period is constrained by the storage volume. Both of these categories are governed by the supply characteristics, which determine the volume received at a particular location during the supply period as well as over the course of a week or month.

Ultimately, the key output of the model is the range of consumer demand satisfaction in the system. Once the model represents a system, different scenarios can be run against the model to assess their impact on demand satisfaction. For example, increasing supply hours, changing supply frequency or ensuring one connection per household. The model will show that changes to the system do not result in uniform impact across all consumer groups. The unique characteristics of each group dictate the manner in which they will be affected.

5 CONCLUSIONS

To understand IWS systems it is important to consider the supply characteristics, network hydraulics and consumer access to their network connection. This allows a full consideration and representation of the inequity in the system. Currently models fail to appreciate the wider system beyond the network boundaries and therefore fail to represent the inequity of the system. By analysing both the hydraulics of a supply period and the wider factors of consumer access and input characteristics, a more representative model can be achieved.

This paper proposes an alternative approach to modelling IWS systems based on representing the distribution of water from source to point-of-use. The conceptual model presented in this paper builds the foundation for more detailed and applicable models. These will allow the inequity of an IWS system to be better represented and will also enable predictive capabilities of a wider range of system interventions.

6 REFERENCES

[1] B. Charalambous and C. Laspidou, “Dealing with the Complex Interrelation of Intermittent Supply and Water Losses,” Water Intelligence Online, vol. 16, Jul. 2017, doi: 10.2166/9781780407074.

[2] E. Kumpel and K. L. Nelson, “Intermittent Water Supply: Prevalence, Practice, and Microbial Water Quality,” Environmental Science and Technology, vol. 50, no. 2. American Chemical Society, pp. 542–553, Sep. 2016. doi: 10.1021/acs.est.5b03973.

[3] C. Calero Preciado et al., “Intermittent Water Supply Impacts on Distribution System Biofilms and Water Quality,” Water Research, vol. 201, 2021, doi: 10.1016/j.watres.2021.117372.

[4] S. Christodoulou and A. Agathokleous, “A study on the effects of intermittent water supply on the vulnerability of urban water distribution networks,” Water Science and Technology: Water Supply, vol. 12, no. 4, pp. 523–530, 2012, doi: 10.2166/ws.2012.025.

[5] A. Ercumen et al., “Upgrading a Piped Water Supply from Intermittent to Continuous Delivery and Association with Waterborne Illness: A Matched Cohort Study in Urban India,” PLoS Medicine, vol. 12, no. 10, 2015, doi: 10.1371/journal.pmed.1001892.

[6] J. J. Erickson, Y. C. Quintero, and K. L. Nelson, “Characterizing supply variability and operational challenges in an intermittent water distribution network,” Water (Switzerland), vol. 12, no. 8, 2020, doi: 10.3390/W12082143.

[7] A. W. Bivins et al., “Estimating Infection Risks and the Global Burden of Diarrheal Disease Attributable to Intermittent Water Supply Using QMRA,” Environmental Science and Technology, vol. 51, no. 13, 2017, doi: 10.1021/acs.est.7b01014.

[8] E. Kumpel, C. Woelfle-Erskine, I. Ray, and K. L. Nelson, “Measuring household consumption and waste in unmetered, intermittent piped water systems,” Water Resources Research, vol. 53, no. 1, Jan. 2017, doi: 10.1002/2016WR019702.

[9] B. Guragai, S. Takizawa, T. Hashimoto, and K. Oguma, “Effects of inequality of supply hours on consumers’ coping strategies and perceptions of intermittent water supply in Kathmandu Valley, Nepal,” Science of the Total Environment, vol. 599–600, pp. 431–441, Sep. 2017, doi: 10.1016/j.scitotenv.2017.04.182.

[10] S. E. Galaitsi, R. Russell, A. Bishara, J. L. Durant, J. Bogle, and A. Huber-Lee, “Intermittent domestic water supply: A critical review and analysis of causal-consequential pathways,” Water (Switzerland), vol. 8, no. 7. MDPI AG, Sep. 2016. doi: 10.3390/w8070274.

[11] A. Ghorpade, A. K. Sinha, and P. Kalbar, “Multi-outlet storage tanks to improve water distribution networks in India,” Urban Water Journal, vol. 18, no. 7, 2021, doi: 10.1080/1573062X.2021.1914117.

[12] S. P. Andey and P. S. Kelkar, “Performance of water distribution systems during intermittent versus continuous water supply,” Journal / American Water Works Association, vol. 99, no. 8. 2007. doi: 10.1002/j.1551-8833.2007.tb08011.x.

[13] J. R. Sánchez-Navarro, D. H. Sánchez, C. J. Navarro-Gómez, and E. H. Peraza, “Multivariate analysis of the pressure variation in intermittent water supply systems and the impact on demand satisfaction,” Water Supply, vol. 21, no. 7, 2021, doi: 10.2166/ws.2021.153.

[14] J. Chandapillai, K. P. Sudheer, and S. Saseendran, “Design of Water Distribution Network for Equitable Supply,” Water Resources Management, vol. 26, no. 2, 2012, doi: 10.1007/s11269-011-9923-x.

[15] R. B. Potter, K. Darmame, and S. Nortcliff, “Issues of water supply and contemporary urban society: the case of Greater Amman, Jordan,” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 368, no. 1931, Nov. 2010, doi: 10.1098/rsta.2010.0182.

[16] M. de Marchis, C. M. Fontanazza, G. Freni, G. la Loggia, E. Napoli, and V. Notaro, “A model of the filling process of an intermittent distribution network,” Urban Water Journal, vol. 7, no. 6, Dec. 2010, doi: 10.1080/1573062X.2010.519776.

[17] S. Mohapatra, A. Sargaonkar, and P. K. Labhasetwar, “Distribution network assessment using EPANET for intermittent and continuous water supply,” Water Resources Management, vol. 28, no. 11, 2014, doi: 10.1007/s11269-014-0707-y.

[18] R. Batish, “A New Approach to the Design of Intermittent Water Supply Networks.” 2004.

[19] A. Battermann and S. Macke, “A Strategy to Reduce Technical Water Losses for Intermittent Water Supply Systems,” 2001.

[20] D. D. J. Taylor, A. H. Slocum, and A. J. Whittle, “Demand Satisfaction as a Framework for Understanding Intermittent Water Supply Systems,” Water Resources Research, vol. 55, no. 7, pp. 5217–5237, 2019, doi: 10.1029/2018WR024124.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14016

CHARACTERIZING THE EFFECTS OF WATER DISTRIBUTION SYSTEM TOPOLOGY MODIFICATIONS ON ITS DYNAMIC BEHAVIOUR THROUGH CONNECTIVITY METRICS

Valentina Marsili[1], Stefano Alvisi[2], Filomena Maietta[3], Caterina Capponi[4], Silvia Meniconi[5], Bruno Brunone[6] and Marco Franchini[7]

[1] Research Assistant, Department of Engineering, University of Ferrara, Ferrara, Italy
[2] Associate Professor, Department of Engineering, University of Ferrara, Ferrara, Italy
[3] PhD Student, Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy
[4] Research Assistant, Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy
[5] Associate Professor, Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy
[6] Professor, Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy
[7] Professor, Department of Engineering, University of Ferrara, Ferrara, Italy

[1] valentina.marsili@unife.it, [2] stefano.alvisi@unife.it, [3] filomena.maietta@studenti.unipg.it, [4] caterina.capponi@unipg.it, [5] silvia.meniconi@unipg.it, [6] bruno.brunone@unipg.it, [7] marco.franchini@unife.it

Abstract

Water distribution networks (WDNs) are complex combinations of nodes and links and their structure has an impact on their behaviour, considering both quantitative (i.e. related to pipe flows and nodal pressures) and qualitative (i.e. related to water age and quality) aspects. The complexity of WDNs has been the basis of several studies that have resorted to the graph theory to relate connectivity properties to system behaviour (e.g. its reliability and water age/quality), evaluated under the assumption of steady-state conditions. Within this framework, in recent years the tendency toward reducing network interconnection through the closure of isolation valves has emerged, mainly to (i) facilitate its monitoring and management, and (ii) increase flow velocity and reduce water age. However, changes in the topology of a network can affect not only aspects evaluated under the assumption of steady-state conditions, but also its dynamic behaviour. Based on these considerations, the present study investigates whether some metrics derived from graph theory, already applied in the context of networks' steady-state analyses, can also provide useful indications for assessing the effects of changes in the topological structure, which could be consequences of branching operations, on the dynamic response of a network subjected to users' activity. The analyses highlight that connectivity metrics can reflect the pressure dynamic behaviour of the hydraulic systems and support in their macroscopic understanding during design and management operations. Thus, their application can be effectively extended from the steady-state to the dynamic framework.

Keywords
Transient analysis, unsteady flow, water demand, topology, graph theory.

1 INTRODUCTION

Water distribution networks' (WDNs) are normally designed and realized with topological structures, or connectivity structures of pipes and nodes, generally characterized by looped layouts that guarantee high hydraulic and mechanical reliability of the system (Elshaboury et al., 2021; Sirsant et al., 2020). However, these benefits from a reliability point of view imply higher costs of the network (Todini, 2000). In this regard, several methods have been developed and proposed in the scientific literature to identify the optimal topological structure ensuring a balance between reliability and cost (e.g. Choi and Kim, 2019; Farmani et al., 2005). As a result, real WDNs are complex combinations of loops and branches (Walski et al., 2003). This complexity has been the basis for the development of a series of studies that have led to the application to WDN of theories and indicators originally developed in the field of graph theory aimed to manage system hydraulic functioning and water quality (e.g. Abokifa et al., 2019; Giudicianni et al., 2018; Torres et al., 2016; Yazdani & Jeffrey, 2012).

Nowadays, along with the network reliability, other issues are becoming of interest for the water utilities, such as (*i*) the observability and controllability of water networks and (*ii*) the compliance with certain operating conditions (e.g. the values of average and minimum velocity able to ensure sufficient water quality).

Concerning the observability, looped systems have proved to be less controllable and, in recent years, a tendency to reduce the interconnection of networks through, for example, sectorisation techniques (i.e. DMA creation), has emerged. This practice consists of dividing a looped system into portions of networks (i.e. districts) that are connected at a limited number of points where flow meters are placed, while all remaining interconnection points are closed through isolation valves (IVs). The closure of such IVs results in an alteration of the topological/connectivity structure of the network and a reduction of the number of loops. These aspects have been extensively investigated in the literature, also resorting to methodologies based on graph theory (e.g. Diao et al., 2013; Di Nardo et al., 2013; Ferrari et al., 2014; Giustolisi, 2020). It is worth noting that procedures for identifying the optimal districts layout often rely on optimization algorithms that require the resolution of a significant number of connectivity configurations (e.g. Zhang et al., 2017).

With reference to the compliance with certain operating conditions, it is worthy of note that the presence of strongly interconnected network structures typically results in a reduction of average water velocity in pipes with consequent (*i*) ageing of the water before it is delivered to the users and consequent reduction of free residual chlorine or formation of chlorination by-products (Quintiliani, 2017), and (*ii*) non-reaching of the daily self-cleaning velocity that guarantees cleaner pipes (Abraham et al., 2017). To ensure optimal conditions of water velocity, several approaches have been developed in order to identify the optimal interventions on the topological structure of the network based, for example, on the closure of IVs and the reduction of the level of interconnection (Brentan et al., 2021). Again, the optimal set of IVs to be closed is often identified through optimization algorithms evaluating a very large number of solutions.

In summary, on the one hand the closure of the IVs present in the network reduces the level of interconnection and produces an alteration of its topological structure, on the other hand these problems have been so far faced mainly under the assumption of steady-state conditions. However, several studies have shown that, in reality, modern WDNs rarely achieve steady states for more than a small fraction of their operational times and are subjected to continuous pressure transients that can be induced by manoeuvres on main devices in the network, whose status is changed to respond to different conditions of water demands (e.g. changes in the setting of pumps). Recently, there has been evidence that even the operation of regulation valves (Changklom and Stoianov, 2017; Meniconi et al., 2015) and users' activity (Marsili et al., 2021), i.e. the opening and closing of domestic devices, can result in the generation of pressure transients.

Within this framework, several studies have shown that the response of the WDN, to these driving forces depends on its characteristics and, in particular, on the topological structure (i.e. network connectivity) and the geometric and mechanical characteristics of the network, i.e. diameters and materials of the pipes (Ellis, 2008) as well as the presence of elements such as dead-ends (Meniconi et al., 2021). With specific reference to the topological structure (i.e. network graph), it is thus important to observe that modifications of the structure through the closure of IVs can also determine a variation in the response of the system to transients' generation and propagation. From an operational standpoint, the evaluation of the dynamic behaviour of a water network with an assigned topological structure and subjected to a given driving force can be conducted through unsteady flow simulation based on numerical approaches as, for example, the Method of the Characteristics (MOC). However, this approach for complex real WDNs can be extremely expensive from a computational point of view, especially when simulations have to be replicated for a significant number of configurations of the system considered.

In the current study, the ability of some indicators belonging to graph theory – already used in the context of network reliability evaluations – to provide useful indications on the effects of the topological structure variations on the dynamic response of a network is evaluated. To this end, four network connectivity indicators are examined: two basic connectivity metrics and two spectral metrics. These indicators are compared in terms of their ability to represent the dynamic behaviour of a simple WDN, the latter evaluated through a numerical model based on MOC, in which the topological structure is modified by closing an increasing number of IVs for reducing the number of loops. In the following, first, the connectivity indicators are introduced. The methodology adopted is then presented and applied to a simple case study. Results are discussed and finally, some conclusive considerations are provided.

2 MATERIALS AND METHODS

In the following, the connectivity indicators considered in the analysis and the methodology proposed aimed at evaluating the ability of these indicators to reflect useful information on the effects of the structure of the network on its dynamic response, the latter obtained through numerical simulations, are introduced.

2.1 The connectivity metrics

In order to study the relationship between the connectivity properties of networks and their dynamic behaviour, four indicators were considered which belong to graph theory. Among the metrics aimed at investigating the structure and the behaviour of complex networks, the basic connectivity metrics and the spectral measures are the most widespread and well-known. On the one hand, the basic connectivity metrics indicate the degree of connectivity of the vertices (i.e. the *nodes*) and the edges (i.e. the *links*) representing the cohesion of the network and its sensitivity to the removal of nodes and links. Spectral metrics, on the other hand, relate the topology of the network to the connectivity strength and cohesion of the graph by analysing the spectrum of the adjacency matrix, A, of the network, which indicates which vertices are connected (i.e. *adjacent*): the generic element of A $a_{ij} = 1$ if a link connects nodes i and j, otherwise $a_{ij} = 0$. These metrics quantify network properties that depend only on the abstract structure of the graph, independently of its representation (Torres et al., 2016; Yazdani et al., 2010).

More in detail, in this study two basic connectivity metrics and two spectral metrics were considered. Concerning basic connectivity metrics, the first metric considered is the average degree, k, defined as:

$$k = \frac{2m}{n} \tag{1}$$

with m and n the number of links and nodes, respectively. The second basic connectivity metric considered is the meshed-ness coefficient, R_m, defined by the ratio between the total number $(m - n + 1)$ and the maximum theoretical number $(2n - 5)$ of independent loops in the network:

$$R_m = \frac{m - n + 1}{2n - 5} \tag{2}$$

A uniform distribution of the average degree indicates the tendency of the network to be invulnerable to faults, whereas the meshed-ness coefficient can effectively describe the redundancy of the network (Yazdani & Jeffrey, 2012). With reference to spectral metrics, the spectral gap, $\Delta\lambda$, and the algebraic connectivity, λ_2, are considered. Specifically, the spectral gap $\Delta\lambda$ is defined as the difference between the first and second eigenvalue of the adjacency matrix A of the network and it is effective in the quantification of *Good Expansion* (GE) properties of the system. In particular, networks characterised by GE are those whose topological structure presents vertices connected robustly to other nodes, even if the graph is not dense with links (Estrada, 2006). The algebraic connectivity, λ_2, corresponds to the second smallest eigenvalue of the Laplacian matrix, $L = D - A$, of the network (Fiedler, 1973), where D is the diagonal matrix of nodal degrees, the latter indicating the number of links converging at each node. Higher values of λ_2 indicate higher network robustness (and higher reliability) and higher fault tolerance (Yazdani & Jeffrey, 2012).

2.2 The methodology

To evaluate the applicability of the indicators previously introduced to characterise the changes of the dynamic response of a WDN when its topological structure is modified, the approach described below was developed. Consider a generic WDN characterised by n nodes that can be connected with a variable number of links (m) for a total of l loops ($l = m - n + 1$). In Figure 1, a simple WDN is reported as an example and its layout is henceforth referred to as original or *reference configuration* (C1). On the one hand, the four indicators previously introduced can be quantifiable for the considered network. On the other hand, in the face of a generic driving force, such as water demand change at a node, a transient is generated and propagates throughout the network, interacting with its topological structure.

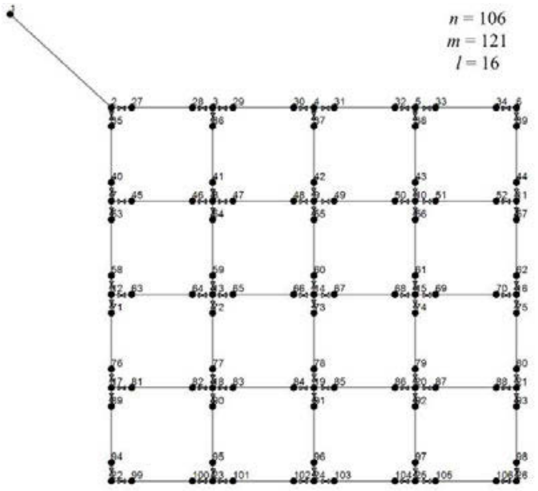


Figure 1 Reference layout of the simple network (i.e. C1).

With reference to the simple network with layout C1, the dynamic response of the system can be obtained through a numerical model and, in the current study, the model introduced in Marsili et al. (2022) and based on MOC is adopted. In greater detail, for the numerical simulation under unsteady flow conditions, it is important to specify that a reservoir located at node 1 fed the network and that all the pipes in the network are characterised by the same diameter (DN125), length (100 m) and wave speed a (500 m/s).

Considering, as an example, an instantaneous closure manoeuvre operated at a generic node, the numerical simulation in a defined time window allows obtaining the value of the interval $\delta p = |p_{max} - p_{min}|$, for each node, representing the range in which pressure oscillates in the considered node, being p_{max} and p_{min} the maximum and minimum values of the pressure. In Figure 2, the interval δp is shown graphically for each node of the network in configuration C1 through a chromatic scale that varies from 0 to 5 m in the face of an instantaneous closure manoeuvre operated, for sake of example, at node 86 highlighted with a red arrow. δp is evaluated considering a simulation time window t_{sim} = 30 s.

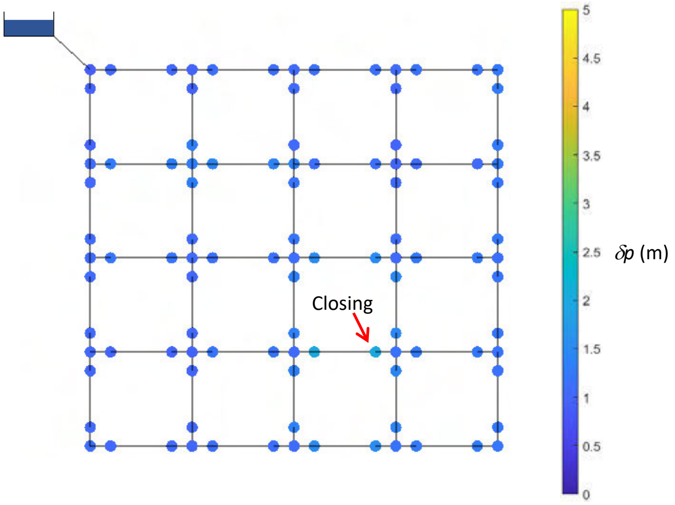


Figure 2 Results of unsteady flow simulation of the network in configuration C1 subjected to an instantaneous closure at node 86 in terms of δp at each node.

More in general, the stress of each node, in terms of the values of δp, can be summarised in the cumulative frequency of δp values. In Figure 3, the cumulative frequency of δp for configuration C1 of the network is shown in blue. Furthermore, the values at the 50th and 90th percentiles of this cumulative distribution – henceforth indicated as δp_{50} and δp_{90}, respectively – can be considered representative of the average and extreme dynamic response of the given network configuration.

If the topological structure, i.e. the connectivity of the system, is modified, for example by closing some IVs to create districts or to improve water quality aspects, this impacts on the connectivity indicators. If *nc* new configurations are considered, *nc* new values for each connectivity indicator is obtained. Moreover, this would lead to different dynamic responses of the system, that can be evaluated in face of the same manoeuvre considered for the *reference configuration* through the numerical model previously introduced. Based on these numerical results, the cumulative frequencies of δp for each new configuration is traceable. The pressure stress variation of each one of the *nc* configuration with respect to the *reference configuration* can be evaluated as $\frac{\delta p_{50}}{\delta p_{50,ref}}$ and $\frac{\delta p_{90}}{\delta p_{90,ref}}$, where $\delta p_{50,ref}$ and $\delta p_{90,ref}$ denote the average and extreme dynamic response of the network in the *reference configuration*.

The ability of the connectivity indicators to provide useful indications on the effects of topological structure variations on the changed dynamic response of a network is evaluated considering their correlation with the corresponding pressure variation indicator obtained through hydraulic simulation. In the following, a numerical example is provided.

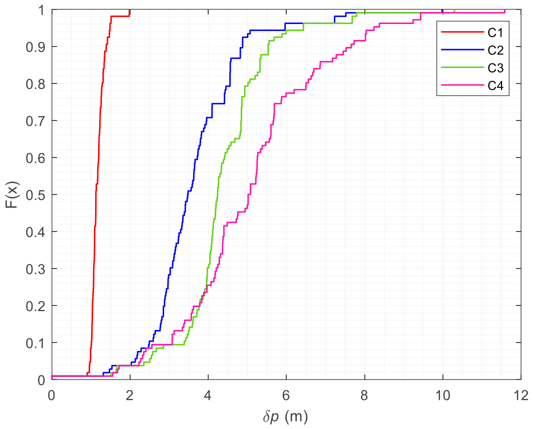


Figure 3 Cumulative frequency of pressure variations δp stressing the network nodes in different configurations (C1-C4).

3 RESULTS AND DISCUSSION

The results of the application of the approach proposed to the simple network considered are hereinafter reported.

Starting from the *reference configuration* C1 (Figure 1), three additional configurations of the network are obtained by closing an increasing number of IVs (12, 15, and 16) and indicated as C2, C3 and C4, for a total of nc = 4 configurations considered, including the reference one. The topological structure of the new configurations is then characterised by the indicators adopted before. The dynamic responses of the simple network in the modified layouts (i.e. C2, C3 and C4) subjected to the same instantaneous manoeuvre at node 86 (which had induced the pressure state in the reference network shown in Figure 2) is obtained through the numerical model and result to be different compared to the response of the network in configuration C1. This is evident in Figure 4, Figure 5 e Figure 6, where the results of the numerical simulations in terms of δp observed in all the nodes are reported when configurations C2, C3 and C4 are considered. Specifically, a greater number of nodes are affected by more significant δp as the number of branches increases and interconnections decrease. In fact, if in the case of the looped configuration (i.e. C1) the nodes are stressed by δp that remains in the order of 1 m (Figure 2), considering a completely branched configuration (i.e. C4), almost all the nodes are stressed by δp of 5 m (Figure 6). Figure 3 compares the pressure dynamic responses of the simple network in the nc configurations in terms of cumulative frequency of δp. As the number of branches in the network increases, the cumulative frequency δp tends to shift to the right. The reciprocal position of the curves - and therefore the values of δp_{50} and δp_{90} - highlights how, geometric and mechanical characteristics being equal, the topological structure influences the dynamic response of the system and, in particular, that the state of pressure stress of the system in terms of δp increases as the number of loops decreases.

BY NC SA

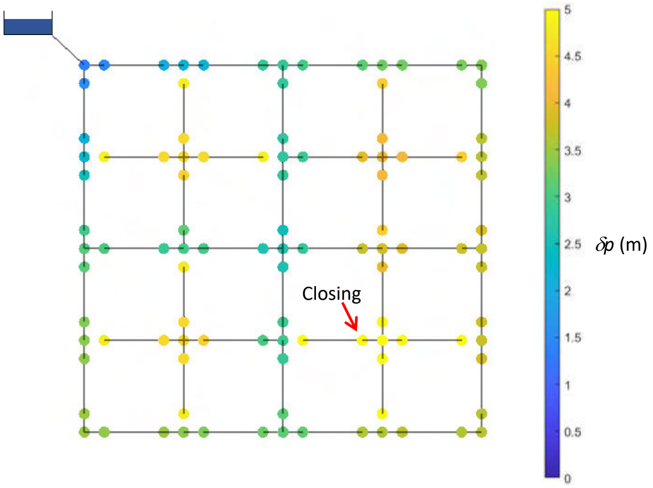


Figure 4 Results of unsteady flow simulation of the network in configuration C2 subjected to an instantaneous closure at node 86 in terms of δp at each node.

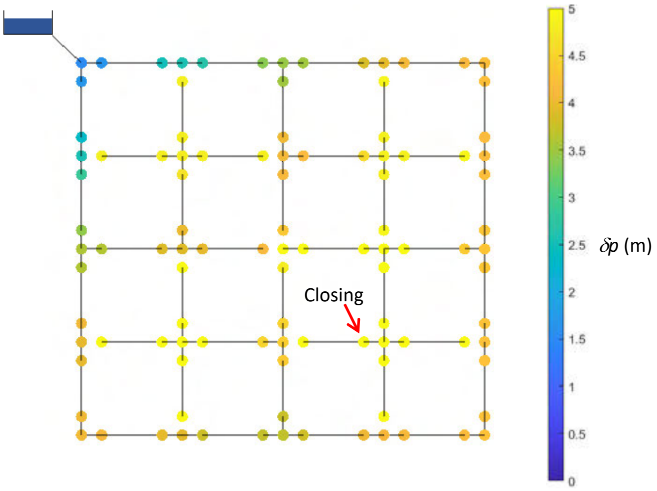


Figure 5 Results of unsteady flow simulation of the network in configuration C3 subjected to an instantaneous closure at node 86 in terms of δp at each node.

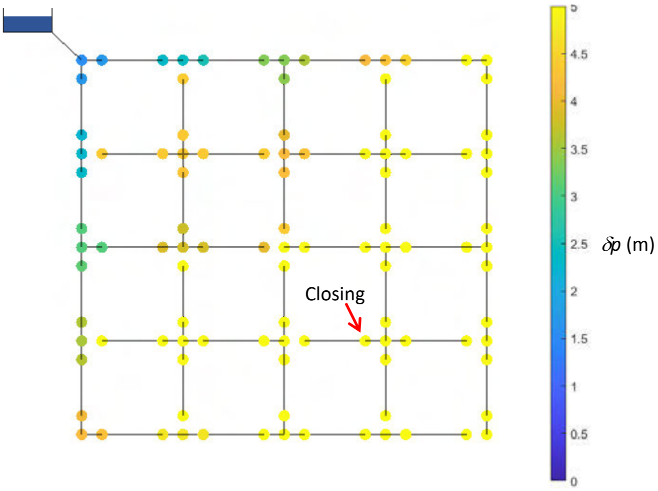


Figure 6 Results of unsteady flow simulation of the network in configuration C4 subjected to an instantaneous closure at node 86 in terms of δp at each node.

The results of the unsteady flow simulations of the nc configurations of the simple network in terms of the dimensionless response of the system at the 50th and 90th percentiles, $\frac{\delta p_{50}}{\delta p_{50,ref}}$ and $\frac{\delta p_{90}}{\delta p_{90,ref}}$, are related to the four connectivity metrics (k, R_m, $\Delta\lambda$ and λ_2) and the result of this evaluation is shown in Figure 7, where the red stars indicate the simple network in its reference configuration whose response in terms of $\frac{\delta p_{50}}{\delta p_{50,ref}}$ and $\frac{\delta p_{90}}{\delta p_{90,ref}}$ is unitary.

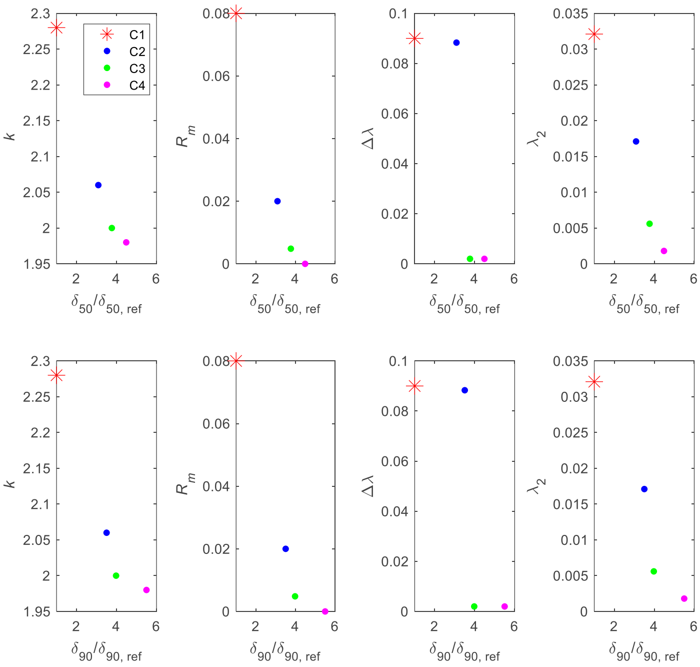


Figure 7 Connectivity metrics (k, R_m, Δλ and λ_2) as a function of the dimensionless dynamic responses of the simple network considered in the nc configurations. The red stars indicate the network in its reference configuration (i.e. C1).

Firstly, it can be observed that as the number of the branches in the network increase (from configuration C1 to C4), k, R_m, $\Delta\lambda$ and λ_2 tend to decrease. Moreover, it is evident that the average degree k, the meshed-ness coefficient R_m and the algebraic connectivity λ_2 are the most effective metrics in reflecting the dynamic behaviour of the system. They indicate, through regular fronts, how as the number of branches increases (so as k, Rm and λ_2 decrease), and therefore as the redundancy of the network decreases (i.e. the network has a lower number of loops), the dynamic response of the network subjected to a single instantaneous manoeuvre is emphasised up to about 4.5 times compared to the average response $\delta p_{50,ref}$ of the reference configuration (i.e. C1) and about 5.5 times compared to the extreme response $\delta p_{90,ref}$. The spectral gap $\Delta\lambda$, on the other hand, is not representative of the dynamic behaviour of the network as its configuration changes and responds with a scattered distribution.

Although it does not emerge in the current case, it is worth stressing the limitation of the metrics k and R_m in distinguishing configurations that are different from the point of view of the adjacency matrix but have the same number of nodes n and links m.

The effectiveness of some connectivity metrics to represent the dynamic behaviour of the network in different configurations could be applied in the evaluation of the impact of a solution of sectioning of the system on the state of stress to which it will be subjected once the modifications have occurred. The use of connectivity metrics could compensate for the execution of a significant number of simulations under unsteady flow conditions, expensive from a computational point of view, since the dynamic response of the network that presents a certain connectivity, and therefore a certain metric, can be easily approximated after the classification of a very small number of exact solutions, perhaps evaluated for "extreme" network configurations, i.e. highly looped or highly branched layouts.

4 CONCLUSIONS

In this study, the effectiveness of some connectivity indicators to represent the effects of the change of the topological structure of a WDN on the resulting dynamic response of the system is evaluated. To this end, four connectivity metrics from graph theory (i.e. average degree k, meshness coefficient R_m, algebraic connectivity λ_2 and spectral gap $\Delta\lambda$) are compared and evaluated in terms of the ability to represent the dynamic behaviour of a WDN. Specifically, this network is subjected to an instantaneous manoeuvre at a node, and its dynamic behaviour is obtained through a numerical model based on MOC, in which the topological structure is modified by closing an increasing number of IVs, thus reducing the number of loops. The results show that k, R_m and λ_2 are the most effective metrics in reflecting macroscopically the dynamic behaviour of the system, while $\Delta\lambda$ is not representative and tends to respond with a scattered distribution. The ability of the metrics considered to represent the dynamic behaviour of a WDN as the configuration changes could be exploited to evaluate the impact of different solutions of sectioning of the system on the stress to which it will be subjected compared to the original configuration. The application of the approach proposed to more complex WDN is the main future objective.

5 REFERENCES

[1] N. Elshaboury, T. Attia, M. Marzouk, Hydraulic Reliability Assessment of Water Distribution Networks Using Minimum Cut Set Method. In "JIC Smart Cities 2019: Design and Construction of Smart Cities. - Sustainable Civil Infrastructures (SUCI)", 1st ed., New York (USA): Springer, 2021.

[2] S. Sirsant and M. J. Reddy, "Assessing the Performance of Surrogate Measures for Water Distribution Network Reliability." J. Water Resour. Plann. Manage., vol. 146, no. 7, July 2000, pp. 04020048.

[3] E. Todini, "Looped water distribution networks design using a resilience index based heuristic approach." Urban Water, vol. 2, no. 2, June 2000, pp. 115-122.

[4] Y. H. Choi, J. H. Kim, "Development of Multi-Objective Optimal Redundant Design Approach for Multiple Pipe Failure in Water Distribution System." Water, vol. 11, no. 3, March 2019, pp. 553.

[5] R. Farmani, G. A. Walters and D. A. Savic, "Trade-off between total cost and reliability for Anytown water distribution network." J. Water Resour. Plann. Manage., vol. 131, no. 3, May 2005, pp. 161–171.

[6] T. Walski, D. V. Chase, D. Savić, W. Grayman, S. Beckwith, E. Koelle, Advanced Water Distribution Modeling and Management., Waterbury, Connecticut (USA): Haestad Press, 2003.

[7] A. A. Abokifa, A. Maheshwari, R. D. Gudi, P. Biswas, "Influence of Dead-End Sections of Drinking Water Distribution Networks on Optimization of Booster Chlorination Systems." J. Water Res. Plan. Manage., vol. 145, no. 12, December 2019, pp. 04019053.

[8] C. Giudicianni, A. Di Nardo, M. Di Natale, R. Greco, G. F. Santonastaso, A. Scala, "Topological Taxonomy of Water Distribution Networks." Water, vol. 10, no. 4, April 2018, pp. 444.

[9] J. M. Torres, L. Duenas-Osorio, Q. Li, A. Yazdani, "Exploring Topological Effects on Water Distribution System Performance Using Graph Theory and Statistical Models." J. Water Res. Plan. Manage., vol. 143, no. 1, January 2017, pp. 04016068.

[10] A. Yazdani, P. Jeffrey, "Applying network theory to quantify the redundancy and structural robustness of water distribution systems." J. Water Res. Plan. Manage., vol. 138, no. 2, March 2012, pp. 153–161.

[11] C. Quintiliani, "Water Quality Enhancement in WDNs through Optimal Valves Setting." PhD Thesis, University of Cassino and Southern Lazio, 2017.

[12] E. Abraham, M. Blokker, I. Stoianov, "Decreasing the Discoloration Risk of Drinking Water Distribution Systems through Optimized Topological Changes and Optimal Flow Velocity Control." J. Water Res. Plan. Manage., vol. 144, no. 2, February 2018, pp. 04017093.

[13] B. Brentan, L. Monteiro, J. Carneiro and D. Covas "Improving Water Age in Distribution Systems by Optimal Valve Operation." J. Water Res. Plan. Manage., vol. 147, no. 8, August 2021, 04021046.

[14] J. Changklom, I. Stoianov, "Fault Detection and Diagnosis for Pressure Control Valves in Water Supply Networks." In Proc. of the Computing and Control for the Water Industry Conference 2017, Sheffield (UK), 5–7 September 2017; F143.

[15] S. Meniconi, B. Brunone, M. Ferrante, E. Mazzetti, D. B. Laucelli, G. Borta, "Hydraulic characterization and transient response of Pressure Reducing Valves." J. Hydroinformatics., vol. 19, no. 6, 2017, pp. 798-810.

[16] V. Marsili, S. Meniconi, S. Alvisi, B. Brunone, M. Franchini, "Experimental analysis of the water consumption effect on the dynamic behaviour of a real pipe network." J. Hydraul. Res., vol. 59, no. 3, 2021, pp. 477-487.

[17] J. Ellis, Pressure transients in Water Engineering: A Guide to Analysis and Interpretation of behaviour. London (UK): Thomas Telford Ltd, 2008.

[18] S. Meniconi, M. Cifrodelli, C. Capponi, H.F. Duan, B. Brunone, "Transient response analysis of branched pipe systems towards a reliable skeletonization." J. Water Res. Plan. Manage, vol. 147, no. 2, 2021, pp. 04020109.

[19] V. Marsili, S. Meniconi, S. Alvisi, B. Brunone, M. Franchini, "Stochastic approach for the analysis of demand induced transients in real water distribution systems." J. Water Res. Plan. Manage., vol. 148, no. 1, January 2022, pp. 04021093.

[20] A. Yazdani, P. Jeffrey, "A complex network approach to robustness and vulnerability of spatially organized water distribution networks." arXiv, August 2010, 1008.1770.

[21] E. Estrada, "Network robustness to targeted attacks. The interplay of expansibility and degree distribution." Eur. Phys. J. B, vol. 52, August 2006, pp. 563-574.

[22] M. Fiedler, "Algebraic Connectivity of Graphs." Czechoslov. Math. J., vol. 23, no. 2, 1973, pp. 298-305.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14024

LOCAL CONTROL SCHEMES FOR REAL-TIME OPTIMIZATION OF VARIABLE SPEED PUMPS

Elad Salomons[1], Mashor Housh[2] and Uri Shamir[3]

[1,2] Department of Natural Resources and Environmental Management, University of Haifa, Israel
[3]Civil and Environmental Engineering, Technion, Haifa, Israel

[1] selad@optiwater.com, [2] mhoush@univ.haifa.ac.il, [3] shamir@technion.ac.il

Abstract

In this work, we have developed local control schemes for optimal operation of a pumping station that includes multiple variable speed pumps. The pumping station must maintain a target pressure for a closed supply network that usually does not contain storage facilities. The control loop includes repeatedly reading the control feedback (pressure) from the control system and then adjusting the operation of the pumps in the station as needed: changing the pumps' speed and on\off status. Firstly, we formulated a control algorithm that simulates the current practice. Next, an optimization algorithm was developed to achieve minimum energy operation considering the efficiency curves of the pumps. The algorithm utilizes the pumps' characteristics curves for ensuring hydraulic feasibility while sustaining the required pressure setpoint. The optimization problem incorporates bound constraints on the speeds, time gap constraints to prevent frequent pump changes, and physical constraints that quantify the operation point in the flow-head domain and the flow-efficiency domain. We considered two operation strategies. The first is the "free strategy", in which each pump can operate with a different speed inside a predetermined speed range, while in the "equal strategy", all active pumps must share the same speed. The methodology was demonstrated using a realistic case study. Our preliminary results indicate potential energy reduction compared to the current practice.

Keywords
Variable speed pumps, real-time control, water distribution systems, optimization.

1 BACKGROUND

Variable speed pumps (VSP) are used to maintain a desired flow or pressure. VSPs are common because they provide several advantages [1], including a) the pump flow can change gradually and give the upstream process (e.g., treatment plant) time to adjust; b) no water storage is required on the demand side, as the pump can adjust to changing demands while maintaining the required pressure in the demand zone; c) the flow can be changed gradually to reduce water hammer, and d) motor life can be extended since fewer starts and stops are needed [2].

On the other hand, Gottliebson et al. [1] argue that VSPs also have disadvantages as compared to fixed speed pumps (FSPs): a) they are more expensive in both installation and maintenance; b) they may be less efficient; c) controlling a VSP is more complex, and d) a VSP may not be suitable for flat H-Q system curves as high efficiency is difficult to maintain over the entire flow range. Despite these disadvantages, VSPs are most popular in systems with no water storage. In these systems, there is a need to regulate the flow using a demand following mechanism.

With the VSPs gaining popularity in practice, they have been modeled in most simulation software, such as EPANET [3], and their modeling and simulation continue to be an active research topic [4–6]. Many studies of VSPs address the operation of WDSs and optimization of pumps scheduling [7–11]. In a recent review, Wu et al. [12], report improved system efficiency due to VSPs and other benefits of increased flexibility in controlling WDSs in real-time. Lima et al. [13] suggested using

CC BY NC SA

VSPs to recover energy and reduce leakage in WDSs. Wu et al. [14] incorporated VSPs in the design stage of water networks and transmission lines. Huo et al. [15] explored using VSPs in deep injection well systems.

The operation of variable speed pumps for pressure control in a closed pipe system is a well-known problem [16–24]. However, a simple and practical methodology is still required for optimal operation of an entire pumping station to utilize the VSPs better and reduce energy costs. One of the most common controllers used is the proportional–integral–derivative (PID) controller [25]. A PID controller continuously calculates the difference (error) between the desired setpoint and the measured variable. Then it applies a correction based on proportional, integral, and derivative terms of the error.

2 METHODOLOGY

2.1 Current controller

In the traditional control loop of the VSP based pumping station, for each time step t, the current state of the system is obtained from the SCADA, which includes the on\off state of the pumps, I, their speed, n, and the last update time of these values, τ_I and τ_n respectively. The reference pressure (setpoint), H_{ref}, and the measured pressure, H are also obtained. The difference (error) between these values is calculated, e, and fed to the proportional control algorithm (PCA). Some parameters for the PCA are predetermined: the minimum and maximum allowed speeds of the pumps, n_{min} and n_{max} respectively, the maximum allowed change in the pump's speed, Δn_{max}, the proportional coefficient, α, which is the change in the pump's speed relative to the calculated error (e). A minimum time, t_{min}, is also set to limit the time between major changes in pumps operations (e.g., start, stop, reduce the pump's speed from n_{max}). The output of the PCA is the adjusted values for I and n, denoted as $\hat{I}$ and $\hat{n}$ respectively. When these new settings are applied to the system, the system will respond with new values of the pressure and the flow, H_{t+1} and Q_{t+1} respectively. At this stage, the control loop is repeated for the next time step.

The above algorithm is traditional in control applications, and it is implemented in many control system use cases. However, it is not uniquely tailored for pumping systems. The control employs control rules to keep the feedback signal at the desired setpoint regardless of the energy efficiency and hydraulics. Relying on a physical model of the pumps' hydraulics and efficiencies, we propose a model-based approach tailored explicitly for pumping stations that work against closed networks without storage. Unlike the traditional approach, we propose optimizing energy cost as a primary objective, while satisfying the setpoint constraints.

2.2 Proposed Controller

The proposed model-based controller solves an optimization problem to minimize energy consumption. The pump curves (i.e., efficiency and flow-head characteristic curves) are used to choose the lowest energy consumptions that meet the target pressure setpoint. The decision variables are the on/off status of the pumps and their speed. The objective is minimization of instantaneous power consumption. The optimization problem incorporates bound constraints on the pumps' speeds, time gap constraints to prevent frequent pump state changes, and physical constraints that place the operation point in the flow-head and flow-efficiency domains. To formulate the constraints, the water demand must be estimated. For the estimation, we use the known operational conditions from the previous time step to estimate the future water demand in the next time step. The model-based controller follows three modules, as detailed below.

Module 1: Demand estimation based on the current system state and all pumps' known flow-head characteristic curves. Figure 1 provides a schematic demonstration of this process, which, algebraically, involves solving a nonlinear equation.

The figure shows two flow-head characteristic curves for two active pumps (green pump and blue pump) at time step t. These are the active pumps (i.e., ON pumps) out of a set of available pumps in the station. The orange curve represents their combined curve, which is the flow-head relationship of the pumping station at time t. Using the pressure measurement at time t, one can estimate the water demand (i.e., flow) at time t using Eq. (1).

$$Q_{est,t} = \sum_{p\in P} I_{p,t-1} \cdot \sqrt{\frac{\left(\frac{n_{p,t-1}}{n_{\max}}\right)^2 \cdot a_p - H_{t-1}}{b_p}} \tag{1}$$

where a_p and b_p are the coefficients of the Q-H pump characteristic curve of pump p.

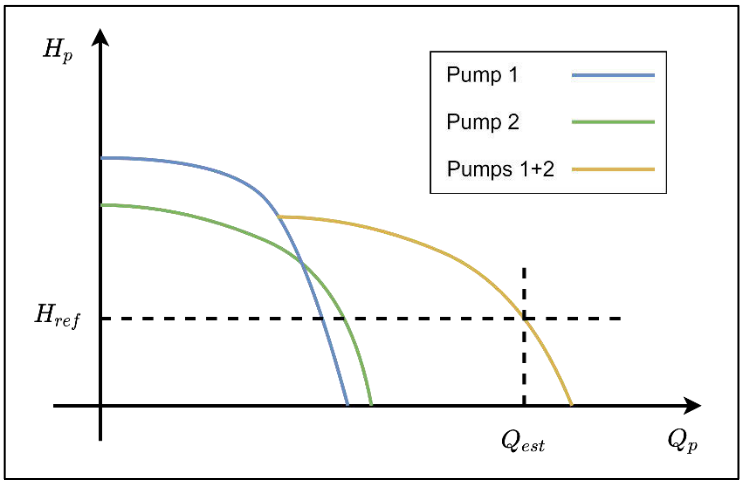


Figure 1: demonstration of flow estimation for two active pumps

Module 2: Determining the feasible speeds for given pumps combinations. Given the estimated flow and a potential pump combination, we seek the set of feasible speeds that reach the operation point of the required pressure setpoint and the estimated flow. While this problem is highly nonlinear, we can rely on a direct search of feasible speeds since the problem dimension is small. That is, we can discretize the allowed speed range and check for the feasibility of each option. Nonetheless, to reach the operation point accurately, we allowed one of the pumps to have continuous speed while the other pumps could have discretized speeds.

Module 3: Direct search for the optimal active pumps combination. In this module, we enumerate all pump combinations in the pumping station. For example, in a pump station with four pumps, there are $2^4 - 1 = 15$ possible combinations of active pumps, C. For each combination module 2 is called, and its output is saved. Two families (i.e., set of sets) are obtained after looping over all the combinations. The first is $N_{feas}^c \quad \forall c \in C$ which contains all the feasible speeds for each combination, and the second is $E_{feas}^c \quad \forall c \in C$, which contains the electric energy. Constraints can be imposed on some of the combinations. For example, some combinations can be ruled out to prevent frequent changes in active pumps. That is, if pump 1 started on time t, we could only consider combinations that have active pump 1 for a predetermined time window $t : t + w$. Similarly, one can filter out combinations with unavailable pumps at any period (e.g., malfunctioned pumps). After ruling out

cc BY NC SA

unwanted combinations, the module outputs the option (i.e., combination and speeds) with the lowest electric power from $c^* \in C$, $n^* \in N^c_{feas}$ and $E^* \in E^c_{feas}$.

The modules above can be easily modified to account for additional operational constraints. To demonstrate this flexibility, we consider two operation strategies. The first is the "free strategy", in which each pump can operate with a different speeds inside a predetermined speed range (presented in modules 1-3). The second is the "equal strategy", in which all active pumps must share the same speed.

3 TEST CASE AND RESULTS

We consider the pressure zone 330P in Mey-Sheva, a water utility in Southern Israel, as a test case. The utilitiy's entire system contains 6 pumping stations, 11 water tanks, and serves a population of 143,000. The total waterpipes' length is 670 km, of which about 100 km are part of the selected pressure zone. The WDS layout of this zone is shown in Figure 2.

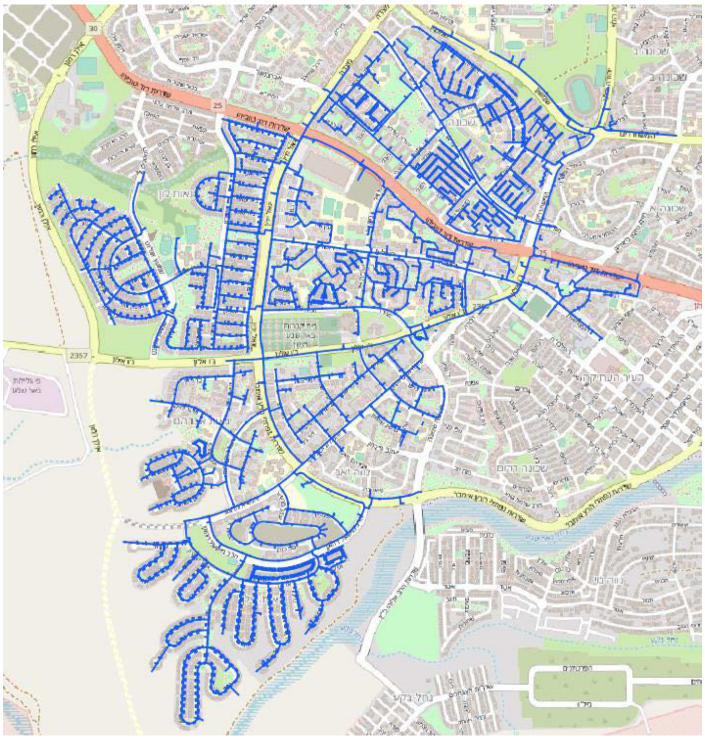


Figure 2: Water Lines- GIS layer for pressure zone 330P

A single pumping station supplies zone 330P without storage tanks. The pumping station has eight pumps, of which four supply water to zone 330P: pumps 1-4 (labeled 11, 21, 31, and 41). The four pumps are variable speed pumps, each operating at a different frequency. SCADA measurements are available at 30-second intervals for the suction and discharge pressures, total flow through the station, and individual pumps frequencies. These frequencies are recorded in percentage [0, 100] for the range of 35-50 Hz. When the value is 0%, it means that the pump may be working at the minimum frequency or it is turned off. We assume the latter. There are no individual pump flow data and no power data, not even for the entire station. There are power meters (SATECs) in the station, but unfortunately, they are not connected to the newly installed SCADA system.

The pumps are operated to maintain the discharge pressure of the station. As demand in the zone increases, this pressure decreases, and the speed of the operating pump is increased to meet the required pressure. First, the speed of one pump is raised to the maximum speed, and then another pump is added at its lowest speed, which can be increased if the pressure continues to drop. The controlled pressure is set to ~47m during day hours (06:30-23:00) and to ~42m during night hours (23:00-0:630), as shown in Figure 3. The pump curves were derived by analyzing the SCADA data of the pump performance, using the methodology in [26].

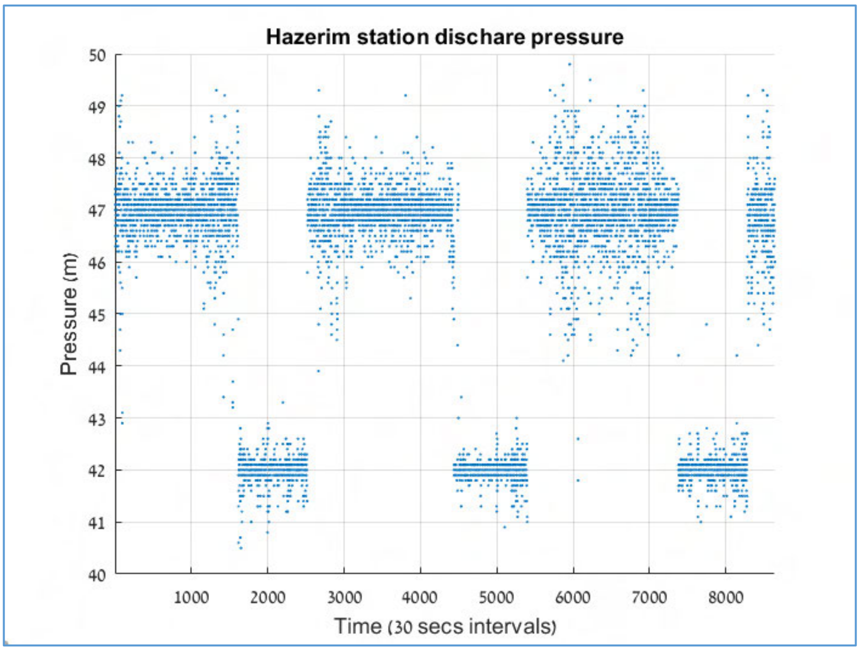


Figure 3: Discharge pressure of Hazerim pumping station

By utilizing the proportional control algorithms presented previously, the current operation of the Hazerim pumping station was simulated. Initial results are shown in Figure 4, in which the simulated pressure is shown in the top figure while the measured pressure is on the bottom. The similarity between the measured and simulated pressure can be easily observed. However, in some time steps, the simulated pressure spikes below the reference pressure, possibly due to the system's (plant) simulations. Further investigation of this issue is required.

The optimization results for two days (out of a representative week) are shown in Figure 5. The results show that the free strategy outperforms the current strategy and the equal speed strategy since it reduced the power consumption in specific periods. This good performance is still achieved while meeting the target pressure, as demonstrated in Figure 6. The figure shows the cumulative probability density function of the three strategies' absolute pressure deviation from the set pressure. The overlapping of the curves indicates that all strategies sustain the required pressure.

BY NC SA

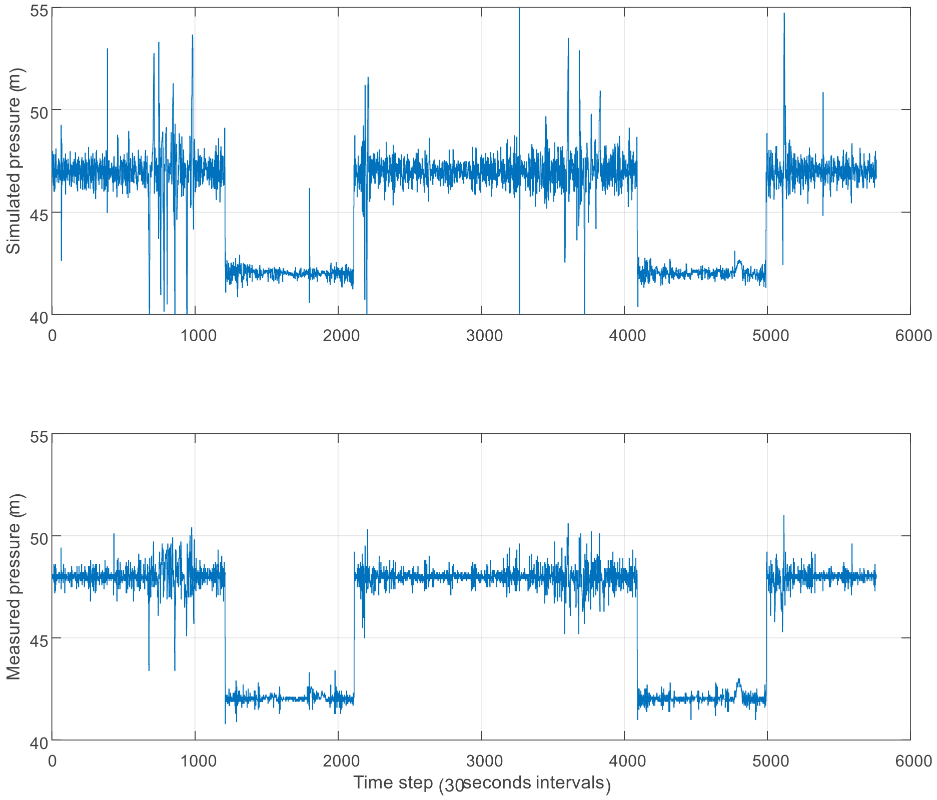


Figure 4: Measured and simulated pressures

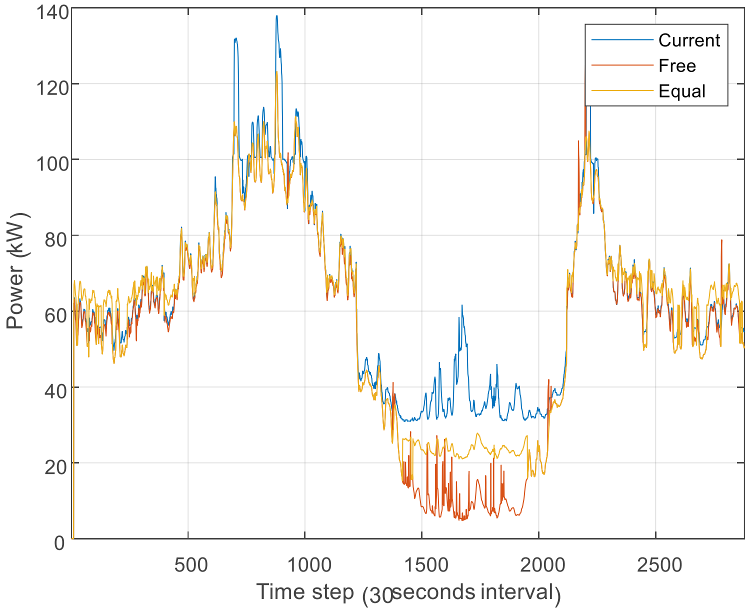


Figure 5: Power consumption for different operational strategies

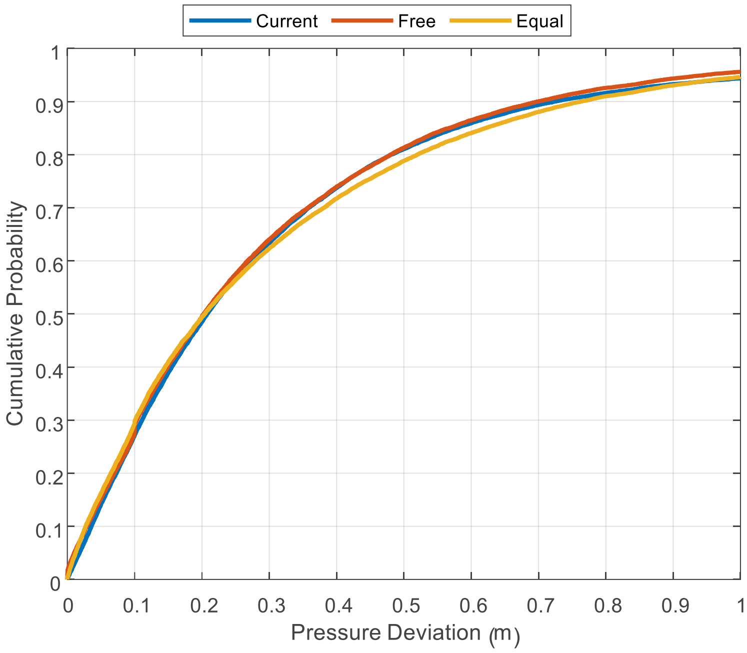


Figure 6: Cumulative probability density function of the absolute pressure deviation from the set pressure for the three strategies

Figure 7 shows that both optimization strategies (free and equal) achieve better balanced operation. Namely, they allocate balanced operation hours for the pumps.

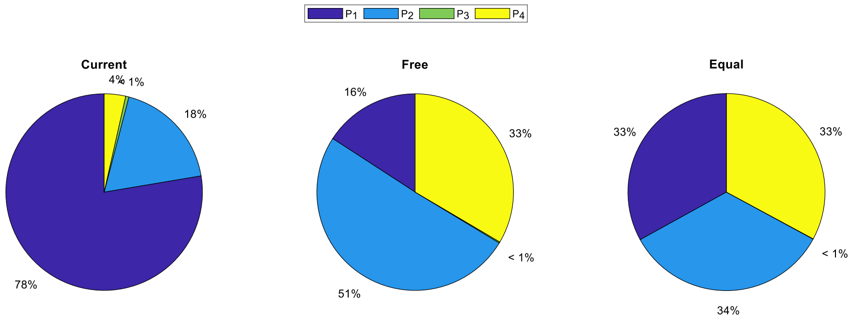


Figure 7: Operation hours for the three strategies

Finally, the energy saving, as compared to the current operation, is 10% for the 'free-strategy' and 5% for the 'equal-strategy'.

4 CONCLUSIONS

This paper presents a new methodology for optimal operation of VSPs. Unlike the traditional approach, in which the physical properties of the pumps are not taken into account, the proposed algorithm uses the pumps' curve to select the optimal pump combinations. The methodology is presented in two variations: a) "free strategy", in which each pump may have a different operating speed, and b) "equal strategy", in which all operating pumps share the same speed. The methodology is demonstrated in a real test case, and the results show a smoother and cost-effective operation compared to the traditional approach. The developed algorithm is planned to be utilized in the pumping station control unit for further testing.

5 REFERENCES

[1] Gottliebson, M.; Sanks, R.L.; Vause, A.; Abelin, S.M.; Benfell, R.S.; Bicknell, H.C.; Densmore, B.; Duncan, R.W.; Hong, S.S.; Jones, C.; et al. Variable-Speed Pumping. In Pumping Station Design; Elsevier Ltd, 2008 ISBN 9781856175135.

[2] Lansey, K.; Awumah, K. Optimal Pump Operations Considering Pump Switches. J. Water Resour. Plan. Manag. 1994, 120, 17–35.

[3] Rossman, L.A. EPANET2 user's manual; U.S. Environmental Protection Agency, Cincinnati, 2000;

[4] Marchi, A.; Simpson, A.R. Correction of the EPANET Inaccuracy in Computing the Efficiency of Variable Speed Pumps. J. Water Resour. Plan. Manag. 2013, 139, 456–459.

[5] Ulanicki, B.; Kahler, J.; Coulbeck, B. Modeling the efficiency and power characteristics of a pump group. J. Water Resour. Plan. Manag. 2008, 134, 88–93.

[6] Marchi, A.; Simpson, A.R.; Ertugrul, N. Assessing variable speed pump efficiency in water distribution systems. Drink. Water Eng. Sci. 2012, 5, 15–21.

[7] AbdelMeguid, H.; Ulanicki, B. Feedback Rules for Operation of Pumps in a Water Supply System Considering Electricity Tariffs. Water Distrib. Syst. Anal. 2010 2018.

[8] Coulbeck, B.; Orr, C.H. Real-Time Optimized Control of a Water Distribution System. IFAC Proc. Vol. 1989, 22, 441–446.

[9] Abdallah, M.; Kapelan, Z. Fast Pump Scheduling Method for Optimum Energy Cost and Water Quality in Water Distribution Networks with Fixed and Variable Speed Pumps. J. Water Resour. Plan. Manag. 2019, 145.

[10] Hashemi, S.S.; Tabesh, M.; Ataeekia, B. Ant-colony optimization of pumping schedule to minimize the energy cost using variable-speed pumps in water distribution networks. Urban Water J. 2014, 11, 335–347.

[11] Bene, J.G.; Hos, C.J. Finding least-cost pump schedules for reservoir filling with a variable speed pump. J. Water Resour. Plan. Manag. 2012, 138, 682–686.

[12] Wu, W.; Maier, H.R.; Dandy, G.C.; Arora, M.; Castelletti, A. The changing nature of the water–energy nexus in urban water supply systems: a critical review of changes and responses. J. Water Clim. Chang. 2020.

[13] Lima, G.M.; Luvizotto, E.; Brentan, B.M.; Ramos, H.M. Leakage control and energy recovery using variable speed pumps as turbines. J. Water Resour. Plan. Manag. 2018, 144, 04017077.

[14] Wu, W.; Simpson, A.R.; Maier, H.R.; Marchi, A. Incorporation of variable-speed pumping in multiobjective genetic algorithm optimization of the design of water transmission systems. J. Water Resour. Plan. Manag. 2012, 138, 543–552.

[15] Huo, J.; Eckmann, D.H.; Hoskins, K.P. Hydraulic simulation and variable speed pump selection of a dual-function pumping station: Pumping treated water to a reclaimed water system or deep injection well system. In Proceedings of the Restoring Our Natural Habitat - Proceedings of the 2007 World Environmental and Water Resources Congress; American Society of Civil Engineers: Reston, VA, 2007; pp. 1–7.

[16] Li, G.; Baggett, C.C. Real-time operation optimization of variable-speed pumping stations in water distribution systems by adaptive discharge pressure control. In Proceedings of the Restoring Our Natural Habitat - Proceedings of the 2007 World Environmental and Water Resources Congress; American Society of Civil Engineers (ASCE), 2007; pp. 1–13.

[17] Page, P.R.; Abu-Mahfouz, A.M.; Yoyo, S. Real-time Adjustment of Pressure to Demand in Water Distribution Systems: Parameter-less P-controller Algorithm. In Proceedings of the Procedia Engineering; Elsevier Ltd, 2016; Vol. 154, pp. 391–397.

[18] Creaco, E.; Franchini, M. A new algorithm for real-time pressure control in water distribution networks. Water Sci. Technol. Water Supply 2013, 13, 875–882.

[19] Campisano, A.; Modica, C.; Vetrano, L. Calibration of Proportional Controllers for the RTC of Pressures to Reduce Leakage in Water Distribution Networks. J. Water Resour. Plan. Manag. 2012, 138, 377–384.

[20] Wu, Z.Y.; Tryby, M.; Todini, E.; Walski, T. Modeling variable-speed pump operations for target hydraulic characteristics. J. / Am. Water Work. Assoc. 2009, 101, 54–64.

[21] Walski, T.; Creaco, E. Selection of Pumping Configuration for Closed Water Distribution Systems. J. Water Resour. Plan. Manag. 2016, 142, 04016009.

[22] Vicente, D.J.; Garrote, L.; Sánchez, R.; Santillán, D. Pressure Management in Water Distribution Systems: Current Status, Proposals, and Future Trends. J. Water Resour. Plan. Manag. 2016, 142, 04015061.

[23] Page, P.R.; Abu-Mahfouz, A.M.; Mothetha, M.L. Pressure Management of Water Distribution Systems via the Remote Real-Time Control of Variable Speed Pumps. J. Water Resour. Plan. Manag. 2017, 143, 04017045.

[24] da Costa Bortoni, E.; Almeida, R.A.; Viana, A.N.C. Optimization of parallel variable-speed-driven centrifugal pumps operation. Energy Effic. 2008, 1, 167–173.

[25] Ang, K.H.; Chong, G.; Li, Y. PID control system analysis, design, and technology. IEEE Trans. Control Syst. Technol. 2005, 13, 559–576.

[26] Salomons, E.; Shamir, U.; Housh, M. Optimization methodology for estimating pump curves using SCADA data. Water (Switzerland) 2021, 13, 586.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14040

EXPERIMENTAL STUDY OF MIXING PHENOMENON IN WATER DISTRIBUTION NETWORKS UNDER REAL-WORLD CONDITIONS

Reza Yousefian[1] **and Sophie Duchesne**[2]

[1] Ph.D. Student, Eau Terre Environnement Research Centre, Institut National de la Recherche Scientifique (INRS), Quebec City, Quebec, (Canada)
[2]Professor, Eau Terre Environnement Research Centre, Institut National de la Recherche Scientifique (INRS), Quebec City, QC, (Canada)
[1] *Reza.Yousefian@inrs.ca* , [2] *Sophie.Duchesne@inrs.ca*

Abstract

Various studies have shown that mixing at the junction of water distribution networks (WDNs) can be neither complete nor instantaneous. Many studies have also been carried out to find the parameters that have a significant effect on the mixing phenomenon by experimental and numerical investigations. The Reynolds ratio of inlet flows, the Reynold ratio of outlet flows, the pipe size of junction's legs (junction configurations), and the type of junction (cross or double-t) are among the most important factors mentioned in these studies. Other studies also focused on developing mixing models. However, these studies and models were based on experimental conditions in which the pipe size was about 25 mm, and the pressure was about 30 kPa. Despite the fact that these models provide acceptable results on the laboratory scale and all of them have been validated based on laboratory results, many researchers still acknowledge that studies on the effects of real conditions of urban WDNs on the mixing phenomenon are lacking [1], [2] and [3]. The pipe size of junctions and the pressure inside the network are distinguishing features of real urban WDNs and laboratory setups that were used in previous studies. Therefore, this research investigates the effect of pipe size and pressure on the mixing phenomenon in a cross junction with the same pipe sizes in all four legs under real-world conditions. Flow rates and pressures were selected based on a statistical study conducted on the Quebec City WDN and previous research work, in which the pressures were 5, 140, 320 and 430 kPa, and the flow rates were a combination of 1.50, 2.00, 2.25, 2.50 and 3.00 LPS in each inlet and outlet pipes. In other words, the Reynolds number in the experiments was between 21,000 and 43,000, and the Reynolds ratio of inlet flows as well as the Reynolds ratio of outlet flows were 0.5, 0.8, 1, 1.25 and 2. In this network, two cross junctions with the same pipe size in each junction's legs of 100 and 150 mm were investigated. Salt was used as a traceable solute, and conductivity meters were used to measure the salt concentration in pipes. In this series of experiments, only one inlet had salty water. The mixing was characterized by the dimensionless concentration in outlets, defined as the observed conductivity in one of the outlet flow divided by the inlet salty water conductivity (after subtracting the conductivity of tap water for both measurements). The results showed that the pressure had only a little effect on mixing, since the dimensionless concentration changed by about 0.05 when considering the low pressure of 5 kPa as compared to the experiments with the other pressure values, while this variation dropped to about 0.02 for all pressures above 140 kPa. Between 100 and 150 mm, the pipe size modified the dimensionless concentration by about 0.035. Taking into account the uncertainty of the experiments, it can be concluded that pressure and pipe diameters (100 and 150 mm) have an insignificant impact on the mixing phenomenon.

Keywords
Water Distribution Networks, Water Quality, Mixing Phenomenon, Numerical Modelling, Junction, Pressure, Pipe Size.

1 INTRODUCTION

Water quality modeling in water distribution systems (WDSs) is one of the main tools to help control and enhance drinking water quality. In the existing commercial software of WDSs' water quality modeling, it is assumed that mixing in junctions occurs completely and instantaneously; however, researchers found that not only is this assumption wrong but also, in many cases, the mixing is totally imperfect in junctions [4], [5] and [6]. Therefore, two series of research were conducted: the first, to find out the effective parameters of the mixing phenomenon and the second, to develop models considering imperfect mixing at junctions.

In the first category of research, O'Hern et al. (2005) studied the mixing phenomenon in cross junctions with the help of a physical model in the laboratory [2]. In their experiments, the Reynolds number (Re) was about 43,000, and two circular 50 and 12.5 mm pipes were used. They observed that the mixing is less complete when the pipe dimension increases. The mixing phenomenon was also investigated in a 3x3 pipes grid network with pipe sizes of 12.5 to 50 mm and a Reynolds number of 40,000 [7]. Ho et al. (2006) realized that the mixing in cross junctions decreases slightly as the velocity increases for a fixed pipe diameter [7]. Ho et al. (2007) used the K-Epsilon turbulent model for their numerical simulations and figured out that the Schmidt Number is the most important factor in RANS turbulent models [8]. Besides, Romero-Gomez et al. (2008) used the value of Schmidt number suggested by Ho et al. (2007) to perform a numerical simulation of mixing with unequal inlet and outlet flow rates [3]. Meanwhile, Webb and van Bloemen Waanders (2006) stated that the LES model is not dependent on the Schmidt number, while they as well believed that the mixing is caused by both bulk flow (advection) and turbulent diffusion at the impinging interface [9]. McKenna et al. (2007) worked on the ratio of Re for inlets in the laboratory with pipe sizes of 12.5, 25, 32, and 50 mm [10]. In addition to considering the turbulent flow, McKenna et al. (2008) used high-speed photography to study the mixing phenomenon for laminar and transient flows ($500 < Re < 5,000$) [11]. van Bloemen Waanders et al. (2005) also considered the mixing in laminar flow and found that the mixing in laminar flow is not complete [6]. Thereafter, Braun et al. (2014) statistically investigated the WDN of Strasbourg, France, and found that laminar flow frequently happens in large areas of the network [12]. Accordingly, the mixing phenomenon in cross junctions was considered under the laminar and transitional conditions ($500 < Re < 7,500$) with pipe sizes of 16, 25, 32, and 50 mm by Shao et al. (2019). Shao et al. (2019) figured out that the pipe diameter has a great influence on the mixing [13].

In the second category of studies, there are two types of models for simulating the mixing phenomenon: empirical and mechanistic. AZRED is the first empirical model that could project the solute concentration based on the inlet concentrations and the Re ratios of inlets and outlets [14]. Austin et al. (2008) carried out experiments in the Re range of 10,000 to 42,000 to develop their model, while they used extrapolation for Re greater than 42,000 [14]. Ho (2008) proposed the model EPANET-BAM as a mechanistic model to study the mixing in cross junctions [15]. This author focused on the inlet and outlet Re ratio by changing the flow rate and only considered the advection term of the transport formula [15]. Yu et al. (2014) experimentally and numerically studied the mixing phenomenon within cross junctions with different pipe sizes [16] and [17]. Shao et al. (2014) proposed a model for the mixing phenomenon in cross junctions based on experimental findings with two configurations: opposing inlets and adjacent inlets [18]. More details of the effective parameters and models were presented in the paper by Yousefian and Duchesne (2022) reviewing the research about the mixing phenomenon in WDSs [1].

As mentioned in the previous paragraphs, most studies on the mixing phenomenon were carried out under the laboratory scale and conditions, while the developed models should be applied to real-world conditions. Besides the fact that the pipe diameter is one of the effective parameters that greatly impact the mixing phenomenon, pipe diameter and pressure are two distinguished differences between laboratory and real-world conditions in WDSs. Therefore, in this research, two pipe diameters of 100 and 150 mm, which are the most used pipes in Quebec City WDN, along with pressures of 5 and140 kPa (used to develop the existing models), and 320 and 430 kPa (encountered in real WDNs) were selected to investigate the effect of real-world pipe diameter and pressure on mixing phenomenon.

2 EXPERIMENTAL SETUP AND PREPARATION

The experiments were carried out in the Mini Water Distribution Network Laboratory of Institut National de la Recherche Scientifique (INRS) in Quebec City, Canada. This laboratory has been built in a hall with dimensions of 15 m x 9 m to hold a 12x5 pipes grid network. All connections in this network are flanged in order to make changing the pipe diameters easy. The network was equipped with two pumps which are a 3 hp (Xylem-AquaBoost) and a 75 hp (Berkeley-B4EPBMS) pumps, and also several pressure probes (Ashcroft-G2) to be able to apply a variety of high and low pressures. In addition, 12 flow control valves with electric actuators (Assured Automation-P2R4 with S4 actuator) and several electromagnetic flow meters (ModMAG-M2000) were installed in this network so that any specific flow can be adjusted in the pipes (Figure 1). In this study, sodium chloride (NaCl) was used as a soluble tracer for the experiments, and 4 conductivity meters (Teledyne-LXT220) were installed in each leg of the cross junction to measure the conductivity (related to the concentration of salts) in each inlet and outlet. All the settings of pumps and valves were set through a central computer (Honeywell-EBI R430 and Honeywell-Controller HC900) to increase the accuracy of applied conditions in experiments. The flow, pressure and conductivity results were also measured and recorded every 5 s by a data logger to the central computer. These data were averaged over a 60 s interval to collect repeatable data by reducing the signal noise of the sensors.

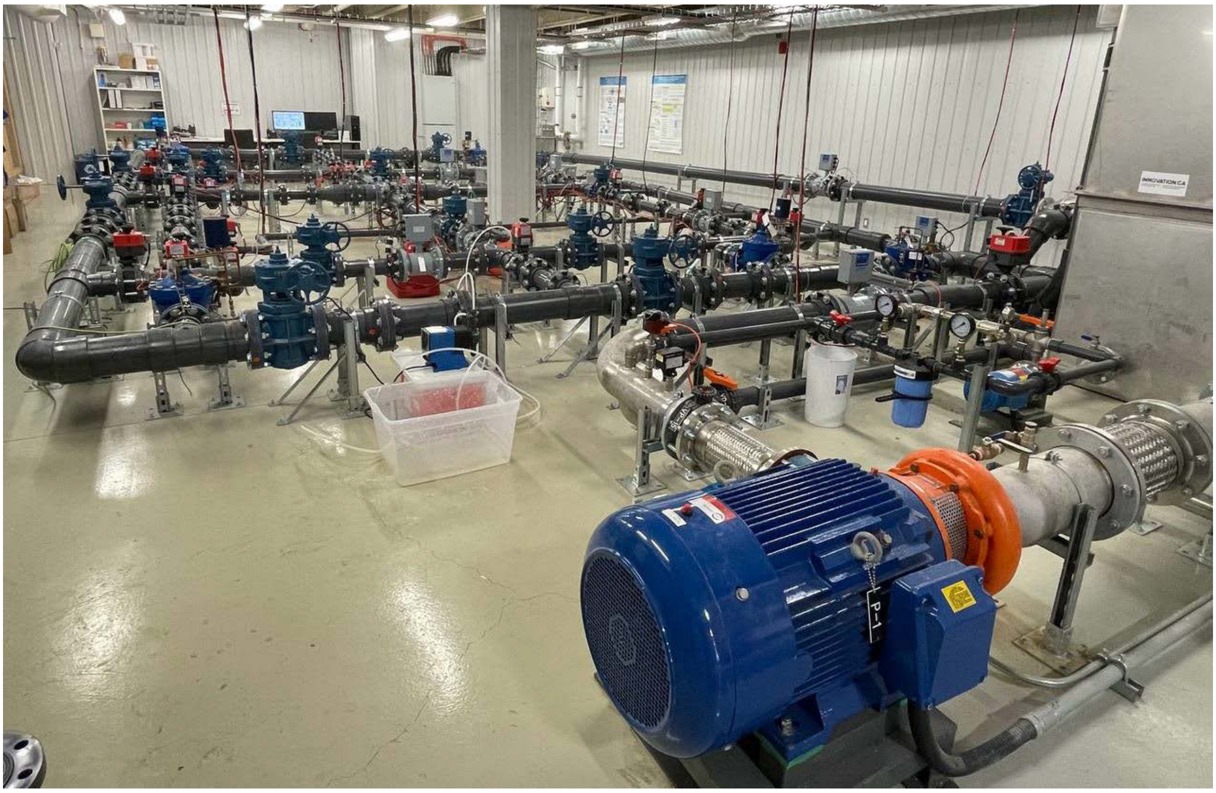


Figure 1. Mini water distribution network Laboratory of INRS

In order to add salt to the flow and network, an injection pump was employed. This injection pump can work under different frequencies, making it possible to have different salt concentrations.

Two cross junctions with the pipe size of 100 and 150 mm were considered. These two cross junctions have the same pipe size in all of their four legs, as shown in Figure 2 for the 150 mm configuration.

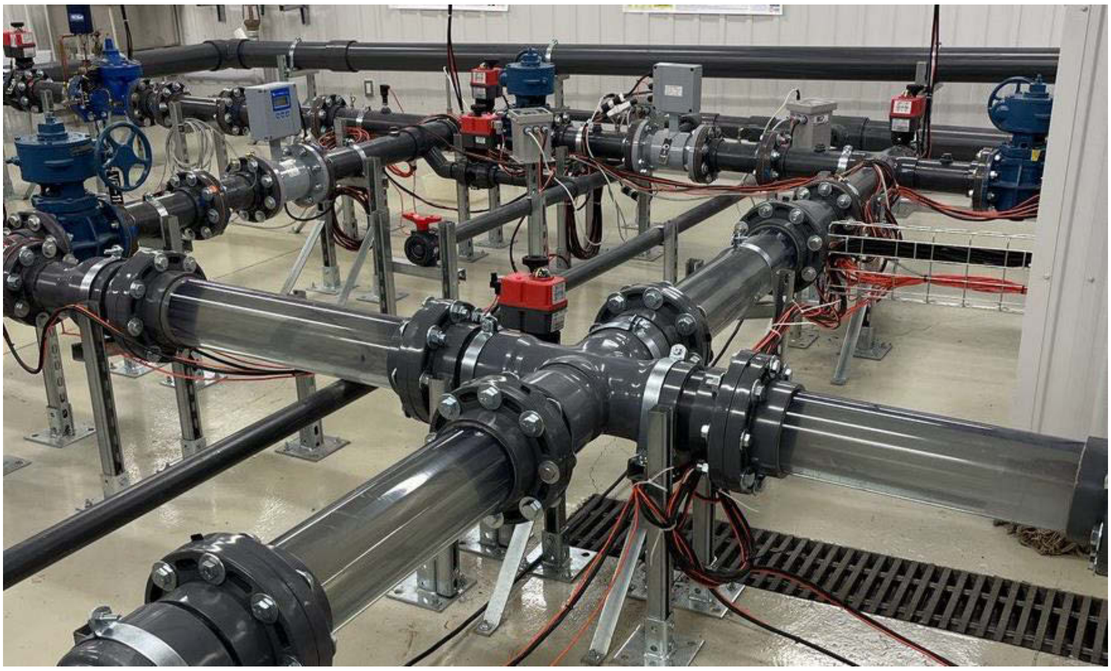


Figure 2. 150-mm cross junction used in mini water distribution network laboratory of INRS

3 DEFINITIONS AND SCENARIOS

In this study, tap water was used in the whole network and salt (NaCl) was injected. The distance between the salt injection point and the cross junction was about 3 m to ensure that the longitudinal mixing occurred entirely in the water before reaching the cross junction. In all the experiments, the salt was injected into the southern pipe (red arrow in Figure 3), tap water was coming from the western pipe (blue arrow in Figure 3), and northern and eastern pipes were the outlets (orange arrows in Figure 3).

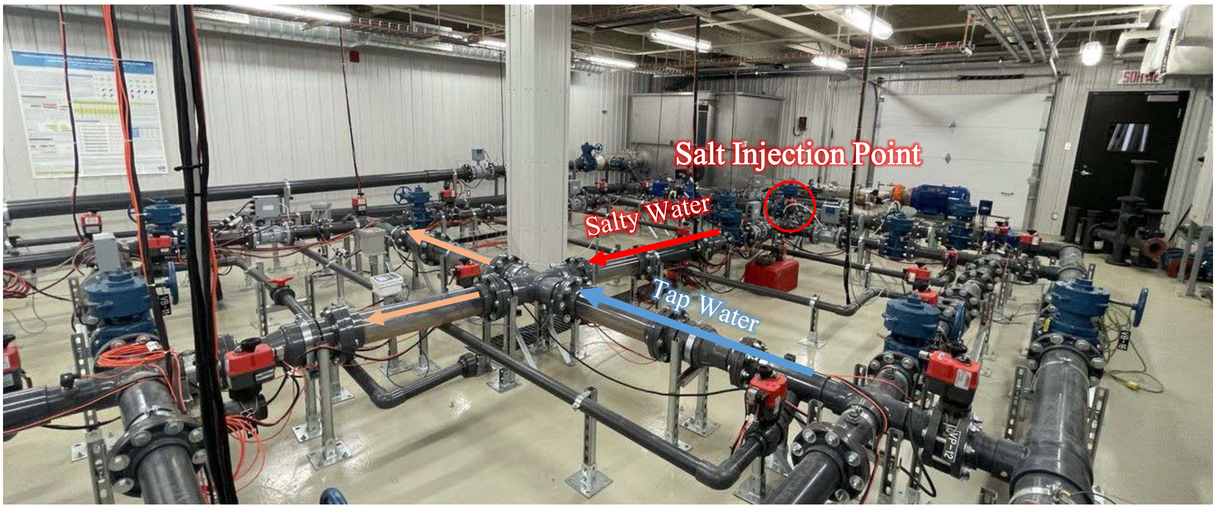


Figure 3. Schematic picture of the experiments to obtain dimensionless concentrations

A statistical analysis was carried out in Quebec City WDN to find the most common real characteristics and conditions of WDNs. In this network, the two pipe sizes of 100 and 150 mm are the most often used pipe size, and flow rates between 1.50 and 3.00 LPS are the most common. Therefore, based on this result, 25 different flow rate scenarios were tested for each cross junction size (Table 1). In each scenario, two pressures, of 320 and 430 kPa, were applied. Ten different concentrations of salt were injected into the southern pipe. In total, 500 experiments were carried out in this phase to find the effect of pressure and pipe diameter on the mixing phenomenon in junctions.

Table 1. Flow rate scenarios for the experiments to consider the effect of pressure and pipe diameter on mixing

South inlet (Salty Water) (LPS)	West Inlet (Tap Water) (LPS)	North Outlet (LPS)	East Outlet (LPS)
1.50	3.00	3.00	1.50
1.50	3.00	2.50	2.00
1.50	3.00	2.25	2.25
1.50	3.00	2.00	2.50
1.50	3.00	1.50	3.00
2.00	2.50	3.00	1.50
2.00	2.50	2.50	2.00
2.00	2.50	2.25	2.25
2.00	2.50	2.00	2.50
2.00	2.50	1.50	3.00
2.25	2.25	3.00	1.50
2.25	2.25	2.50	2.00
2.25	2.25	2.25	2.25
2.25	2.25	2.00	2.50
2.25	2.25	1.50	3.00
2.50	2.00	3.00	1.50
2.50	2.00	2.50	2.00
2.50	2.00	2.25	2.25
2.50	2.00	2.00	2.50
2.50	2.00	1.50	3.00
3.00	1.50	3.00	1.50
3.00	1.50	2.50	2.00
3.00	1.50	2.25	2.25
3.00	1.50	2.00	2.50
3.00	1.50	1.50	3.00

The dimensionless concentration of injected salt in each outlet, which is obtained from equation (1), was chosen to present the level of mixing:

$$C^* = \frac{C - C_w}{C_s - C_w} \tag{1}$$

where C^* is dimensionless concentration of injected salt in the studied outlet (north or east), C is the concentration of salt in the studied outlet, and C_s and C_w are concentrations of salt in the southern (salty water) and western (tap water) inlets, respectively. Finally, since for each flow scenario, 10 different quantities of salt were injected, the dimensionless concentration of injected salt for each scenario was obtained through taking the average of dimensionless concentration for 10 quantities of injected salt:

$$C^* = \frac{\sum_{i=1}^{10} C_i^*}{10} \tag{2}$$

where C_i^* is the dimensionless concentration in any outlet for the i^{th} of 10 experiments carried out in each flow rate scenario.

3.1 Uncertainty of dimensionless concentration in outlets

The accuracy of conductivity meters is ±0.1% of full scale. Based on our calibration data, we selected a conductivity range of [0, 1.41] mS/cm. Therefore, ±0.1% of the maximal conductivity (1.41 ms/cm) would be ±0.00141 ms/cm. Then, to find the dimensionless concentration of salt in each outlet, the following equations of absolute error propagation were used.

For addition and subtraction ($Z = X + Y$ or $Z = X - Y$):

$$\Delta Z = |\Delta X| + |\Delta Y| \tag{3}$$

and for multiplication and division ($Z = X \times Y$ or $Z = X/Y$):

$$\frac{\Delta Z}{Z} = |\frac{\Delta X}{X}| + |\frac{\Delta Y}{Y}| \tag{4}$$

where X and Y are measurements with absolute uncertainty of $|\Delta X|$ and $|\Delta Y|$, respectively. Finally, the uncertainty of the dimensionless concentration of injected salt in each outlet was estimated to be about $|\pm 0.05|$.

4 RESULTS

4.2 Pressure impact on mixing phenomenon

To study the effect of pressure on the mixing phenomenon, 25 different flow scenarios (see Table 1) within two pipe sizes cross junctions were studied. Detailed results will be presented here for three of these cases. In the first case, the simplest flow scenario, i.e. with equal flow rate (2.25 LPS) in all cross junction legs, was investigated. In this case, the cross junction with pipe diameters of 100 mm was used, and three different pressures were applied: 5, 140, and 320 kPa. Different flow rates in inlets and outlets were then considered, in Case 2, to study the effect of pressure within a more complicated flow rate scenario. The same cross junction and pressures as in Case 1 were used for this experiment. In the last experiment, Case 3, the same flow scenario as in Case 2 was applied but this time in a 150-mm cross junction. For this experiment, four pressures were considered. More details about each case are presented in Table 2 and the results of these 3 cases are shown in Figure 4 and 5.

BY NC SA

Table 2. Specifications of some of the experiments conducted to study the pressure impact

		South	West	North	East
Case 1	Pipe Diameter	100 mm	100 mm	100 mm	100 mm
	Flow Rate	2.25 LPS	2.25 LPS	2.25 LPS	2.25 LPS
Case 2	Pipe Diameter	100 mm	100 mm	100 mm	100 mm
	Flow Rate	1.50 LPS	3.00 LPS	1.50 LPS	3.00 LPS
Case 3	Pipe Diameter	150 mm	150 mm	150 mm	150 mm
	Flow Rate	1.50 LPS	3.00 LPS	1.50 LPS	3.00 LPS

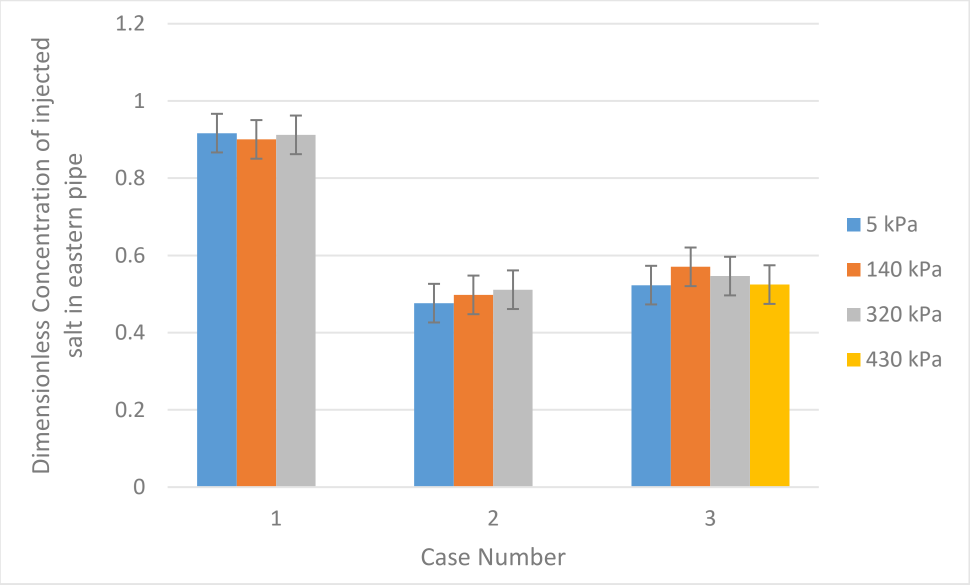


Figure 4. Dimensionless concentration of injected salt in the eastern pipe for the experiments conducted to study the pressure impact

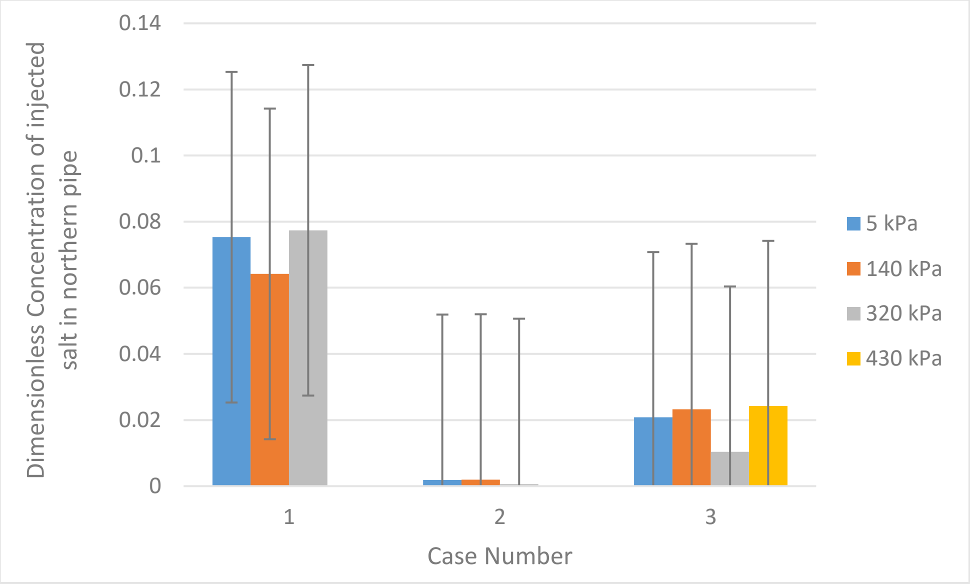


Figure 5. Dimensionless concentration of injected salt in the northern pipe for the experiments conducted to study the pressure impact

In general, pressure can affect the intensity of flow turbulence and molecular diffusion. However, as shown in Figure 4, when the flow rates or Reynolds numbers are equal in all cross junction legs (Case 1), the effect of pressure results in a maximal difference of about 0.016 in dimensionless concentrations in the eastern pipe, which happens between 5 and 140 kPa. In comparison, this difference decreases to about 0.011 in dimensionless concentration in the eastern pipe for the pressures of 140 and 320 kPa. When considering the uncertainty (shown by error bars in each figure), it can be concluded that for this flow scenario (Case 1), the pressure has a negligible impact on mixing. For Case 2, with different flow rates or Reynolds numbers in the four cross junction legs, the difference between the dimensionless concentration of eastern pipe with 5 and 320 kPa is about 0.035. However, this difference decreases to 0.013 when the pressures of 140 and 320 kPa are considered. For the last experiment, Case 3, also, the difference in dimensionless concentration in the eastern pipe is about 0.047 between 5 and 140 kPa. However, when pressures greater than 140 kPa are compared to each other, the difference in dimensionless concentration in the eastern pipe is about 0.024. Therefore, it can be concluded that when the flow rates are different in the cross junction legs, the pressure can change the dimensionless concentration up to 0.047 in the eastern pipe when a low pressure like 5 kPa (which was previously used to develop the existing empirical models) is considered; however, this difference is lower than the uncertainty related to the measurements. For pressures greater than 140 kPa, the pressure can change the dimensionless concentration in the eastern pipe by about 0.02. Since such low pressures do not usually happen in real WDNs and since for pressures greater than 140 kPa the impact of pressure in comparison with the uncertainty of experiments is insignificant, it can be concluded that the impact of pressure on mixing phenomenon can be neglected. It should be added that this conclusion was also obtained from the results of experiments carried out under the 25 flow scenarios (see Table 1) and the pressures of 320 and 430 kPa in both 100 and 150 mm cross junctions.

4.3 Pipe diameter impact on mixing phenomenon

To find the impact of pipe diameter on mixing in cross junctions, all 25 flow scenarios (see Table 1) were tested in both cross junctions (100 and 150 mm); the results of three of those scenarios are presented here as examples. The following table (Table 5) shows the specifications of these scenarios, while the results are shown in Figures 7 and 8.

Table 3. Specifications of the experiments to study the pipe diameter impact

	South	West	North	East
Case 4-Flow Rates	2.25 LPS	2.25 LPS	2.25 LPS	2.25 LPS
Case 5-Flow Rates	3.00 LPS	1.50 LPS	3.00 LPS	1.50 LPS
Case 6-Flow Rates	1.50 LPS	3.00 LPS	1.50 LPS	3.00 LPS

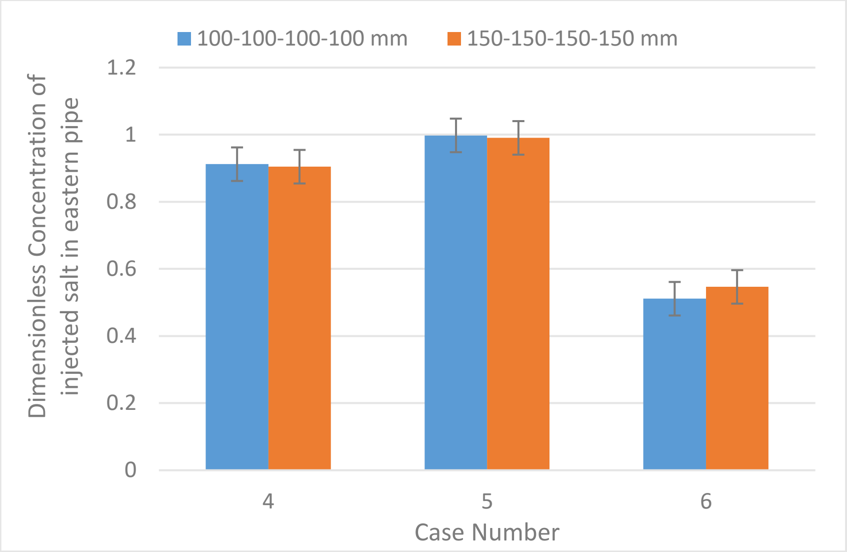


Figure 6. Dimensionless concentration of injected salt in the eastern pipe for the experiments conducted to study the pipe diameter impact

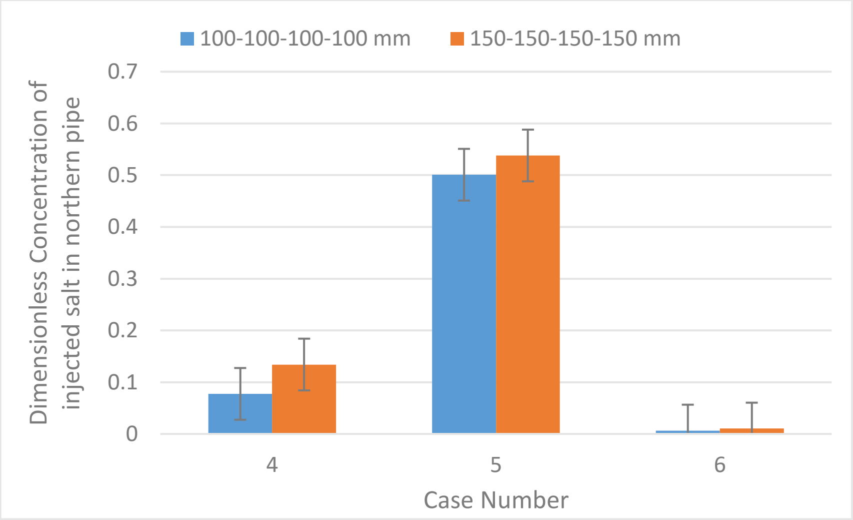


Figure 7. Dimensionless concentration of injected salt in the northern pipe for the experiments conducted to study the pipe diameter impact

As shown in Figure 6, the difference in dimensionless concentration in the eastern pipe for the two pipe diameters, when the flow rates are equal in all four legs of the cross junction (Case 4), is less than 0.01. This difference was also observed in the Case 5, where the flow rates were different in the cross junction legs. However, for the last case (Case 6), in which the flow rate in the inlet

cc BY NC SA

with salty water was less than the flow rate in the inlet with tap water, this difference increases to about 0.035. Considering the uncertainty, which is shown by error bars in all charts, it can be concluded that when the flow rates are equal in all four cross junction legs (Case 4) or when the flow rate of salty water is different from the one of tap water (Case 5 and Case 6), the mixing is the same in the cross junctions with 100 and 150 mm diameter, for the conditions that were simulated in the laboratory. This was also observed for all other 22 scenarios in both 100 and 150 mm cross junctions.

5 CONCLUSIONS

In this research, the effects of pressure and pipe size on mixing phenomenon in the cross junctions of WDNs were studied. For this purpose, two cross junctions with 100 and 150 mm pipe diameters were used. For each size, 25 flow rate scenarios were tested under two different pressures in the mini WDN laboratory of INRS, Canada. In all experiments, the salty water was injected into the southern inlet and tap water was coming from western inlet and two northern and eastern pipes were outlets. Each experiment was repeated ten times with varying concentrations of salt. By comparing the results of these 500 experiments, the following conclusions were acquired:

1. When the flow rates are the same in all four legs of cross junctions, the pressure does not impact the mixing phenomenon.
2. When the flow rates are different in the legs of the cross junction, the pressure can change the mixing or dimensionless concentration in the eastern pipe by about 0.05, which is of the same order as the uncertainty related to measurements, if low pressures like 5 kPa, which are not encountered in real word conditions, are considered.
3. For pressures greater than 140 kPa, the effect of pressure can change the dimensionless concentration of injected salt in the eastern pipe by about 0.02, which considered as non-significant since the uncertainty related to measurements is about 0.05.
4. Since low pressure like 5 kPa do not happen in real WDNs and since the effect of high pressure is lower than the uncertainty of the experiments, it is concluded that in real WDNs, the pressure does not have a significant impact on the mixing phenomenon and can be neglected.
5. The impact of pipe diameter (100 and 150 mm) on mixing within different flow rate scenarios (Table 1) can change the dimensionless concentration by 0.035, which is considered to be negligible since it is lower than the uncertainty of the laboratory measurements,

Since in real WDNs, the pressures are mostly above 140 kPa and since the effect of pipe size for 100 and 150 mm on the dimensionless concentration in the eastern pipe in our experiments was found to be less than the uncertainty related to measurements (±0.05), it leads to the conclusion that in real WDNs, the mixing for cross junctions with four equal pipe sizes is the same for different pressures and for pipe sizes of 100 and 150 mm. However, cross junctions with different pipe sizes (e.g. 100-150-100-150 mm) still needs to be studied in the future.

6 REFERENCES

[1] R. Yousefian and S. Duchesne, "Modeling the Mixing Phenomenon in Water Distribution Networks: A State-of-the-Art Review," Journal of Water Resources Planning and Management, vol. 148, no. 2, 2022, pp. 03121002.

[2] L. Orear, G. Hammond, S.A. McKenna, P. Molina, R. Johnson, T. O'Hern, and B. van Bloemen Waanders, "Physical Modeling of Scaled Water Distribution System Networks." USDOE (U.S. Department of Energy

Office of Scientific and Technical Information): Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2005, pp. 1-37.

[3] P. Romero-Gomez, C.K. Ho, and C.Y. Choi, "Mixing at cross junctions in water distribution systems. I: Numerical study," Journal of Water Resources Planning and Management, vol. 134, no. 3, 2008, pp. 285-294.

[4] A.G. Fowler and P. Jones, "Simulation of water quality in water distribution systems," in Proc. Water Quality Modeling in Distribution Systems. Cincinnati, Ohio, United States: AwwaRF/EPA, 1991, pp. 4-5.

[5] N. Ashgriz, W. Brocklehurst, and D. Talley, "Mixing mechanisms in a pair of impinging jets," Journal of propulsion and power, vol. 17, no. 3, 2001, pp. 736-749.

[6] B. van Bloemen Waanders, G. Hammond, J. Shadid, S. Collis, and R. Murray, "A comparison of Navier Stokes and network models to predict chemical transport in municipal water distribution systems," in World Water and Environmental Resources Congress. Anchorage, Alaska, United States: American Society of Civil Engineers, 2005, pp. 1-10.

[7] C.K. Ho, L. O'Rear Jr, J.L. Wright, and S.A. McKenna. "Contaminant mixing at pipe joints: Comparison between laboratory flow experiments and computational fluid dynamics models." in Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, United States: American Society of Civil Engineers, 2006. pp. 1-18.

[8] C.K. Ho, C.Y. Choi, and S.A. McKenna. "Evaluation of complete and incomplete mixing models in water distribution pipe network simulations." in World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, Tampa, Florida, United States: American Society of Civil Engineers, 2007. pp. 1-12.

[9] S.W. Webb and B.G. van Bloemen Waanders. "High fidelity computational fluid dynamics for mixing in water distribution systems." in Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, United States: American Society of Civil Engineers, 2006. pp. 1-15.

[10] S.A. McKenna, L. Orear, and J. Wright. "Experimental determination of solute mixing in pipe joints." in World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, Tampa, Florida, United States: American Society of Civil Engineers, 2007. pp. 1-11.

[11] S.A. McKenna, T. O'Hern, and J. Hartenberger. "Detailed Investigation of Solute Mixing in Pipe Joints Through High Speed Photography." in Water Distribution Systems Analysis 2008, Kruger National Park, South Africa: American Society of Civil Engineers, 2008. pp. 1-12.

[12] M. Braun, T. Bernard, H. Ung, O. Piller, and D. Gilbert, "Model based investigation of transport phenomena in water distribution networks for contamination scenarios," Procedia Engineering, vol. 70, no. 2014, pp. 191-200.

[13] Y. Shao, L. Zhao, Y.J. Yang, T. Zhang, and M. Ye, "Experimentally Determined Solute Mixing under Laminar and Transitional Flows at Junctions in Water Distribution Systems," Advances in Civil Engineering, vol. 2019, no. 2019, pp. 1-10.

[14] R. Austin, B. van Bloemen Waanders, S. McKenna, and C.Y. Choi, "Mixing at cross junctions in water distribution systems. II: Experimental study," Journal of Water Resources Planning and Management, vol. 134, no. 3, 2008, pp. 295-302.

[15] C.K. Ho and S.S. Khalsa, "EPANET-BAM: water quality modeling with incomplete mixing in pipe junctions," in Water Distribution Systems Analysis 2008. Kruger National Park, South Africa: American Society of Civil Engineers, 2008, pp. 1-11.

[16] T. Yu, L. Tao, Y. Shao, and T. Zhang, "Experimental study of solute mixing at double-Tee junctions in water distribution systems," Water Science and Technology: Water Supply, vol. 15, no. 3, 2014, pp. 474-482.

[17] T.C. Yu, Y. Shao, and C. Shen, "Mixing at Cross Joints with Different Pipe Sizes in Water Distribution Systems," Journal of Water Resources Planning and Management, vol. 140, no. 5, 2014, pp. 658-665.

[18] Y. Shao, Y.J. Yang, L. Jiang, T. Yu, and C. Shen, "Experimental testing and modeling analysis of solute mixing at water distribution pipe junctions," Water research, vol. 56, no. 2014, pp. 133-147.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14044

OPTIMAL OPERATION IN SECTORIZED NETWORKS WITH INTERMITTENT WATER DISTRIBUTION

Rui Gabriel Souza[1], Bruno Brentan[2], Gustavo Meirelles[3] and Daniel Bezerra[4]

[1, 2, 3, 4]Department of Hydraulic Engineering and Water Resources, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[1]Deptartment of Civil Engineering, Polytechnical Institute of Pontifical Catholic University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[1] *rui.g182@gmail.com;* [2] *brentan@ehr.ufmg.br;* [3] *gustavo.meirelles@ehr.ufmg.br;*
[4] *danielbezerrab@gmail.com*

Abstract

The deterioration of water distribution infrastructure over the years, associated with an increase in demand and seasonal droughts can lead to an intermittent operation of the system, as the consumers cannot be supplied with a minimum pressure. The lack of investments to rehabilitate the system can be limited, and in this scenario, the intermittent water supply become a normal operation. This scenario difficulties even more the recovery of the continuous supply, as the expenses for water production increase and the revenue decrease. Thus, strategies to achieve an optimal operation in these conditions are fundamental to overcome this critical period in the best way possible. Two main objectives have to be set: minimize the operational costs, that can be described by the energy consumption in pumping stations and the volume of water lost in leakages, and, maximize water demand supplied to the consumers. Avoiding the supply during night periods, when the pressure remains high, naturally reduces the leakage volume lost. However, as the system is not operated 24h per day, it is expected a different pattern of consumption, with higher peaks during the supply period. This can lead to a significant increase in the power required by the pumps, as its head have to be higher to overcome the increased headlosses. Thus, this paper proposes an optimal operation of water distribution networks based on the scheduling of supply to different sectors of the network. This strategy aims to control the increase of the headlosses, as only part of the consumers will be supplied during a period of the day. Thus, the main pipes will not be overloaded and the power required for the pumping stations will remain low. The proposed procedure first divides the network into sub-systems using a k-means algorithm. Then, with the number of sub-systems defined, an optimal scheduling of their supply will be done. Each sub-system can have different time periods of supply, as bigger sub-system will require a higher water volume to be supplied. In addition, the pumps will be select to optimize the operation, and for each period, their rotational speed will be optimized to minimize the operational costs. The same number of sub-systems will be considered for the number of pumps in the pumping station, so adequate pumps can be selected to supply each sector. PSO algorithm will be used to optimize the operation.

Keywords
Water Distribution Network, Intermittent Operation, Sectorization, Optimization.

1 INTRODUCTION

Water Distribution Networks (WDNs) are designed to operate continuously to guarantee comfort and security for the consumers. However, the population growth, the deterioration of the infrastructure and periodically droughts lead to a situation where the system is not capable to attend the demand. When this condition is established, the Intermittent Water Supply (IWS) is commonly used to mitigate the problem. In this case, the consumers will only have water supplied during a few hours of the day, and currently, one-third of people are affected by this problem [1].

Despite the attenuation of the water supply problem, [2] highlight that this approach can lead to other issues, such as pipe bursts, water contamination and increase in energy consumption.

The IWS can be the result of a pressure deficit caused by increased headlosses or a diminished water source compared to the consumption. Leakages are a significant issue in these situations, as it contributes both for the pressure and water deficits. As the WDNs becomes larger, its operation, especially in IWS conditions, becomes more complex, as the adjustment of several hydraulic components are added to the decision-making process [3]. Thus, as described by [4], the partition of the network into District Metered Areas (DMAs) can simplify its operation, as each DMA, with less complexity, can be individually studied. As a result, [5] highlight the improvements in leakage control, identification of pipe bursts, water quality and security. In addition, the DMAs can be useful for the IWS, as it allows an equitable water supply for each DMA separately [6].

However, in a WDN supplied by a pump station, the alteration of the operation pattern from continuous to intermittent can significantly increase the energy consumption, as the pumps will not operate in their Best Efficiency Point (BEP). [7] shows the importance of the correct selection of pumps according to the systems characteristics to achieve energy and hydraulic efficiency. Thus, new pumps can be selected to avoid this issue. However, the change in operation pattern will also change the consumption pattern. [8] found that IWS leads to a reduction in consumption per capita, while [9] describes the behaviour to store water in emergency situations and discard unused "old" water when the supply cycle restarts, increasing the consumption. The number of operating hours, the average pressure and the social conditions of the consumers are aspects to consider when evaluating this issue and then, select an appropriate pump to the case. [2] proposes the full supply of water demand in shorter times. The authors found that this can be achieved, but with a high operational cost, since the headlosses are significantly increased. Posteriorly, [10] shown that the rehabilitation of main pipes could solve this issue.

As mentioned above, leakages are a major problem for IWS. Even if the total volume lost is reduced due to the lower number of hours operating, especially during the night period, when the supply cycle starts, pumps need to operate with higher head to try attend the demand, and leakage at some points can increase. Once again, the use of DMAs could be well used in this case, using smaller pumps at the entrance of each sector, or a pressure reducing valve, in case the DMA is located at a lower elevation.

This paper proposes the optimal operation of WDN under IWS conditions. First, DMAs are created using a Social Network Community Detection algorithm. Each sector will be individually supplied during a certain time, with the goal to attend the total water volume demanded in a shorter period. The duration and the period of the day will be defined trough an optimization process, as bigger system can require more time to attend the demand, and the operation of pump stations with higher power can be avoid during periods with higher energy tariffs. The demand pattern will be adjusted accordingly to each case. In addition, the scheduling (rotational speed at each operating hour) and selection (head and flow at BEP) of the pumps located at the entrance of each sector will be also done in this optimization process, where the Particle Swarm Optimization (PSO) will be used. The proposed methodology is applied into the OBCL-1 network [11].

2 METHODOLOGY

The proposed methodology to optimize the IWS operation is based on the creation of sectors (DMAs) and the operation of each one during different periods along the day. The operational optimization of each sector is based on the minimization of the cost function compound by two parts: i) energy cost; and ii) water leakage cost. The first is based on the energy consumption of the pump station to supply the water volume required by the customers of each sector at a minimal pressure level. The second aims to improve the pressure of the operated sector to values closer to the minimum required in order to reduce the leakage volume during the operation period of the sector. Figure 1 summarizes the process to achieve an optimized operation and the following sections describe each process presented.

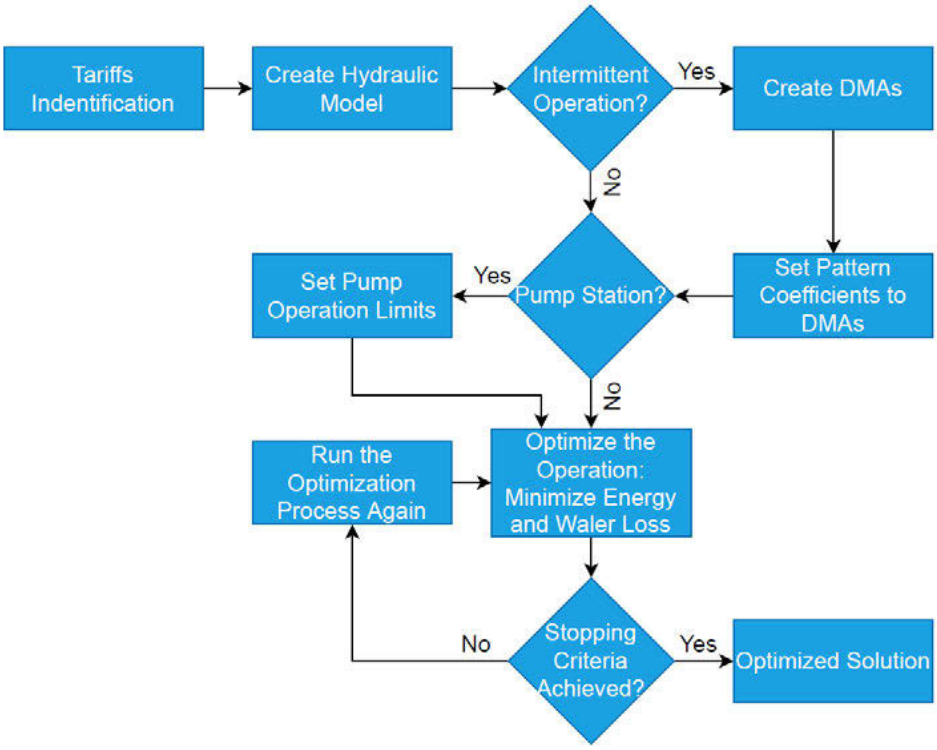

Tariffs Indentification
Create Hydraulic Model
Intermittent Operation?
Yes
Create DMAs
No
Set Pattern Coefficients to DMAs
Pump Station?
Yes
Set Pump Operation Limits
No
Optimize the Operation: Minimize Energy and Waler Loss
Run the Optimization Process Again
Stopping Criteria Achieved?
No
Yes
Optimized Solution

Figure 1. Flowchart of the proposed optimization for intermittent sectorization operation

2.1. Case Study

The benchmark network OBCL-1, firstly presented by [11] and further studied by [2] for an optimized intermittent operation, is used as a case study. As presented in Figure 2, it has 269 nodes and 294 pipes, with a daily demand of 14,270 m^3. The supply is made by a pump station composed by four pumps. Each pump will be used to supply a specific DMA to try to operate as close as possible to its best efficiency point. Finally, the leakage Q_L is modelled trough emitters, using Equation 1, with the coefficients α and β set for each node of the network model as 0.03 and 0.5 respectively, the same adopted by [2] to be able to compare the results and visualize the improvements. Then, the daily leakage volume (DL) can be calculated multiplying the Equation 1 by each time step to be simulated (Δt) until reaching the total operating time (t) each day from the first node up to the n^{th} node of the system with emitter as described in Equation 2.

$$Q_L = \alpha \cdot h^{\beta} \tag{1}$$

$$DL = \sum_{j=1}^{n} \sum_{i=1}^{t} Q_{L_{n,i}} \cdot \Delta t \tag{2}$$

CC BY NC SA

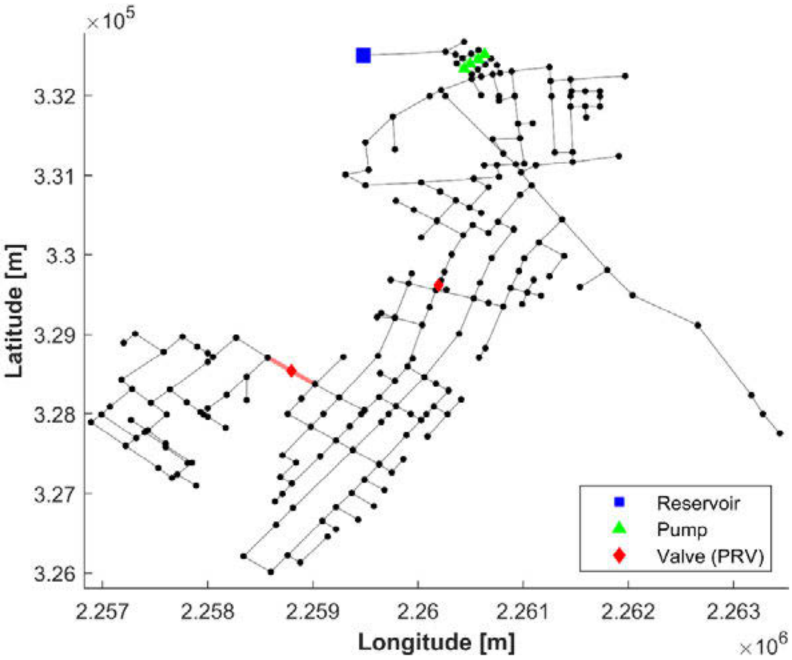


Figure 2. OBCL-1 network

For the economic analysis, the energy and power tariffs, and the cost of water production considered are presented in Table 1. As there is a significant difference in the water production costs in Brazil, three different scenarios were evaluated to verify its impact on the operation. The peak hours (PH) considered are between 17 h and 20 h and the other periods are nonpeak hours (NPH).

Table 1. Energy and water tariffs

Energy tariff - nonpeak [R$/kWh]*	0.3567
Energy tariff - peak [R$/kWh]*	0.5342
Power tariff - nonpeak [R$/kW]*	13.950
Power tariff - peak [R$/kW]*	43.850
Water production cost – High [R$/m^3]**	7.820
Water production cost – Low [R$/m^3]**	0.300
Water production cost – Average [R$/m^3]**	3.570

Source: [12]* and [13]**.

2.2. District metered area (DMA) creation

In order to generate the DMAs for intermittent operation, this work applies the methodology presented by [14]. This methodology is based on data mining technique applied to the water distribution features for clustering the nodes of the hydraulic model. As data mining algorithm, k-means [15] is used. K-means is a similarity-based technique for clustering non-labelled data. The similarity of samples is measured as an Euclidian distance $d_{i,j}$ between a normalized input vector, x, and the clusters centers c (Equation 3).

cc BY NC SA

$$d_{i,j} = \sqrt{(x_i - c_j)^2} \quad (3)$$

The clustering process starts randomly selecting points in the feature space, according to the number of pre-defined clusters. In this work, the number of clusters (i.e., the number of DMAs) is defined as four (4), considering the topology of the water distribution system. After defining the clusters centers, k-means algorithm calculates the distance among all samples and centers (Equation 4). Each sample is attributed to a cluster according to that distance. Then, the new center position is recalculated as the average value of all samples belonging to a defined cluster j.

$$c_j = \frac{\sum x_i}{N_j} \; \forall \, x_i \, \in j \quad (4)$$

where N_j is the number of samples belonging to the cluster j. After recalculate the clusters centers, the similarity matrix is recalculated and the samples are re-clustered. This process finishes when the changes on clustering center is lower than a defined limit.

In this work, following [14], the input feature vector is built based on geographic coordinates of nodes, base demand and elevation. Geographic coordinates are responsible to give information about the closeness of nodes, regarding those nodes belonging to a DMA should be interconnected by pipes and control elements. Base demand and elevation can bring to data mining analysis the hydraulic similarity of the nodes, resulting in a more controllable DMA. This because more similar hydraulic nodes require similar controls for improving the hydraulic efficiency of the system.

After the clustering process, the DMAs are almost defined. Nevertheless, a post-processing algorithm is required, since k-means is not able to catch connectivity features of the system. Eventually, non-connected nodes can belong to the same cluster, that is hydraulic impossible. In this sense, the post-processing algorithm evaluate the connectivity of all nodes and rearrange each cluster according to the connectivity of the system.

Finally, boundary pipes should be identified. Those pipes connect two DMAs and they are responsible for isolation and control of each DMA. Usually, flow meters and control valves are installed at open boundary pipes, while isolation valves are installed at closed boundary pipes.

2.3. Pattern Distribution

The water consumption (D_t), according to Equation 5, is the product of the base demand (D_m) set for each node of the hydraulic model – that represents the average daily consumption of customers connected to the system – by the dimensionless pattern coefficients (q_t), that adjust the base demand for a pattern consumption for a specific hour (t) of the day.

$$D_t = D_m q_t \quad (5)$$

As proposed by [2] and assumed in this work, the daily water consumption in CWS is equal to IWS, being necessary to modify only the pattern coefficients (q_t) according to the number of operation hours of each sector, which results in higher demand flows. So, the difference in the operational costs is the result exclusively of the improvements in the management of the system, since the total water volume supplied in CWS and IWS is the same. The pattern for each sector and time operation are presented in Figure 3.

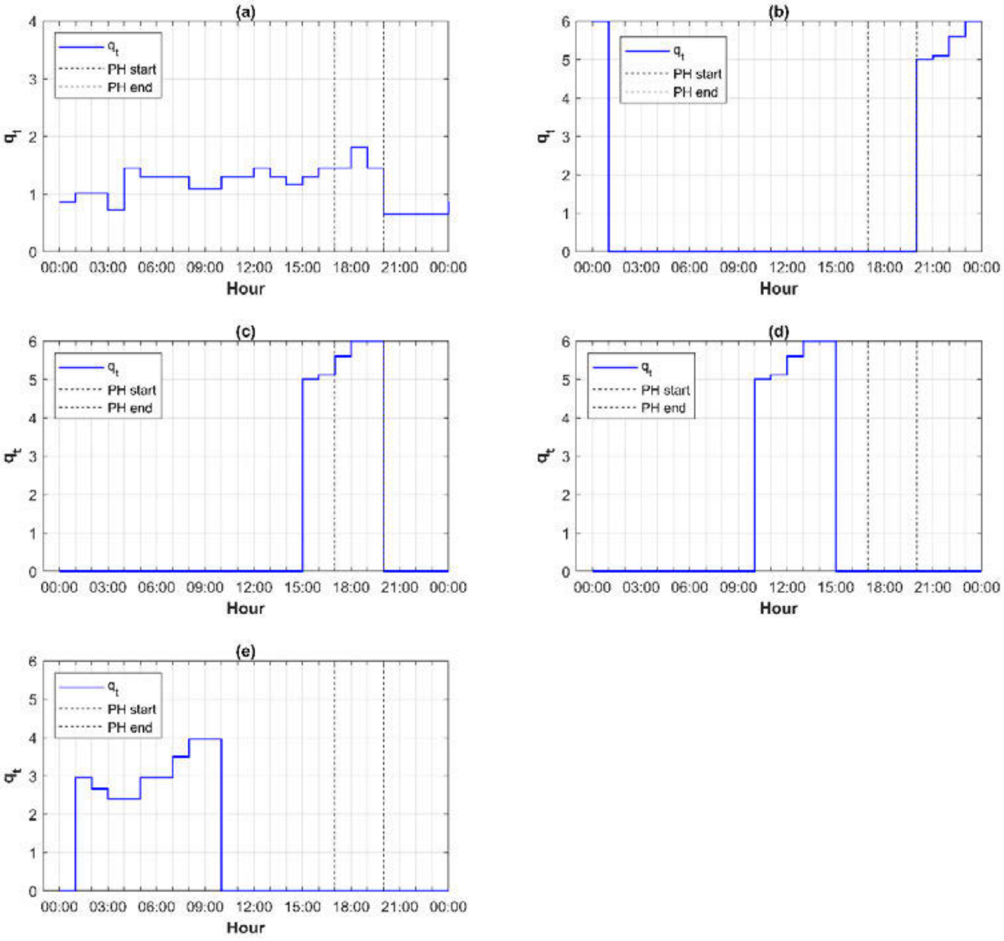


Figure 3. Demand pattern (q_t) for OBCL-1 Network: (a) continuous operation;(b) sector 1 operation; (c) sector 2 operation; (d) sector 3 operation; (e) sector 4 operation

2.4. Optimal Operation

The optimization of the system consists of selecting an optimized pump (Q_{BEP} and H_{BEP}) and the respective relative rotational speed for every hour of operation (N_t) for each sector. Hence, the number of variables to be optimized at each sector are t+2, where t is total time of operation. Then, the variables are evaluated by the objective function (OF) that represent the operational cost to be minimized, evaluated by the costs with energy and power consumption for each pump and the leakage volume of the systems, as defined by Equation 6.

$$OF = \sum_{p=1}^{Np} \sum_{i=1}^{t} [\frac{\gamma \cdot Q_i \cdot H_i}{1000 \cdot \eta_i} \cdot te_i] + \sum_{p=1}^{Np} P_{max} \cdot tp_i + \sum_{i=1}^{t} [Q_{L,i} \cdot \Delta t \cdot tw] + Pen \tag{6}$$

where:
OF [R$] – objective function to be minimized, describing the operational costs of a day;
Np [dimensionless] – number of pumps operating in the network;
t [h] – time simulation;
γ [N/m^3] – specific weight of water;
Q_i [m^3/s] – pump flow at time *i*;
H_i [m] – pump head at time *i*;
η_i [dimensionless] – pump efficiency at time *i*;
te_i [R$/kWh] – energy tariff at time *i*;

CC BY NC SA

P_{max} [kW] – maximum power;
tp_i [R$/kW] – power tariff at time *i*;
$Q_{L,i}$ [$\mathrm{m^3/h}$] – flow leakage at time *i*;
Δt [h] – time simulation step;
tw [R$/kW] – water production tariff;
Pen [R$] – penalty function.

The energy tariffs and water production costs are obtained from the Energy Company of Minas Gerais – Brasil (CEMIG) [12] and Brazilian Sanitation Information System (SNIS) [13], both presented in the section 2.1. The penalty factor (Pen), as presented in Equation 6, is added to consider the restriction problem, as the optimization method is an unconstrained method. The constraint considered is the minimum operational pressure, p_{min}, in demand nodes, set as 10 m to respect the minimal pressure established by the Brazilian standards [16]. Thus, the penalty added to the objective function is calculated with Equation 7, where p is the pressure in a demand node and β is the penalty coefficient set as 10^8.

$$if(p < p_{min}) \rightarrow Pen = \beta . |p_{min} - p| \tag{7}$$

The variables are evaluated to achieve an optimized solution by Particle Swarm Optimization (PSO) proposed by Eberhart and Kennedy in 1995, an algorithm based on the collective response of flocks of birds [17]. The initial possible solution (X) is randomly initialized and the next solution is defined by the experience of each particle (or birds) searching the space for the best solution with a velocity (V) and the collective experience of all the particle (flocks of birds) by three components, namely: i) inertia coefficient (ω); ii) cognitive coefficient (c_1); and iii) social coefficient (c_2). Equations 8 and 9 describe mathematically the process to achieve an optimized solution.

$$V_i^{t+1} = \omega . V_i^t + c_1 . rand_1 . \frac{(Xp_i - X_i^t)}{\Delta t} + c_2 . rand_2 . \frac{(Xg - X_i^t)}{\Delta t} \tag{8}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} . \Delta t \tag{9}$$

The search for an optimized solution continues until a criterion is met, assumed as 1,000 iterations or a relative change in the objective function in 20 consecutive iterations below 10^{-10}. The default values defined in the MATLAB® software for the search components are adopt, where inertia coefficient (ω) is 1.1 and the cognitive (c_1) and social (c_2) coefficients are 1.49.

3 RESULTS

The benchmark network OBCL-1, firstly presented by [11] and further studied by [2] for an optimized intermittent operation, is used as a case study. Figure 4 presents the four DMAs created using the proposed methodology. To define the number of hours that each DMA would be supplied, the total demand of each of them was evaluated. As the DMA 4 comprises almost 40 % of the total demand, it was defined a longer period of supply to avoid excessive flows in the pipes, and consequently higher headlosses, which would significantly increase the power required for the pump. Thus, DMA 4 is supplied during 9 h of the day, while the remaining three are supplied in equal periods of 5 h, as there is no significant difference in their demand. Figure 4 shows the demand pattern adopted for each DMA, considering that the same water volume will be consumed.

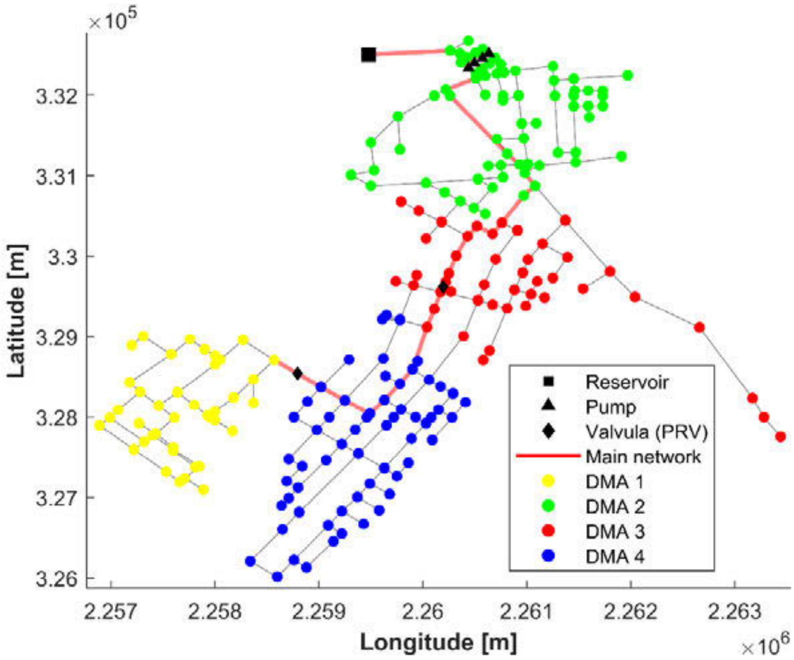


Figure 4. DMAs created for the intermittent operation

Tables 2, 3 and 4 shows the results for the different scenarios of water production costs. As expected, the energy consumption increased in all scenarios, as the pumps required more power to attend the same demand in a shorter period. When the water production cost is high, this increase in energy costs is even more significant, as economically, it is more relevant to minimize leakages than to avoid the operation during peak hours. On the other hand, the supply schedule proposed for each DMA significantly reduced the leakages, as some nodes that would have high pressures, and therefore, high leakage flow, could be isolated during a certain period. From the economic point of view, it can be seen that the energy tariffs and the water production costs play an important role, as for the scenario with low water production cost the continuous operation is 75.7 % better, and for the high and average scenarios, the intermittent operation is 27.6 % and 44.0 % better respectively. These results indicate that systems with a lower relation between energy tariffs and water productions costs will greatly benefit from the intermittent operation.

Table 2. Results for the optimized operation for the low water production cost

Description	**Scenario 1: *tw* = 0.30 R$/m³**				
	24 h*	**DMA 1**	**DMA 2**	**DMA 3**	**DMA 4**
Sector Energy Consumption [kWh]	-	688	59	377	1,468
Daily Energy Consumption [kWh]	581	2,592			
Sector Leakage (m³)	-	252	240	272	952
Daily Leakage (m³)	5,900	1,717			
Daily Leakage (%)	29.3	10.7			
Sector Energy Cost (R$)	-	2,195	720	1,206	2,917
Sector Leakage Cost (R$)	-	76	72	82	286
Daily Energy Cost (R$)	2,500	7,039			
Daily Leakage Cost (R$)	1,800	515			
Daily Operation Cost (R$)	4,300	7,554			
Economic Efficiency (%)	-	-75.7			

Source: [2]*.

Table 3. Results for the optimized operation for the high-water production cost

Description	**Scenario 2: *tw* = 7.82 R$/m³**				
	24 h*	**DMA 1**	**DMA 2**	**DMA 3**	**DMA 4**
Sector Energy Consumption [kWh]	-	688	59	377	1,494
Daily Energy Consumption [kWh]	608	2,618			
Sector Leakage (m³)	-	252	240	271	940
Daily Leakage (m³)	5,800	1,704			
Daily Leakage (%)	28.9	10.7			
Sector Energy Cost (R$)	-	2,197	721	1,206	7,354
Sector Leakage Cost (R$)	-	1,970	1,878	2,122	2,956
Daily Energy Cost (R$)	2,500	11,479			
Daily Leakage Cost (R$)	45,600	8,926			
Daily Operation Cost (R$)	48,100	20,406			
Economic Efficiency (%)	-	57.6			

Source: [2]*.

Table 4. Results for the optimized operation for the average water production cost

Description	**Scenario 3: $tw = 3.57$ R$/m^3**				
	24h*	**DMA 1**	**DMA 2**	**DMA 3**	**DMA 4**
Sector Energy Consumption [kWh]	-	687	59	377	1,482
Daily Energy Consumption [kWh]	556	2,605			
Sector Leakage (m^3)	-	252	240	271	940
Daily Leakage (m^3)	5,800	1,704			
Daily Leakage (%)	29	10.7			
Sector Energy Cost (R$)	-	2,196	721	1,206	2,932
Sector Leakage Cost (R$)	-	901	858	970	3,373
Daily Energy Cost (R$)	2,700	7,054			
Daily Leakage Cost (R$)	20,800	6,101			
Daily Operation Cost (R$)	23,500	13,156			
Economic Efficiency (%)	-	44.0			

Source: [2]*.

4 CONCLUSIONS

This paper presented a methodology to operate a WDN under an intermittent cycle, trying to supply the same volume for the consumer but in a shorter period. For this, DMAs were created and the supply period of each of them was defined according to their total demand, with the bigger DMA being supplied for a longer period to avoid excessive headlosses. Then an optimization was made to select a specific pump to supply each DMA, so its operation is as close as possible form its best efficiency point, reducing the energy consumption. The results showed that the intermittent operation significantly increase the energy consumption to attend the demand in a shorter period, but also reduces significantly the water losses, as high pressures points can be isolated during some periods. Thus, it is expected that systems with a low relation between energy tariffs and water production costs are the ones that could achieve a greater economic benefit from the intermittent operation.

5 ACKNOWLEDGMENTS

The authors acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq).

6 REFERENCES

[1] Preciado, C. C., Husband, S., Boxall, J., Del Olmo, G., Soria-Carrasco, V., Maeng, S. K., & Douterelo, I. (2021). Intermittent water supply impacts on distribution system biofilms and water quality. Water Research, 201, 117372.

[2] Souza, R. G., Meirelles, G., & Brentan, B. (2022). Energy and Hydraulic Efficiency in Intermittent Operation of Water Distribution Networks. Journal of Water Resources Planning and Management, 148(5), 04022017.

[3] Lifshitz, R., & Ostfeld, A. (2018). Clustering for analysis of water distribution systems. Journal of Water Resources Planning and Management, 144(5), 04018016.

[4] Di Nardo, A., Di Natale, M., Santonastaso, G. F., Tzatchkov, V. G., & Alcocer-Yamanaka, V. H. (2014). Water network sectorization based on graph theory and energy performance indices. Journal of Water Resources Planning and Management, 140(5), 620-629.

[5] Khoa Bui, X., S Marlim, M., & Kang, D. (2020). Water network partitioning into district metered areas: A state-of-the-art review. Water, 12(4), 1002.

[6] E. Ilaya-Ayza, A. E., Martins, C., Campbell, E., & Izquierdo, J. (2017). Implementation of DMAs in intermittent water supply networks based on equity criteria. Water, 9(11), 851.

[7] Móller, D. S., Lima, G. M., Brentan, B. M., & Barros, D. B. (2020). Optimal pump selection for variable speed operation in water distribution network. RBRH, 25.

[8] Andey, S. P., & Kelkar, P. S. (2009). Influence of intermittent and continuous modes of water supply on domestic water consumption. Water Resources Management, 23(12), 2555-2566.

[9] Speight, V., Collins, R., & Weston, S. (2020) Transitioning from IWS, WaterAid

[10] Souza, R. G., Meirelles, G., Brentan, B., & Izquierdo, J. (2022). Rehabilitation in Intermittent Water Distribution Networks for Optimal Operation. Water, 14(1), 88.

[11] Hernandez, E.; Hoagland, S.; Ormsbee, L.E. WDSRD: A Database for Research Applications. 2016. Available online: http://www.uky.edu/WDST/database.html (accessed on 28 December 2020).

[12] CEMIG (Companhia Energética De Minas Gerais). 2020. "Valores de tarifas e serviços." Accessed December 13, 2020. https://www.cemig.com.br/atendimento/valores-de-tarifas-e-servicos/.

[13] Brazil. 2019. "Ministério de Desenvolvimento Regional. Sistema Nacional de Informações sobre Saneamento (SNIS)." In Diagnóstico dos Serviços de Água e Esgotos 2018. Brasília, Distrito Federal, Brazil: SNS/MDR.

[14] Brentan, B., Meirelles, G., Luvizotto Jr, E., & Izquierdo, J. (2018). Hybrid SOM+ k-Means clustering to improve planning, operation and management in water distribution systems. Environmental modelling & software, 106, 77-88.

[15] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), 100-108.

[16] ABNT (Associação Brasileira de Normas Técnicas). 2017. Projeto de Rede de Distribuição de Água para Abastecimento Público—Procedimentos. NBR 12218. Rio de Janeiro, Brazil: ABNT

[17] Eberhart, R., and J. Kennedy. 1995. "Particle swarm optimization." In Proc., IEEE Int. Conf. on Neural Networks, 1942–1948. New York: IEEE.

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14050

IS WATER QUALITY BASED STORMWATER MANAGEMENT ACTUALLY FEASIBLE? A SWMM-BASED STUDY ON THE TRADE-OFFS OF VARIOUS STORMWATER MANAGEMENT APPROACHES

Kristjan Suits[1], Anatoli Vassiljev[2], Ivar Annus[3] and Nils Kändler[4]

[1,2,3,4] Tallinn University of Technology
Tallinn, Estonia

[1] *kristjan.suits@taltech.ee*, [2] *anatoli.vassiljev@taltech.ee*, [3] *ivar.annus@taltech.ee*,
[4] *nils.kandler@taltech.ee*

Abstract

Extreme weather events caused by climate change are becoming more frequent all across the world, particularly in the Northern Hemisphere, where these events manifest themselves as more severe or long-lasting rainfalls. Because the typical response, infrastructure refurbishment, is costly and time consuming, it is necessary to develop and implement alternative, more resource efficient, stormwater management methods. The methods must benefit both the urbanites as well as the natural environment, but they must be grounded on a comparative examination of alternatives.

The purpose of this paper is to evaluate the trade-offs between several stormwater control strategies in the context of climate change, such as no intervention, water quality-based intervention, water quantity-based intervention, and a combined approach. Each of the aforementioned approaches is considered to have intrinsic advantages and disadvantages that should be examined and quantified. Analysis of these trade-offs is critical because it provides some insight into the feasibility of implementing various stormwater management strategies for protecting the environment.

The SWMM software was used to analyse the various scenarios, and the model was built using data from multiple national databases as well as data sets provided by the Municipality of Viimsi in Estonia.

This study confirmed that water quality protection can be prioritized even in densely populated areas without jeopardizing citizens' well-being or causing unnecessary floods. This can be accomplished even with simple rule-based control if the necessary hardware, such as sensors and flow devices for digitalizing and combining inter-linking stormwater infrastructure are installed on-site. Through the SWMM simulations it was determined that digitalization of the stormwater solutions allows for a 30-92% reduction in flow, and consequently a 50-90% reduction in pollution load entering the Baltic Sea.

Keywords
SWMM, stormwater water quality, rule-based control, e-monitoring, stormwater management, feasibility.

1 INTRODUCTION

In recent years various agencies, consultancies, and international organizations have produced a plethora of reports covering the state of global water infrastructure [1, 2, 3]. Despite the fact that these documents have been composed by different organizations, they all have come to the same conclusion: globally there exists a massive infrastructure investment gap (drinking water, wastewater, and stormwater systems), which according to the Global Infrastructure Outlook (A G20 Initiative) could be worth up to $700 billion [1]. The bulk of this money is required to replace outdated pipes, pumps, storage facilities, and ensure the proper operation of sewage and stormwater treatment facilities, and thus protect the people and environment from the harmful effects of floods and pollution. If this growing investment gap is not bridged, then in the future, the reliability of water infrastructure shall decrease, and the frequency of emergencies caused by the infrastructure breaks and failures shall increase and the effects of these malfunctions may cascade and cause wide scale and severe disruptions in the urban environment.

Undoing the effects of years of chronic underfunding in the water sector is a challenging task, particularly if the goal is to bridge the gap without jeopardizing stakeholders' long-term financial stability. Achieving such a goal requires deliberate planning and foresight, as well as a shift in stakeholder engagement. It is necessary to change the way that we think about, organize, and manage stormwater, our urban environments and infrastructure [4]. Today's typical stormwater management solution is technical, relying on a network of engineered systems, however, such systems are designed for specific scenarios and modifying or refurbishing such systems in response to climate change is costly and time-consuming. A good attempt for reconciling these inherent weaknesses is done through the implementation of nature-based solutions. These systems, although requiring extensive planning and large initial investments may offer significant long-term benefits through increasing the resiliency of the urban environments to climatic fluctuations [5]. A plausible approach for leveraging the strengths and weaknesses of technical and natural systems is digitalization as it may allow to further reduce the volume of required infrastructure investments, lower the long-term operating and maintenance costs, and improve the provision of environmental and climate services. Overall, these hybrid systems are expected to be superior in every aspect in comparison to the individual systems as they enhance the strengths of individual systems by mitigating the weaknesses of them. Thus, these systems provide improved capacity to attenuate, buffer, retain, treat, and route stormwater flow and they are seemingly an adequate solution for future proofing the urban environment in face of climate change [6, 7].

Finding an optimal stormwater management solution requires balancing three objectives: cost (energy consumption, land use, construction, maintenance, rehabilitation, future proofing related to climate change), stormwater quantity and stormwater quality. It is necessary to strike a balance between these objectives while accounting for model-related uncertainties (stochastic or deterministic models) and the type of infrastructure on site (e.g., combined sewers or separate sewers) [8]. Many researchers have investigated the digitalization of stormwater systems and the various control approaches for their management and the control approaches have fallen broadly into two categories: static and dynamic [8. 9, 10, 11, 12]. Both of these approaches can be implemented for either the goal of stormwater quantity or stormwater quality management. The quantity-based approach typically focuses on the prevention of flooding and reduction of combined sewer overflow events through the control of runoff, and the quality-based approach typically focuses on pollutant reduction through the manipulation of hydraulic retention time and limiting the first flush effect. These approaches, however, are not inseparable from one another, as for example managing water quantity through peak reduction may also improve the water quality, however, it is typically not the primary goal of such approach.

Dynamic control in hybrid stormwater systems has long been regarded as an appealing stormwater management approach, however, its viability has so far been limited due to a lack of appropriate technologies and/or large costs associated with implementing the solutions [8, 9, 13]. The development of increasingly reliable, accurate, and low-cost sensors, along with advancements in data storage, processing, and transfer technologies has enabled concepts such as internet of things (IoT) to emerge, which has enabled the development of large-scale, decentralized solutions. Another catalyst for the widespread application of dynamic systems is the increase in the availability of off-grid solutions as their costs have dwindled and their energy production and storage capacities have increased [14, 15]. The primary advantage of a dynamic system over a static system is its efficiency in responding to a variety of highly stochastic and time critical environmental conditions. Although dynamic systems are considered to be more beneficial, they are not without their shortcomings as they are typically very complex, data intensive and costly to set up and maintain [11, 16].

The advent IoT solutions, the dwindling sensor and off-grid solution costs, and an annual increase in the body of knowledge on stormwater quality and quantity aspects gradually prepare us for the transition from static to dynamic systems. The purpose of this study was to determine the feasibility of implementing various stormwater management strategies on a small scale at a site in Viimsi Parish, Estonia. The research used SWMM modelling software and looked at the viability of three different control strategies: quality-based control, quantity-based control, and a combined control approach, as well as their viability in various climate scenarios.

2 MATERIAL AND METHODS

2.1 Data and modelling environment

The first step of any modelling task is to define the minimum data accuracy and requirements. The data requirements for the given study can be divided into three categories:

- Information needed to define the design storm and climate scenarios.
- Information needed for water quantity modelling.
- Information needed for water quality modelling.

Modelling software

SWMM was chosen as the modelling software because it is open source and it has a large user base in the scientific community, and it can simulate changes in both water quality and quantity. The input file was generated by utilizing GIStoSWMM5 [17], this tool required the preparation of 23 input files composed of data on the atmospheric, land surface, sub-surface, and conveyance compartment, and running them concurrently as a batch file. Although not all data had to be detailed, it was necessary to provide as much information as possible for at least the following: physical characteristics of the existing stormwater system, rainfall characteristics, elevation, land use, and flow directions, in order to create a model with a high level of utility. High utility in the context of modelling refers to a model that is as accurate as possible in terms of the interactions between flow routes, sub-catchments, and the stormwater network.

Digital Elevation and Land use data

The majority of the data needed to create the input files was obtained through consultations with Viimsi Municipality and various open-source databases in Estonia. The Municipality contributed the stormwater infrastructure data, which included historical geodetic surveys, as-built projects, and other digitalized data.

The Estonian Land Board website was used to obtain land use and elevation data (5x5 m resolution) [18]. The land use data was composed of various layers, such as waterbodies, buildings, green areas, roads and many more, and it was used to define the percentage of imperviousness of sub catchments. The elevation data was processed with an open-source toolbox TauDEM [19, 20] in ArcMap, as a result the shape, slope, area, and width of each catchment was obtained, and this information was used to create the flow direction file (.dir). The direction file was one of the main files required for running GIStoSWMM5 tool [17].

Water quality model and data requirements

Total suspended solids (TSS) were chosen as a proxy for stormwater quality estimation. TSS was chosen because it has been found to be frequently associated with several known contaminants, including metals (Cu, Cd, Zn, Pb, Cr, As, Ni), nutrients, persistent organic pollutants, and petroleum

CC BY NC SA

hydrocarbons, and thus it is frequently regarded as one of the main routes of contaminant transmission into urban waterbodies. Thus, TSS may be considered a source of concern for long-term human and ecological well-being. The risks related to suspended solids bound fraction of metals/metalloids are primarily related to the potential for phase shift into more bioavailable forms, resulting in chronic toxicity [12, 21, 22, 23, 24].

Another important factor in selecting TSS as a proxy for water quality status is its frequently discovered relationship with water turbidity, which is a parameter that can be monitored in situ in real time [21, 24, 25]. Because the TSS load reaching the waterbody is frequently linked to hydraulic retention time, time was chosen as a water quality control parameter. The modelling goal was to keep the water in the system as long as possible in order to limit the amount of TSS reaching the outfall and thus limit the export of pollutants from the urban environment to the aquatic environment. According to Gaborit et. al. 20 hours of hydraulic retention time is sufficient for approximately 50% of TSS removal and 40 hours of hydraulic retention time is sufficient for about 90% of TSS removal [26].

Another important factor in water quality was the point of occurrence of pollutant transport. According to the general body of literature, the pollution load is typically limited by the accumulation of contaminants from the previous dry days, and they are mobilized by the kinetic energy of rain drops and the turbulence created by stormwater runoff [27]. A common phenomenon observed in stormwater quality measurements is first flush, which occurred within the first 30 minutes at a nearby site in Estonia and ended up flushing at least 50% of pollutant load. A unitless concentration-volume curve represents this relationship (Figure 1).

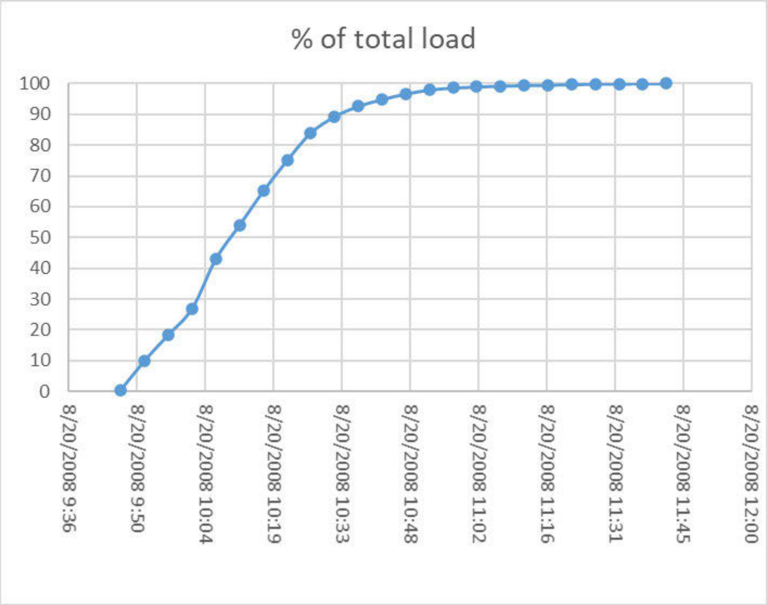


Figure 1 Pollutant concentration curve (unitless)

The authors anticipate that by retaining stormwater at appropriate times, it will be possible to reduce the amount of pollutant load entering the Baltic Sea during storm event while also reducing the number of storage units and thus the amount of money spent on water quality management.

2.2 Design storm and climate scenarios

The design storm is the acceptable hazard threshold for stormwater systems. This rigid, frequency-based approach (e.g., based on the probability of the occurrence of a 1-in-10-year event or 1-in-100-year event occurring) assumes that climate is stationary, and it is heavily reliant on the accuracy of the previously collected rainfall data. However, it is now widely accepted that our climate is constantly changing, and the intensity and the variability of climate hazards is increasing. This means that the past may no longer be representative of the future and a large part of our previously designed infrastructure may become overwhelmed. To offset the uncertainties related to the futureproof design of stormwater infrastructure it is necessary to emulate the effect of various climate scenarios on the stormwater infrastructure [28].

A design rainfall suggested by Estonian Design Standard (EVS848:2021) [29] with a return period of 10 years was used to build the NULL scenario of the rainfall. This was achieved by utilizing the following equation (1):

$$q_s = \frac{aP^b}{t^c} \tag{1}$$

Where q_s refers to average rainfall intensity (mm/h), t refers to the duration of the rainfall (in minutes), P refers to the return period (years) and b, c are empirical factors which depend on the location of the site (a= 325.7, b = 0.324, c = 0.77). Following the aforementioned formula an average precipitation intensity of 68.24 mm/h was calculated.

The runoff distribution acquired from outfall flow measurements was utilized to distribute the calculated average precipitation across a synthetic rainfall event of two hours. Thus, the total precipitation was assumed to be 136.48 mm over two hours and the distribution time step was 5 minutes. The distribution may be seen under the Null Scenario in Figure 2.

The NULL Scenario was then adjusted to account for climate change related precipitation increase by implementing the IPCC's Representative Concentration Pathway (RCP) methodology. Two alternative scenarios were chosen – RCP4.5 (intermediate scenario) and RCP8.5 (worst case scenario). In their report [30], the Estonian Environmental Agency translated the RCP methodology for Estonian climatic conditions and found that RCP4.5 forecasts an increase of rainfall intensity by 30% from the baseline conditions (approximately 88.72 mm/h) and RCP8.5 forecasts an increase of 80% from the baseline conditions (approximately 122.84 mm/h). The RCP4.5 and RCP8.5 scenarios are represented on the same hyetograph as the Null scenario (Figure 2).

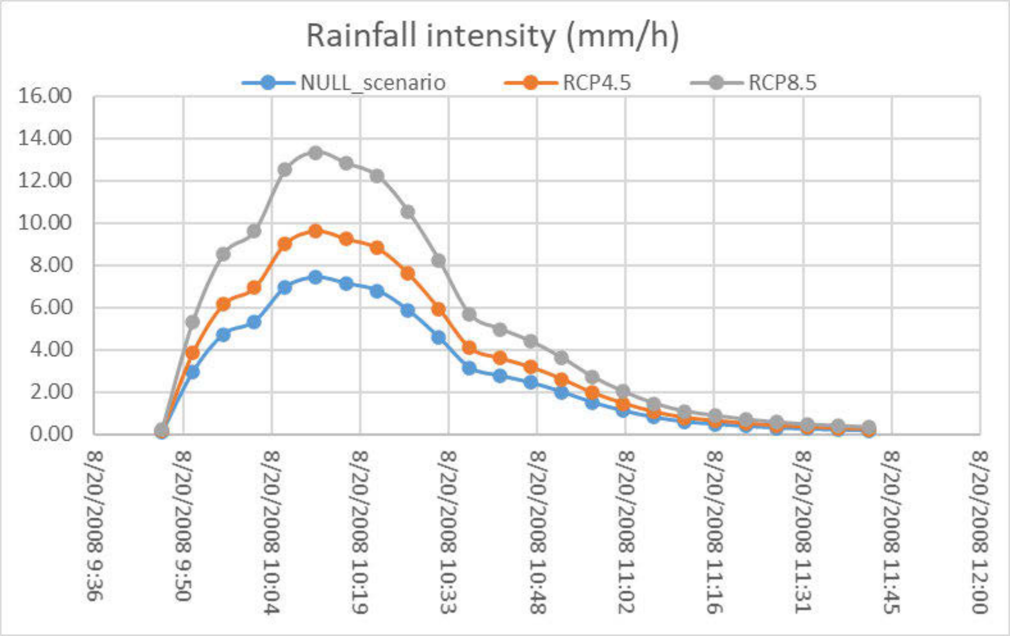


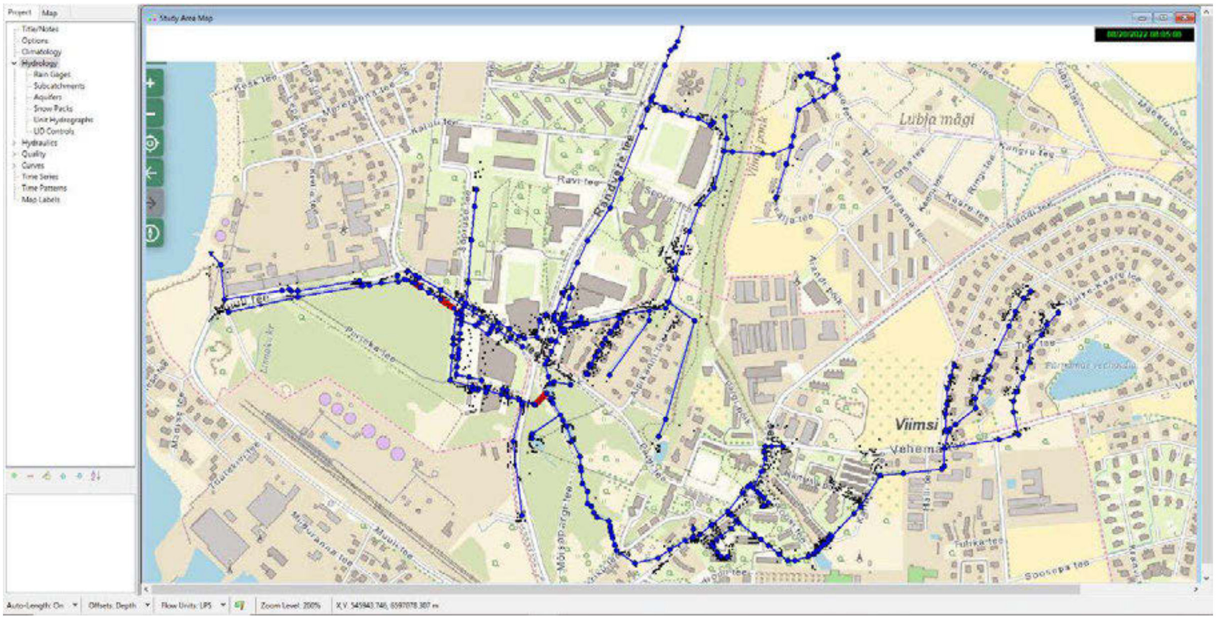
Figure 2 Design rain intensity

BY NC SA

2.3 Pilot site description

The study site is located in Viimsi Parish, northeast of Tallinn, on the Viimsi Peninsula. The climate in this area is considered to be temperate and mild and characterized by warm summers and cold winters. The average summer temperature is about 20 °C and in the winter about -8°C and the average yearly precipitation is about 700 mm [31]. The municipality is regarded as one of the fastest developing areas, and it is actively seeking opportunities to mitigate the negative effects of development, such as increased runoff and deterioration of stormwater quality. The deterioration of stormwater quality is expected to contribute to further deterioration of Baltic Sea water quality and the emergence of potential health hazards for residents who use the Haabneeme Beach for recreational purposes. The study site size is approximately 271 ha, and it is composed of large chunks of green areas, such as Laidoner Park and Haabneeme-Klindiastangu landscape protection area and some still un-developed land parcels near the urban centre. The Parish also has all of the typical land use types that are commonly in the city, such as residential areas, as well as commercial, industrial, and recreational space.

The stormwater infrastructure at the study site is currently a mix of typical structural (pipes, culverts, and manholes) and non-structural (ditches, detention ponds) infrastructure, and there is no capacity to control this system. The total length of the system is about 57214 meters, with non-structural infrastructure accounting for 18% of it and structural infrastructure about 82% of it. Plans are in place to augment this system with modern technologies such as sensors, weirs, and actuators in the future to enable control. However, there is currently no preliminary assessment of the optimal placement and potential catchment scale effects of using these devices. Figure 3 depicts an overview of the site.



Auto-Length: On
Offsets: Depth
Flow Units: LPS
Zoom Level: 200%
X,Y: 545943.746, 6597078.307 m

Figure 3 SWMM model of the study site

2.4 Identifying control locations

The proper placement of inundations is critical for stormwater quality and quantity control because it can increase the efficiency of the control system while significantly reducing potential economic, environmental, and health hazards. There are several methodologies for selecting control sites [e.g., 8, 32], but these are beyond the scope of this paper.

A typical dynamic stormwater management system is composed of the following hardware: sensors, actuators, power supply, data storage and transfer technology, and a shut-off device, such as a weir or an orifice. All of this hardware as well as local conditions, such as the slope, inundation capacity and regulatory constraints, impose concrete limitations on the placement of control devices. In this study the most appropriate locations were identified based on the following criteria: possibility of electrification to avoid data loss, ease of access for the maintenance crew, and no backflow is caused by the inundation (only controlled floods are allowed). Figure 4 depicts the final locations chosen based on these criteria.

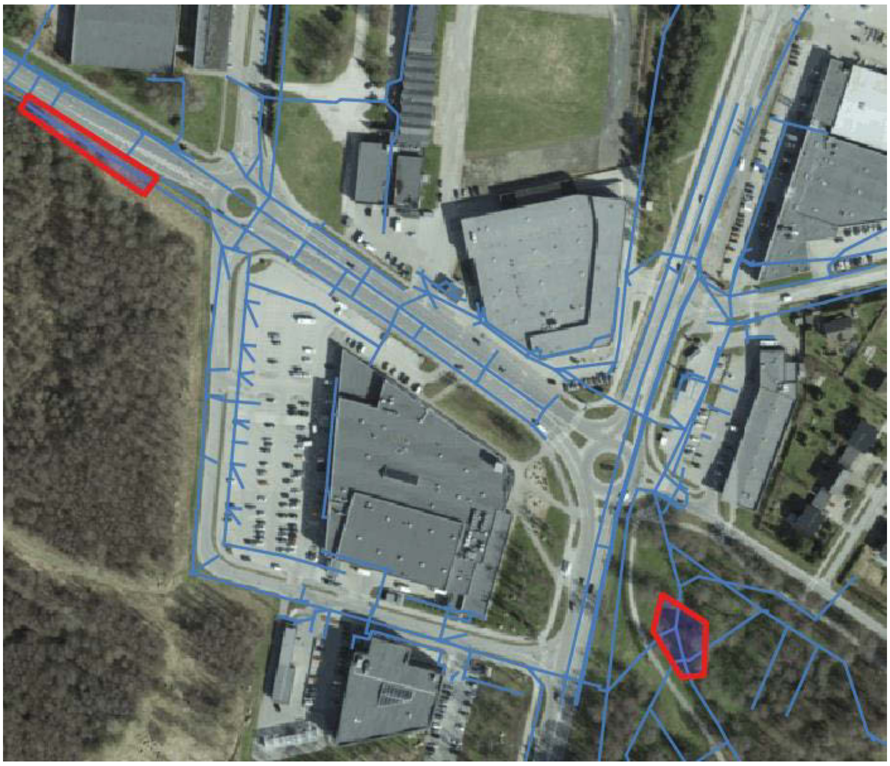


Figure 4 Points of inundation (red lines)

2.5 Control strategies

The points shown in Figure 4 were chosen for stormwater system control as they matched the expectations of the Municipality and were relatively important from an engineering perspective. The control points were relatively spacious (pond around 250 m^3 and ditch around 350 m^3), assuming that the water level at both of the waterbodies is at 0,5 meters. If the total depth is to be considered, then the inundation capacity would increase by about 2-3 times.

SWMM model was used to confirm the suitability of the chosen sites and to determine their actual maximal retention capacity. The modelling compared the following control strategies:

- Option 1: NULL scenario; no control was imposed, and the catchment behaviour was simulated during design rainfalls of varying intensities. Capacity of the infrastructure to withstand intense rainfall events was assessed.
- Option 2: Quantity based scenario; controls were imposed based on water quantity – the priority was to avoid flooding at all costs. Maximum extent of local flooding was compared with the capacity of the foreseen inundations.
- Option 3: Quality based scenario; controls were imposed based on water quality and it was assumed that 40 hours of water retention is sufficient to provide at least 90% of TSS reduction in the outfall.

- Option 4: Combined scenario; both water quality and quantity were assessed, the goal was to retain as much water as possible, while keeping the pollutant load (TSS) as low as possible. It was assumed that 20 hours of water retention is sufficient to provide at least 50% of TSS reduction in the outfall.

The control logic is shown in Figure 5.

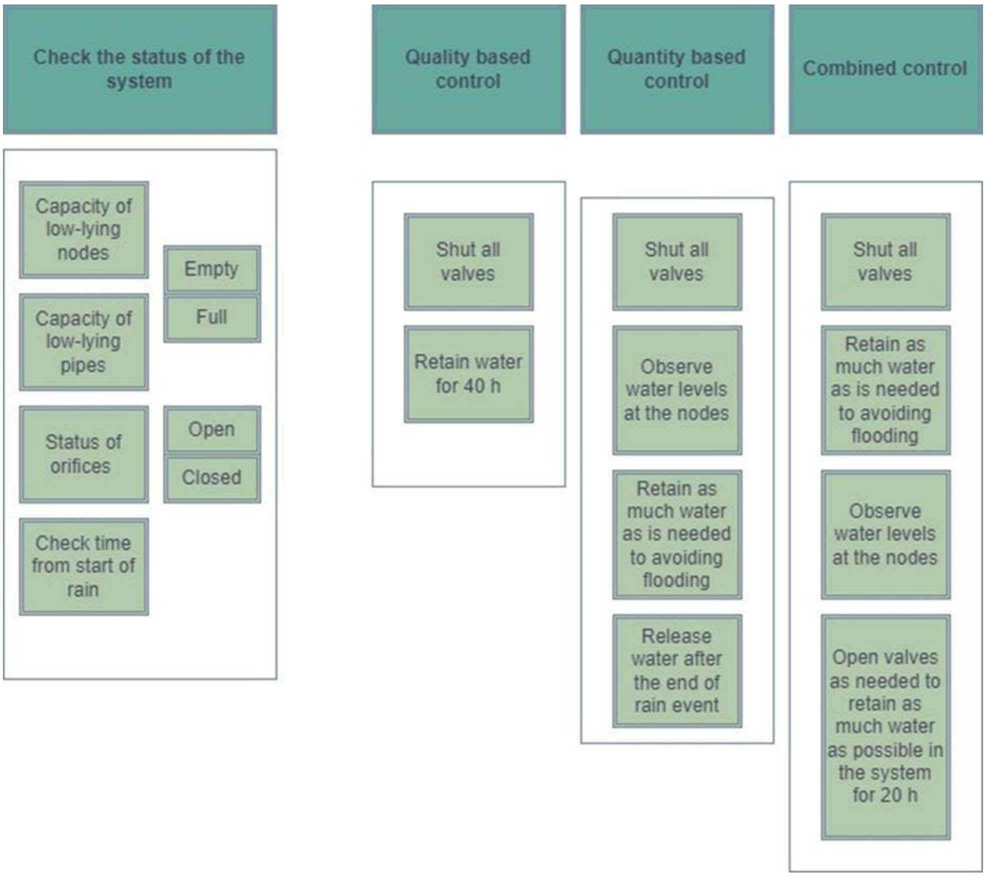

Check the status of the system
Capacity of low-lying nodes
Empty
Capacity of low-lying pipes
Full
Status of orifices
Open
Closed
Check time from start of rain
Quality based control
Shut all valves
Retain water for 40 h
Quantity based control
Shut all valves
Observe water levels at the nodes
Retain as much water as is needed to avoiding flooding
Release water after the end of rain event
Combined control
Shut all valves
Retain as much water as is needed to avoiding flooding
Observe water levels at the nodes
Open valves as needed to retain as much water as possible in the system for 20 h

Figure 5 Control Scheme

3 RESULTS AND DISCUSSION

Several rainfall scenarios were simulated to assess their impact on the performance of Viimsi's stormwater system, and controls were designed to implement various stormwater management scenarios, including water quality-based, water quantity-based, and combined. These scenarios were compared using a 2-hour rain event as a baseline, without considering the effect of consecutive rainfalls.

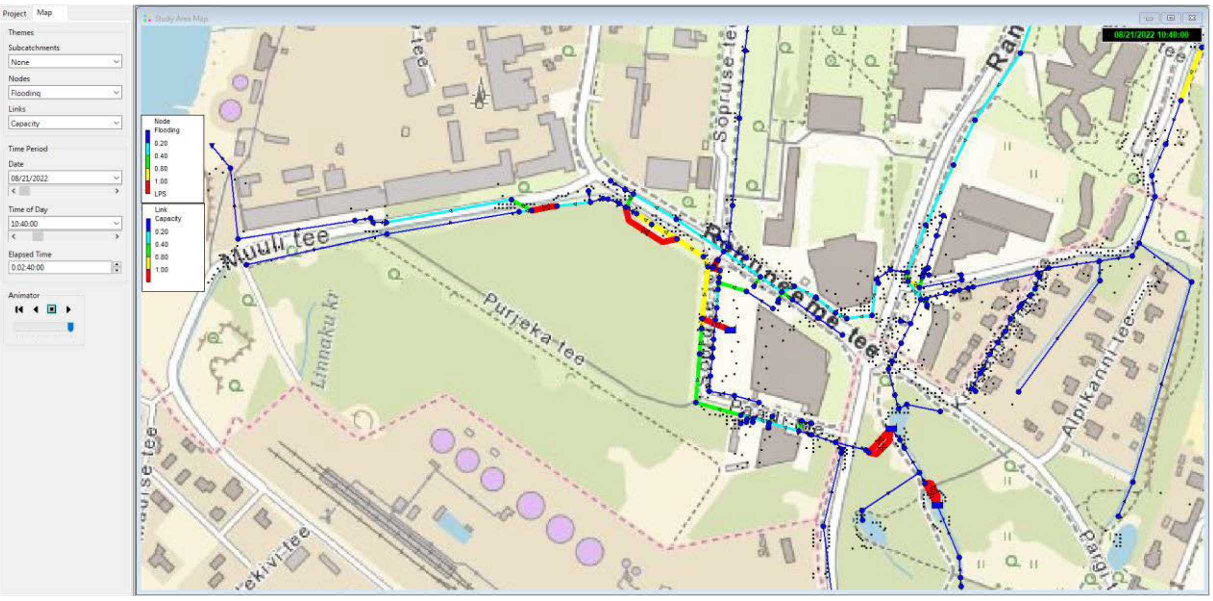
Because the stormwater system is not a hypothetical one, the emphasis was on utilizing the existing space within the system and the environment as a means to increase the hydraulic retention time of the system. It was assumed that increasing retention time would help to restore the natural hydrological cycle (through increasing infiltration, evaporation, groundwater recharge) and improve water quality (through increasing sedimentation, biotic degradation, and physico-chemical transformations). The stormwater system as a whole was found to be quite well designed, as the 2-year rainfalls did not cause any flooding within the catchment area, even when accounting for climate change scenarios RCP4.5 and RCP8.5, which predict 30% and 80% increase in rainfall intensity. As a result, the resiliency of the stormwater system had to be further tested with rainfalls of various return periods, such as 5-years and 10-years, which were also adjusted in accordance with the aforementioned climate scenarios.

The 10-year return period scenario (RCP8.5) was chosen as the scenario for further modelling because it had the greatest impact on the stormwater system performance. The model was initially run without any interventions to determine the behaviour of the system at the critical nodes and links, as well as the total volume of water leaving the catchment via the outfall. The baseline scenario, in which no control was implemented, predicted that a 2-hour storm event with a rainfall intensity of 122 mm/h would result in 3652 m^3 of stormwater reaching the Baltic Sea. Because such a scenario lacks controls, it was assumed that significant improvements could be made with minor adjustments.

The first approach aimed to reduce the volume of water reaching the Baltic Sea while avoiding flooding in the urban areas. Similarly, to water quality-based management, orifices were added at critical nodes and links of the system. During the simulation it was monitored that the capacity of the system's nodes was not exceeded. The exceedance of capacity of the nodes refers to avoidance of floods of the manholes and maintaining the fill of pipes below 0.8. The quantity-based control demonstrated that by controlling the water level at a few key nodes, the volume of water reaching the Baltic Sea could be reduced by about 30%. Overflows were used in this approach to avoid flooding while still utilizing the system's full capacity. In this case, approximately 2563 m^3 of stormwater was discovered to make its way to the outfall. Because the purpose of this scenario was not water quality control, the system was emptied as soon as the rainfall event ended.

The second scenario tested on the system was centred on water quality, with the goal of reducing the pollutant load reaching the Baltic Sea by increasing the volume of water retained in the system. It was assumed that at least 40 hours hydraulic retention time would suffice to reduce total suspended solids (TSS) load by approximately 90%. It was determined that by simply maximizing the utilization of capacity of the existing stormwater pond, pipelines, and ditches within the catchment by installing orifices in key locations and implementing a simple rule-based control that focused on retention time and ignored the occurrence of floods in the urban environment, the volume of water reaching the Baltic Sea could be reduced by about 92%. Only 279 m^3 of stormwater reaches the Baltic Sea in this case. This method ignored the effect of multiple sequential rainfall events, and the authors acknowledge that significant flooding may occur if a rainfall even of sufficient intensity and/or total volume occurs.

The third scenario (Figure 6) was a hybrid scenario that aimed to reduce both the volume of water reaching the Baltic Sea and improve the runoff water quality. The results were obtained by implementing both time-based control for capturing first flush, observing water level at key nodes and links, and attempting to keep the stormwater within the system for approximately 20 hours in total to achieve a 50% reduction in TSS load. Overflows were extensively used in this approach, resulting in 43% reduction in stormwater volume reaching the Baltic Sea. The estimated total volume reaching the outlet was about 2092 m^3.



Themes

Subcatchments: None

Nodes: Flooding

Links: Capacity

Time Period

Date: 08/21/2022

Time of Day: 10:40:00

Elapsed Time: 0.02:40:00

Project
Map
Animator
Study Area Map
08/21/2022 10:40:00
Node Flooding
0.20
0.40
0.80
1.00
LPS
Link Capacity
0.20
0.40
0.80
1.00
Muuli tee
Purjeka tee
Linnaku kr
Sõpruse tee
Alpikannikese tee

Figure 6 Snip of the model (combined scenario)

An overview of the results may be seen in Table 2 and the dynamics of the outfall are presented in Figures 7, 8, 9, 10.

Table 1 Control results

	Baseline scenario	Water quantity	Water quality	Combined
Total volume of runoff (m^3)	3652	2563	279	2092
Reduction (%)	-	30	92	43
Max flow (l/s)	285.72	455.58	28.83	202.97
Min flow (l/s)	58.54	202.97	3.79	32.73

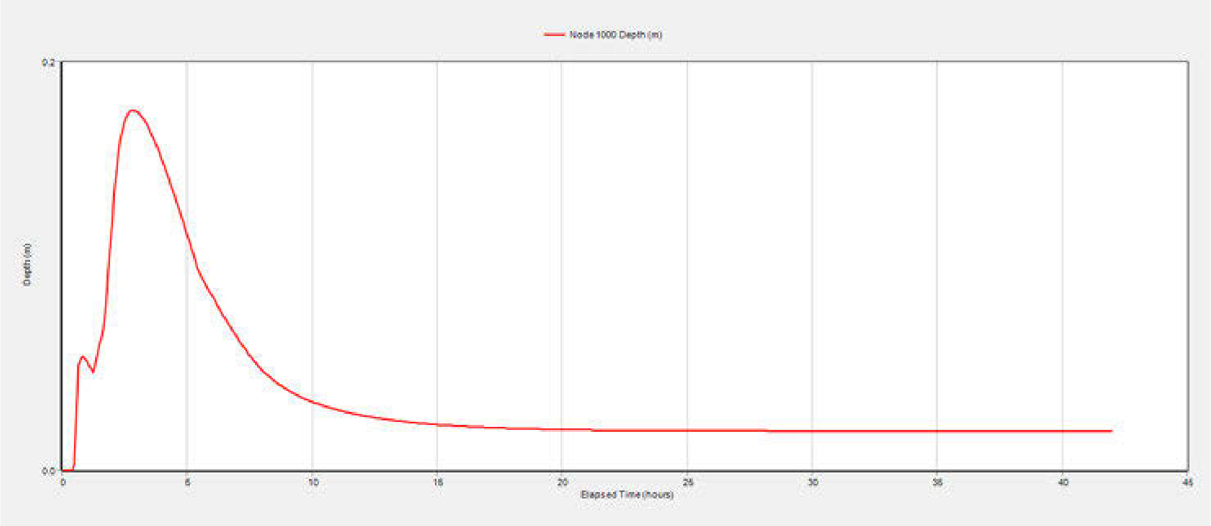


Figure 7 Baseline scenario (no control)

CC BY NC SA

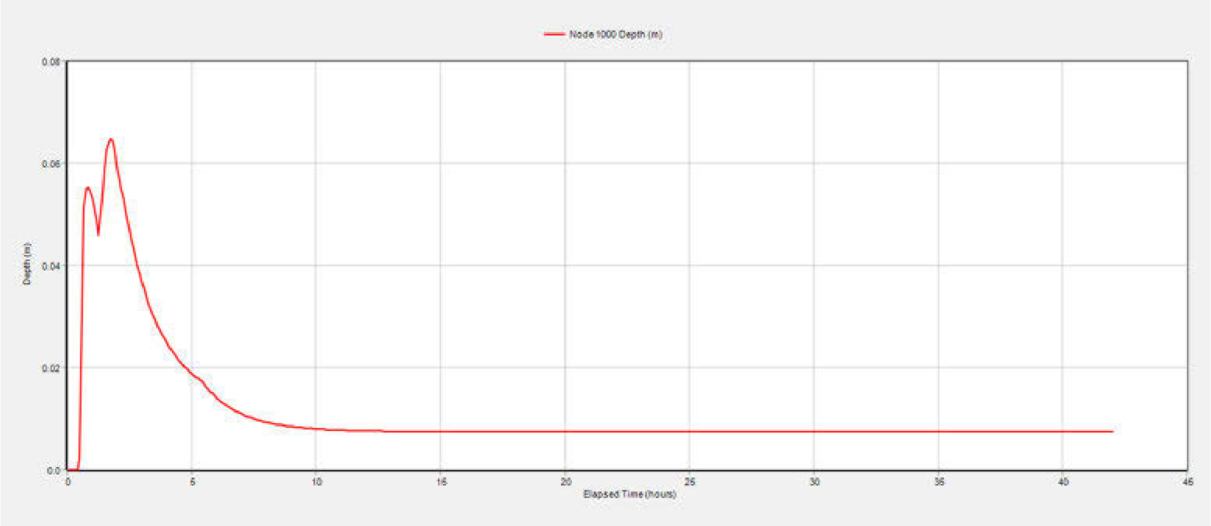


Figure 8 Water quantity control scenario

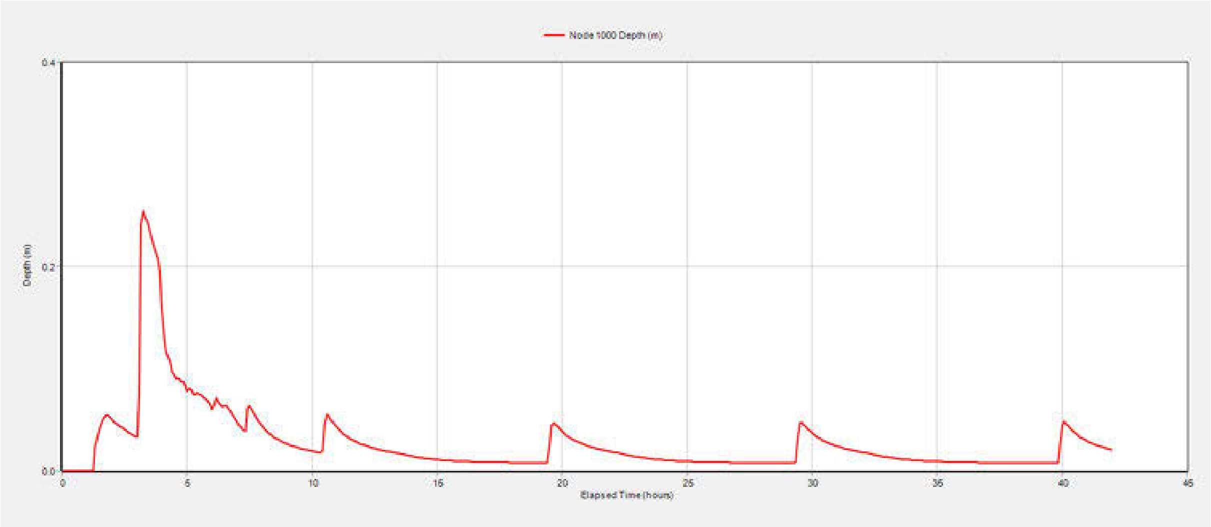


Figure 9 Water quality control scenario

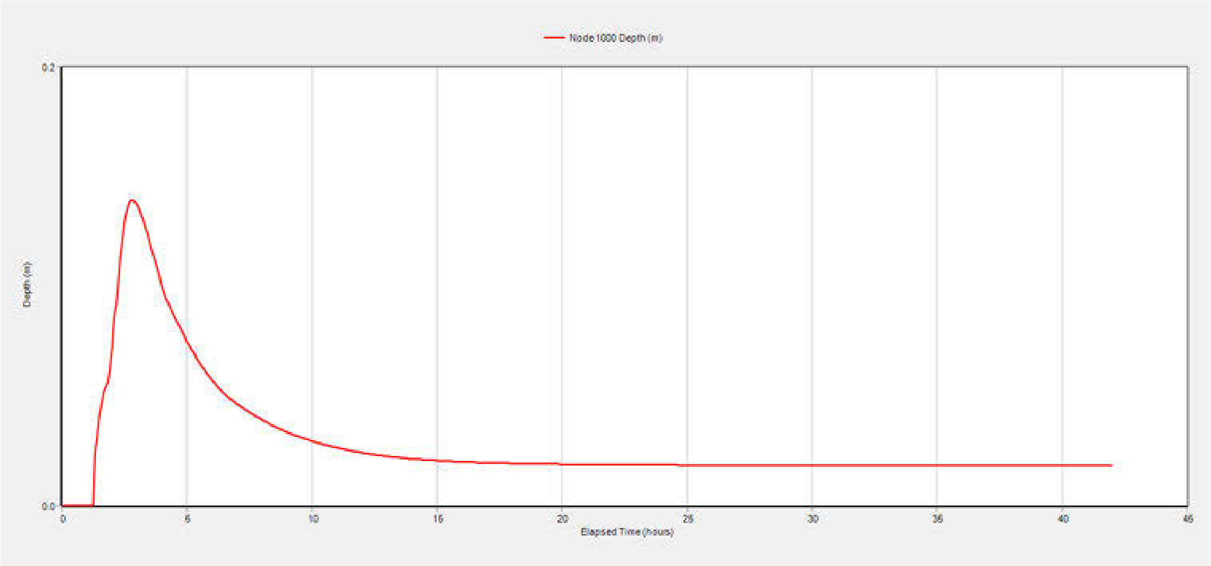


Figure 10 Combined control scenario

This simulation demonstrated that even small-scale interventions can reduce the total volume of polluted water reaching the Baltic Sea, as well as optimize the use of the existing stormwater system and the urban environment, avoiding costly investments in various stormwater infrastructure such as tanks, pipes, and pumping stations, and limiting the energy spent on water pumping. In our case, system digitalization enable us to create 1089 to 3373 m^3 of relatively low-cost storage space. The capacity of the entire stormwater network is even greater, so further system adjustments in appropriate locations promise to increase utilization of the system's free capacity even further.

This work was only a preliminary assessment of the potential for digitalization of stormwater systems in Viimsi, Estonia. Further steps will be taken in the future to assess the feasibility and practicability of this approach. Simple rule-based control could be expanded by increasing the number of control nodes and links to which various terms are assigned, as well as water quality measurements and modelling could be undertaken to confirm the TSS reduction. The water quality-based control could aim to keep the TSS load at the outfall below 35 mg/l by regulating and slowing the flow and allowing enough time for sedimentation to occur. A criterion of keeping the node and link capacity below a certain threshold value (e.g., 0.8) could be tested in the case of the links. These tests could begin with rule-based control and then be expanded to include proportional-integral-derivative (PID) and even model-predictive controls (MPC). Furthermore, a concept of traffic data-based stormwater management could be explored, in which a rule is added to the control algorithm, that some sites within the urban environment can be used as extra storage space if the data indicates a low volume of movement through the area at certain times of the day.

4 CONCLUSIONS

The study case based on the Viimsi Parish urban drainage system (UDS) successfully demonstrated how, by installing sensors and orifices and implementing RTC (even simple rule based), it is possible to significantly improve the operation of a stormwater system. The investigation revealed that all types of control strategies allow for a closer relationship to the natural hydrological cycle by retaining water in the system and reducing the volume of water released into the Baltic Sea. It also revealed that stormwater management focused on water quality is feasible; however, caution must be exercised in terms of observing rainfall events –

BY NC SA

various scenarios should be preliminarily investigated, and proper overflows should be installed to avoid extreme flooding events in the system. An optimal solution appeared to be a combined solution that seeks to extend the retention period while avoiding floods; this is also the approach that was deemed to have the most potential for optimization, but this is a matter future research. In any case, there are trade-offs between various scenarios, and the suitability of a control solution is heavily dependent on the characteristics of the site – such as the availability of space, the size and goodness of design of the existing stormwater system, the topography of the area, and the percentage of impervious area on site.

Replicating this assessment of the stormwater system with rainfalls of different return periods and for different climate conditions is a promising first step toward further developing a more complex control logic for a stormwater system of a particular catchment. So far, the results indicate that realizing RTC may help to create a triple bottom line, this is accomplished through i) reducing the need for additional infrastructure investments by better utilizing the existing system's capacity, ii) reducing the risk of urban flooding, and iii) significantly improving water quality.

This research confirmed that even a densely populated catchment may be improved so that water quality protection is prioritized without jeopardizing citizens' well-being or causing unnecessary floods. In order to achieve this water levels at the most critical nodes must be monitored and the most low-impact sites for inundation must be identified. It was also observed that RTC has the potential to realize the benefits even in future conditions, where rainfall intensity may increase by 30% (RCP4.5) or by 80% (RCP8.5).

Acknowledgements

This work was supported by the Estonian Research Council, grant number PRG667, and European Union (European Regional Development Fund) Interreg Central Baltic Programme, grant number CB878.

5 REFERENCES

[1] Oxford Economics. (2017). Global infrastructure outlook. Global Infrastucture Hub: Sydney, Australia, 64. https://www.oxfordeconomics.com/recent-releases/99f4fa86-a314-4762-97c6-fac8bdcbe40a%0Ahttps://outlook.gihub.org/%0Ahttps://www.oxfordeconomics.com/recent-releases/Global-Infrastructure-Outlook

[2] United Nations. (2018). Water infrastructure and development. United Nations, 10(9). https://sustainabledevelopment.un.org/content/documents/hlpwater/08-WaterInfrastInvest.pdf

[3] The American Society of Civil Engineers. (2020). The Economic Benefits of Investing in Water Infrastructure: How a Failure to Act Would Affect the US Economic Recovery. 1–40. www.downstreamstrategies.com

[4] Monstadt, J., Colen Ladeia Torrens, J., Jain, M., Macrorie, R. M., & Smith, S. R. (2022). Rethinking the governance of urban infrastructural transformations: a synthesis of emerging approaches. Current Opinion in Environmental Sustainability, 55, 101157. https://doi.org/10.1016/j.cosust.2022.101157

[5] Andersson, E., Grimm, N. B., Lewis, J. A., Redman, C. L., Barthel, S., Colding, J., & Elmqvist, T. (2022). Urban climate resilience through hybrid infrastructure. Current Opinion in Environmental Sustainability, 55, 101158. https://doi.org/10.1016/j.cosust.2022.101158

[6] Browder, G., Ozment, S., Rehberger Bescos, I., Gartner, T., & Lange, G.-M. (2019). Integrating Green and Gray: Creating Next Generation Infrastructure. World Resources Institute. https://doi.org/10.46830/wrirpt.18.00028

[7] Mullapudi, A., Wong, B. P., & Kerkez, B. (2017). Emerging investigators series: Building a theory for smart stormwater systems. In Environmental Science: Water Research and Technology (Vol. 3, Issue 1, pp. 66–77). https://doi.org/10.1039/c6ew00211k

[8] Shishegar, S., Duchesne, S., & Pelletier, G. (2018). Urban Water Journal Optimization methods applied to stormwater management problems: a review Optimization methods applied to stormwater management problems: a review. H

[9] Sadler, J. M., Goodall, J. L., Behl, M., Bowes, B. D., & Morsy, M. M. (2020). Exploring real-time control of stormwater systems for mitigating flood risk due to sea level rise. Journal of Hydrology, 583(April). https://doi.org/10.1016/j.jhydrol.2020.124571

[10] Saliba, S. M., Bowes, B. D., Adams, S., Beling, P. A., & Goodall, J. L. (2020). Deep reinforcement learning with uncertain data for real-time stormwater system control and flood mitigation. Water (Switzerland), 12(11), 1–19. https://doi.org/10.3390/w12113222

[11] Xu, W. D., Burns, M. J., Cherqui, F., & Fletcher, T. D. (2021). Enhancing stormwater control measures using real-time control technology: a review. Urban Water Journal, 18(2), 101–114. https://doi.org/10.1080/1573062X.2020.1857797

[12] Khiem Ly, D., Maruéjouls, T., Binet, G., & Bertrand-Krajewski, J.-L. (2019). Urban Water Journal Application of stormwater mass-volume curve prediction for water quality-based real-time control in sewer systems Application of stormwater mass-volume curve prediction for water quality-based real-time control in sewer systems. https://doi.org/10.1080/1573062X.2019.1611885

[13] Xu, C., Tang, T., Jia, H., Xu, M., Xu, T., Liu, Z., Long, Y., & Zhang, R. (2019). Benefits of coupled green and grey infrastructure systems: Evidence based on analytic hierarchy process and life cycle costing. https://doi.org/10.1016/j.resconrec.2019.104478

[14] Oberascher, M., Rauch, W., & Sitzenfrei, R. (2022). Towards a smart water city: A comprehensive review of applications, data requirements, and communication technologies for integrated management. Sustainable Cities and Society, 76(October 2021), 103442. https://doi.org/10.1016/j.scs.2021.103442

[15] Nižetić, S., Šolić, P., López-de-Ipiña González-de-Artaza, D., & Patrono, L. (2020). Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. Journal of Cleaner Production, 274. https://doi.org/10.1016/J.JCLEPRO.2020.122877

[16] Shrestha, A., Mascaro, G., & Garcia, M. (2022). Effects of stormwater infrastructure data completeness and model resolution on urban flood modeling. Journal of Hydrology, 607, 127498. https://doi.org/10.1016/J.JHYDROL.2022.127498

[17] GitHub - AaltoUrbanWater/GisToSWMM5: An automated subcatchment generator for SWMM5 model. (n.d.). Retrieved March 28, 2022, from https://github.com/AaltoUrbanWater/GisToSWMM5

[18] Download Topographic Data | Geoportal | Estonian Land Board. (n.d.). Retrieved March 28, 2022, from https://geoportaal.maaamet.ee/index.php?lang_id=2&page_id=618

[19] David Tarboton: Hydrology Research Group-Terrain Analysis Using Digital Elevation Models (TauDEM) Version 5. (n.d.). Retrieved March 28, 2022, from https://hydrology.usu.edu/taudem/taudem5/

[20] GitHub - dtarb/TauDEM: Terrain Analysis Using Digital Elevation Models (TauDEM) software for hydrologic terrain analysis and channel network extraction. (n.d.). Retrieved March 28, 2022, from https://github.com/dtarb/TauDEM

[21] Rügner, H., Schwientek, M., Beckingham, B., Kuch, B., & Grathwohl, P. (2013). Turbidity as a proxy for total suspended solids (TSS) and particle facilitated pollutant transport in catchments. Environmental Earth Sciences, 69(2), 373–380. https://doi.org/10.1007/s12665-013-2307-1

[22] Nasrabadi, T., Ruegner, H., Schwientek, M., Bennett, J., Valipour, S. F., & Grathwohl, P. (2018). Bulk metal concentrations versus total suspended solids in rivers: Time-invariant & catchment-specific relationships. PLoS ONE, 13(1), 1–15. https://doi.org/10.1371/journal.pone.0191314

[23] Sun, C., Romero, L., Joseph-Duran, B., Meseguer, J., Guasch Palma, R., Martinez Puentes, M., Puig, V., & Cembrano, G. (2021). Control-oriented quality modelling approach of sewer networks. Journal of Environmental Management, 294, 113031. https://doi.org/10.1016/j.jenvman.2021.113031

[24] McDonald, S., Holland, A., Simpson, S. L., Gadd, J. B., Bennett, W. W., Walker, G. W., Keough, M. J., Cresswel, T., & Hassell, K. L. (2022). Metal forms and dynamics in urban stormwater runoff: New insights from diffusive gradients in thin-films (DGT) measurements. Water Research, 209, 117967. https://doi.org/10.1016/J.WATRES.2021.117967

[25] Wang, Q., Zhang, Q., Dzakpasu, M., Lian, B., Wu, Y., & Wang, X. C. (n.d.). Development of an indicator for characterizing particle size distribution and quality of stormwater runoff. https://doi.org/10.1007/s11356-017-1074-z

[26] Gaborit, E., Muschalla, D., Vallet, B., Vanrolleghem, P. A., & Anctil, F. (2012). Urban Water Journal Improving the performance of stormwater detention basins by real-time control using rainfall forecasts https://doi.org/10.1080/1573062X.2012.726229

[27] Wijesiri, B., Egodawatta, P., Mcgree, J., & Goonetilleke, A. (2016). Understanding the uncertainty associated with particle-bound pollutant build-up and wash-off: A critical review. https://doi.org/10.1016/j.watres.2016.06.013

[28] Markolf, S. A., Chester, M. V., Helmrich, A. M., & Shannon, K. (2021). Re-imagining design storm criteria for the challenges of the 21st century. Cities, 109 (July 2019), 102981. https://doi.org/10.1016/j.cities.2020.102981

[29] EVS 848:2021 - Estonian Centre for Standardisation and Accreditation. (n.d.). Retrieved May 5, 2022, from https://www.evs.ee/en/evs-848-2021

[30] Luhamaa, A., Kallis, A., Mändla, K., Männik, A., Pedusaar, T., & Rosin, K. (n.d.). Keskkonnaagentuur Eesti tuleviku kliimastsenaariumid aastani 2100.

[31] Sademed |Keskkonnaagentuur. (n.d.). Retrieved May 5, 2022, from https://www.ilmateenistus.ee/kliima/kliimanormid/sademed/

[32] Kändler, N., Annus, I., & Vassiljev, A. (2022). Controlling peak runoff from plots by coupling street storage with distributed real time control. Urban Water Journal, 19(1), 97–108. https://doi.org/10.1080/1573062X.2021.1958235

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14054

COMPARISON OF SECOND AND THIRD ORDER ALGORITHMS FOR STEADY STATE RESOLUTION OF WATER DISTRIBUTION NETWORKS

Enrico Creaco[1], Armando Di Nardo[2], Michele Iervolino[3] and Giovanni Francesco Santonastaso[4]

[1]Dipartimento di Ingegneria Civile e Architettura, Università degli Studi di Pavia, Via Ferrata 3, 27100 Pavia.
[2,3,4] Dipartimento di Ingegneria, Università degli Studi della Campania Luigi Vanvitelli, via Roma 29, 81031 Aversa (CE)

[1] *creaco@unipv.it,* [2] *armando.dinardo@unicampania.it,*
[3] *michele.iervolino@unicampania.it,* [4] *GiovanniFrancesco.Santonastaso@unicampania.it*

Abstract

This paper presents the comparison of second and third order algorithms for steady state resolution of water distribution networks (WDNs). The algorithms are obtained by using the direct outflow/pressure relationship and linearizing the global equations using the Newton Raphson method. The increase in the order of convergence from quadratic to cubic is obtained by refining system matrices at half Newton Raphson step. Two variants are considered for the third order algorithm, differing in the evaluation of the matrix expressing the derivative of the outflow/pressure relationship at WDN nodes: the derivative is evaluated analytically and numerically for the first and second versions, respectively. Specifically, the numerical evaluation is obtained by using outflow and head values that are available at the half Newton Raphson step. The results of applications to five case studies of increasing complexity point out that the third order algorithm converges in a smaller number of iterations than the second order algorithm. The third order algorithm with numerical evaluation of the derivative of the outflow/pressure relationship gives significant benefits in terms of convergence performance when the service pressure range for passing from no outflow to full outflow conditions at WDN nodes is small. All the algorithms developed in this work will be considered for implementation inside the SWANP version 4.0 software.

Keywords
Water distribution networks (WDNs), Pressure-driven modelling, Resolution algorithm; High-order convergence, Matrix numerical approximation.

1 INTRODUCTION

Simulation models are traditionally used by water utility operators to replicate the nonlinear behaviour of water distribution networks (WDNs), in both off-line and real-time applications. Off-line applications concern the use of WDN models calibrated based on historical data collected from the field for specific managerial objectives, such as contingency planning, network optimization, and strategy planning [1]. Thanks to the increasing adoption of smart sensors and smart water metering, the real-time modelling of WDNs has recently started to catch on [2, 3], with the main aim to proactively simulate WDN behaviour in emergency and other situations not encountered during the calibration period. Between unsteady flow modelling and extended period simulation, i.e., WDN resolution in a sequence of steady states, the latter seems to offer better applicability in the context of real time modelling, considering the trade-off between consistency of results and computational burden, as long as it is applied with sufficiently long temporal steps to ensure dampening of hydraulic transients in the WDN [4].

The real-time modelling and management of WDNs requires use of fast, stable, and robust solvers, which must be used for both simulation and optimization purposes. While the convergence of

WDN resolution algorithms is very stable in the demand driven modelling, i.e., in the case of no dependence of nodal outflow on service pressure, the implementation of the pressure driven modelling is well known to create difficulties for convergence. To tackle this issue, four main approaches have been used in the scientific literature. The first approach lies in transforming the pressure driven resolution into an iteration of demand driven resolutions, in which nodal outflows at the generic iteration are calculated based on the head values obtained at the previous iteration [5-9]. Among these authors, Alvisi et al. [7] proposed updating nodal outflows across the algorithm iterations and implemented a relaxation procedure, later refined by Ciaponi and Creaco [9], on nodal outflows to facilitate convergence. The second approach consists of modifying the WDN resolution algorithms, by substituting the preferred pressure-driven equation into the mass conservation equations at WDN nodes. Based on this concept, Giustolisi et al. [10], Wu et al. [11], Siew and Tanymboh [12] and Elhay et al. [13] used different methods to improve the convergence behaviour of pressure driven modelling. Giustolisi et al. [10] used a heuristics-based relaxation to correct both pipe water discharges and nodal heads. Siew and Tanyimboh [12] proposed a heuristic algorithm based on backtracking and line search to correct only nodal heads. Elhay et al. [13] proposed a mathematically well-posed damping scheme based on Goldstein's algorithm, to be applied on both nodal heads and pipe water discharges. In the third approach, e.g., [14-16], the inverse outflow-pressure relationship, namely expressing the pressure as a function of the outflow, is used to eliminate the problem of oscillations. The fourth, and last, approach was recently proposed by Creaco et al. [17] and consists of using high order algorithms in the direct outflow-pressure relationship, namely expressing the outflow as a function of the pressure. Starting from the traditional second order algorithms, the high order algorithms, such as the third order ones, are obtained by refining the evaluation of system matrices at the generic iteration of WDN resolution.

The present paper is the follow-up of the paper of Creaco et al. [17] and aims to present some additional results on the comparison of second and third order WDN resolution algorithms based on the direct outflow-pressure relationship, as well as to provide insights on the treatment of system matrices expressing the outflow/pressure relationship, to improve the algorithm convergence performance.

2 ALGORITHMS FOR WDN RESOLUTION

2.1 Pressure driven modelling

For a WDN with p pipes and n nodes, including n_1 nodes with unknown head (demanding nodes) and n_0 nodes with known head (source or tanks), the steady state modelling of WDNs includes the following system of p energy balance equation and n_1 mass balance equations, written in the compact vector form:

$$\begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{pmatrix} \begin{pmatrix} \mathbf{Q} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} -\mathbf{A_{10}}\mathbf{H_0} \\ \mathbf{0} \end{pmatrix}, \tag{1}$$

in which $\mathbf{Q}$ (p,1) and $\mathbf{H}$ (n_1,1) are the unknown vectors, i.e., the vectors of water discharges at pipe and heads at unknown head nodes, respectively. $\mathbf{H_0}$ (n_0,1) is the vector of heads at known head nodes. $\mathbf{A_{10}}$ (p, n_0) and $\mathbf{A_{12}}$ (p, n_1) are matrices obtaining by extracting the n_0 and n_1 columns associated with the known and unknown head nodes, respectively, from the topological incidence matrix A (p, n). This matrix is constructed in such a way that the generic i-th row helps identifying the upstream and downstream end node, according to the arbitrarily defined positive direction in the generic i-th pipe. In the i-th row, the element $A(i.j)$ associated with the j-th node is equal to −1 or 1 if the j-th node is the upstream or downstream node of the i-th pipe, respectively. Otherwise, $A(i.j)$ = 0 if the j-th node does not belong to the i-th pipe. Finally, $\mathbf{A_{21}}$ (p, n_1), $\mathbf{A_{11}}$ (p, p) and $\mathbf{A_{22}}$ (n_1, n_1) are the transpose matrix of $\mathbf{A_{12}}$, a diagonal matrix expressing the resistance of the WDN pipes and a diagonal matrix expressing the ratio of outflow to head for the unknown head nodes, respectively. The various pressure driven relationships for the evaluation of user consumption and leakage can be easily considered for the construction of $\mathbf{A_{22}}$. As an example, for the Wagner et al. [18] pressure driven formulation, the expression of the element A_{22} (i, i) associated with the generic i-th node with unknown head is:

$$A_{22}(i,i) = \begin{cases} 0 & H \le H_{min} \\ \left(\dfrac{H - H_{min}}{H_{des} - H_{min}}\right)^{\alpha} \dfrac{d}{H} & H_{min} \le H \le H_{des} \\ \dfrac{d}{H} & H \ge H_{des} \end{cases} \tag{2}$$

In which d and H are the demand and head of the generic i-th node, respectively. H_{min} = z + h_{min} and H_{des} = z + h_{des}, in which h_{min} and h_{des} are the minimum head for having a positive outflow and the desired head for full demand satisfaction, respectively.

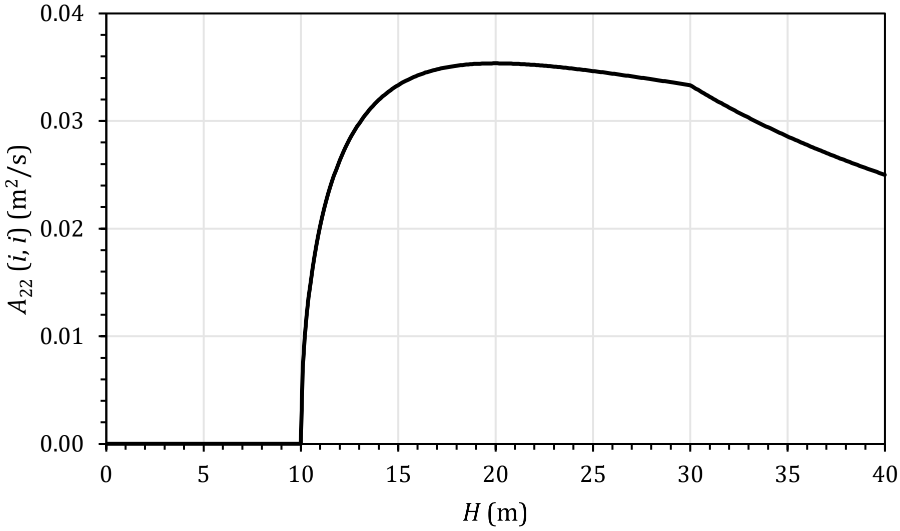


Figure 1. Pattern of the generic diagonal element of $\mathbf{A_{22}}$ *as a function of H for* H_{min}=10 m, H_{des}=30 m *and* α=0.5.

2.2 Second order Newton Raphson Method

The system of equations (1) can be solved iteratively by applying the Newton Raphson method, as explained by Todini and Rossman [19]. If $\mathbf{H}^k$ and $\mathbf{Q}^k$ are the vectors $\mathbf{H}$ and $\mathbf{Q}$, respectively, at the generic k-th iteration, the vectors $\mathbf{H}^{k+1}$ and $\mathbf{Q}^{k+1}$ at the new iteration can be obtained by solving the two following vector equations (3) and (4), respectively:

$$(\mathbf{A_{21}D_{11}^{-1}A_{12}} - \mathbf{D_{22}})\mathbf{H^{k+1}} = \{\mathbf{A_{21}D_{11}^{-1}}[(\mathbf{D_{11}} - \mathbf{A_{11}})\mathbf{Q^k} - \mathbf{A_{10}H_0}] + \mathbf{A_{22}H^k} - \mathbf{D_{22}H^k}\} \tag{3}$$

$$\mathbf{D_{11}Q^{k+1}} = \mathbf{D_{11}Q^k} - (\mathbf{A_{11}Q^k} + \mathbf{A_{12}H^{k+1}} + \mathbf{A_{10}H_0}) \tag{4}$$

in which $\mathbf{D_{11}}$ and $\mathbf{D_{22}}$ are diagonal matrices that can be calculated analytically as $\mathbf{D_{11}}$=d($\mathbf{A_{11}Q}$)/d$\mathbf{Q}$ and $\mathbf{D_{22}}$=d($\mathbf{A_{22}H}$)/d$\mathbf{H}$. As an example, for the Wagner et al. [18] pressure driven formulation, the analytical expression of the element D_{22} (i, i) (Figure 2) associated with the generic i-th node with unknown head is:

$$D_{22}(i,i) = \begin{cases} 0 & H \le H_{min} \\ \frac{\alpha}{(H_{des} - H_{min})^{\alpha}} (H - H_{min})^{\alpha-1}\, d & H_{min} \le H \le H_{des} \\ 0 & H \ge H_{des} \end{cases} \tag{5}$$

All matrices $\mathbf{A}_{11}$, $\mathbf{A}_{22}$, $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ are evaluated based on the values of $\mathbf{Q}$ and $\mathbf{H}$ at iteration k.

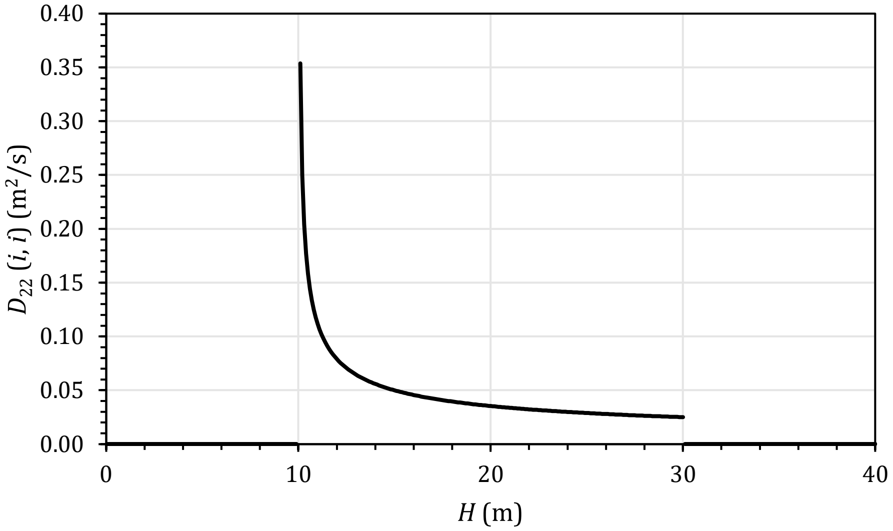


Figure 2. Pattern of the generic diagonal element of $\mathbf{D}_{22}$ *as a function of H for* H_{min}=10 m, H_{des}=30 m *and* α=0.5.

From equation (3), the new vector $\mathbf{H}^{k+1}$ of heads at demanding nodes can be obtained through the solution of the linear system of n_1 equations. From equation (4), the new vector $\mathbf{Q}^{k+1}$ of pipe water discharges can be obtained through the solution of p independent linear equations.

To speed up convergence, the second order Newton Raphson method can be dampened by applying the following underrelaxation to the heads:

$$\mathbf{H}^{\mathbf{k+1}} = \mathbf{H}^{\mathbf{k}} + \lambda^{k}(\mathbf{H}^{\mathbf{k+1}} - \mathbf{H}^{\mathbf{k}}) \tag{6}$$

in which λ^k is a number between 0 and 1, to be evaluated at each iteration as explained by Creaco et al. [17]. Therefore, the second order dampened Newton-Raphson method is applied by first solving the vector equation (3), then applying underrelaxation (6) and finally solving the vector equation (4).

2.3 Third order Newton Raphson Method

As explained by Creaco et al. [17], the increase in the order of convergence is obtained by refining the evaluation of matrices $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ at the generic iteration at half Newton Raphson step. To accomplish this, the second order Newton Raphson is initially applied to obtain first estimates for the vectors $\mathbf{H}$ and $\mathbf{Q}$ of nodal heads and pipe water discharges respectively. These first estimates are indicated as $\mathbf{H}^{k+1,ie}$ and $\mathbf{Q}^{k+1,ie}$, respectively. Then, the nodal head and pipe water discharge vectors at half Newton Raphson step are derived as:

$$\mathbf{H}^{\mathbf{k+1/2}} = \frac{\mathbf{H}^{\mathbf{k}} + \mathbf{H}^{\mathbf{k+1,ie}}}{\mathbf{2}} \tag{7}$$

$$\mathbf{Q}^{\mathbf{k+1/2}} = \frac{\mathbf{Q}^{\mathbf{k}} + \mathbf{Q}^{\mathbf{k+1,ie}}}{\mathbf{2}} \tag{8}$$

After evaluation of the refined matrices $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ based on the vectors $\mathbf{H}^{k+1/2}$ and $\mathbf{Q}^{k+1/2}$ at half Newton Raphson step, the sequence of vector equations (3), (6) and (4) is repeated to obtain the vectors $\mathbf{H}^{k+1}$ and $\mathbf{Q}^{k+1}$ at the new iteration. As Creaco et al. [17] proved, the increase in the order of convergence from quadratic to cubic yields significant benefits in terms of convergence performance under pressure driven modelling conditions.

2.4 Numerical approximation of $\mathbf{D}_{22}$

The presence of derivative discontinuities in the outflow/pressure relationship, which makes matrix $\mathbf{D}_{22}$ discontinuous, is known to slow down the convergence of WDN resolution algorithms under pressure driven modelling conditions. As an example, this happens in the Wagner et al. [18] formulation for $H=H_{min}$ and $H=H_{des}$ (see equation 5 and Figure 2). While Creaco et al. [17] proposed the regularization of matrix $\mathbf{D}_{22}$ for the third order Newton Raphson algorithm, this work proposes its numerical evaluation to obtain its smoothening. Remembering that $\mathbf{D}_{22}=\mathrm{d}(\mathbf{A}_{22}\mathbf{H})/\mathrm{d}\mathbf{H}$, this can be done by calculating the generic element of $\mathbf{D}_{22}$ in the second step of the third order Newton Raphson algorithm as:

$$D_{\mathbf{22}}(i,i) = \frac{A_{22}^{+}H^{+} - A_{22}^{-}H^{-}}{H^{+} - H^{-}} \tag{9}$$

In which H^+ and H^- can be set equal to $H^{k+1,ie}$ and H^k, respectively. Furthermore, A_{22}^{+} and A_{22}^{-} are the elements A_{22} (i,i) evaluated at H^+ and H^-, respectively. In the first step of the third order Newton Raphson algorithm, except for the first iteration in which equation (5) is applied, $\mathbf{D}_{22}$ is set equal to its value in the second step of the previous iteration.

3 APPLICATIONS

Five case studies of increasing complexity (Figure 3) were considered in this paper to show the comparison of the second order algorithm (SO), third order algorithm with analytically calculated matrix $\mathbf{D}_{22}$ (third order variant 1 TO1) and third order algorithm with numerically calculated matrix $\mathbf{D}_{22}$ (third order variant 2 TO2). The first case study is the branched WDN of Gupta and Bhave [20] with n_0=1, n_1=4 and p=4. The second case study is the 2-looped WDN of Deuerlein et al. [14] with n_0=1, n_1=4 and p=6. The third case study is the 3-looped WDN of Hanoi [21] with n_0=1, n_1=31 and p=34. The fourth case study is the 49 looped WDN of Modena [22] with n_0=4, n_1=268 and p=317. Finally, the fifth case study is the 11-looped WDN of Balerma [23] with n_0=4, n_1=443 and p=454. The data concerning the features of WDN nodes and pipes can be found in the referenced works or in [16].

The algorithms SO, TO1 and TO2 of the present work were tested against the five case studies considering various values of h_{min} and h_{des}. Specifically, the values h_{min} = 0 m and h_{des} = 20 m were considered for the first and second case studies. For the three remaining case studies, four pairs of h_{min} and h_{des} were analyzed, namely h_{min} = 10 m - h_{des} = 40 m, h_{min} = 10 m - h_{des} = 30 m, h_{min} = 10 m - h_{des} = 20 m, h_{min} = 10 m - h_{des} = 10.1 m, to create increasingly challenging pressure driven conditions. In fact, the closer h_{min} and h_{des}, the smaller the service pressure variation required for increasing the generic nodal outflow from 0 to the desired demand d.

While the three algorithms analysed always converged to the same solution, the number of iterations required for convergence varied a lot. In case studies 1, 2, 4 and 5, the simple algorithm

SO proved capable of converging in less than 10 iterations in the case of sufficiently large service pressure range [h_{min}, h_{des}], i.e., h_{des} -h_{min}≥10 m. Problems of convergence were always observed in case study 3, in which large head oscillations were noticed during the iterations, and in all case studies when h_{des} -h_{min}=0.1 m. The use of the third order algorithm yielded benefits in terms of convergence performance, in comparison with SO. However, TO1 still needed more than 30 iterations to converge in case studies 4 and 5 for h_{des} -h_{min}=0.1 m. This problem was totally fixed in TO2. Except for case studies 4 and 5 in the case of h_{des} -h_{min}=0.1 m, featuring a number of iterations equal to 14 and 13, respectively, TO2 always converged in a number of iterations lower than or equal to 6.

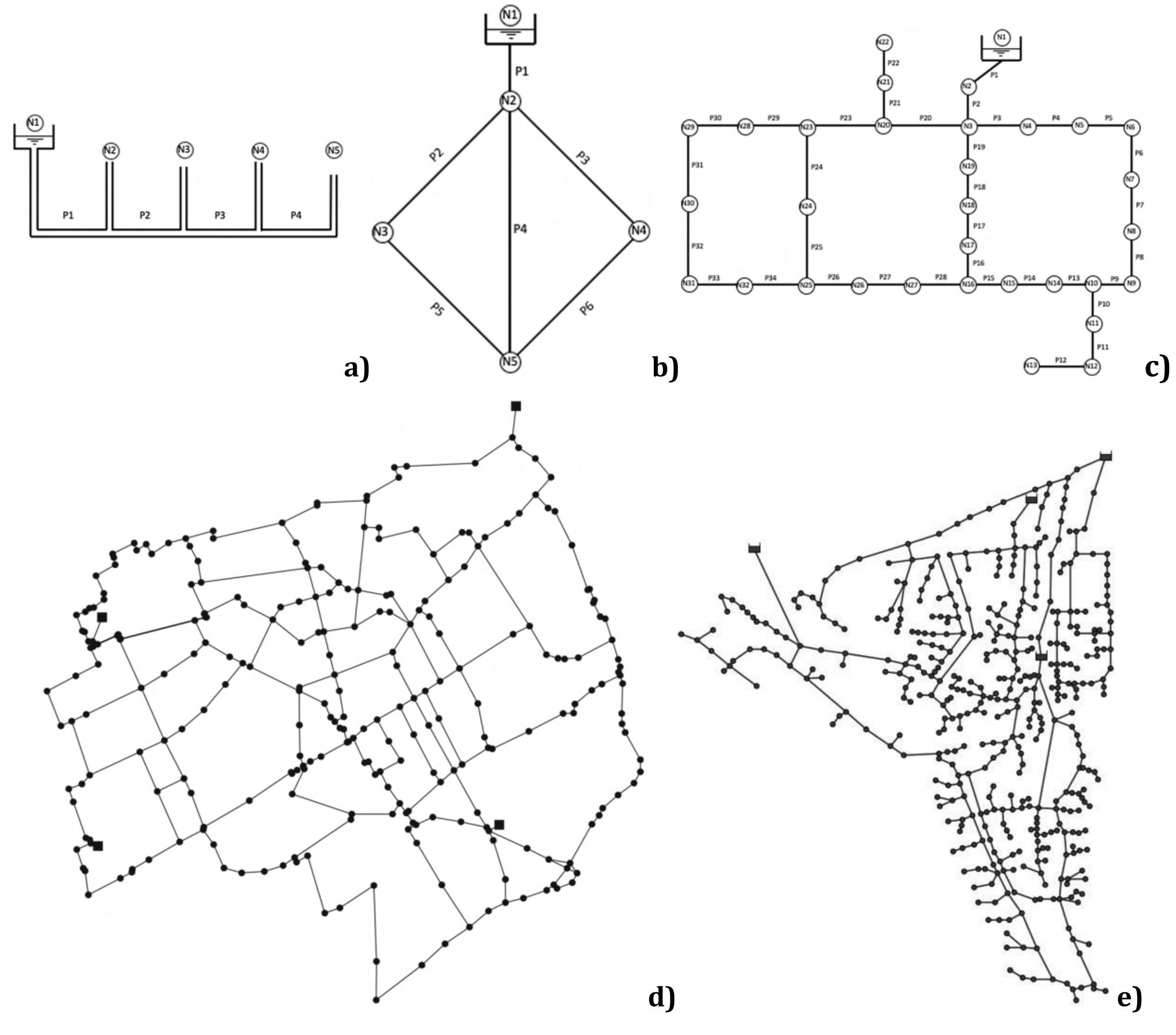

N1
N2
N3
N4
N5
P1
P2
P3
P4
a)
N1
N2
N3
N4
N5
P1
P2
P3
P4
P5
P6
b)
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
N11
N12
N13
N14
N15
N16
N17
N18
N19
N20
N21
N22
N23
N24
N25
N26
N27
N28
N29
N30
N31
N32
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
P19
P20
P21
P22
P23
P24
P25
P26
P27
P28
P29
P30
P31
P32
P33
P34
c)
d)
e)

Figure 3. WDNs of case studies a) 1, b) 2, c) 3, d) 4 and e) 5.

Table 1. Convergence performance of second order (SO) and third order (TO1 and TO2) Newton Raphson algorithms for WDN resolution in the five case studies.

Network	h_{min} (m)	h_{des} (m)	SO Iterations	TO1 Iterations	TO2 Iterations
1	0	20	4	3	3
2	0	20	6	5	5
3a	10	40	56	5	4
3b	10	30	56	12	5
3c	10	20	56	15	5
3d	10	10.1	103	9	6
4a	10	40	4	4	4
4b	10	30	4	4	4
4c	10	20	5	4	4
4d	10	10.1	42	38	14
5a	10	40	5	5	4
5b	10	30	5	5	4
5c	10	20	7	5	4
5d	10	10.1	60	34	13

4 CONCLUSIONS

In this work, a comparison of second and third order algorithms for steady state resolution of WDNs was carried out. These algorithms were obtained by using the direct outflow/pressure relationship and linearizing the global equations using the Newton Raphson method. The increase in convergence order from quadratic to cubic was obtained by refining system matrices at half Newton Raphson step. The numerical approximation of the matrix expressing the derivative of the outflow/pressure relationship was proposed as a novel aspect of the present work. Globally, the results of the applications to five case studies of increasing complexity pointed out that:

- All the algorithms analysed converge to the same solution.
- The convergence of the second order algorithm is observed to slow down in case studies where nodal heads tend to oscillate and when the service pressure range for passing from no outflow to outflow is small.
- The third order algorithm features better convergence performance, especially when the matrix expressing the derivative of the outflow/pressure relationship is numerically approximated.

The algorithms developed in the present work are being considered for implementation inside the SWANP version 4.0 software [24], which enables tackling various kinds of modeling, design, and managerial problems for WDNs

5 REFERENCES

[1] Walski, T., D. Chase, D. Savic, W. Grayman, S. Beckwith, and E. Koelle. Advanced water distribution modelling and management, Waterbury, CT: Haestad, 2003.

[2] Machell, J., S. R. Mounce, and J. B. Boxall, "Online modelling of water distribution systems: A UK case study, drinking water." Eng. Sci., vol. 3, no. 1, 2010, pp. 21–27.

[3] Okeya, I., Z. Kapelan, C. Hutton, and D. Naga, "Online modelling of water distribution system using data assimilation." Procedia Eng., vol. 70, no. 1, 2014, pp. 1261–1270. https://doi.org/10.1016/j.proeng.2014.02.139.

[4] Creaco, E., G. Pezzinga, and D. Savic, "On the choice of the demand and hydraulic modeling approach to WDN real-time simulation." Water Resour. Res., vol. 53, no. 7, 2017, pp. 6159–6177. https://doi.org/10.1002/2016WR020104.

[5] Bhave, P. R., "Node flow analysis distribution systems." Transp. Eng. J., vol. 107, no. 4, 1981, pp. 457–467. https://doi.org/10.1061/TPEJAN.0000938.

[6] Tabesh, M. Implications of the pressure dependency of outflows on data management, mathematical modelling and reliability assessment of water distribution systems. Ph.D. thesis, Dept. of Civil Engineering, Univ. of Liverpool, 1998.

[7] Alvisi, S., and M. Franchini "Near-optimal rehabilitation scheduling of water distribution systems based on a multi-objective genetic algorithm." Civ. Eng. Environ. Syst., vol. 23, no. 3, 2006, pp. 143–160. https://doi.org/10.1080/10286600600789300.

[8] Jun, L., and Y. Guoping "Iterative methodology of pressure dependent demand based on EPANET for pressure-deficient water distribution analysis." J. Water Resour. Plann. Manage., vol. 139, no. 1, 2013, pp. 34–44. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000227.

[9] Ciaponi, C., and E. Creaco "Comparison of pressure-driven formulations for WDN simulation." Water, vol. 10, no. 4, 2018: 523. https://doi.org/10.3390/w10040523.

[10] Giustolisi, O., D. Savic, and Z. Kapelan "Pressure-driven demand and leakage simulation for water distribution networks." J. Hydraul. Eng., vol. 134, no. 5, 2008, pp. 626–635. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:5(626).

[11] Wu, Z., R. Wang, T. Walski, S. Yang, D. Bowdler, and C. Baggett "Extended global-gradient algorithm for pressure-dependent water distribution analysis." J. Water Resour. Plann. Manage., vol. 135, no. 1, 2009, pp. 13–22. https://doi.org/10.1061/(ASCE)0733-9496(2009)135:1(13).

[12] Siew, C., and T. Tanyimboh. "Pressure-dependent EPANET extension." Water Resour. Manage., vol 26, no. 6, 2012, pp. 1477–1498. https://doi.org/10.1007/s11269-011-9968-x.

[13] Elhay, S., O. Piller, J. Deuerlein, and A. Simpson. "A robust, rapidly convergent method that solves the water distribution equations for pressure-dependent models." J. Water Resour. Plann. Manage., vol. 142, no. 2, 2016: 04015047. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000578.

[14] Deuerlein, J. W., O. Piller, S. Elhay, and A. R. Simpson. "Contentbased active-set method for the pressure-dependent model of water distribution systems." J. Water Resour. Plann. Manage., vol. 145, no. 1, 2019: 04018082. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001003.

[15] Water Infrastructure Division. EPANET 2.2 users manual. U.S. Environmental Protection Agency Cincinnati, OH 45268. Accessed May 3, 2020. https://epanet22.readthedocs.io/en/latest/index.html.

[16] Todini, E., S. Santopietro, R. Gargano, and L. A. Rossman. "Pressure flow-based algorithms for pressure-driven analysis of water distribution networks." J. Water Resour. Plann. Manage., vol. 147, no. 8, 2021: 04021048. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001401.

[17] Creaco, E., A. Di Nardo, M. Iervolino, and G. Santonastaso. "High-Order Global Algorithm for the Pressure-Driven Modeling of Water Distribution Networks" J. Water Resour. Plann. Manage., vol. 148, no. 3, 2022: 04021109. https://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0001524.

[18] Wagner, J., U. Shamir, and D. Marks. "Water distribution reliability: Simulation methods." J. Water Resour. Plann. Manage., vol. 114, no. 3, 1988, pp. 276–294. https://doi.org/10.1061/(ASCE)0733-9496(1988)114:3(276).

[19] Todini, E., and L. A. Rossman. "Unified framework for deriving simultaneous equation algorithms for water distribution networks." J. Hydraul. Eng., vol. 139, no. 5, 2013, pp. 511–526. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000703.

[20] Gupta, R., and P. R. Bhave. "Comparison of methods for predicting deficient-network performance." J. Water Resour. Plann. Manage., vol. 122, no. 3, 1996, pp. 214–217. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:3(214).

[21] Fujiwara, O., and D. B. Khang. "A two-phase decomposition method for optimal design of looped water distribution networks." Water Resour. Res., vol. 26, no. 4, 1990, pp. 539–549. https://doi.org/10.1029/WR026i004p00539.

[22] Artina, S., A. Bolognesi, C. Bragalli, C. D'Ambrosio, and A. Marchi. "Comparison among best solutions of the optimal design of water distribution networks obtained with different algorithms." In: Vol. 3 of Proc., CCWI2011 Urban Water Management: Challenges and Opportunities, 2011, pp. 985–990. Exeter, UK: Centre for Water Systems.

[23] Reca, J., and J. Martinez. "Genetic algorithms for the design of looped irrigation water distribution networks." Water Resour. Res., vol. 42, no. 5, 2006, pp. 1–9. https://doi.org/10.1029/2005WR004383.

[24] Di Nardo, A., M. Di Natale, A. Di Mauro, E. Martínez Díaz, J. A. Blázquez Garcia, G. F. Santonastaso, and F. P. Tuccinardi. "An advanced software to design automatically permanent partitioning of a water distribution network." Urban Water J., vol. 17, no. 3, 2020, pp. 259–265. https://doi.org/10.1080/1573062X.2020.1760322.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14055

SENSITIVITY ANALYSIS OF WATER DISTRIBUTION NETWORKS DETERIORATION IN HYDRAULIC AND ECONOMIC PARAMETERS

Leandro Alves Evangelista[1], Gustavo Meirelles Lima[2] and Bruno Melo Brentan[3]

[1,2,3] Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[1] *leandroalves1989@gmail.com*, [2] *gustavo.meirelles@ehr.ufmg.br*, [3] *brentan@ehr.ufmg.br*

Abstract

Water Distribution Networks (WDNs) are responsible for the majority of the costs in a water supply system. Thus, it is reasonable to carefully evaluate the possible interventions to achieve the one with the best cost benefit. Usually, WDN's projects are designed for a life cycle of 20 years, but many systems are already operating longer than this, up to 100 years. Obviously, several maintenance services are necessary during this period, but the knowledge of how the system deteriorate can help the management of these services and keep the operational efficiency high, Therefore, in this work the deterioration of the components of a WDN is evaluated trough the economic losses generated and the pressure deficit. The pipe deterioration is modelled considering the roughness increases at a certain rate. Pumps deterioration are similarly modelled, considering the formation of internal roughness that increases the internal recirculation. Finally, the demand increase is also modelled considering the population growth rate and a minimum leakage flow rate. In addition, pressure dependent leakages are modelled as an orifice, maintaining its characteristics constant during the simulation. The results shown that maintenance services are essential for the efficient operation of WDNs, with leakages representing the most impactful problem.

Keywords
Water distribution networks, energy efficiency, life cycle, maintenance.

1 INTRODUCTION

Water Distribution Networks (WDNs) represent the major cost of the infrastructure of water supply systems. In addition, the performance of the networks directly reflects on operational costs, since pumping stations are designed to attend the consumers demand, in both quantity (flow) and quality (pressure). In addition, leakages are an imperative problem of WDNs, and the leaking flow impacts both in the energy costs of pumping stations as in the treatment of raw water. Therefore, the design of WDNs is one of the most important problems in water supply systems, and different approaches have been proposed to achieve solutions economic and technically feasible.

Firstly, the focuses was on the most cost effective solution, where the pipes diameter were minimized to achieve the lowest cost, while maintaining the minimum required pressure on the network [1]. As good as this approach sounds, as highlighted by different authors [2–4] this strategy can be harmful for other parameters, as water quality and the network resilience. Thus, multiobjective optimization and multicriteria analysis were studied trying to balance operational parameters with a reasonable performance in all aspects. [5] presented a methodology to improve the design of the WDN by increasing the diameter of main pipes to increase the energy recovery potential through a pump operating as turbine, thus reducing the net operational costs.

Even if an optimal solution is achieved, the operation of the WDN can create several conditions that deteriorate the infrastructure during its life cycle. Water quality, soil condition, laying methods, pipe material and pressure surges are just a few parameters that can create failure

CC BY NC SA

conditions of the system [6]. As highlighted by [7], age or the remaining life of the infrastructure are a good criterion for the management of the WDN. Although a pipe failure is very harmful to the system, it can also be quickly fixed, since it is easy to identify the problem. On the other hand, leakages can occur in several points of the infrastructure as small holes in pipes, joints and fittings, reservoirs floor and walls and consumers metering [8]. These small leakages are hard to identify, since there is no visual signal, and they tend to increase during the WDN life, as highlighted by [9] and reinforced by the results found by [10], that shown the change of old pipes was more effective for leakage control than pressure management.

When the WDN requires the use of pump stations, the design becomes even more complex. Pipes costs are inversely proportional to the energy costs for pumping, larger diameters reduce the head losses, and consequently, the required power by the pumps. [11] proposes a joint optimization for pipes diameters and pump selection, while [12] also includes the pump location as a problem in the WDN design. As can be expected, pumps also deteriorate during the WDN life, reducing its capacity and efficiency. In addition, its operational point can be drastically affected by the deterioration of other components, as pipes roughness and leakages.

Therefore, in this paper a sensitivity analysis is made to evaluate the energy, hydraulic and economic aspects of the deterioration of a WDN. Pipe deterioration is modelled considering roughness increase at a certain rate, as proposed by [13]. Pumps deterioration are similarly modelled, considering the formation of an internal roughness that increases the internal recirculation [14]. Finally, the demand increase is also modelled considering the population growth rate [15] and a minimum leakage flow rate, economically unfeasible to be fixed. According to the model proposed by [16], the minimum leakage can be calculated using the total length of the WDN and the population. A pressure dependent leakage is also modelled as an orifice. During the simulation time, all orifice features are kept constant. The results show that maintenance services are essential for the efficient operation of WDNs, with leakages representing the most impactful problem.

2 WDN DETERIORATION

2.1 General features of deterioration scenarios

WDNs are subject to deterioration over their lifetime in several ways. Pipes increase internal roughness due to factors such as water quality, pipe diameter, age, and the material [17]. Pipe roughness increases leads to an increase also to the system's head losses, and consequently, the operation characteristics of the system should be changed due to the pressure deficit created. The system's water demand is another important factor to be evaluated. Along the life cycle of the WDN, the water demand of the consuming population in addition to leakage tends to present changes. These changes in demand will directly impact energy consumption, as they can drastically modify the operating point of pumping stations. In addition, the pumps can also deteriorate over time due to misalignment, excessive vibration, and corrosion. As the pumps deteriorates, the hydraulic power provided is compromised and the WDN's overall operating efficiency is harmed.

Thus, this paper proposes to evaluate the parameters that can deteriorate the WDN efficiency over its lifetime. These are the roughness of the pipes, the water demand (consumption and leakages), and the deterioration of the pumps.

For this, the Matlab R2021 software is used for the elaboration of algorithms, which coupled to the Epanet hydraulic simulator [18] and using the Epanet-Matlab Toolkit developed by [19], performs the hydraulic simulations and the modifications to evaluate the proposed deterioration scenarios. For the analyses, 3 scenarios are prepared, as discussed in the following sections, in which certain parameters are changed over the life cycle of the WDN. In addition, a scenario

BY NC SA

considered ideal is elaborated to be used as a benchmark for comparison with the data observed in the other scenarios.

Some features used in hydraulic simulations are the same for all scenarios and are summarized as:

- All developed scenarios are simulated in 20-year life cycle periods. As the deterioration processes discussed in this paper occur slowly, hourly hydraulic simulations are performed in a period of 24 hours, and observed data in this period are considered to represent the average of a typical day of the year. In this way, the simulation of the entire life cycle of the WDN is reduced to 20 consecutive daily hydraulic simulations.
- The costs of water losses due to leaks and the costs of electrical energy with the operation of the pumps are considered. For the electricity costs, the tariff differences between on-peak and off-peak hours are considered, observing the current tariffs of the [20]. The costs of leakage are considered through the costs of water production [21].
- In order to evaluate the hydraulic performance of the WDN, with regard to the operating pressures, the critical nodes of the network are analysed for all scenarios. Critical pressure is given as the lowest value observed at any consumption node of the WDN.

2.2 Ideal scenario

In the ideal scenario, the hydraulic components of the WDN remain as the initial characteristics along the entire life cycle. However, the water demand and a given value of minimum leakages (economically unfeasible to be fixed) increase proportionally with population growth, as both features are not manageable through the WDN improvement. The method of demographic components adjusted to a 3rd degree polynomial is used to estimate the population in each year of the simulation. This method is used by the Brazilian Institute of Geography and Statistics and the United Nations [15,22]. The 3rd polynomial used is given by Equation 1, in which the parameters adopted were the values assigned to the state of Minas Gerais in Brazil [15].

$$Y = aX^3 + bX^2 + cX + d \tag{1}$$

Where:

Y [n°]: Population;
X: Year; and
a, b, c and d: adjustment coefficients of the polynomial equation (a=0.0000004335; b=-0.0000023355; c=-0.0007653779; d=1.0003835392)

The minimum leakages considered in the WDN are calculated according to Equation 6, described in more detail in section 2.4. This ideal scenario is used as a benchmark for comparison, as this is the best situation among the scenarios evaluated.

2.3 Scenario for evaluation the deterioration of pipes

The process of deterioration of pipes can be classified in two categories: i) structural deterioration, which reduces the ability of pipes to withstand mechanical stress; and ii) functional deterioration, associated with the increase in the roughness of the internal surface of the pipes, reducing their hydraulic capacity [17,23]. Pipes age, material, diameter, the characteristics of the surrounding soil, external load and water quality are among the main factors related to the pipe deterioration rate [17].

Since in this paper the risk of failure is not considered, only the functional deterioration is evaluated. This deterioration relates to the pressure drop in the pipeline Δh, and can be calculated using the Hazen-Williams equation (Equation 2).

$$\Delta h = 10{,}653\left(\frac{Q}{C}\right)^{1{,}85} D^{-4{,}87} L \tag{2}$$

Where:

Q [m^3/s]: is the flow in the pipe;
L [m]: is the length of the pipe;
D [m]: is the diameter of the pipe; and
C: is the Hazen-Williams head loss coefficient.

The increase in pipe roughness changes the head loss of the WDN. This change reduces the available pressure for the consumers and, in critical conditions, can lead to an intermittency in the supply. To maintain the pressures above the minimum required, it is necessary to change the operational rules of the system and increase the hydraulic power by starting more pumps or increasing its speed when variable speed drives are available. In both cases, the adopted measures will reflect in an increase of electric energy consumption.

To consider its impact a new roughness is calculated each year using the methodology presented in [13], where the roughness increase at a constant rate as shown in Equation 3.

$$C = 18{,}0 - 37{,}2 \times log\left(\frac{e_0 + at}{D}\right) \tag{3}$$

Where:

C: Hazen-Williams Parameter;
e_0 [mm]: Initial absolute roughness;
a [mm/year]: Roughness increase rate;
t [years]: Time; and
D [m]: pipe diameter.

The values of initial roughness and rate of increase are respectively 0.18 mm and 0.094488 mm/year [13]. However, these values contain many associated uncertainties, due to direct measurement difficulty [17,24]. Thus, as a way to evaluate the uncertainties and the impact of the roughness increase in the WDN operation, different rates re used as shown in Table 1.

Table 1. Rates of increase in the internal roughness of pipes

Rates [mm/year]				
50%	80%	Ref:100%	120%	150%
0.047244	0.07559	0.094488	0.113386	0.141732

2.4 Scenario for evaluating the leakage rate

Water losses correspond to the water volume distributed but not accounted [25]. At WDN, this volume encompass actual losses, such as leakages, and apparent losses, such as clandestine connections. To calculate the flow of water lost through a single leakage, the formulation given by equation 4 is often used [26–28].

$$q = C_d A (2gh)^{0{,}5} \tag{4}$$

Where:

q [m^3/s]: is the leakage flow;
C_d : is the emission coefficient;
A [m^2]: is the leakage area;
g [m/s^2]: is the gravitational acceleration; and

h [m]: is the pressure load on the leakage.

However, for application in real WDN and hydraulic modelling through software, such as EPANET, Equation 4 can be rewritten in a more general way, according to Equation 5 [18,26]. In that case, the values of C_d and y must be calibrated to the WDN under study.

$$q = C_d h^y \tag{5}$$

Where:

C_d : is the emission coefficient; and
y: is the emission exponent.

By default, the EPANET software uses the value of the emission exponent equal to 0.5, often used in the modelling of holes and nozzles. In addition, studies suggest similar values for WDN [26,29]. This leakage modelling is used to represent water losses that can be fixed, as they have a great economic impact. However, small leakages, as observed in domiciliary connections and pipe junctions, are hard to identify, resulting in a cost/benefit relation that is not attractive to fix it. Thus, the proposal given by [16] is used to estimate the minimum losses in the WDN. This formulation considers the network length and the number of consumers to estimate the minimum value of leakages, as shown in Equation 6.

$$q_{min} = 54 + 2.7\,(\frac{N_p}{L_r}) \tag{6}$$

Where:

q_{min} [m^3/km]: is the minimum volume of leakages to be considered;
N_p : is the number of customers served; and
L_r [km]: is the total length of the network pipes.

In the equations proposed for modelling the leakages, it is observed that the volume of leakages changes with the variation of pressure at the node when modelled through Equation 5 (daily variation), and with the variation of the population when modelled by Equation 6 (yearly variation). Finally, as done for the pipe roughness, the emission coefficient is changed to different values to evaluate the sensitivity of the WDN to different water losses percentages, as shown in Table 2.

Table 2. Values used for the emission coefficient

Values used for the C_d				
25%	70%	Ref: 100%	130%	175%
0.125	0.350	0.500	0.650	0.875

2.5 Scenario for evaluating the deterioration of the pump

Pumps used in WDN are subject to deterioration over life cycle, especially if maintenance plans are not carried out properly. Several factors, such as cavitation, corrosion, incrustations, misalignments and excessive vibrations can interfere in this deterioration process, which reflects on the pump performance. [30] suggest that the two main mechanisms lead to pump performance decrease: ii) the development of an internal flow due to gaps; and ii) increase in the roughness of the internal surface of the pumps.

Following the methodology initially proposed by [30], and used by [14], the change in pump head over its life cycle can be calculated according to Equation 7.

$$H'_P = \omega \left(a \left(\frac{Q+R}{\omega}\right)^2 + b \left(\frac{Q+R}{\omega}\right) + c - K_T t \left(\frac{Q}{\omega}\right)^2 \right) \tag{7}$$

Where:

H'_P [m]: Corrected pump head;
ω: relative speed;
a, b and c: Pump curve coefficients;
Q [l/s]: Pump flow;
R [l/s]: Pump internal recirculation flow;
K_T : Internal roughness increase rate; and
t [h]: Cumulative operating time.

The pump internal recirculation flow R, varies with the initial head and with the clearance in the pump impeller wear ring, calculated according to Equation 8.

$$R = 2\gamma D_a H_P^{0,5} \left(\sqrt{\frac{c^3}{75 \times c + L}} - \sqrt{\frac{c_0^3}{75 \times c_0 + L}} \right) \tag{8}$$

Where:

γ [N/m^3]: Specific gravity of water;
D_a [mm]: Wear ring diameter;
H_p [m]: Initial pump head;
L [mm]: Axial length of wear ring;
c [mm]: Wear ring clearance; and
c_0 [mm]: Initial wear ring clearance.

The clearance in the pump impeller wear ring (c) increases with a deterioration parameter β, given according to Equation 9.

$$c = c_0 \times ln(\beta t + e) \tag{9}$$

Where:

c [mm]: Wear ring clearance;
c_0 [mm]: Initial wear ring clearance;
β: Parameter of wear ring deterioration;
t [h]: Cumulative operating time; and
e: Euler number.

There are no reference values for the deterioration parameters K_T and β. Ideally, both should be obtained through calibrations made for each situation. However, in the elaboration of this proposal, the values given by [14], respectively 1.0×10^{-9} and 5.0×10^{-3}, are used as reference. To analyze the impact of severe and mild deteriorations, other rates are used. The rates are changed in values between 10 times lower and 10 times higher than the reference values, as shown in Table 3.

Table 3. Values used in the parameters K_T and β

Parameter change rates K_T and β

coefficient	10%	20%	50%	100%	200%	500%	1000%
$K_T \times 10^{-9}$	0.1	0.2	0.5	1.0	2.0	5.0	10
$\beta \times 10^{-3}$	0.5	1.0	2.5	5.0	10.0	25	50

2.6 Case study

For the development of the case study, the WDN called Anytown is used, conceived by [31], shown in Figure 1. The WDN consists of 40 pipes and 19 nodes supplied by two tanks and a pump station connected to a reservoir used as the only water source for the system.

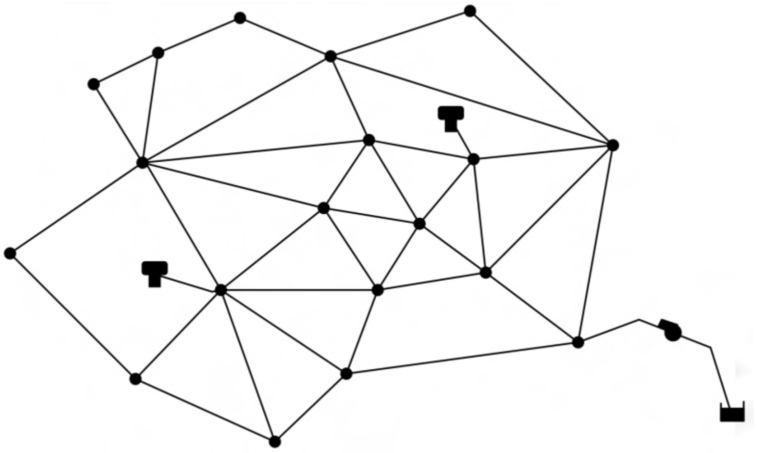


Figure 1. Water distribution network Anytown

All proposed scenarios and their respective variations are applied to this WDN. The annual total costs are calculated according to Equation 10. In addition, for each year of operation the minimum hourly pressure of the network is determined.

$$C_{year} = C_{typical} \times D_{year} \tag{10}$$

Where:

C_{year} [R\$]: is the total cost during the year of operation;

$C_{typical}$ [R\$]: is the cost during a typical day of the year of operation; and

D_{year}: is the number of days in the year.

Typical day costs are calculated as the sum of electricity costs during the day of operation, summed to the costs related to leakage during the same period, according to Equation 11.

$$C_{typical} = (D_{ener} \times P_{max}) + \sum_{i=1}^{24} (P_{(i)} \times C_{ener(i)}) + (Vol_{(i)} \times C_{vol(i)}) \tag{11}$$

Where:

$C_{typical}$ [R\$]: is the total cost during the typical day;

$P_{(i)}$ [kwh]: is the energy consumed in the hour *i*;

$C_{ener(i)}$ [R\$/kwh]: is the cost of electricity in the hour *i*;

D_{ener} [R\$/kW]: is the cost with electrical energy demand;

P_{max} [kW]: is the maximum power achieved in the period considered;

$Vol_{(i)}$ [m^3]: is the volume of water lost in the hour *i*; and

$C_{vol(i)}$ [R\$/m^3]: is the cost of producing water in the hour *i*.

For electricity costs, power demand and energy tariffs were considered at peak times (period between 5 and 7 PM), where tariffs are more expensive, and at off-peak times, when tariffs are cheaper, according to the current rates of the Energy Company of Minas Gerais in Brazil [32]. Leakage costs are calculated as the product of the cost of water production and the volume of water lost [21]. Table 2.4 summarizes the costs considered.

Table 4 – Values used to calculate total costs

Off-peak times		Peak times		Water production cost [\$/m^3]
Electricity [\$/Kwh]	Demand [\$]	Electricity [\$/Kwh]	Demand [\$]	
0.35666	13.95	0.53425	43.85	0.30

3 RESULTS

3.1 Pipe deterioration

Figure 2 shows the total annual costs, for each rate of increase in the internal roughness of the pipes evaluated. The overlap of the lines indicates that the different rates of increase in roughness do not cause direct changes in operating costs, since the pump operating point is not altered by changes in roughness. This occurs because the modeling of the WDN is demand driven, so the water consumption remains the same despite any pressure variation.

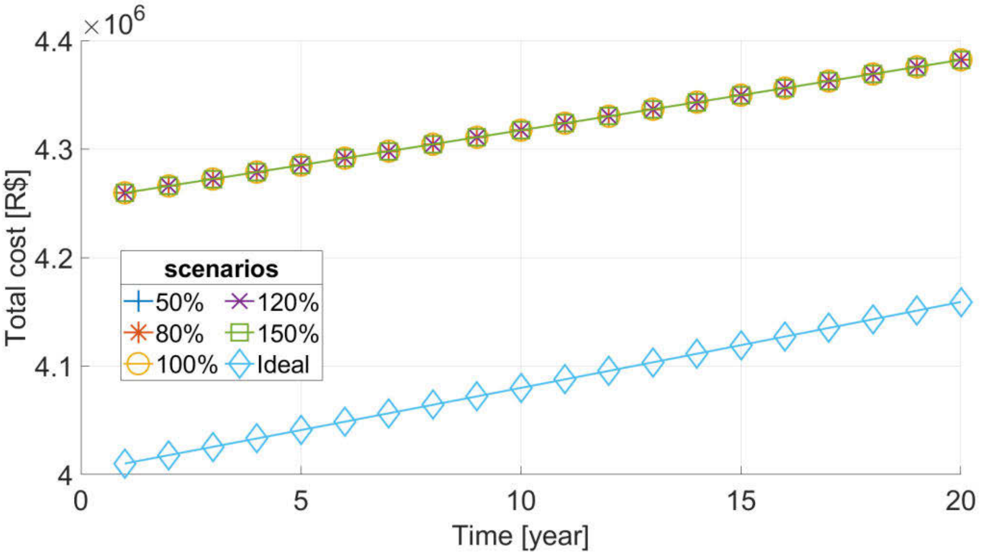


Figure 2. Total costs for each roughness increase rate

However, the operating pressures of the WDN are constantly reducing, as shown in Figure 3, both due to population growth (water demand) and the increase in pipe roughness, which increase the head losses. Thus, to maintain a minimum pressure required, it would be necessary to change the operating point of the pump, or add another one, which would change the operating costs. It's worth noticing in Figure 2 that, compared to the ideal scenario, the operating costs are 6.2% higher in every year, result of the addition of the minimum leakage.

Figure 3 shows the critical pressures of the WDN each year during its life cycle. The pressure observed in the initial year is the same for all scenarios, however for the different rates of increase in roughness there is a significant deviation in the behavior of pressures over time. In the scenario where the average rate (0.09448 mm/year) is used, the critical pressure in the WDN decreases until the value of 20 m (minimum required pressure according to [13]) in the 11th year and 10 m (minimum required pressure according to the [33]) in the 20th year of operation. When the roughness increase rate is 50% higher than the average value (0.141732 mm/year), the worst case scenario considered, the pressure of 20 m is reached in the 8th year and the 10 m in the 15th year of operation, earlier compared to the average scenario. In the case of the best scenario, with the roughness increase rate 50% lower than the average value (0.047244 mm/year), the pressure of 20 m is reached only in the 18th year of operation and the pressure of 10 m is always satisfied.

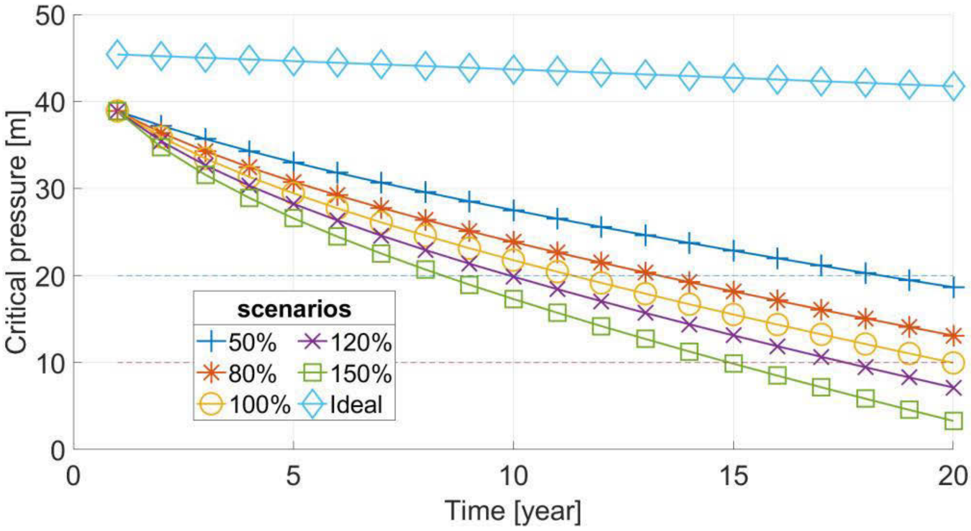


Figure 3. Critical pressures for each roughness increase rate

With the roughness increase rate 50% higher than the average value, the critical pressure at the end of the WDN lifecycle is 67.1% lower than the average rate scenario. When the roughness increase rate is 50% less than the average value, the critical pressure at the end of the WDN life cycle is 86.5% greater than the average rate scenario. The critical pressure values presented show how harmful the pipe deterioration can be for the WDN operation. On the other hand, when the deterioration is not severe, the pipe roughness has a very small impact, which shows the importance of the knowledge of system characteristics, especially water quality parameters, pipe material and soil characteristics, which can be an indicative of potential problems. Table 5 presents the critical pressures and the percentages in relation to the ideal scenario, for the pressure value at the beginning and end of the WDN life cycle.

Table 5. Critical pressures at the beginning and end of the life cycle for each roughness increase rate and for the ideal scenario

	Scenarios					
Scenarios	Ideal	50%	80%	Ref: 100%	120%	150%
Rates [mm/year]	-	0.047244	0.07559	0.094488	0.113386	0.141732
Pressure at the beginning of the life cycle [m]	45.4	38.9	38.9	38.9	38.9	38.9
End of life pressure [m]	41.8	18.7	13.1	10.0	7.1	3.3
Difference in relation to the ideal	-	-55.3%	-68.7%	-76.1%	-83.0%	-92.1%

3.2 Leakages

Figure 4 shows the total costs for each year when only changes in the emission coefficient were evaluated. Operating costs are significantly impacted by leakages, with the highest total values observed among all the scenarios evaluated in this work. This because, in addition to increasing the costs of water lost through the modelled orifices, the pump operation point is considerably altered due to the greater volume pumped to meet the demand for leakages.

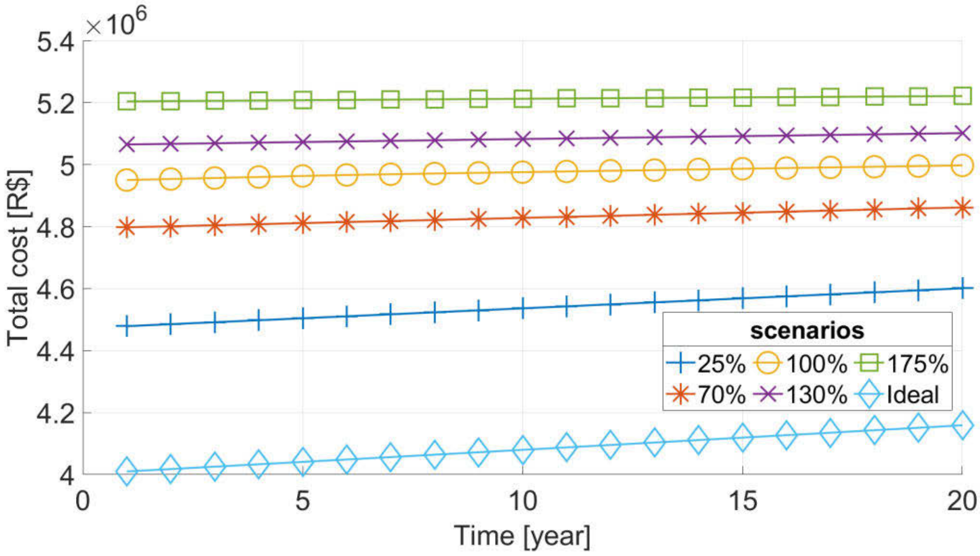


Figure 4. Total for different emission coefficients

When the emission coefficient is 75% higher than the average value (0.875), the total costs in the final year of the WDN life cycle are 4.5% higher, and when the emission coefficient is 75% lower (0.125), the total costs is 7.9% lower. Compared to the ideal scenario, the operating costs for the different leakage percentages are always higher as can be seen in Figure 4. Table 6 presents the total costs for the end of life cycle of the WDN and the percentage relationship with the ideal scenario.

Table 6. Total costs in the final year of the WDN life cycle for each emission coefficient compared to the ideal scenario

Scenarios	Ideal	25%	70%	Ref: 100%	130%	175%
Coefficients	0	0.125	0.350	0.500	0.650	0.875
Leakage index	0	9.0%	14.4%	18.9%	21.8%	25.5%
Cost [R$]	4,159,130	4,600,970	4,861,670	4,997,300	5,101,350	5,220,960
Difference in relation to the ideal	-	+10.6%	+16.9%	+20.2%	+22.7%	+25.5%

Critical pressures in this scenario vary for each emission coefficient considered, as shown in Figure 5. With the emission coefficient equal to 0.125, the critical pressure is 56.0% greater than the average value at the end of the life cycle. With the emission coefficient equal to 0.875, the critical pressure is 42.5% lower than the average value. The critical pressures in the initial year vary from 14.4 m to 34.6 m, respectively being the worst and best cases evaluated.

BY NC SA

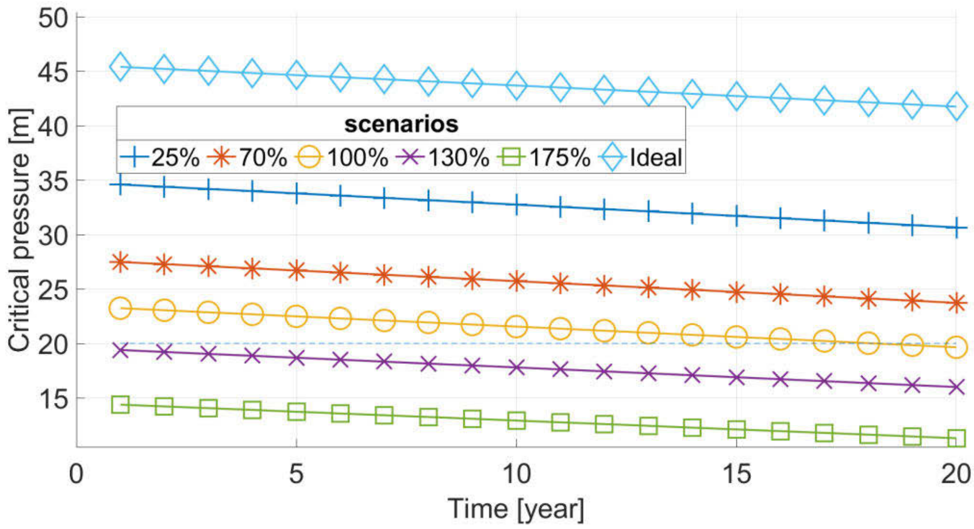


Figure 5. Critical pressures each year for the different emission coefficients

When the pressure of these scenarios is compared with the ideal scenario, it is observed that it is always lower within the evaluated life cycle. The higher the emission coefficient, the worse it is.

It is also worth mentioning that the leakage percentages observed in the simulations are still lower than those observed in Brazil, which presents an average of 39.2 %, but with cities operating with values well above [21]. This corroborates the importance of leakage control strategies in WDN.

3.3 Pump deterioration

Figure 6 shows the total costs each year over the life cycle of the WDN, for each pump deterioration rate. Contrary to what would be intuitively expected, as pump deterioration rates increase, operating costs decrease. A maximum increase of 1.3% is observed in the final year of the life cycle of the WDN.

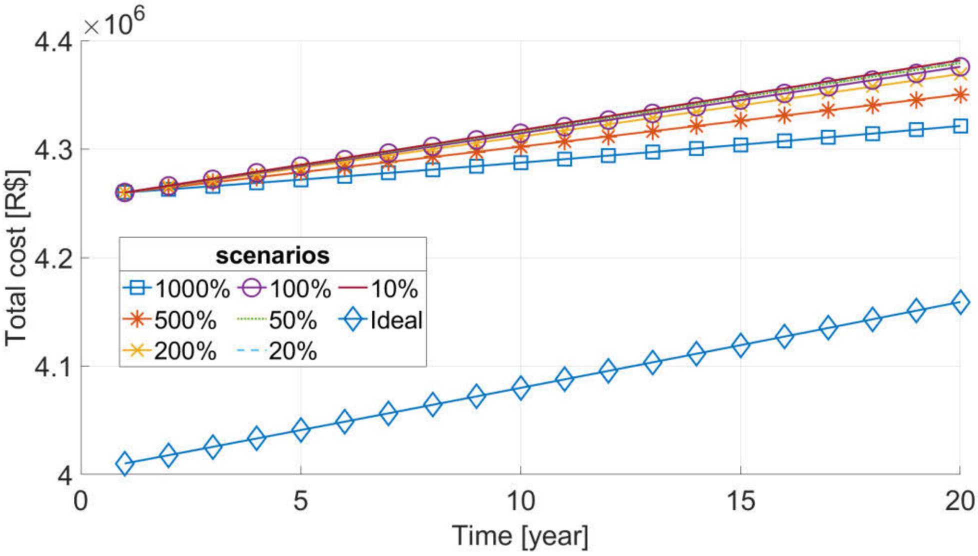


Figure 6. Total costs each year for different pump deterioration rates

This because, as the rate of deterioration of the pump increases, the head decrease according to the formulation used. Once again, as the model is demand driven, the power required to deliver

BY NC SA

water to the consumers is lower, as the head for the same flow is now lower. This behaviour reflects on the pressure of the WDN as will be following shown. However, it is important to emphasize that in this case the operating costs should not be evaluated individually, as the deterioration of the pump can also lead to supply failures due to mechanical problems resulting from the deterioration of the pump.

Figure 7 shows the critical pressures each year in the WDN, for each pump deterioration rate. Critical pressures decrease in all scenarios over time, but the decrease is more pronounced as the rate of deterioration increases. When the pump deterioration rate is 10 times the average value, the critical pressure at the end of the WDN life is 3.5% lower. When the pump deterioration rate is 10 times lower than the average rate, the system pressure remains practically the same. Although these scenarios present variations in critical pressures, with the methodology and rates adopted for the pump deterioration scenarios in this work, the values suggest that these changes are not very expressive for the time considered. Thus, if the pumps are capable to deliver the required flow with a minimum pressure, they should not be the main focus for the system improvement, since, as already seen in items 3.1 and 3.2, pipes deterioration and leakages have a greater impact in the operation. However, it is emphasized that maintenance plans must take place to avoid sudden shortages caused by mechanical failures, which can become more frequent due to the deterioration process.

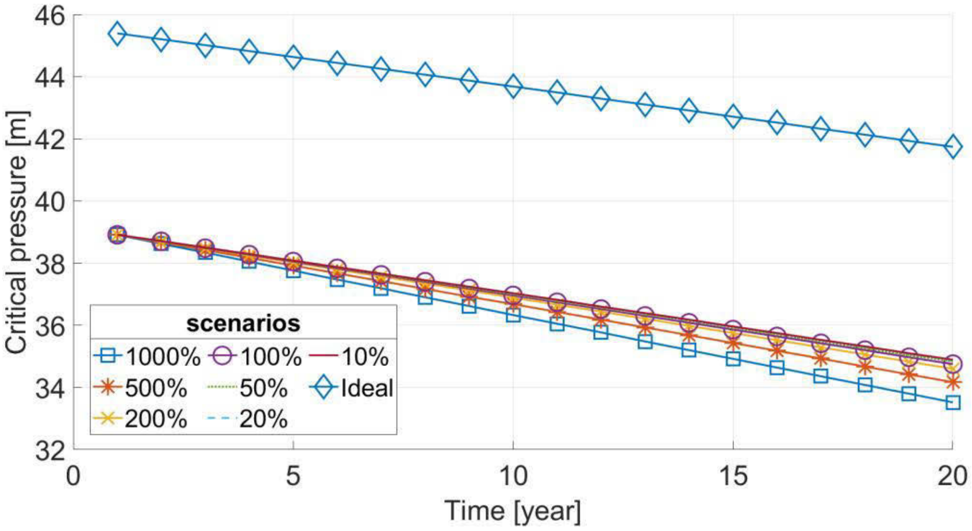


Figure 7. Critical pressures for each rate of deterioration increase

3.4 Evaluation between the average scenarios

Figure 8 shows the total annual costs for the three average scenarios evaluated. In addition, it shows the ideal scenario, presented earlier, and a new intermediate scenario. In this intermediate scenario all three average parameters analyzed were considered simultaneously. The main impact on total costs is clearly caused by leakage, as seen by the different amplitude in the curves. Relative to the ideal scenario, the total costs each year are 23.4% higher in the leakage scenario and 6.2% higher in the pump deterioration and roughness change scenarios. The growth of total costs in each of the scenarios follows a similar rate, given by the slope of the curve, which is related to the population growth rate. The intermediate scenario presents total costs close to the values observed in the leakage scenario. This is because in the intermediate scenario, leakage is also considered in the simulations, in addition to the other parameters.

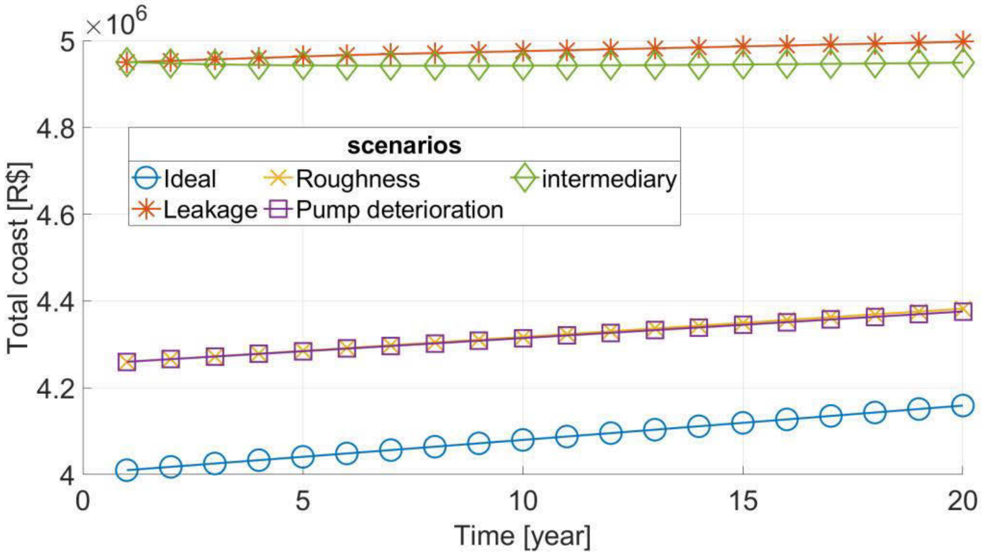


Figure 8. Total costs for each year of the evaluated scenarios

Figure 9 shows the critical pressures for the three average scenarios evaluated above, in addition to the ideal scenario and intermediary. Until the middle of the WDN life cycle, the scenario with the lowest critical pressures is observed in the leakage scenario. However, from the 10th year onwards, the pressures are lower in the scenario where the roughness of the pipes changes. The increase in the roughness of the pipes with the adopted rate causes the pressures to decrease significantly with time. When associated with the additional demand over the years, due to population growth, the head losses observed in the WDN significantly increase, reaching levels where intervention would be necessary for the WDN to continue operating satisfactorily. In relation to the ideal scenario, the pressures of the roughness change scenario are 4.2 times lower at the end of the life cycle. In the leak and pump deterioration scenarios, the pressures are 2.1 and 1.2 times lower, respectively.

When the increase in the roughness of the pipes is added to the other parameters evaluated, in the intermediate case, we observe the worst scenario in terms of pressure in the network. In this case, the pressure is always lower than all the scenarios evaluated, reaching values that would require intervention in the first years of operation.

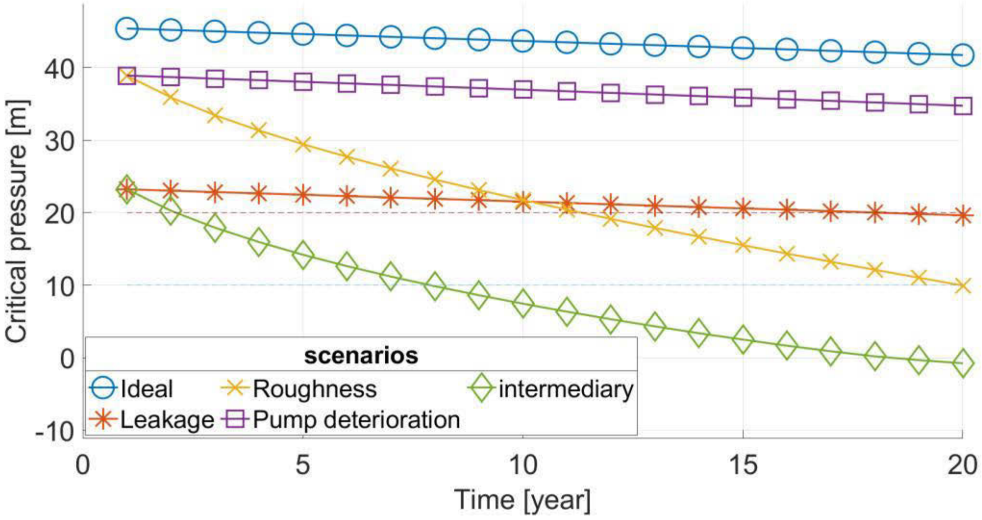


Figure 9. Critical pressures each year for the evaluated scenarios

The individual results shown are for evaluating the sensitivity of each parameter. However, in real systems the deterioration of the parameters occurs simultaneously, as in the intermediate scenario, and can cause significant impacts. Decisions about network maintenance and rehabilitation plans need to be made based on the impact (cost/benefit) of each of the parameter, to keep the WDN operating according to the standards.

4 CONCLUSIONS

In this paper, it was evaluated how the deterioration of the WDN, through the alteration of the internal roughness of the pipes, the deterioration of the pump and the change in the volume of leakages, impacted the WDN throughout its life cycle, in terms of operating costs and operating pressures. With the scenarios evaluated, the following conclusions could be drawn:

- The results show that the deterioration of the network has significant impacts on its performance, which corroborates what is observed in practice. This shows the importance of monitoring the functioning of the WDN to verify its efficiency, in addition to assisting in decision-making on the implementation of rehabilitation plans.
- In terms of total costs, leakages were the main responsible for the direct impacts on cost growth. The increase in water demand due to leakages significantly alters the operating point of the pumping station, consequently increasing the energy consumption. In addition, a large percentage of the water produced is wasted.
- WDN operating pressures are significantly altered by increasing pipe roughness. The observed decrease is a consequence of the increase in the head losses of the system. Considering that WDN work for life cycles many times longer than 20 years, the rehabilitation or replacement of pipe sections must be properly planned in order to maintain efficiency in the network operation.

5 REFERENCES

[1] Mala-Jetmarova H, Sultanova N, Savic D. Lost in optimisation of water distribution systems? A literature review of system operation. Environmental Modelling & Software. 2017;93:209–254.

[2] Kapelan ZS, Savic DA, Walters GA. Multiobjective design of water distribution systems under uncertainty. Water resources research. 2005;41.

[3] Shokoohi M, Tabesh M, Nazif S, et al. Water quality based multi-objective optimal design of water distribution systems. Water Resources Management. 2017;31:93–108.

[4] Tanyimboh TT, Kalungi P. Multicriteria assessment of optimal design, rehabilitation and upgrading schemes for water distribution networks. Civil Engineering and Environmental Systems. 2009;26:117–140.

[5] Meirelles GL, Luvizotto E, Brentan BM, et al. Leakage Control and Energy Recovery Using Variable Speed Pumps as Turbines. Journal of Water Resources Planning and Management. 2018;144:04017077.

[6] Malm A, Ljunggren O, Bergstedt O, et al. Replacement predictions for drinking water networks through historical data. Water research. 2012;46:2149–2158.

[7] Burn S, Marlow D, Tran D. Modelling asset lifetimes and their role in asset management. Journal of Water Supply: Research and Technology—AQUA. 2010;59:362–377.

[8] Trow S, Farley M. Developing a strategy for leakage management in water distribution systems. Water Science and Technology: Water Supply. 2004;4:149–168.

[9] Engelhardt MO, Skipworth PJ, Savic DA, et al. Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. Urban Water. 2000;2:153–170.

[10] Lambert AO. International report: water losses management and techniques. Water Science and Technology: Water Supply. 2002;2:1–20.

[11] Geem ZW. Harmony search optimisation to the pump-included water distribution network design. Civil Engineering and Environmental Systems. 2009;26:211–221.

[12] Wang B, Liang Y, Zhao W, et al. A continuous pump location optimization method for water pipe network design. Water Resources Management. 2021;35:447–464.

[13] Sharp WW, Walski TM. Predicting internal roughness in water mains. Journal-American Water Works Association. 1988;80:34–40.

[14] Nault J, Papa F. Lifecycle Assessment of a Water Distribution System Pump. Journal of Water Resources Planning and Management. 2015;141:A4015004.

[15] IBGE, editor. Projeções da população: Brasil e unidades da Federação, revisão 2018. 2a edição. Rio de Janeiro: IBGE, Instituto Brasileiro de Geografia e Estatística; 2018.

[16] Ahopelto S, Vahala R. Cost–Benefit Analysis of Leakage Reduction Methods in Water Supply Networks. Water. 2020;12:195.

[17] Abd Rahman N, Muhammad NS, Abdullah J, et al. Model Performance Indicator of Aging Pipes in a Domestic Water Supply Distribution Network. Water. 2019;11:2378.

[18] Rossman LA. EPANET 2 USERS MANUAL. 2000;200.

[19] Eliades DG, Kyriakou MS, Vrachimis SG, et al. EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB. 2016.

[20] ANEEL. Agencia Nacional de Energia Elétrica - ANEEL [Internet]. Tarifa de Energia Elétrica. 2021 [cited 2021 Jul 25]. Available from: https://www.aneel.gov.br/tarifas.

[21] SNIS. Sistema Nacional de Informações sobre Saneamento [Internet]. Brasília: Ministério do Desenvolvimento Regional; 2019 [cited 2021 Jul 18]. p. 183. Report No.: 25. Available from: http://www.snis.gov.br/diagnosticos.

[22] Borges A da S, Marques C da S, Brito LPG de, et al. Projeções populacionais no brasil: subsídios para seu aprimoramento. Anais. 2016;1–26.

[23] Berardi L, Giustolisi O, Kapelan Z, et al. Development of pipe deterioration models for water distribution systems using EPR. Journal of Hydroinformatics. 2008;10:113–126.

[24] Kang D, Lansey K. Demand and Roughness Estimation in Water Distribution Systems. Journal of Water Resources Planning and Management. 2011;137:20–30.

[25] Pereira SF, Tinôco JD. Avaliação das perdas de água em sistema de abastecimento de água: estudo no setor parque das nações, Parnamirim/RN. Revista Eletrônica de Gestão e Tecnologias Ambientais. 2021;32–45.

[26] Boian RF, Macedo DO, Oliveira PJA de, et al. Comparação entre as equações de FAVAD e geral para avaliação da vazão perdida por meio de vazamentos em sistemas urbanos de distribuição de água. Eng Sanit Ambient. 2019;24:1073–1080.

[27] Braga AS, do Nascimento G, Fernandes CVS. Vazamentos por fendas longitudinais: uma abordagem experimental. Florianópolis/SC: ABRH; 2007. p. 8.

[28] Piller O, van Zyl JE. Incorporating the FAVAD Leakage Equation into Water Distribution System Analysis. Procedia Engineering. 2014;89:613–617.

[29] Macedo DO, Alves P, Veríssimo Gonçalves F, et al. Efeito de fatores geométricos e hidráulicos sobre a vazão perdida e o expoente de vazamento em sistema de distribuição de água. Revista AIDIS de Ingeniería y Ciencias Ambientales Investigación, desarrollo y práctica. 2018;11:238–250.

[30] Richardson S, Hodkiewicz M. Modeling Tool to Support Budgeting and Planning Decisions for Pump Overhauls. Journal of Water Resources Planning and Management. 2011;137:327–334.

[31] Walski TM, Brill ED, Gessler J, et al. Battle of the Network Models: Epilogue. Journal of Water Resources Planning and Management. 1987;113:191–203.

[32] CEMIG. Valores de tarifas e serviços [Internet]. Cemig. [cited 2022 May 1]. Available from: https://www.cemig.com.br/atendimento/valores-de-tarifas-e-servicos.

[33] ABNT. Estudos de concepção de sistemas públicos de abastecimento de água. NBR 12211/1992 1992 p. 14.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14056

MULTI-LEAK DETECTION AND ISOLATION IN WATER DISTRIBUTION NETWORKS

Débora Alves[1], Joaquim Blesa[2], Eric Duviella[3], Lala Rajaoarisoa[4]

[1] Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, 08222 Terrassa, Barcelona, Spain
[2] Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain
[1,3,4] IMT Nord Europe, Univ. Lille, CERI Digital Systems, F-59000Lille, France
[2] Serra Húnter Fellow, Universitat Politècnica de Catalunya (UPC), Automatic Control Department (ESAII), Eduard Maristany, 16 08019, Barcelona, Spain

[1] *adeboracris@gmail.com*, [2] *joaquim.blesa@upc.edu*, [3] *eric.duviella@imt-nord-europe.fr*,
[4] *lala.rajaoarisoa@imt-nord-europe.fr*

Abstract

Water Distribution Networks (WDNs) are complex systems that faces the challenge of detecting and locating water leaks in the system as quickly as possible due to the need for an efficient operation that satisfies the growing world demand for water. This paper introduces an entirely data-driven leak detection and localization method based on flow and pressure analysis. The method can be divided into leak detection when the fusion data of the flow and pressure measurements are studied, thus obtaining the instant where the leak starts and if there is more than one simultaneous leak (multi-leak) occurring in the network. The second part is the leak localization using the fusion of the pressure residues by applying the radial base function (RBF) interpolation to obtain the network zone with the highest leak probability. The method is validated using the L-TOWN benchmark proposed at the Battle of the Leakage Detection and Isolation Methods (BattLeDIM) 2020 challenge.

Keywords
Leakage detection, Leakage localization, Water distribution networks (WDN), Radial base function (RBF), Interpolation.

1 INTRODUCTION

The worldwide growing demand for water generates a constant concern about the proper functioning of the Water Distributions Networks (WDNs). Therefore, the search for new strategies for detecting, estimating, and locating leaks is an important topic, as water leaks are one of the main factors in water loss. In addition, it can produce substantial economic losses, infrastructure damage, and health risks. Because of this, many studies have been carried out to develop WDN leak detection and location methods. Some of the techniques are based on model-based approaches, which provide adequate performance. Still, they rely on the calibration of accurate models and data availability for all possible complex scenarios that some networks are not available. At the same time, data-driven techniques combine standard operation data and topological information to detect and locate the presence of the leak, although they may produce less accurate results.

Works on leak localization applying model-based approaches compare simulated hydraulic information with actual measurements from the WDN; an example, the research in [1] is based on the analysis of pressure residues. In another work in [2], the authors use hydraulic models with AI methods. Moreover, in [3], performing sensitivity analysis and a search space reduction approach to find the leak's location. In [4,5], combine the use of standard operation data and topological information. The particular method in [6] studies the effect of the extra flow when a leak occurs in the pressure sensors presented in the network. It aims at developing a relative incidence of a leak using network topology correlated with the flow and pressure measurement. In [7] has more details about the model-based and data-driven methods.

Another important fact is that in real WDNs, the system can instantly have more than one leak. The Battle of the Leakage Detection and Isolation Methods (BattLeDIM) [8] has raised this concern by presenting the L-Town network representing a small hypothetical town with 782 inner nodes, two reservoirs, and one tank. Several challenges were presented in this challenge. One of them was the rapid detection of leaks and the fact that the system had multiple leaks during the year. The research [9] presented a method of leak detection and estimation using information from flow sensors installed in the reservoir. The technique can give an estimate of the magnitude of the leak, and with a presence of a second leak, the estimation is the sum of these two leaks, being necessary for a human intervention to evaluate the presence of a multi leak.

This work presents a complementary study of leak detection of work [9]. Presenting an entirely data-driven technique to leak detection and localization that tackle multi leaks problems that require minimal topological knowledge of the network and measurements from pressure sensors distributed at a set of inner nodes and flow sensors installed in the inlets. The case study of the L-Town network is analyzed to display the improvement of the method.

The rest of the document is organized as follows: Section 2 presents the leak detection and localization methodology. Section 3 shows the application and the results obtained in the L-TOWN benchmark proposed at the BattLeDIM. Finally, Section 4 concludes this work.

2 METHODS

An overview of the two steps of leak detection and location and the order in which they are applied is illustrated in Fig.1—describing the steps for obtaining the leak initiation time information and calculating the most likely zone to contain a leak. The first leak detection phase descends from the base of sensor fusion theory using the inlet flow and the pressure measurements of the WDN to generate virtual measurements, able to detect the start time of the leaks in a multi-leak scenario. In the second phase, the fused pressure residual of all sensors and the longitude, latitude, and elevation of each node is applied in the radial base function (RBF) interpolation method to determine a network zone with the fault. The two steps of leakage identification and leakage localization are described in detail.

3 LEAK IDENTIFICATION

The fundamental aspect of the detection phases represents the WDN inlet flow and pressure, approximating the current and historical data. Therefore, the demand and pressure forecasts in the WDN are out of the scope of this work. However, it can be assumed that a demand forecast method is calibrated using historical data of the WDN [10] and leak-free pressure estimations that can be computed through available historical data.

The first step of leak identification, LI-1, is the development of the fusion of flow and pressure data. This step transforms each hour of the day into different features, having 24 features, and

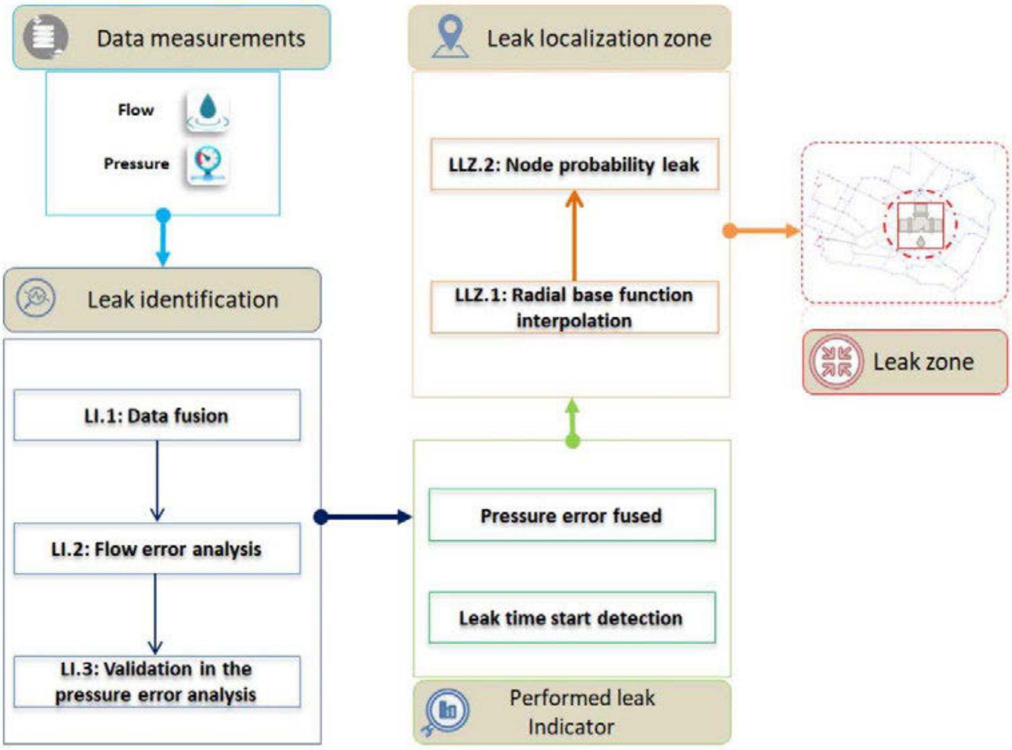

Data measurements
Flow
Pressure
Leak identification
LI.1: Data fusion
LI.2: Flow error analysis
LI.3: Validation in the pressure error analysis
Leak localization zone
LLZ.2: Node probability leak
LLZ.1: Radial base function interpolation
Pressure error fused
Leak time start detection
Performed leak Indicator
Leak zone

Figure 1. Flowchart of the leak detection and localization proposed method

their fusion improves leak detection thanks to reducing the uncertainties and noise in the measurement.

The first fusion data addressed will be the flow measurement, introduced in [9]. The current inflow y at time k is given as:

$$y(k) = \hat{y}(k) + e(k) \tag{1}$$

where $k = 0,1,2,3,\ldots$ denotes the discrete time corresponding to time $0, T_s, 2T_s, 3T_s, \ldots$, being T_s the sample time of demand forecasting model, $\hat{y}(k)$ is the demand forecast and $e(k)$ is the error that for this study is considered adjusted by a normal distribution (Gaussian) [10], represented by the notation $\mathcal{N}(\mu, \sigma^2(k+T))$ with mean μ and standard deviation $\sigma^2(k+T)$, where T is a periodic variation in time representing the different accuracy of the incoming demand in the periods of the days. In the case of the presence of a leak, i.e.,$l(k) > 0$, equation (1) leads to:

$$y(k) = \hat{y}(k) + e(k) + l(k) \rightarrow \hat{l}(k) = y(k) - \hat{y}(k) = l(k) + e(k) \tag{2}$$

where $\hat{l}(k)$ approximation of the leak size given by the difference between the actual and the estimated inlet flow, with a leak estimation error equal to the demand forecasting error. It is possible to generate different leak estimations using a time window, W considering the current inlet flow value and the previous values:

$$l(k) \approx \bar{l}(k) = \sum_{i=0}^{W-1} \frac{l(k-i)}{W} \tag{3}$$

an average leak estimation $\hat{\bar{l}}(k)$ can be computed at instant k applying the maximum Likelihood estimation method to the joint probability distribution of the W estimations fused in $\bar{l}(k)$

$$\hat{\bar{l}}(k) = \frac{\sum_{i=0}^{W-1} \frac{\hat{l}(k-i)}{\sigma^2(k-i)}}{\sum_{i=0}^{W-1} \frac{1}{\sigma^2(k-i)}} \tag{4}$$

Leak detection can be formulated as a change detection problem, $\hat{\bar{l}}$ will lead to small (but non-zero) values due to demand estimation errors in a no-leakage scenario. In contrast, its value will increase in a leakage scenario. Therefore, a threshold ∇ can be calculated to determine the value of $\hat{\bar{l}}$ above it, which can be assumed to be a leak in the WDN.

The value of ∇ can be calculated by applying equation (4) for leak-free historical data, considering the worst-case scenario ∇ equal to the maximum value of $\hat{\bar{l}}$ calculated for all leak-free historical data, referent the LI-2 step. Furthermore, once $\hat{\bar{l}}$ is above ∇ is considered a disturbance in the system alarming to a probable presence of a leak that needs to be validated with the study of data fusion of pressure measurements, which will be explained in the next topic.

Data fusion of pressure measurements is performed by analyzing pressure residues generated by comparing internal pressure measurements and leak-free pressure for each sensor, installed in the WDN, estimates such as:

$$r_i(k) = \hat{p}_i(c(k)) - p_i(c(k)) \qquad i = 1, \dots, s \tag{5}$$

where $r_i(k)$, $\hat{p}_i(c(k))$ and $p_i(c(k))$ are the residual, leak-free pressure estimation, and pressure measurement at inner node i, $c(k)$ is the operating condition at given instant k defined by inlet measurements and s is the number of inner sensors installed in the WDN. In the same way as equation (3), it is possible to generate different residuals analyses using a time window, W, (the same value of the leak estimations) considering the current residual pressure value and the previous values. The average pressure residuals $\bar{r}_i$ can be computed at instant k applying the maximum Likelihood estimation method to the joint probability distribution of the W residuals analyses fused in:

$$\bar{r}_i(k) = \frac{\sum_{i=0}^{W-1} \frac{r_i(k-i)}{\sigma^2(k-1)}}{\sum_{i=0}^{W-1} \frac{1}{\sigma^2(k-1)}} \qquad i = 1, \dots, s \tag{6}$$

The finite difference will be applied to demarcate the beginning and end of a leak in the system to the daily data of residuals fused $\bar{r}_i$. Being analyzed, the maximum value in every 24 hours, k_{day}. The finite difference corresponds to differential operation, an important concept in calculus commonly used to smooth nonstationary time series [12] expression of the form $f(x)$ to $f(x + b) - f(x + a)$. In this study, the difference value $\Delta\bar{r}_i$ is calculated as follows:

$$\Delta\bar{r}_i(k_{day}) = \max(\bar{r}_i(k_{day})) - \max(\bar{r}_i(k_{day} - 24)) \qquad i = 1, \dots, s \tag{7}$$

When a leak occurs in the WDN, all the measurements of the pressure sensors will be affected; nevertheless, if the sensors closest to the failure show more disturbance. Knowing that the network will be divided into α groups $G = \{g_1, \dots, g_\alpha\}$, with the region and the neighbouring sensors as a parameter. Moreover, the sum of the $\Delta\bar{r}_i$ of each group will be performed, being normalized in a range of [0,1].

In these analyses, a peak is produced in the signal when has a disturbance in the sensors, for example, when a leak starts or when it is fixed. To proceed with the leak detection method, a threshold, th, for each group $g_1, \dots, g_\alpha$, is calculated with the number of the sensors of the group divided by 3, the leak detection method can be computed by:

$$\Delta\bar{r}_{gi}(k_{day}) = \left\{\begin{matrix} \Delta\bar{r}_{gi}(k_{day}) & \Delta\bar{r}_{gi}(k_{day}) > th_{gi} \\ 0 & otherwise \end{matrix}\right. \qquad \text{i=1}, \dots, \alpha, \tag{8}$$

The $\Delta\bar{r}_{gi}$calculated in equation (8) is set to only present disturbances when a failure is similar to a leak in the system. To set the analysis for disturbances like a leak repair signature, the threshold of the first line must be set to $\Delta\bar{r}_{gi}(k_{day}) < -th_{gi}$.

With the study of equation (4), it is possible to analyse whether the WDN leaks, but it is limited to when there is only one leak in the system or when there are more leaks with time spaces of more than time window W. In other words, if multiple leaks co-occur or with a period smaller than W, the information from equation (4) will only show the sum of the magnitude of all leaks. However, with the validation of the information with the equation (8), it is possible to know when multiple leaks happen because it will present a peak in the analysis data, having a better result if the locations of the leaks are in different groups.

Leak localization zone

The interpolation of data for the WDN has already been studied in other works [2,5]. Still, as questioned in work [5], the interpolation of measured pressure to the nodes that do not have sensors trying to identify the fault at a node-level still has a long way to develop. However, the interpolation of leak indicators to determine the zone close to the sensors that have a fault is of great help for water companies as it will reduce the system zone for the leak's location.

To predict zones with unmeasured nodes the method will use the following information: (i) the average pressure residuals of equation (6) available from the installed sensors, (ii) the topological information of the nodes in the network, and (iii) the Radial basis function (RBF) interpolation technique.

RBF provides a very general and flexible way of interpolation in multidimensional spaces, even for unstructured data, where it is often impossible to apply polynomial or spline interpolation, see for more explanation [14-16]. Due to its good approximation properties, it was chosen in this work.

The method usually works in d dimensional Euclidean space which is $\mathbb{R}^d$ fitted with the Euclidean norm $||\cdot||$. The interpolation space consists of all functions of the form:

$$f(\underline{x}) = \sum_{j=1}^{N} \lambda_j \emptyset(||\underline{x} - \underline{x}_j||) \tag{9}$$

where $\underline{x}$ is a point in $\mathbb{R}^d$, $\underline{x}_j$ are the centre points for the RBFs (equation (6)), λ_j are coefficients to determine, N are points in this space at which the function to be approximated is known, and $\emptyset(r)$ is a radial basis function, set as a multiquadric problem:

$$\emptyset(r) = \sqrt{1 + \varepsilon^2 r^2} \tag{10}$$

where ε is the shape parameter (see [13]). The RBF interpolation can be used in any dimension; in this work, the dimensions used are the latitude, longitude, and elevation of each node in the WDN, and the average pressure residuals of equation (6) are the values to be interpolated.

4 CASE STUDY

The Battle of the Leakage Detection and Isolation Methods is a challenge provided by the organizers of the BattLeDIM [8]. The aim is to detect and locate several leaks in a hypothetical city created with this intent, as depicted in Fig. 2. the city is located in the Northern hemisphere and regroups a population of about 10,000 people. Thus, higher water usage is expected around July/August and lower in December/January. The network is divided into three distinct areas:

Area A is supplied by two reservoirs, each containing flow sensors; Area B that was installed with a pressure reduction valve (PRV) to help reduce background leakages; and Area C was installed with a pump and a water tank, with a flow sensor in this pump to control the flow that enters in the tank. In addition, has been installed in Area C 82 Automated Metered Readings (AMRs), which is a technology used in utility meters for collecting data that does not require physical access or visual inspection. The data can be transmitted to a central database, in this area, only ten regular sensors were distributed. Area C has a significant quantity of AMRS installed in the zone. Because of that, a model-based approach is a good option to solve the leak localization problem in this area.

In this challenge, the network can be divided into two distinct parts with different challenges: the first, Area A and Area C containing simultaneous leakage, and the second, Area C containing the AMR devices.

The leaks in Area A and B of the 2018 year will be addressed in this work. The data set of the BattLeDIM for this year contains the time and repair location of 9 pipe bursts that were fixed. Three types of leaks exist:

- Small background leaks with 1%–5% of the average inflow
- Medium pipe breaks with 5%–10%
- Large pipe bursts with leakage flow of more than 10% of the average system inflow (≈50l/s)

The water utility corrects significant leaks with a flow rate above 4.5 l/s after a reasonable time within two months. The leakages have two different time profiles: either abrupt bursts with

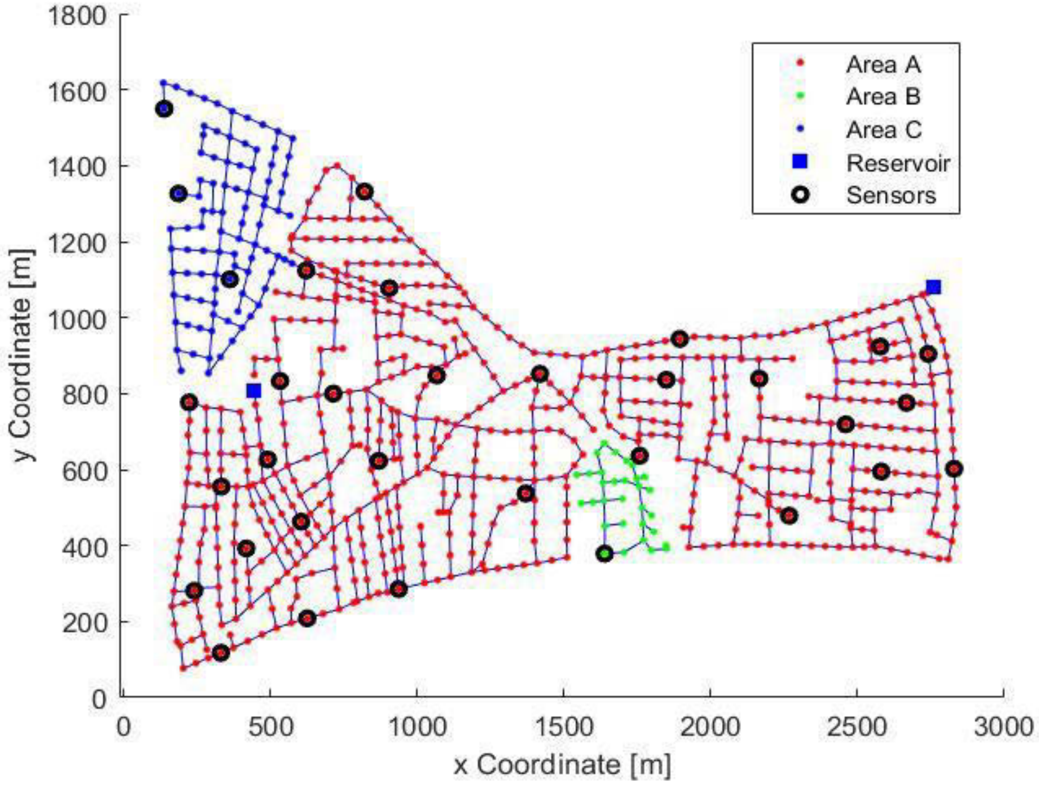


Figure 2. Overview of L-town water distribution network

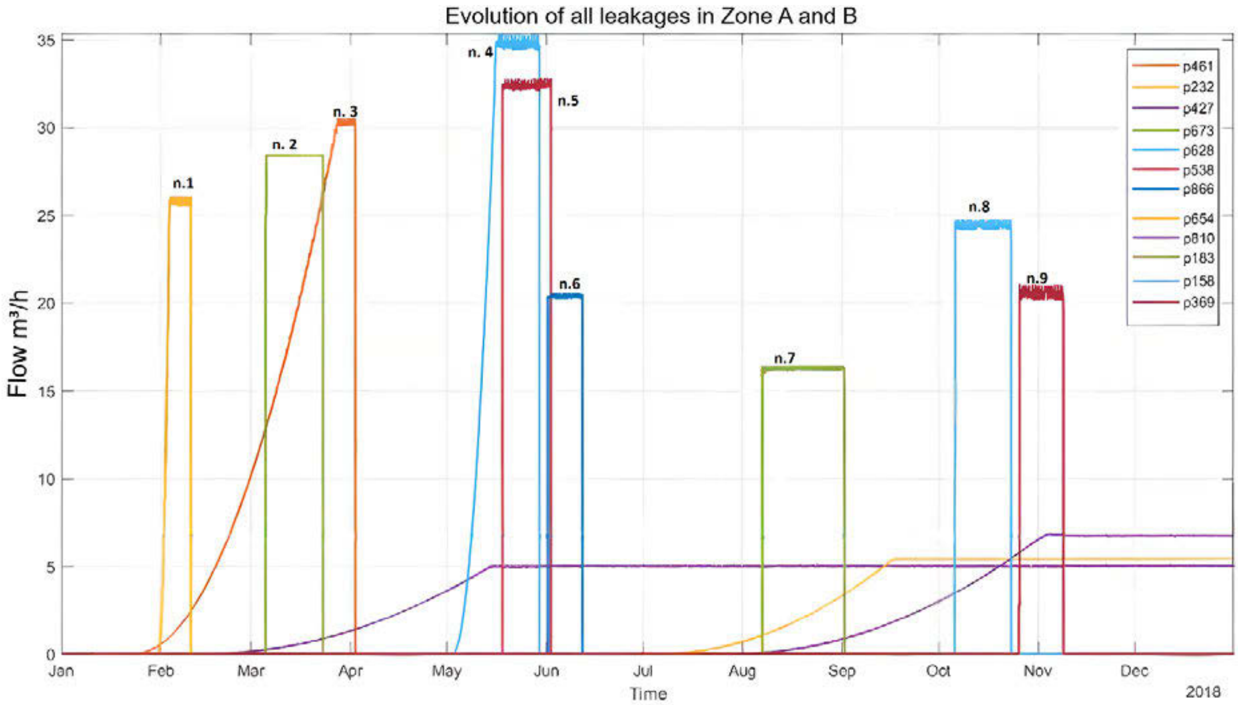


Figure 3. Evolution of leaks (m³/h) in Area A and B during 2018

constant leak flow rates or incipient leaks that evolve until significant outflow rates at which they remain constant. Fig.3 shows the 12 leaks in 2018, with outflow rates between 1.4 and 9.7 l/s (5 and 35 m ^{3}/h). Three leaks are not fixed, and nine leaks are repaired throughout the year that will be analysed in this paper in the highlighted order of n.1 to n.9.

To perform the first step of the proposed approach, it is necessary to define the sensors belonging to group G. In this work, the groups were obtained by the heuristic approach considering the neighboring sensors and the distance between them. The groups do not have the same number of sensors since group 1 has more sensors concentrated in the same area. Another factor is the use of the pressure sensor data in more than one group because if a leak happens in the border zone between groups, the fault will be identified in more than one group analysis.

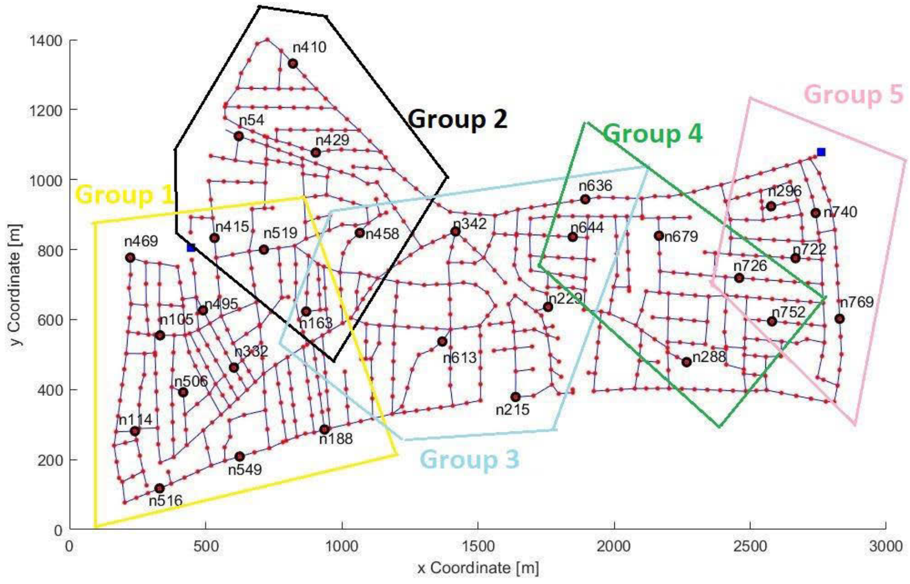

Group 1
Group 2
Group 3
Group 4
Group 5
n410
n54
n429
n415
n519
n458
n342
n469
n495
n105
n163
n332
n506
n114
n188
n549
n516
n613
n215
n229
n636
n644
n679
n296
n740
n722
n726
n769
n752
n288
y Coordinate [m]
x Coordinate [m]
1400
1200
1000
800
600
400
200
0
500
1000
1500
2000
2500
3000

Figure 4. Division of sensors into G groups

Six signals are used to leak detection: first, the $\hat{\bar{l}}$ is calculated with the inner flow measurement, equation (4), and the five-group signals $\Delta\bar{r}_g$ are calculated with the pressure data in equation (8).

Fig.5 presents the result of these six signals, the fig.5 (a) is the $\hat{\bar{l}}$ analysis, which is the first step to detecting a leak. A red circle is highlight for every time k that the threshold is transposit with a red line limited in the pressure analysis $\Delta\bar{r}_{gi}$, fig. 5 (b-f).

When these flow detections happen, it is necessary to validate with the $\Delta\bar{r}_g$ study. When abrupt bursts faults begin in the WDN, it is possible to remark a peak in the $\Delta\bar{r}_g$ analysis in the group more affected by the leak. This is the case of leaks number 1 and 7, and it is possible to point out that leak number 7 started hours before the analysis of the $\hat{\bar{l}}$ alarm the fault. A careful analysis needs to be made in cases where a multi-leak exists, that is leaks number 2-3, leaks number 4-6, e leaks number 8-9.

In leaks number 2-3, and an incipient leak begins in the pipe p427, which was not repaired, but the size magnitude is smaller than the other two, and it is impossible to detect it. The other two

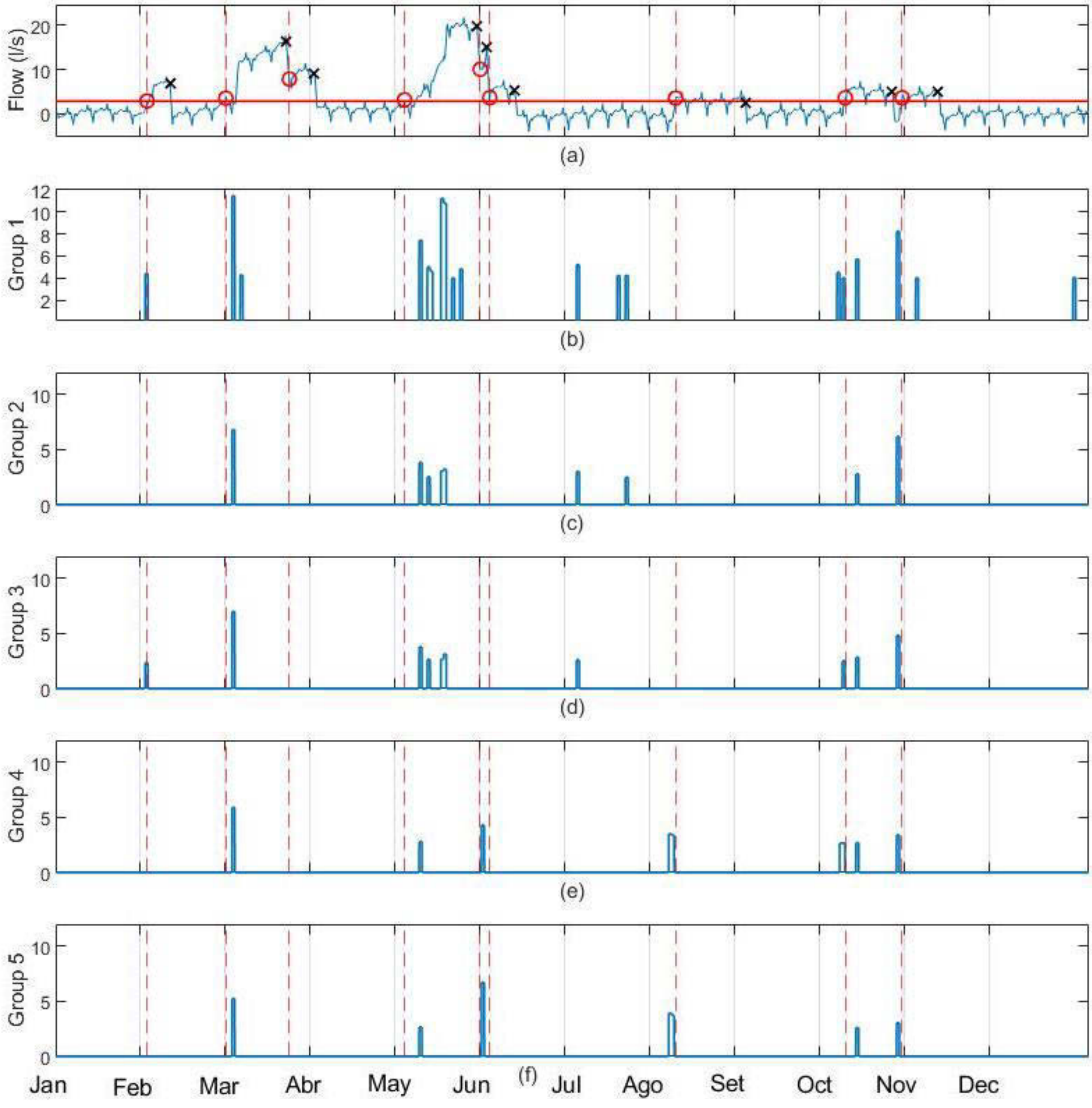


Figure 5. Result of leak detection, red line is the time when $\hat{\bar{l}}$ exceeds the defined threshold

are the types of incipient and bursts. The bursts occur in Area B of the network and affect all Δr_g signal groups. However, this zone is an isolated area with just one sensor, and a study of it can be done, see [9]. A second peak can be detected that happens only in Group 1, indicating a probable second fault in this area: the incipient leak.

Leaks number 4-6 have an extra incipient leak in the pipe p427 that saturates in the meanwhile. The study of $\Delta\bar{r}_g$ of these times instant needs to have more attention because the leaks 4 and 5 are situated near each other in groups 1 and 3, and the leak in the pipe p427 is in group 2. Group 1 has five peaks at this period, with the two most prominent peaks identifying leaks 4 and 5. The other peaks are due to saturation in the pipe p427 and the proximity in the time when leaks start. Leak number 6 is in group 5, and it is easy to identify the start time because it only affects groups 5 and 4.

The leaks 8-9 are not occurring together. However, the system has three saturated leaks in pipes p427, p654, and p610 that achieve the saturation moment during the leak 9. In the analysis of $\hat{\bar{l}}$ in this instant is possible only to identify the leaks 8 and 9. In the Δr_g examination, group 1 is the

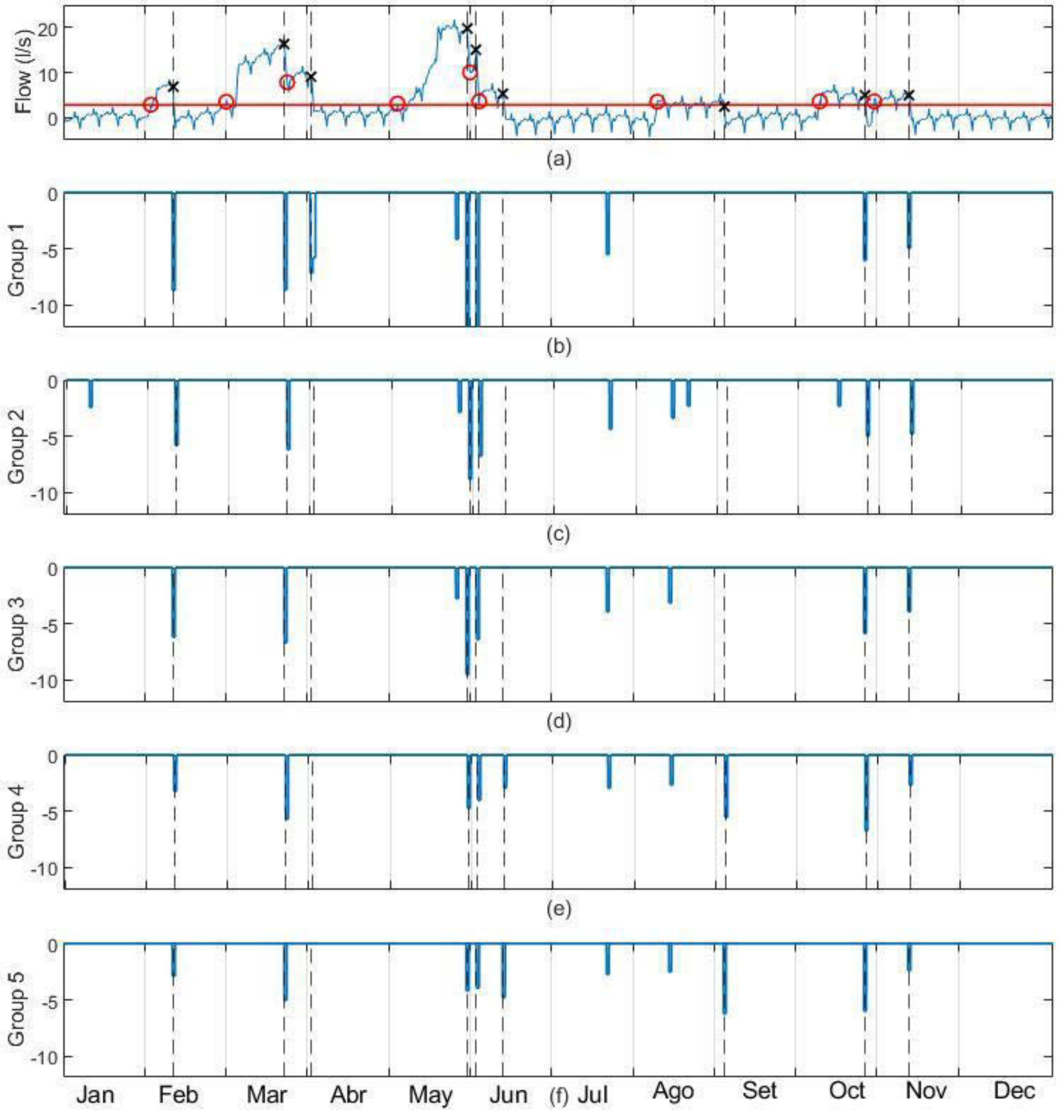


Figure 6. Result of leak detection, red line is the time when $\hat{\bar{l}}$ exceeds the defined threshold

BY NC SA

more affected, having five peaks, not making it clear at which time leaks 8 and 9 started but indicating a fault in the WDN.

The same analysis can be done when a leak is fixed. Figure 6 shows these results. Fig. 6 (a) is the same study as $\hat{\bar{l}}$ of Fig. 5(a) but the black line that propagation to the other $\Delta\bar{r}_g$ signal is when a leak is fixed in the zone. In all $\Delta\bar{r}_g$ analyses, a peak negative occurs due to a leak repair; the signal has more than 4 negative peaks caused by some uncertainties of measurements and their estimations.

To perform the second step of the proposed approach, the time instant of each leak begins more than the time they are repaired was used to calculate an average of the residues in equation (5) to apply the RBF interpolation method. Fig. (7) shows the results of the nine fixed leaks. The zones quoted to have a leak vary according to the location of the fault and how it affects the surrounding sensors, but for all leaks retaining the apex in red in the region of the leak.

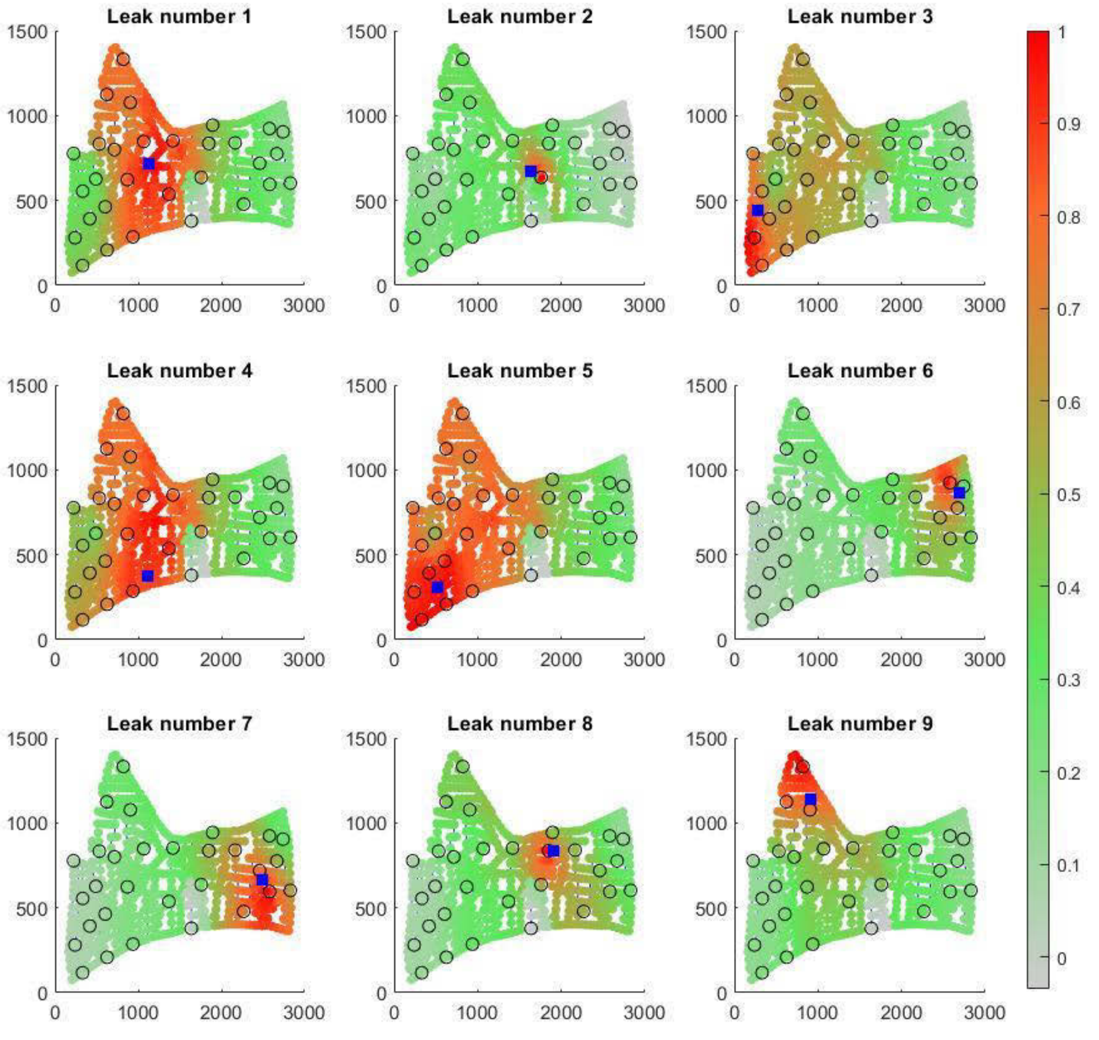


Figure 7. Graphical comparison of the interpolated states for the nine leaks in the WDN

5 CONCLUSIONS

In this work, we present a new complete data-driven method utilizing flow and pressure measurements and the information of longitude, latitude, and elevation of all the nodes in the WDN to leak detection and location of overlapping leakages purposes. The methodology has been explained. It has mainly two phases: first, the leak detection, which converts every hour of the day into features and fuses them to obtain an average flow and pressure measurement signal. The leak detection method is a multi-validate problem that starts with a study of the fused average flow and validates with the analysis of the fused average residual pressure divided by groups made by neighbour sensor and the area of the WDN. The second phase is the leak localization zone that applies the Radial basis function to interpolate the average residual pressure for each sensor to all the nodes in the network, resulting in the zone most likely to have the fault.

The L-town network utilized in the Battle of the Leakage Detection and Isolation Methods has been used as a case study. The data studied were from the year 2018 with 12 leaks and only 9 repaired, having two different temporal profiles: burst pipe and incipient leaks that stature in some instant. The result of the leak detection demonstrates a good result when the leak is of the bursts type leak. On the other hand, detecting when the leak is incipient with a low growth rate is difficult because the method evolves with the data. Moreover, the method can detect simultaneous leaks.

The “leak localization zone” phase is satisfactory, even using only data information and without resorting to hydraulic models. Also, it was possible to locate the leakage area, limiting it to a single leak at a time in the WDN. If simultaneous leaks happen, the leak location zone will be the region closest to the leak, thus increasing the result area. Future work will investigate a clustering method to obtain the groups in the leak detection phase to define the most appropriate leak detection and location procedure.

6 ACKNOWLEDGEMENTS

This work has been partially funded by the project PID2020-115905RB-C21 (L-BEST) funded by MCIN/ AEI /10.13039/501100011033.

7 REFERENCES

[1] Soldevila, A., Fernandez-Canti, R.M., Blesa, J., Tornil Sin, S., and Puig, V. (2016). Leak localization in water distribution networks using model-based bayesian reasoning. In 2016 European Control Conference (ECC),1758–1763. IEEE.

[2] Javadiha, M., Blesa, J., Soldevila, A., and Puig, V. (2019). Leak localization in water distribution networks using deep learning. In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), 1426–1431. IEEE.

[3] Sophocleous, S., D. Savic, and Z. Kapelan. (2019). “Leak localization in a real water distribution network based on search-space nreduction.” J. Water Resour. Plann. Manage.145

[4] Soldevila, A., Boracchi, G., Roveri, M. et al. (2021), “Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models”. Neural Comput & Applic

[5] Sun, Congcong, et al. 1 (2020) "Leak localization in water distribution networks using pressure and data-driven classifier approach." Water 12.: 54

[6] Alves, D., Blesa, J., Duviella, E., & Rajaoarisoa, L. (2021). Robust Data-Driven Leak Localization in Water Distribution Networks Using Pressure Measurements and Topological Information. Sensors, 21(22), 7551

[7] Hu, Z., Chen, B., Chen, W., Tan, D., & Shen, D. (2021). Review of model-based and data-driven approaches for leak detection and location in water distribution systems. Water Supply, 21(7), 3282-3306.

[8] Vrachimis, S. G., Eliades, D. G., Taormina, R., Ostfeld, A., Kapelan, Z., Liu, S., ... & Polycarpou, M. M. (2020). BattLeDIM: Battle of the Leakage Detection and Isolation Methods.

[9] Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., & Duviella, E. (2022). Leak Localization in Water Distribution Networks Using Data-Driven and Model-Based Approaches. Journal of Water Resources Planning and Management, 148(5), 04022016.

[10] Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban water demand forecasting: review of methods and models. Journal of Water Resources Planning and Management, 140(2), 146-159.

[11] Malik, O. (2016). Probabilistic leak detection and quantification using multi-output Gaussian processes (Doctoral dissertation, University of Southampton).

[12] Flajolet, P., & Sedgewick, R. (1995). Mellin transforms and asymptotics: Finite differences and Rice's integrals. Theoretical Computer Science, 144(1-2), 101-124.

[13] Mongillo, Michael. "Choosing basis functions and shape parameters for radial basis function methods." SIAM undergraduate research online 4.190-209 (2011): 2-6.

[14] Buhmann, M. D. (2003). Radial basis functions: theory and implementations (Vol. 12). Cambridge university press.

[15] Fasshauer, G. E. (2007). Meshfree approximation methods with Matlab. vol. 6 World Scientific Publishing Co Inc.

[16] Wendland, H. (2004). Scattered data approximation (Vol. 17). Cambridge university press.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14062

SMART WATER APPLICATIONS VS. INFORMATION AND COMMUNICATION TECHNOLOGIES – AN INTEGRATIVE SELECTION FRAMEWORK

Martin Oberascher[1], Wolfgang Rauch[2] and Robert Sitzenfrei[3]

[1,2,3] Unit of Environmental Engineering, Faculty of Civil Engineering, Institute of Infrastructure, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria

[1] *martin.oberascher@uibk.ac.at*, [2] *wolfgang.rauch@uibk.ac.at*, [3] *robert.sitzenfrei@uibk.ac.at*

Abstract

In urban water infrastructure, information and communication technology are presently concentrated on central facilities (e.g., treatment plant) or are installed at main points in the urban drainage or water distribution network, e.g., inlet points of district meter areas or combined sewer overflow structures. In this regard, the Internet of Things concept as part of smart city development enables a large-scale implementation of measuring equipment and allows the integration of decentralised elements into an overall controlled system. Consequently, reliable and suitable information and communication technology is a key element for the exchange of measurement and control data as well as for the success of these systems. From a water engineering perspective, it is often difficult to choose the right information and communication technology for the intended urban water infrastructure application. Likewise, from an Internet of Things perspective, it is often unclear what kind of urban water infrastructure applications are feasible, making them difficult to efficiently implement.

Aim of this work is to develop a first-decision making tool, which can be used by network operators, researcher, and stakeholders to support supervisory control and data acquisition development and to realise an efficient information and communication technology system in the field of network-based urban water infrastructure. In contrast to existing recommendations, our approach is based on a comprehensive review of required spatial and temporal resolution of measurement and control data for a wide range of different network-based urban water infrastructure applications. Subsequently, this enables a targeted coordination with the properties of communication technologies (e.g., data rate, range, and quality of service) and leads to a significantly improved and integrative decision-making tool.

Subsequently, we tested the functionality of the framework on two exemplary applications in urban water infrastructure, namely (1) determining suitable communication technologies for an early warning system for leakage detection and localisation in water distribution networks (e.g., (Wireless)Meter-Bus for water meters, long range wide area network for water pressure sensors, and global system for mobile communications at inlet points of district meter areas, and (2) identifying feasible applications for an existing long range wide area network (e.g., monitoring micro-climate and automatic irrigation at nature-based solutions). Results from the framework application have been evaluated through a literature review on used communication technologies, and are found to be consistent with real-word applications. As conclusion, different communication technologies are necessary to satisfy different requirements associated with an integrative management of urban water infrastructure.

Keywords
Decision making tool, ICT, integrative management, SCADA system, urban water infrastructure

cc BY NC SA

1 INTRODUCTION

In urban water infrastructure (UWI), information and communication technology (ICT) are widely found in central facilities [1], e.g., treatment plants, while the implementation in the networks is mainly concentrated at the main points, e.g., combined sewer overflow (CSO) structures [2] or inlet points of district metering areas (DMAs) [3]. In this regard, the Internet of Things (IoT) concept as part of smart city development enables a large-scale implementation of measuring equipment even at remote and underground structure [4], thereby increasing the data availability significantly. Additionally, the IoT concept supports the integration of decentralised network elements like nature-based solutions (NBS) into an overall controlled system.

Consequently, reliable and suitable ICT is a key element for the exchange of measurement and control data as well as for the success of these systems. Thereby, the following two challenges can be identified: (1) from a water engineering perspective, the limitations and benefits of different IoT concepts are usually unclear and thus it is difficult to choose suitable communication technologies and (2) from an IoT technology perspective, it is often unclear what kind of UWI applications are feasible, making them difficult to efficiently implement. To tackle these challenges, there are several frameworks outlined in literature, e.g., [5-7], in which suitable communication technologies are suggested based on the area of interest (e.g., smart metering). However, as concluded in the review of [8], it requires a coordination of the usable communication technology and the required temporal and spatial resolution of the measurement and control data to implement an efficient monitoring and controlling network.

To overcome this limitation, we present a first decision making tool in this work. In contrast to the existing recommendations, our approach is based on a comprehensive review of required spatial and temporal resolution of measurement and control data as well as used communication technologies for a wide range of different network-based UWI applications. The decision-making tool can be used by network operators, researcher, and stakeholders to support supervisory control and data acquisition (SCADA) development and to realise an efficient ICT system in the field of network-based UWI.

2 METHODS

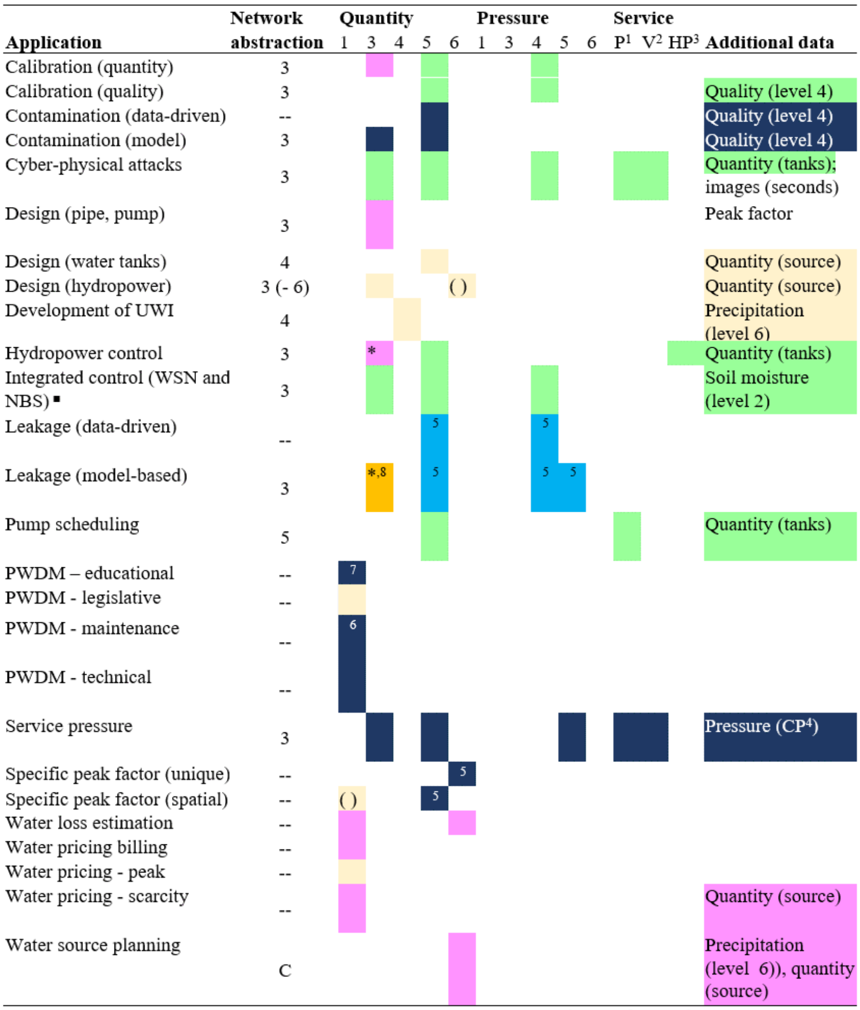
The development of the decision making tool is based on the results of the comprehensive review of [8] as mentioned above. Therefore, the main outcomes are shortly summarised in this chapter.

2.1 Data requirements of application

Figure 1- 3 gives an overview over the temporal and spatial resolution of different applications in water distribution and urban drainage networks including nature-based solutions identified from literature. For the abbreviations used, refer to Table 1. As can be seen, each application is characterised by different requirements regarding data resolution, which also differ between the applications.

Table 1. Abbreviation for used spatial and temporal resolution

Colour	Temporal resolution	System level	Spatial resolution
⬛	1 s – 5 min	1	Household scale
🟦	5 min – 10 min	2	Nature-based solution
🟩	10 min – 1 h	3	Network nodes
🟨	1 h – 1 d	4	Grid with 100 – 500 m
🟧	1 d – 1 m	5	Examination area (e.g., DMA, CSO)
🟪	1 m – 1 y	6	Total area (e.g., city)

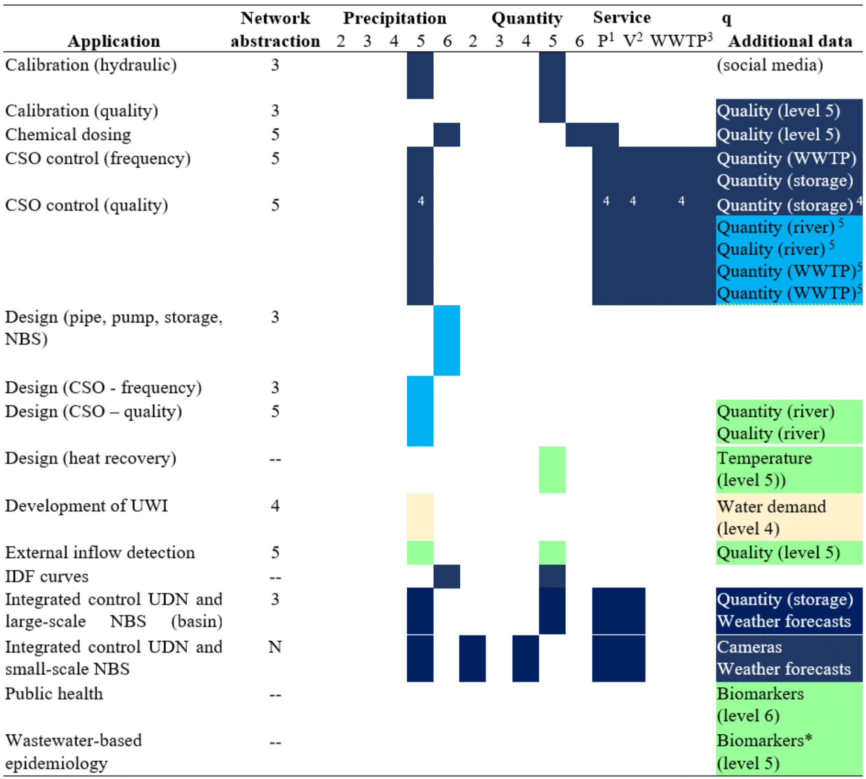


	Network	Quantity					Pressure					Service			
Application	**abstraction**	1	3	4	5	6	1	3	4	5	6	P^1	V^2	HP^3	**Additional data**
Calibration (quantity)	3		🟪		🟩				🟩						
Calibration (quality)	3				🟩				🟩						🟩 Quality (level 4)
Contamination (data-driven)	--				⬛										⬛ Quality (level 4)
Contamination (model)	3		⬛		⬛										⬛ Quality (level 4)
Cyber-physical attacks	3		🟩		🟩				🟩			🟩	🟩		🟩 Quantity (tanks); images (seconds)
Design (pipe, pump)	3		🟪												Peak factor
Design (water tanks)	4				🟨										🟨 Quantity (source)
Design (hydropower)	3 (- 6)		🟨			🟨 ( )									🟨 Quantity (source)
Development of UWI	4			🟨											🟨 Precipitation (level 6)
Hydropower control	3		🟪 *		🟩									🟩	🟩 Quantity (tanks)
Integrated control (WSN and NBS) ■	3		🟩		🟩				🟩						🟩 Soil moisture (level 2)
Leakage (data-driven)	--				🟦 5				🟦 5						
Leakage (model-based)	3		🟧 *,8		🟦 5				🟦 5	🟦 5					
Pump scheduling	5				🟩							🟩			🟩 Quantity (tanks)
PWDM – educational	--	⬛ 7													
PWDM - legislative	--	🟨													
PWDM - maintenance	--	⬛ 6													
PWDM - technical	--	⬛													
Service pressure	3		⬛		⬛					⬛		⬛	⬛		⬛ Pressure (CP^4)
Specific peak factor (unique)	--					⬛ 5									
Specific peak factor (spatial)	--	🟨 ( )			⬛ 5										
Water loss estimation	--	🟪				🟪									
Water pricing billing	--	🟪													
Water pricing - peak	--	🟨													
Water pricing - scarcity	--	🟪													🟪 Quantity (source)
Water source planning	C					🟪									🟪 Precipitation (level 6)), quantity (source)

( ) = possibility, *downscaled to same temporal resolution as other data; ^{1}pump, ^{2}valve, ^{3}hydropower unit, ^{4}critical point (e.g., area with low pressure); ^{5}also used with 5 min – 10 min in literature, ^{6}also used with 10 min – 1 h in literature, ^{7}also used with 1 d – 1 m in literature, ^{8}also used with 1 m – 1 y in literature

Figure 1. Spatial and temporal resolution for applications related to water distribution networks (This figure is reproduced with small alterations (deletion of the reference column) from [8] under an Attribution 4.0 International license (CC BY 4.0)).

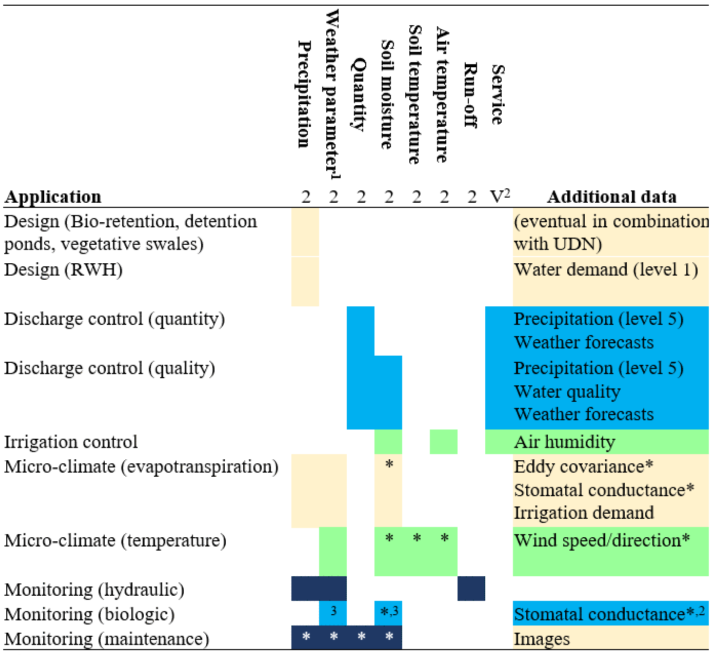
CC BY NC SA



Application	Network abstraction	Precipitation 2	3	4	5	6	Quantity 2	3	4	5	6	Service P[1]	V[2]	WWTP[3]	q Additional data
Calibration (hydraulic)	3				■					■					(social media)
Calibration (quality)	3									■					■ Quality (level 5)
Chemical dosing	5					■					■	■			■ Quality (level 5)
CSO control (frequency)	5				■							■	■	■	■ Quantity (WWTP) ■ Quantity (storage)
CSO control (quality)	5				■ 4							■ 4	■ 4	■ 4	■ Quantity (storage)[4] 🟦 Quantity (river)[5] 🟦 Quality (river)[5] 🟦 Quantity (WWTP)[5] 🟦 Quantity (WWTP)[5]
Design (pipe, pump, storage, NBS)	3					🟦									
Design (CSO - frequency)	3				🟦										
Design (CSO – quality)	5				🟦										🟩 Quantity (river) 🟩 Quality (river)
Design (heat recovery)	--									🟩					🟩 Temperature (level 5))
Development of UWI	4				🟨										🟨 Water demand (level 4)
External inflow detection	5				🟩					🟩					🟩 Quality (level 5)
IDF curves	--					■				■					
Integrated control UDN and large-scale NBS (basin)	3				■					■		■	■		■ Quantity (storage) ■ Weather forecasts
Integrated control UDN and small-scale NBS	N				■		■		■			■	■		■ Cameras ■ Weather forecasts
Public health	--														🟩 Biomarkers (level 6)
Wastewater-based epidemiology	--														🟩 Biomarkers* (level 5)

[1]pump, [2]valve, [3]wastewater treatment plant; [4]also used with 5 min – 10 min in literature, [5]also used with 10 min – 1 h in literature

Figure 2. Spatial and temporal resolution for applications related to urban drainage networks (This figure is reproduced with small alterations (deletion of the reference column) from [8] under an Attribution 4.0 International license (CC BY 4.0)).



Application	Precipitation 2	Weather parameter[1] 2	Quantity 2	Soil moisture 2	Soil temperature 2	Air temperature 2	Run-off 2	Service V[2]	Additional data
Design (Bio-retention, detention ponds, vegetative swales)	🟨								🟨 (eventual in combination with UDN)
Design (RWH)	🟨								🟨 Water demand (level 1)
Discharge control (quantity)			🟦					🟦	🟦 Precipitation (level 5) 🟦 Weather forecasts
Discharge control (quality)			🟦	🟦				🟦	🟦 Precipitation (level 5) 🟦 Water quality 🟦 Weather forecasts
Irrigation control				🟩		🟩		🟩	🟩 Air humidity
Micro-climate (evapotranspiration)	🟨	🟨		🟨 *					🟨 Eddy covariance* 🟨 Stomatal conductance* 🟨 Irrigation demand
Micro-climate (temperature)		🟩		🟩 *	🟩 *	🟩 *			🟩 Wind speed/direction*
Monitoring (hydraulic)	■	■					■		
Monitoring (biologic)		🟦 3		🟦 *,3					🟦 Stomatal conductance*,2
Monitoring (maintenance)	■ *	■ *	■ *	■ *					🟨 Images

[1]air temperature, air humidity, atmospheric pressure, solar radiation, wind speed, wind direction wind, [2]valve; [3]also used with 10 min – 1 h in literature; * = multiply measurement points;

Figure 3. Temporal resolution for applications related to nature-based solutions (This figure is reproduced with small alterations (deletion of the reference column) from [8] under an Attribution 4.0 International license (CC BY 4.0)).

2.2 Communication technologies

Figure 4 gives an overview over the communication technologies with transmission ranges and data rates as characteristic properties. The communication technologies can be subdivided into wired and wireless communication, using a cable and electromagnetic waves for the exchange of data, respectively. For more information about communication technologies, refer to relevant literature [6,9-11].

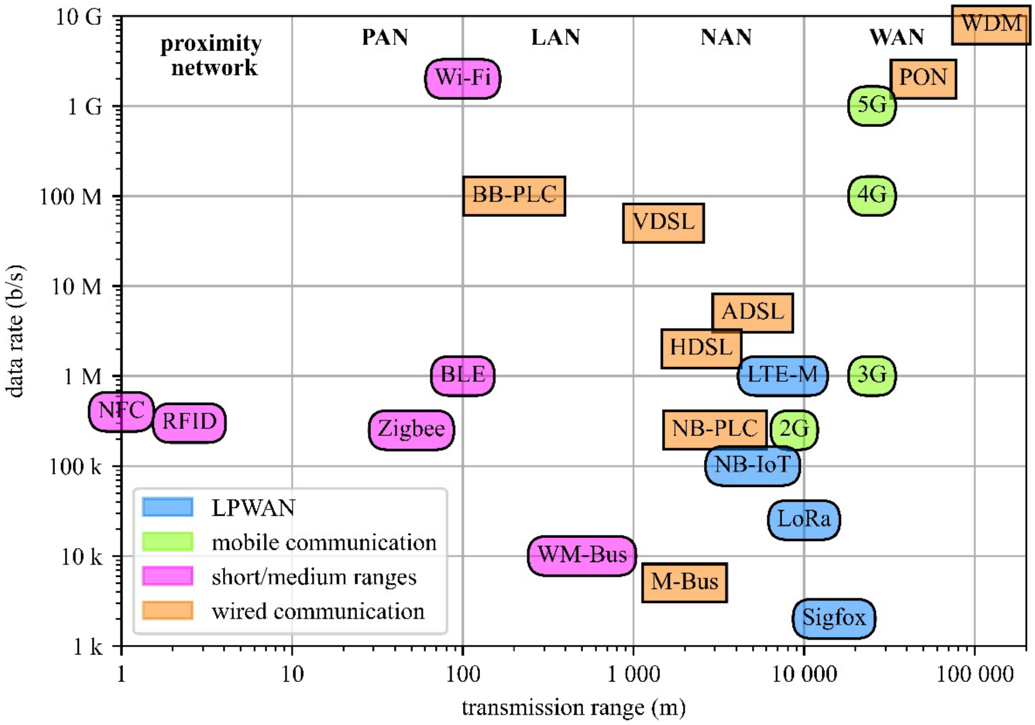


Figure 4. Overview of different communication technologies based on transmission range and data rate (This figure is reproduced from [8] under an Attribution 4.0 International license (CC BY 4.0)).

For UWI, the wired communication technologies fibre optic as the backbone of the internet and Meter-Bus (M-Wus) for the remote-readout of water meters are of importance. Wired communication technologies provide a high quality of service as packet losses are low and allow nearly almost any transmission interval. In contrast to wireless communication technologies, the installation of wires requires considerable efforts and is less flexible.

Short and medium ranges include a wide range of technologies with a transmission range up to 500 m. For interests are ZigBee (smart home applications), Wireless-Fidelity (Wi-Fi) (public hot spots) and WM-Bus (alternative for water meters). These communication technologies are primarily applied in the surroundings of buildings, as the transmission range is limited.

Mobile communication uses radio waves for the transmission of data, and cellular communication networks cover a large area. For UWI, GSM/GPRS (2G) are widely applied for monitoring and controlling approaches in UWI. Furthermore, cellular communication operates in licenced frequency bandwidths, thereby providing a high quality of service with low packet losses. However, the energy consumption is high, requiring additional approaches for a long-maintenance operation, e.g., to decrease transmission interval or to include additional energy sources.

Low power wide area networks (LPWANs) support the large-scale implementation of multiple devices through long transmission ranges and ultra-low-power operation. Leading technologies are LoRa and Sigfox in the unlicensed frequency bands and NB-IoT in the licensed frequency bands. Devices using LPWANs for data transmission are most of the time unreachable, as the transceivers are turned off to save energy. Therefore, these communication technologies are usable for delay-tolerant applications. Additionally, as Sigfox and LoRa are operating in the unlicensed bands, they are subject to fair-use policy (max. number of packets and limitation of packet length) including packet losses, which are depending on number of connected devices and connection quality.

Summarised, each communication technology has different properties and limitations, influencing the spatial and temporal resolution of transmittable measurement and control data. Therefore, it requires a coordination between intended applications and communication technologies to implement an efficient monitoring and controlling network.

3 RESULTS AND DISCUSSION

The information about required spatial and temporal resolution of measurement data and communication technologies was used to develop a first-decision making tool to support network operators, researcher, and stakeholders with the implementation of ICT. This work is primarily based on real-world implementation, of course, different approaches are also feasible. Additionally, field test should be carried out before implementation to assess the functionality under local conditions. The framework is organised into two categories: (1) (near) real-time transmission of data and (2) transmission of aggregated and historical data.

3.1 Real-time Transmission

Figure 5 gives an overview about recommended communication technology for real-time transmissions. The first decision criterion is the intended purpose. Real-time monitoring describes the exchange of monitoring data from the sensor nodes to the central system and therefor a unidirectional connection is sufficient. In contrast, real-time operation requires a bidirectional connection for the exchange of status notifications (from the control organ to the central system) and control commands (from the central system to the control organ). Nearly all communication technologies support a bidirectional communication, but the number of downlinks can be limited as with LPWANs.

Second, the desired properties influence the choice of the communication technology. Real-time monitoring can be distinguished between text packet (e.g., an identification number, a time point, and a measurement value) and images, having a packet size in bytes and mega-bytes, respectively. Therefore, the data rate and the maximum packet length influences the choice of suitable communication technologies. Real-time operation differs between delay-sensitive applications (control process should take place immediately) and delay-tolerant applications (delays have only a limited influence on system performance). Following, the reachability is the decisive factor for the selection of suitable communication technologies.

Finally, it requires a trade-off between maintenance and installation efforts and the reliability. Cellular and wired communication provide a high reliability and allow nearly every transmission interval but require either a wire for data communication or power supply due to high energy consumption. Therefore, these communication technologies are suitable for only few installation places. In contrast, LPWANs enable an easy large-scale implementation of sensor nodes with relative low investment costs. However, they are not suitable for high transmission intervals due to an increased energy demand. Additionally, Sigfox and LoRa operates in the public frequency band widths which include also packet losses. In this regard, short and medium range technologies can be used for both high spatial and temporal resolution, but the limited range will require the inclusion of the public to cope with the high installation and maintenance efforts.

Application purpose		**Maintenance and installation efforts**		**Communication technology**									
				Wired communication		Mobile communication		Short and medium ranges			Low power wide area networks		
Purpose	Properties	Transmission interval	Scale of implementation	M-Bus	Fibre optic	GSM/GPRS	LTE / 5G	Wi-Fi	WM-Bus	Zigbee	LoRa	NB-IoT	Sigfox
Real-time monitoring	Text	1s - 5min	1 - 4	✓[1,2]				✓[1]	✓[1,2]	✓[1]	✓[3]		
			5 - 6		✓	✓[4]						✓[3]	
		5min - 10min	1 - 4	✓[1,2]				✓[1]	✓[1,2]	✓[1]	✓		
			5 - 6		✓	✓[4]						✓	
		10min - 1d	1 - 4	✓[1,2]				✓[1]	✓[1,2]	✓[1]	✓[5]	✓[6]	✓[5]
			5 - 6		✓	✓[4]						✓	
	Images	1s - 5min	1 - 4		✓			✓[1]					
			5 - 6		✓		✓[4]						
		5min - 10min	1 - 4		✓			✓[1]					
			5 - 6		✓		✓[4]						
		10min - 1d	1 - 4		✓			✓[1]					
			5 - 6		✓		✓[4]						
Real-time operation	Delay-sensitive	1s - 5min	1 - 4					✓[1]		✓[1]			
			5 - 6		✓	✓[4]							
		5min - 10min	1 - 4					✓[1]		✓[1]			
			5 - 6		✓	✓[4]						✓[3]	
		10min - 1d	1 - 4					✓[1]		✓[1]		✓	
			5 - 6		✓	✓[4]						✓	
	Delay-tolerante	1s - 5min	1 - 4					✓[1]		✓[1]			
			5 - 6		✓	✓[4]							
		5min - 10min	1 - 4					✓[1]		✓[1]			
			5 - 6		✓	✓[4]						✓[3]	
		10min - 1d	1 - 4								✓		✓[3]
			5 - 6									✓	

[1]requires involvement of public; [2]alternative for water meters; [3]possible to a limited extent - requires feasibility study on site (not recommended); [4]requires additional energy sources; [5]peak -hour pricing – not recommended; [6]peak -hour pricing – alternative;

Figure 5. Recommended communication technologies for real-time applications.

3.2 Historical and aggregated data

Figure 6 gives an overview about the recommended communication technologies for applications based on historical data. In contrast to real-time applications, multiple measurement values can be aggregated and transmitted periodically. As stated before, the maximum packet length is limited for LPWANs. Therefore, the first decision criterion is the number of measurement values per transmitted data packet. For example, the maximum number of measurement values is 6, 120, and 800 for Sigfox, LoRa, and NB-IoT, respectively, by assuming a storage size of 2 bytes.

Another decision criterion is the reliability of data transmission. Using technologies operating in the public frequency band widths, data gaps must be expected. In contrast, if a continuous time series without data gap is needed, wired and cellular communication technologies are recommended. Additionally, short and medium range technologies represent an alternative for high spatial resolution.

BY NC SA

Data size		Communication technology									
		Wired		Mobile		Short and medium			Low power wide area		
Transmission interval	Measurement interval	M-Bus	Fibre optic	GSM/GPRS	LTE / 5G	Wi-Fi	WM-Bus	Zigbee	LoRa	NB-IoT	Sigfox
30min	10s		✓	✓	✓	✓				✓	
	1min	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	5min	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	10min	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1h	10s		✓	✓	✓	✓				✓	
	1min	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	5min	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	10min	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	1h	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1d	10s		✓	✓	✓	✓					
	1min		✓	✓	✓	✓					
	5min		✓	✓	✓	✓				✓	
	10min		✓	✓	✓	✓				✓	
	1h	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	1d	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1m	10s		✓	✓	✓	✓					
	1min		✓	✓	✓	✓					
	5min		✓	✓	✓	✓					
	10min		✓	✓	✓	✓					
	1h		✓	✓	✓	✓				✓	
	1d	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	1m	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 6. Recommended communication technologies for historical and aggregated data.

3.3 Usability of the developed first decision making tool

The developed first decision making tool can be used for following aims: (1) to determine suitable communication technologies for a selected application in the field of UWI and (2) to identify feasible applications in the field of UWI for an existing communication network. Therefore, Figure 7 gives an overview of how the first decision making tool can be applied.

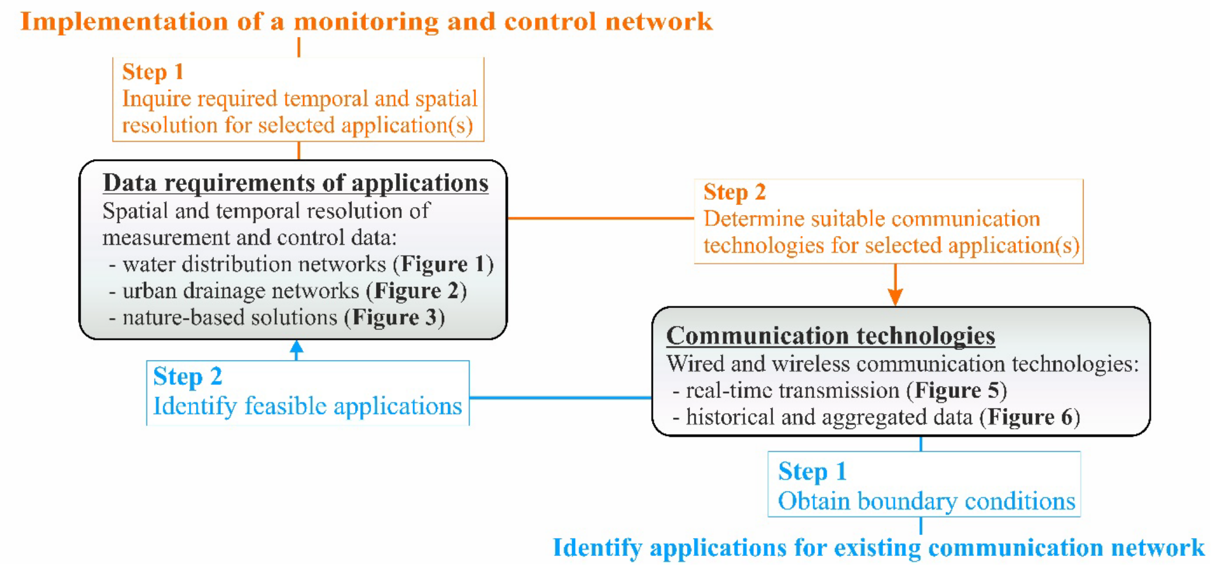

Implementation of a monitoring and control network
Step 1
Inquire required temporal and spatial resolution for selected application(s)
Data requirements of applications
Spatial and temporal resolution of measurement and control data:
- water distribution networks (Figure 1)
- urban drainage networks (Figure 2)
- nature-based solutions (Figure 3)
Step 2
Determine suitable communication technologies for selected application(s)
Communication technologies
Wired and wireless communication technologies:
- real-time transmission (Figure 5)
- historical and aggregated data (Figure 6)
Step 2
Identify feasible applications
Step 1
Obtain boundary conditions
Identify applications for existing communication network

Figure 7. Developed first decision making tool.

In the following two examples for applying the developed decision making tool are outlined:

- A network operator decides to install an early warning system for model-based leakage detection and localisation in the water distribution networks. The workflow is as follows: First, the spatial and temporal resolution of the required measurement data can be obtained in Figure 1, and second, this information is used afterwards in Figure 5 to determine suitable communication technologies. As the temporal resolution of the required measurement data is 5 to 10 min and text-based data packets are transmitted, following communication technologies are selected: (1) (W)M-Bus for water meters installed in households, (2) LoRaWAN for water pressure sensors in a grid arrangement of 100 to 500 m as packet losses are acceptable due to the high number of sensors, and (3) GPRS for water inflow and water pressure at the inlet points of the DMA for a high reliability of data transmission.

- An infrastructure operator deploys an existing LoRa network and searches for real-time applications in the field of UWI. The workflow can be described as follows: First, as shown by Figure 5, LoRa is recommended for text-based data packets with measurement points ranging from household scale to grid arrangement and a temporal resolution of 5 min to 1 day. Subsequently, this information can be used in Figure 1 -3 to identify feasible real-time applications. Exemplary applications are monitoring of micro-climate and an automatic irrigation at nature-based solutions.

The Figures 5 and 6 are based on an extensive literature [8]. As validation, real-world implementations for the two examples are shown: (W)M-Bus is used for read-out of water meters [12,13], LoRa is applied for large-scale monitoring of urban drainage and water distribution networks [13-15] as well as for smart rainwater harvesting [16]; and GPRS is commonly utilised for single measurement points [17,18], which is in concordance with the recommendations of the presented framework.

4 CONCLUSION

Reliable and suitable information and communication technologies (ICT) are a key factor for the implementation of efficient monitoring and controlling systems in the field of urban water infrastructure (UWI). Additionally, the Internet of Things (IoT) concept allows the installation of low-cost sensors even at remote and underground structures, thereby providing new possibilities in the management of UWI. Subsequently, the challenges for both, water engineers (the limitations and benefits of different IoT concepts are usually unclear) and ICT operator (what kind of UWI applications are feasible) are increasing.

To tackle these challenges, we presented a detailed first decision making tool for the realisation of an efficient ICT system in the field of network-based UWI. In contrast to existing recommendations, our approach is based on a comprehensive review of required spatial and temporal resolution of measurement and control data for a wide range of network-based UWI applications. This information was used for a targeted coordination with the properties of different communication technologies, leading to a significantly improved and integrative decision-making tool.

The developed framework can be used by network operators, researchers, and stakeholders for following aims: (1) to determine suitable communication technologies for a selected application in the field of UWI and (2) vice versa to identify feasible applications in the field of UWI for an existing communication network. The functionality of the framework was tested by using two exemplary applications in the field of network-based UWI. The recommended communication technologies are consistent with real-word implementations, thereby demonstrating the applicability of our approach. Additionally, the combination of different communication technologies is necessary to satisfy the requirements for an integrative management of UWI.

5 FUNDINGS

This publication was produced as part of the "Smart Water City" project. This project is funded by the Climate and Energy Fund and is part of the programme "Smart Cities Demo - Living Urban Innovation 2018" (project 872123).

6 REFERENCES

[1] Yuan, Z.; Olsson, G.; Cardell-Oliver, R.; van Schagen, K.; Marchi, A.; Deletic, A.; Urich, C.; Rauch, W.; Liu, Y.; Jiang, G. Sweating the assets - The role of instrumentation, control and automation in urban water systems. *Water Res* 2019, *155*, 381-402, doi:https://doi.org/10.1016/j.watres.2019.02.034.

[2] van Daal, P.; Gruber, G.; Langeveld, J.; Muschalla, D.; Clemens, F. Performance evaluation of real time control in urban wastewater systems in practice: Review and perspective. *Environ Model Softw* 2017, *95*, 90-101, doi:https://doi.org/10.1016/j.envsoft.2017.06.015.

[3] Creaco, E.; Campisano, A.; Fontana, N.; Marini, G.; Page, P.R.; Walski, T. Real time control of water distribution networks: A state-of-the-art review. *Water Res* 2019, *161*, 517-530, doi:https://doi.org/10.1016/j.watres.2019.06.025.

[4] Mohanty, S.P.; Choppali, U.; Kougianos, E. Everything you wanted to know about smart cities: The Internet of things is the backbone. *EEE Consum. Electron. Mag.* 2016, *5*, 60-70, doi:https://doi.org/10.1109/mce.2016.2556879.

[5] Buurman, B.; Kamruzzaman, J.; Karmakar, G.; Islam, S. Low-Power Wide-Area Networks: Design Goals, Architecture, Suitability to Use Cases and Research Challenges. *IEEE Access* 2020, *8*, 17179-17220, doi:https://doi.org/10.1109/access.2020.2968057.

[6] Lalle, Y.; Fourati, M.; Fourati, L.C.; Barraca, J.P. Communication technologies for Smart Water Grid applications: Overview, opportunities, and research directions. *Computer Networks* 2021, *190*, doi:https://doi.org/10.1016/j.comnet.2021.107940.

[7] Malik, H.; Kandler, N.; Alam, M.M.; Annus, I.; Moullec, Y.L.; Kuusik, A. Evaluation of low power wide area network technologies for smart urban drainage systems. In Proceedings of 2018 IEEE International Conference on Environmental Engineering (EE), 12-14 March 2018; pp. 1-5, doi:https://doi.org/10.1109/EE1.2018.8385262.

[8] Oberascher, M.; Rauch, W.; Sitzenfrei, R. Towards a smart water city: A comprehensive review of applications, data requirements, and communication technologies for integrated management. *Sustain. Cities Soc.* 2022, *76*, doi:https://doi.org/10.1016/j.scs.2021.103442.

[9] Ding, J.; Nemati, M.; Ranaweera, C.; Choi, J. IoT Connectivity Technologies and Applications: A Survey. 2020, *IEEE Access*, 8, 67646-67673, doi:https://doi.org/10.1109/ACCESS.2020.2985932.

[10] Sikimić, M.; Amović, M.; Vujović, V.; Suknović, B.; Manjak, D. An Overview of Wireless Technologies for IoT Network. In Proceedings of 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH), 18-20 March 2020; pp. 1-6, doi:https://doi.org/10.1109/INFOTEH48170.2020.9066337.

[11] Abdelwahab, R.H.; El-Habrouk, M.; Abdelhamid, T.H. Survey of Communication Techniques in Smart Grids. In Proceedings of 2019 21st International Middle East Power Systems Conference (MEPCON), 17-19 Dec. 2019; pp. 550-555, doi:https://doi.org/10.1109/MEPCON47431.2019.9007961.

[12] Cherukutota, N.; Jadhav, S. Architectural framework of smart water meter reading system in IoT environment. In Proceedings of 2016 International Conference on Communication and Signal Processing (ICCSP), 6-8 April 2016; pp. 0791-0794, doi:https://doi.org/10.1109/ICCSP.2016.7754253.

[13] Antzoulatos, G.; Mourtzios, C.; Stournara, P.; Kouloglou, I.O.; Papadimitriou, N.; Spyrou, D.; Mentes, A.; Nikolaidis, E.; Karakostas, A.; Kourtesis, D., et al. Making urban water smart: the SMART-WATER solution. *Water Sci Technol* 2020, *82*, 2691-2710, doi:https://doi.org/10.2166/wst.2020.391.

[14] Ebi, C.; Schaltegger, F.; Rust, A.; Blumensaat, F. Synchronous LoRa Mesh Network to Monitor Processes in Underground Infrastructure. *IEEE Access* 2019, *7*, 57663-57677, doi:https://doi.org/10.1109/access.2019.2913985.

[15] Saravanan, M.; Das, A.; Iyer, V. Smart water grid management using LPWAN IoT technology. In Proceedings of 2017 Global Internet of Things Summit (GIoTS), 6-9 June 2017; pp. 1-6, doi:https://doi.org/10.1109/GIOTS.2017.8016224.

[16] Oberascher, M.; Kinzel, C.; Kastlunger, U.; Kleidorfer, M.; Zingerle, C.; Rauch, W.; Sitzenfrei, R. Integrated urban water management with micro storages developed as an IoT-based solution – The smart rain barrel. *Environ Model Softw* 2021, *139*, doi:https://doi.org/10.1016/j.envsoft.2021.105028.

[17] Perez, R.; Sanz, G.; Puig, V.; Quevedo, J.; Escofet, M.A.C.; Nejjari, F.; Meseguer, J.; Cembrano, G.; Tur, J.M.M.; Sarrate, R. Leak Localization in Water Networks: A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona. *IEEE CONTR SYST MAG* 2014, *34*, 24-36, doi:https://doi.org/10.1109/MCS.2014.2320336.

[18] Quevedo, J.; Puig, V.; Cembrano, G.; Blanch, J.; Aguilar, J.; Saporta, D.; Benito, G.; Hedo, M.; Molina, A. Validation and reconstruction of flow meter data in the Barcelona water distribution network. *Control Eng. Pract.* 2010, *18*, 640-651, doi:https://doi.org/10.1016/j.conengprac.2010.03.003.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14067

COMPARISON BETWEEN THE TOP-DOWN AND BOTTOM-UP APPROACH FOR THE DIFFUSE-DISPERSIVE PHENOMENON ANALYSIS

Stefania Piazza[1], E. J. Mirjam Blokker[2], Mariacrocetta Sambito[3], Gabriele Freni[4]

[1,3,4] School of Engineering and Architecture, University of Enna "Kore", Cittadella Universitaria,
94100 Enna (Italy)
[2]KWR Water Research Institute, P.O. Box 1072, 3430 BB Nieuwegein (The Netherlands)

[1] *stefania.piazza@unikore.it,* [2] *Mirjam.Blokker@kwrwater.nl,* [3] *mariacrocetta.sambito@unikore.it,*
[4]*gabriele.freni@unikore.it*

Abstract

In order to detect deliberate or accidental contamination in drinking water distribution systems (DWDS), typically water quality sensors need to be installed in this system, and the data need to be analysed in order to feed alert systems and prevent the harm of contamination. This requires numerical (hydraulic and water quality) models that are as realistic as possible to support monitoring systems. Currently, water quality models used in the literature adopt an advective approach and simplified reaction kinetics, such as EPANET, which neglect diffusion-dispersion phenomena that are relevant in the presence of laminar and transient flow regimes. Another important aspect providing relevant uncertainty is related to the simplified estimation of sub-daily water demands that are commonly estimated from highly aggregated consumption data.

The present study aims to analyse diffusion-dispersive phenomena in a realistic DWDS model, which shows turbulent, transitional and laminar flows, and compare this to how such a DWDS would typically be modelled with a coarse estimate of demands. We are therefore considering two different demand allocation approaches (Top-down and Bottom-up).

In this paper the EPANET advective model and the diffusive-dispersive model, developed in a previous study, were used to better understand what the effect using the latter approach has within the DWDS as a function of two different types of demand allocation. To do this, the models results were compared to numerical tests that were performed on the real network of Zandvoort (the Netherlands) using a conservative tracer. For the 4 locations considered, it was noted that the diffusive-dispersive model responds well when using the bottom-up approach compared to the top-down approach. We found that in order to predict the tracer pattern, the Top-Down approach of demand allocation does not work well, even when an optimized diffusive-dispersive model is used. The bottom-up approach of demand allocation leads to far better results in predicting the tracer patterns, and with the diffusive-dispersive model the prediction improves even more. This means that in order to model water quality in a DWDS the first step should be to improve water demand models for this DWDS. This leads to an improved representation of flow regimes, and will most likely include laminar flows.

Keywords
Dispersion, TD and BU demand allocation, WDN, Water quality, Numerical analysis.

1 INTRODUCTION

Water distribution systems are made up of elements at risk, such as valves, pipes and tanks that could represent a preferential way for the intrusion of contaminants [1]. To better identify the occurrence of the contamination, the monitoring system must be able to reduce this risk [2], maximizing the detection efficiency and minimizing equipment costs [3]. To do this, simulation tools that manage to represent reality must support it.

Currently literature studies are based on hydraulic simulation tools, such as EPANET [4], which adopt a simplified approach regarding water quality, based on advective transport and some simplified reaction kinetics ([5], [6], [7]), neglecting the diffusive phenomena.

Although the use of simplified advective approach does not produce significant errors in the event that the DWDS is subject to a purely turbulent flow regime, they are not able to effectively model the behaviour of contaminants for laminar and transitional flow regimes, as shown by Piazza et al. (2020) [8].

Furthermore, the Top-down approach is the one conventionally used for allocating demand within DWDS and it is used in real applications. It consists of assigning a demand pattern of multiplier factors (typically taken from the drinking water production station) and correction factors (base demands) to all demand nodes (typically based on measured annual demands).

To ensure that a model is as representative of reality as possible, as well as being able to solve adequately the diffusive-dispersive processes [9], it must also be equipped with an accurate hydraulic model [10].

In this study the EPANET model and EPANET-DD (Dynamic-Dispersion) model, developed by Piazza et al. 2022 [9], were applied to the really existing network in Zandvoort (Netherlands), considering two different types of demand allocation (Top-down and Bottom-up). The Bottom-Up approach, unlike the Top-Down approach, consists in assigning different demand models to all nodes in the DWDS, according to the characteristics of the nodes. This was possible through the use of the SIMDEUM [11] which allows to generate stochastic demands according to the types of users analysed.

The aim of the present study is to analyse how the diffusive-dispersive process is influenced by the two different types of demand allocation. To do this, the numerical results were compared with those obtained from the experimental monitoring campaign carried out on the Zandvoort network in 2008 using a conservative tracer such as sodium chloride. Four locations were monitored, two of which are located near apartment buildings (Burg. Fennemaplein, De Ruyterstraat), the third is located in the basement of the hotel (NH hotel) and the fourth is located in the basement of a small apartment buildings of 15 residences (Sterflat Friedhoffplein).

2 MATERIALS AND METHODS

2.1 Case Study

The Zandvoort network was built in the 1950–1960s and consists of 5.7 km of Ø100 mm lined cast iron pipes and 3.5 km of Ø100 mm PVC pipes; it supplies 1000 homes, 2 hotels and 30 beach clubs. The area is supplied from one point with a fixed head through a booster pump; there are no tanks in the network (Figure 1). Inflows are monitored with an electromagnetic flowmeter with similar characteristics to those used in laboratory experiments and supplied flows are monitored with turbine flow meters compliant with MID directive (maximum error lower than 5%), variable depending on water meter age, diameter and installation. The water use in the network was determined by the historical flow patterns at the stimulation station, measured by the Provincial Water Company Noord-Holland (PWN), and the domestic water demand accounted for 70% of the total demand. Drinking water is distributed without any disinfectant, as is common in the Netherlands.

A tracer study with NaCl was performed between 2 September and 20 October 2008. During the study, the water service was guaranteed to the users and sodium chloride (NaCl) was used as a tracer, as it does not cause inconvenience or risks to the health of users and provides results of good precision and is low cost. The quantities of tracer used were such as not to compromise the characteristics of the water for normal use. The numerical simulations were carried out considering the same conditions used in Blokker et al. 2010 [11]. Electrical conductivity (EC) values were measured at 4 locations. The modeled results were compared with the experimental data obtained by the tracer study.

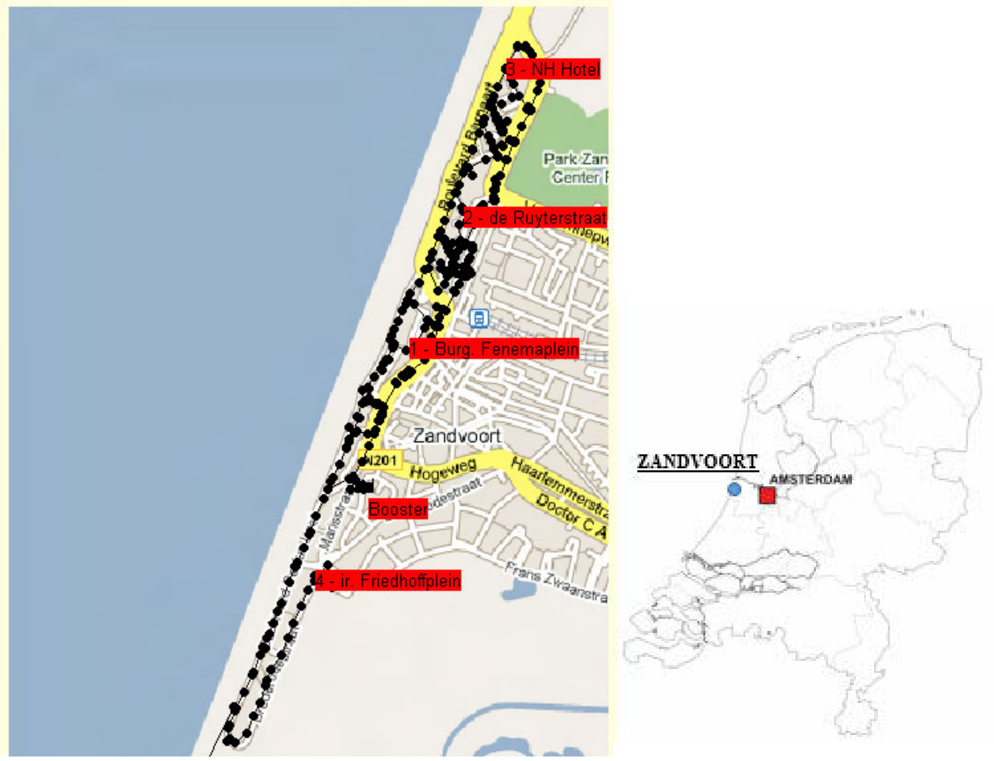

3 - NH Hotel
Park Zan Center
2 - de Ruyterstraat
1 - Burg. Fenemaplein
Zandvoort
N201
Hogeweg
Haarlemmerstr.
Booster
4 - ir. Friedhoffplein
ZANDVOORT
AMSTERDAM

Figure 1. Layout of the Zandvoort water distribution network.

2.2 Numerical Model

The upgraded version of the EPANET model, used in Piazza et al. 2022 [9], which solved the dispersion / diffusion equations proposed by Romeo-Gomez and Choi (2011) [12] in quasi-stationary flow conditions, solving the hydraulic problem in steady-state flow conditions with the EPANET-MATLAB-Toolkit [13] and the equation of advection-diffusion-dispersion in dynamic flow conditions in the two-dimensional case with the classical random walk method [14].

The model allows to determine the position of the solute particles in the x and y directions using equations (1) and (2) as a function of the different flow regimes that occur inside the network, and the tracer concentration using equation (3).

$$x = x + \frac{3}{2} u_x \left(1 - \left(\frac{y}{\frac{d}{2}} \right)^2 \right) dt + \sqrt{2 \cdot E_{f\ or\ b} \cdot dt} \tag{1}$$

$$y = y + u_y dt + \sqrt{(E_f + E_b) \cdot dt} \tag{2}$$

$$C = \frac{C \cdot n}{\left(\frac{L}{\Delta x} \cdot \pi \frac{d^2}{4} \right)} \tag{3}$$

where

u_x and u_y	are the velocities along the two x and y axes respectively;
dt	is the duration of the contamination event;
d	is the pipe diameter;
E_f and E_b	are the forward and backward diffusion coefficients, respectively, as defined by Romero-Gomez and Choi (2011). In equation (1) they are a function of positive or negative flow direction.
$(C \cdot n)$	is the concentration per unit of particles;
L	is the length of the pipe;
Δx	is the section number of the pipe;
$\pi \frac{d^2}{4}$	is the cross-sectional area of the pipe.

The advective-diffusive-dispersive model, suitably calibrated (Eb = 0.05 m^2 / s and Ef = 0.3 m^2 / s), was applied to the real Zandvoort network (near Haarlem) in order to compare the results of the model with the values of the monitoring campaign [11] conducted between 2 September and 20 October 2008.

2.3 Experimental and Numerical Setup

Solute transport monitoring was carried out near two apartment buildings (locations 1 and 2), in the basement of a hotel (location 3) and in the basement of a small apartment building of 15 residences (location 4) (Figure 1). Monitoring was enabled by dosing sodium chloride (NaCl) within a booster location nearby the network inlet, raising the electrical conductivity (EC) from about EC=57 mS/m without dosage to about EC=68 mS/m. Short intermittent tracer events (3 hours) were performed with an inter-event time of 20 hours. The tracer study was carried out for 7 weeks, but here we reported only a few days (the event of 3 September 2008).

Electrical conductivity (EC) values were measured at 4 locations in the system, and two models were constructed that are distinguished by demand allocation: ModelTD (top-down) and ModelBU (bottom-up).

The first was allocated to all demand nodes with a correction factor DMP. This correction factor is the base demand and has been assigned based on the same demand category for all demand nodes, having a pattern time step of 15 min. The bottom-up demand allocation was done with the use of the end-use model SIMDEUM that considers a stochastic water demand pattern, obtaining a specific demand pattern as a function of the different types of demand nodes, having a pattern time step of 5 min [11]. In order to validate the ModelBU, in the study of Blokker et al. 2010 [11], 10 different SIMDEUM models were used and, in this work, only one was shown as an example. As leakage in the Netherlands is generally very low (2–4%) ([15]; [16]), no leakage is assumed in this network.

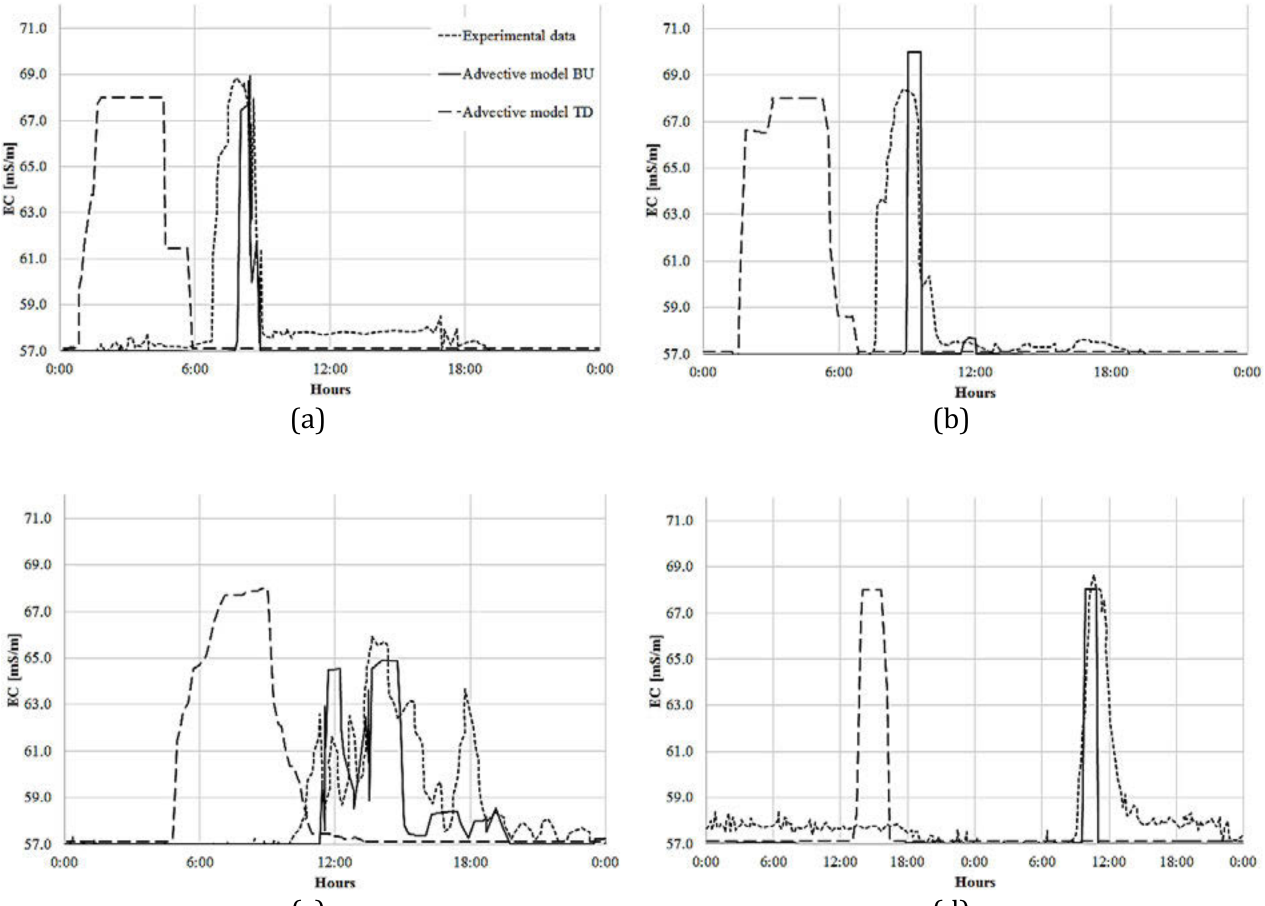
3 RESULTS

In Figure 2 and Figure 3 the numerical results obtained from the resolution of the advective and advective-diffusive-dispersive model respectively were compared, considering both the demand allocation approaches (Top-Down and Bottom-Up), with the electroconductivity (EC) measurements collected during a tracer experiment of 3 September 2008, present in Blokker et al. 2010 [11].

Note that the tracer inserted at the booster location on the 3rd September 2008 reached the locations Burg. Fennemaplein (Figure 2a), De Ruyterstraat (Figure 2b), NH hotel (Figure 2c) a few hours later, while it reached the Sterflat Friedhoffplein location (Figure 2d) one day late, around 9:00 am.

As known from the Figure 2 and the Table 1, in which the values of the Nash-Sutcliffe coefficient have been reported for the four monitored locations, it is observed that the EPANET advective model is not at all representative of the measured data, using both demand allocation models.

In fact, it is observed that the tracer event simulated using the Top-Down approach is significantly anticipated with respect to the real data and, once exceeded, is completely cancelled. In particular, for the location of Sterflat Friedhoffplein (Figure 2d) it should be noted that the advective model detects the event one day in advance. Furthermore, considering the Bottom-Up approach, the advective model tends to provide a much shorter event with a much smaller mass of the tracer reaching the user.



(a) (b)

(c) (d)

Figure 2. Comparison of EC measurements (Blokker et al., 2010) and simulated with only advection considering Top-down and Bottom-up approach for demand allocation for the 3rd September 2008 tracer event at Burg. Fennemaplein (a), De Ruyterstraat (b), NH hotel (c), Sterflat Friedhoffplein (d).

The inadequacy of the advective model in the real data representation is also highlighted by the values of the Nash-Sutcliffe coefficients which in this case are negative for both models (Table 1).

Table 1. Nash – Sutcliffe coefficient for the Top-down and Bottom-up approach for only advective (adv) and advective – dispersive – diffusive model (disp) for $E_b = 0.05\ m^2/s$ $E_f = 0.30\ m^2/s$ for the 3 September 2008 tracer event.

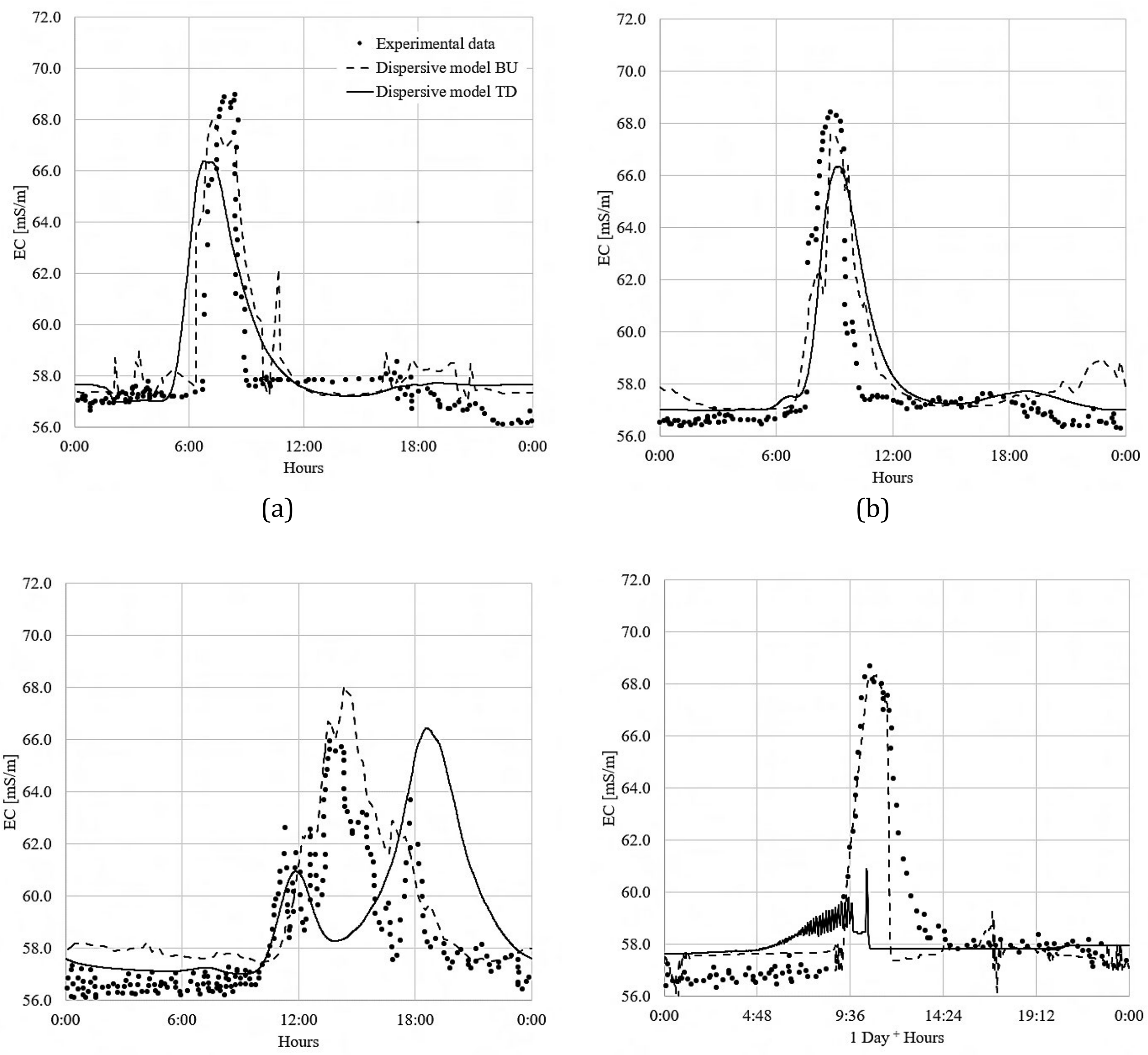
	Burg. Fennemaplein	De Ruyterstraat	NH Hotel	Sterflat Friedhoffplein
Top-Down N – S (adv)	-0.09	-0.12	-1.07	-0.50
Top-Down N – S (disp)	-0.04	0.78	-0.62	-0.44
Bottom-Up N – S (adv)	0.28	0.24	0.22	0.50
Bottom-Up N – S (disp)	0.79	0.88	0.68	0.71

In Figure 3 it is observed that the advective-diffusive-dispersive model is much more efficient in representing the experimental data in the case in which use the Bottom-Up model of allocation of demand.

Using the previously calibrated dispersion coefficients (E_b = 0.05 m^2/s E_f = 0.30 m^2/s), the model not only manages to centre the peak concentration, but is able to best represent the descending traits of the pollutogram. This can be seen from Table 1 in which the values of the Nash-Sutcliffe (N-S) coefficient, used to evaluate the adaptability between the simulated and measured data, are high and in some cases close to unity.

By comparing the experimental results with the data obtained using the Top-Down model of demand allocation, it is observed that the model adequately reproduces only the data relating to the location De Ruyterstraat (Figure 3b). In fact, in this case the model, although with lower peak values, reproduced the real event with an efficiency of 78%.

In all other cases, the model differs from the experimental data as it not only anticipates the real event (Figure 3a and Figure 3d), but is also unable to reproduce the shape of the event. This is more evident in the locations of Figure 3c, in which a double peak occurs, and Figure 3d, in which the curve is remarkably flattened. To confirm this, it is observed from Table 1 that the values of the Nash-Sutcliffe coefficient for the above locations are negative and therefore the model is not suitable for reproducing the real data.



(a) (b)

(c) (d)

Figure 3. Comparison of EC measurements (Blokker et al., 2010) and simulated with dispersion considering Top-down and Bottom-up approach for demand allocation (backward dispersion coefficient=0.05 m^2/s and forward dispersion coefficient=0.30 m^2/s) for the 3rd September 2008 tracer event at Burg. Fennemaplein (a), De Ruyterstraat (b), NH hotel (c), Sterflat Friedhoffplein (d).

4 CONCLUSIONS

The present study applied the EPANET and EPANET-DD model to the real network of Zandvoort (Netherlands), considering two different demand allocation models (Top-Down and Bottom-Up). The models were suitably calibrated from a hydraulic (EPANET and EPANET-DD) and quality (EPANET-DD) point of view. Different demand patterns were used depending on the demand allocation model considered: in the first case, a single pattern that was the same for all demand nodes was chosen with a pattern time step equal to 15 min; in the second case, the demand patterns were obtained through a SIMDEUM stochastic model as a function of the features of the demand node (residential house, hotel, etc.) having a pattern time step equal to 5 min.

Furthermore, the backward and forward dispersion coefficients respectively equal to of 0.05 m^2/s and 0.30 m^2/s were calibrated, using the trial-and-error heuristic method through the Nash-Sutcliffe efficiency coefficient.

The model results were compared with numerical tests performed on the network from 2 September to 20 October 2008 using a conservative tracer.

The analysis showed that by varying the demand allocation model, dispersive and diffusive processes are relevant in the simulation of solute propagation in water networks.

In fact, using the Top-Down approach for the demand allocation, considering both the advective model and the advective-diffusive-dispersive model, it is not able to represent real data in terms of time, as it produces an impulse that arrives several hours before the measured event. This effect is less accentuated by using the complete model. Furthermore, considering the latter model, it underestimates the peak concentration of the tracer and also generates a double peak at the NH Hotel location. Using this approach, only one of the monitored locations (De Ruyterstraat) was adequately modelled with an efficiency of 78%.

However, in all other cases, the ineffectiveness of the model was highlighted by the negative values of the Nash-Sutcliffe coefficient, which allows us to evaluate the predictive power of the models with respect to the measured data.

On the other hand, using the Bottom-Up approach for the demand allocation, the advective-diffusive-dispersive model is much more performing. In fact, not only it is able to centre the pollutogram peaks, but it is able to adequately model the shape of the event, with the Nash-Sutcliffe efficiencies coefficient ranging between 68% and 88%.

The study has highlighted that the use of diffusive-dispersive models coupled with demand patterns as close as possible to the real demand of the users, allows to obtain models that are very representative of reality (in terms of time and concentration peaks) and this it can be a valid modeling tool to be used upstream of optimization models to prevent illegal intrusions into water distribution networks.

REFERENCES

[1] M. Blokker, P. Smeets and G. Medema, "Quantitative microbial risk assessment of repairs of the drinking water distribution system," Microbial Risk Analysis, vol. 8, pp. 22-31, 2018.

[2] R. Murray, T. Haxton, R. Janke, W. E. Hart, J. Berry and C. Phillips, “Sensor network design for drinking water contamination warning systems: A compendium of research results and case studies using TEVA-SPOT,” EPA/600/R-09/141, Office of Research and Development, National Homeland Security Research Center, 2009.

[3] M. Weickgenannt, Z. Kapelan, M. Blokker and D. A. Savic, “Optimal Sensor Placement for the Efficient Contaminat Detection in Water Distribution Systems,” Water Distribution Systems Analysis, 2008.

[4] L. Rossman, “EPANET 2 Users Manual,” Environmental Protection Agency, 2000.

[5] A. Ostfeld and E. Salomons, “Securing Water Distribution Systems Using Online Contamination Monitoring,” Journal of Water Resources Planning and Management, vol. 131, pp. 402-405, 2005.

[6] O. N. Ozdemir and M. E. Ucaner, “Success of booster chlorination for water supply networks with genetic algorithms,” Journal of Hydraulic Research, vol. 43, no. 3, pp. 267-275, 2005.

[7] J. Guan, M. M. Aral, M. L. Maslia and W. M. Grayman, “Identification of Contaminant Sources in Water Distribution Systems Using Simulation–Optimization Method: Case Study,” Journal of Water Resources Planning and Management, vol. 132, pp. 252-262, 2006.

[8] S. Piazza, E. J. M. Blokker, G. Freni, V. Puleo and M. Sambito, “Impact of diffusion and dispersion of contaminants in water distribution networks modelling and monitoring,” Water Supply, vol. 20, no. 1, pp. 46-58, 2020.

[9] S. Piazza, M. Sambito and G. Freni, “Impact of water-quality modelling simplifications on the optimal positioning of sensors in looped water distribution networks,” Preprint (version1) available at Research Square, 06 April 2022.

[10] E. J. M. Blokker, J. H. G. Vreeburg, S. G. Buchberger and J. C. van Dijk, “Importance of demand modelling in network water quality models: a review.,” Drink. Water Eng. Sci., vol. 1, no. 1, pp. 27-38, 2008.

[11] E. J. M. Blokker, J. H. G. Vreeburg, H. Beverloo, M. Klein Arfman and J. C. van Dijk, “A bottom-up approach of stochastic demand allocation in water quality modelling,” Drinking Water Engineering and Science, vol. 3, pp. 43-51, 2010.

[12] P. Romeo-Gomez and C. Y. Choi, “Axial Dispersion Coefficients in Laminar Flows of Water-Distribution Systems,” Journal of Hydraulic Engineering, vol. 137, no. 11, pp. 1500-1508, 2011.

[13] D. G. Eliades, M. Kyriakou, S. Vrachimis and M. M. Polycarpou, EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB, The Netherlands: Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI), 2016, pp. 1-8.

[14] F. Delay, P. Ackerer and C. Danquigny, “Simulating Solute Transport in Porous or Fractured Formations Using Random Walk Particle Tracking: A Review,” Vadose Zone Journal, vol. 4, no. 2, p. 360–379, 2005.

[15] R. H. S. Beuken, C. S. W. Lavooij, A. Bosch and P. G. Schaap, “Low leakage in the Netherlands confirmed,” Ohio, USA, 2006.

[16] P. J. J. G. Geudens, “Water supply statistics 2007,,” Vewin, 2008.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14074

COSMOS – A FRAMEWORK FOR CONTAINERISED, DISTRIBUTED EXECUTION AND ANALYSIS OF HYDRAULIC WATER DISTRIBUTION SYSTEM MODELS

Georg Arbesser-Rastburg[1], David Camhy[2] and Daniela Fuchs-Hanusch[3]

[1,2,3] Institute of Urban Water Management and Landscape Water Engineering, Graz University of Technology, Graz, Austria

[1] *georg.arbesser-rastburg@tugraz.at,* [2] *camhy@tugraz.at,* [3] *fuchs-hanusch@tugraz.at,*

Abstract

Many scientific problems related to water distribution systems like optimization problems or sensitivity analysis require the creation and execution of a large number of hydraulic models. To reduce computation times, different approaches have been used in the past, often by employing multiple CPU cores to solve the hydraulic equations of a single model or to simulate multiple models in parallel on a single computer. However, these approaches often cannot make use of distributed computing. Furthermore, using these approaches in applications with a (web-based) graphical user interface (GUI) often requires the development of tailored software solutions and application programming interfaces (API) to link GUI and model execution backend.

To tackle these issues, we propose COSMOS (Containerised Model Simulator), a highly scalable Python-based framework which allows for the modification of hydraulic models using OOPNET, an API between Python and the hydraulic solver of the modelling software EPANET and can run model simulations. Simulation results can be then further analysed while all these tasks run encapsulated in containers in a cluster. It also allows to easily link the described functionality with other applications by providing a REST API.

A standard-based OpenAPI allows for passing hydraulic models and running scientific workflows via HTTP and generating clients based on the provided OpenAPI schemas, which simplifies the creation of web-based user interfaces. Python was chosen because of its growing spread in the scientific community, the availability of data processing and optimization packages and its high code readability.

Prefect, a data workflow orchestration framework, is employed to create workflows, starting with the transformation of hydraulic models into JSON representations for further use in web applications. The models can then be executed and simulated distributed over the available CPU cores (locally or in a cluster). Further tasks for doing analysis in the cluster can be easily added if necessary. Hydraulic models and simulation artifacts are stored on S3-compatible storage and can be easily retrieved.

A main advantage of this approach is the use of containers, which allows for reproducible workflows. Compared to other high performance computing approaches and container-based systems, Prefect has the advantage of being able to keep dedicated worker nodes available for use. This comes in handy especially when dealing with relatively short computation times where the start of a container might take longer than the actual simulation. Additionally, simulation data post-processing can be easily added to workflows in Prefect. Furthermore, as the structure of COSMOS is highly scalable, it can be used for different levels of problem complexity and simulation runtimes.

Keywords
Hydraulic modelling, EPANET, containerisation, model simulation backend.

1 INTRODUCTION

Hydraulic models have a wide variety of applications, from optimization problems like optimal design or optimal sensor placement, to simulating different kinds of operating conditions like system failures, to tasks like leak localization. Several application programming interfaces between the hydraulic modelling software EPANET [1] and different programming interfaces have been developed to simplify these tasks [2]–[5]. While there are several examples of web applications that use simulation results from water distribution system models [6]–[9], there appears to be only one framework that is targeted at linking web applications with simulating hydraulic models, epanet-js [3].

To link such a web interface with simulating hydraulic models, an interface to a hydraulic solver is necessary and the library epanet-js, written in JavaScript, provides such an interface. However, it is not does not provide a framework for running more complex workflows and that supports the fast simulation of many hydraulic simulations in parallel [3]. Speeding up the execution of hydraulic models in general has already been of interest to researchers in the past.

One possible way to speed up model execution is to employ several CPU cores in parallel to simulate a model. Wu and Elsayed developed a parallelization algorithm to concurrently run hydraulic and quality simulations [10]. Some employed parallelization to compute the individual models faster by distributing the computational load across several CPU cores [11], [12]. This approach's scalability however is limited, since the execution is constrained by the number of processor cores available at the used workstation. Additionally, Burger et al. found that they were not able to develop a solver that outperformed EPANET's solver when using real-world hydraulic models and even raised the question, if any solver will ever be faster than EPANET's original solver [11].

Another approach is to parallelize the computation of a set of hydraulic models. In this case, every processor core handles the computation of a single model. This approach however is again limited by the number of processor cores available at the used workstation [13].

Instead of doing calculations on a local workstation, computations can also be outsourced to a dedicated server infrastructure. Using distributed computing the computational load is spread across many computers and therefore the number of available CPU cores is increased. This approach has already been employed in the field of water distribution systems in the past. Alonso et al. used the Message Passing Interface (MPI) to distribute the calculation of hydraulic equations across several PCs with a custom hydraulic solver [14]. Wu and Zhu also used MPI to distribute the optimization of pump schedules [15]. Hu et al. developed a genetic algorithm for sensor placement that is based on the cluster computing framework Apache Spark [16]. Additionally, several frameworks written in the popular programming language Python have been developed for distributing calculations in a cluster in a simple manner. Examples for such frameworks are Celery, Apache Airflow, Prefect, and Dask.

Any results generated in a scientific context should be reproducible by others to validate conclusions or develop new methods based on existing research. Containers that package the environment and software required for running code can help facilitate reproducibility [17].

Containers are similar in their functioning to virtual machines (VMs) as both concepts rely on virtualization. In contrast to VMs, containers however virtualize software while virtual machines also virtualize the underlying hardware. Containers in contrast to virtual machines share the host's kernel and offer almost the same performance as the host's operating system, decreased starting times and a reduced storage footprint on the host machine [14].

Containerizing model execution has several advantages. First, containers can help with analysis reproducibility. When using a suitable container image repository (e.g., Docker Hub or GitLab),

BY NC SA

versioned images can be kept as an archive and later be reused. Second, it leads to the possibility to quickly scale the number of available worker nodes by using a suitable orchestration software (e.g., Kubernetes or Docker Swarm). Third, containers can be easily deployed on workstations locally, to develop and test the containers. This can be further enhanced with continuous integration and continuous delivery (CI/CD) systems that automatically test analysis tasks for their correctness [15] and build the images for the containers. A container image is a blueprint for containers that include the entire environment and include all code necessary for running the code in the container.

In this paper we present COSMOS (Containerised Model Simulator), a framework for containerised hydraulic simulations that employs cloud computing and can be accessed by web-apps via a standard-based API. Section 2 describes the requirements that were determined while developing a frontend for hydraulic model simulations. In section 3 different available cloud computing frameworks are analysed regarding their suitability as web-app backends for scientific applications. COSMOS itself is described in section 4 and section 5 finally gives an outlook into further possible enhancements and use cases of COSMOS.

2 REQUIREMENTS FOR HYDRAULIC MODEL SIMULATION WEB-APP BACKENDS

Requirements for a hydraulic model simulation backend were derived during the development of an interactive web-based application that allows users to execute complex scientific workflows that are based on hydraulic simulations (e.g., sensitivity analysis or calibration).

The frontend should provide users with the possibility to manage stored models and their simulation and analysis results and run pre-defined algorithms or tasks via a REST API. Optionally, a graphical user interface (GUI) should provide a platform for easy execution of workflows.

First, the requirements for the backend were derived. They can be grouped into general, web application specific and scientific requirements. Below is a list and description of the requirements that were identified during an internal co-creation process:

Common requirements:

- Scalability

  The system should be able to cope with both very short tasks as well as longer-running and more complex tasks that require the simulation of many hydraulic models at the same time. This required a framework that allows for scaling from a small number of computing nodes to a large-scale computing cluster. The web application and the distributed computing framework should finally be deployed to a Kubernetes cluster.

- Python-based or existing Python client

  A well-established programming language was required to reach a wide audience. The choice fell on Python due to its simple syntax, its many existing libraries (by May 2022 the Python Package Index listed more than 375.000 projects) in general and especially the libraries tailored towards scientific usage like NumPy, SciPy or pandas.

- Usage of open-source and free software

  Open-Source provides a transparent view on the implementation of the underlying algorithms and delivers an easy way to communicate about issues in implementations or get helpful support from the community. Using free software in addition mitigates financial obstacles for reproducing results.

- Stability and support

  A mature and well supported framework was required to guarantee long-term support. Also, the documentation should be extensive and well written to enable new users to get into the framework more easily.

- General Data Protection Regulation (GDPR) conformity

  Scientific analysis is sometimes based on personal data that must be treated according to GDPR requirements. An on-premises solution was sought after to keep all data on internal servers in a controlled environment.

- Easy-to-use

  One of the most important requirements was the usability for users. The platform should provide convenient and easy-to-use entry points for both experienced developers via a standard-based API as well as for users without a dedicated IT background via a web GUI.

- Centralized and findable

  To help other scientist in getting insights in already processed research topics, a history of executed workflows and their metadata should be centrally stored together with used parameters and obtained results.

- Monitoring and alerting

  Users should be alerted about failed tasks and workflows via multiple channels (e.g., email or different messengers) and querying the current state of running workflows as well as their results should be possible. Keeping track of computing resources requires easy-to-use monitoring that allows for assessing the computing resources in use.

Web application specific requirements:

- Low-latency Execution

  Users should be able to interactively explore algorithms and their results in a responsive environment. Short running tasks should provide immediate feedback, which lead to the requirement of “low-latency” workflow executions. This means that when executing a relatively small number of models with a short runtime, the system should return results as quickly as possible. This requires a framework that adds little overhead to the executed analysis and simulation tasks.

- Easy API Access

  Integrating the platform into other web apps should be possible via an easily accessible API. Creating clients in different programming languages should facilitate the integration in other apps, for example by providing an OpenAPI or GraphQL API.

- Result and model storage

  Results and models should be stored centrally and in an easy-to-use fashion. Versioning of results and models should guarantee reproducibility. Hosting the storage on-premises should be possible as well as access via APIs in different programming languages.

Scientific requirements:

- Reproducible and repeatable complex workflows

  Main goals for research tasks and workflows are reproducibility and repeatability, so that others can evaluate and reproduce any generated results. This requires management of input and output data, the corresponding metadata and which programming code or workflow description was used to generate the results. Easy exchange and versioning of workflow descriptions also contributes to the openness of the methods used and makes it easier for other researchers to reproduce the results. This also includes any used software in the workflow run's environment, for instance EPANET if its hydraulic solver is being used.

Depending on the algorithms used (e.g., evolutionary algorithms) the need for complex workflows can arise, where one or many steps can be dependent on the previous ones. Therefore, only frameworks that already support separating tasks in terms of small units of code and task dependencies were considered.

- Easy integration into existing scientific software packages

  Since Python and its accompanying scientific stack offer a variety of scientific tools, a solution that provides a similar interface was required. Users familiar with those tools should be able to transition seamlessly into the new distributed computing environment.

- Easy local development, testing and debugging

  Developing workflows locally should allow researchers to test and debug their code. Being able to run and test algorithms locally before running them in a computing cluster was deemed necessary to support the scientific workflow.

- Integration with already existing workflows

  While the execution of models is the main use case discussed in this paper, an integration of other already existing scientific tasks like machine learning or measurement data pipelines would provide a great benefit. The focus however lies on running hydraulic simulations.

3 DISTRIBUTED COMPUTING FRAMEWORKS

One of the most important aspects of the development of COSMOS was the evaluation of distributed computing frameworks with Python bindings.

Based on the requirements stated above, several frameworks were evaluated. Frameworks that did not fulfil all requirements but where the missing features could be implemented with low effort were also considered. Exchange and versioning of workflow descriptions if not already integrated into the framework can for instance be provided by git or other versioning systems. The evaluation was based on the framework documentations and small test runs to get to know the frameworks. Additionally, some of the software packages listed below have been in use at the Institute for several years so limitations and features were already clear.

All the frameworks were open-source and freely available. They all provided enough documentation and support to get to know the frameworks well enough to assess their features. Only frameworks where scalability according to the requirements listed above was given and which enable GDPR conformant workflows were considered. Support for containers should provide a reproducible environment and while the actual implementation between the frameworks is different, they all provide a way to use containers as execution layers.

Besides full-blown workflow scheduling solutions, GitLab as advocate for classical DevOp platforms and Jenkins as a more general automation platform were taken into account. Celery and Dask, while being more low-level in their abstractions, were also evaluated especially because the web-app approach requires a low-latency execution of certain workflows and tasks. Argo Workflows was considered because it provides support for running tasks in Kubernetes clusters as well. HTCondor was added to the mix as a more classical batch system which is often readily available on super-computers in scientific infrastructures. Prefect as a rather new competitor was considered because of its included abstractions and the good documentation. Finally, because of its widespread usage for cloud computing tasks, Apache Airflow was considered.

- Apache Airflow [18]

  Apache Airflow is a workflow scheduling and monitoring solution. It provides dynamic workflow descriptions written in Python, which helps users who already have experience in Python to generate more formalized workflow descriptions. Versioning of workflow descriptions is therefore very easy using git or other version control systems. It provides modular executors which allow scaling to different infrastructures like Kubernetes. It can act as layer over Celery and Dask which provides great flexibility. Integrated monitoring and logging as well as great expandability would make this a great solution for many of the requirements. However, Apache Airflow uses a central scheduling loop and jobs require a distinct execution date and time, which does not cover the use case of interactive web applications very well. First tests also showed that it did not behave according to the low-latency requirement when many workflows with small fast returning tasks are executed.

- Argo Workflows [19]

  Argo Workflows is a Kubernetes based and container native workflow-engine with good documentation and support for complex workflows. Workflows are created using Kubernetes manifests and can therefore easily be versioned and integrated in Kubernetes native GitOps frameworks like Argo CD. By using Argo Events the scheduling of workflows can be abstracted and there are many event-sources supported. Protocols like MQTT or NATS could then be used to send events that trigger a workflow execution.

  Being (only) Kubernetes based can be seen as a plus or minus depending on the use-cases stated in the requirements. While the main execution platform for the web-application is a Kubernetes cluster that can be easily scaled, reusing the workflow description in another infrastructure would not be possible.

  The main disadvantage however is the fact that Argo Workflows usually starts containers in the cluster only when needed and not permanently. While this provides great reproducibility and repeatability, it also adds significant overhead to the workflow execution and seems more suited to long running tasks. In our trials the container start time often exceeded the model execution time, especially when simulating small hydraulic models with run-times of less than a second. Therefore, the low-latency requirement is not fulfilled.

- Celery [20]

  Celery is a distributed task queue system with a large community of users. By building on a message broker like RabbitMQ or Redis and by deploying long-running workers, it adds very little. It provides a result backend abstraction which allows for keeping results connected to the task executions and therefore fulfils some of the centralization and findability requirements. In comparison to Apache Airflow and Prefect it appears to be a more low-level framework (for instance Apache Airflow has its dedicated Celery executor). Celery does not expose a standard based API for starting workflow runs, although the tool Flower provides API endpoints for monitoring Celery [21]. Triggering workflows via an API would therefore require the implementation of a custom API with a web framework like Flask, FastAPI or Django.

- Dask [22]

  Dask provides readily available larger-than-memory data structures built on common interfaces like NumPy, pandas or Python iterators making it well-suited for many scientific

BY NC SA

workflows. Switching between a local and a distributed scheduler is easily possible and does not impact the basic algorithms' design. The distributed scheduler adds very little overhead and seems similar in performance to Celery with its message broker approach. Also of interest are the multiple ways of deploying Dask clusters. It is possible to execute tasks on Kubernetes, via SSH and even on high performance computing (HPC) resources. This allows users to design scientific algorithms independent of the infrastructure it is running on. However, Dask does not include an abstraction layer for workflow definitions and task execution monitoring via an API.

- HTCondor [23]

  HTCondor is a batch software which was already used extensively at our Institute for large scale model execution and well documented. Being a more classical batch system, the tool DAGMan adds support for workflow definitions. An advantage of HTCondor is the possibility to use free computing resources from user workstations when they are not in use. It is very well suited for an extremely large number of models and monitoring can be performed over the command line. Preliminary tests showed that the low-latency execution of models was slower than the other approaches. Also there seems to be no already available monitoring solution that can be integrated easily in a web app.

- GitLab [24]

  GitLab as a representative for git implementations with support for continuous integration and continuous delivery (CI/CD) pipelines was considered as well. Some of the scaling requirements are fulfilled by the concept of GitLab runners and GitLab also provides an easy-to-use API for querying pipeline runs and their status. However, it seems not to be very well suited for the low-latency requirement and first tests showed a considerable delay between starting, scheduling and running pipeline executions.

- Jenkins [25]

  Jenkins is an open-source automation server and is used for CI/CD pipelines. While being mainly built for CI/CD pipelines, Jenkins can also be used for more general automation tasks. Using Jenkins pipelines, reproducibility and repeatability requirements can be fulfilled by using a version control system. Running pipeline steps either in a Kubernetes cluster, via SSH on Linux worker nodes and even Windows workstation, the scalability requirement was met as well. However, as with GitLab, low-latency execution of tasks could not be achieved.

- Prefect [26]

  Prefect is a data pipeline orchestration and runtime system that can use Dask for executing complex workflows. It offers abstractions for tasks, workflows, storage and executors. Among the supported storages are Docker images, Git, AWS Simple Storage Service (S3) and Bitbucket. Workflow definitions and result storages can be defined for workflows and tasks individually. A web interface can be used to trigger workflow runs. Workflow runs are then started by Agents. Agents provide the environment that is needed to start the workflow, i.e., they contain all the necessary code and dependencies and keep track of the workflow run's status. KubernetesAgents for instance start Kubernetes jobs that first pull the newest workflow definition from the designated storage and then start the workflow run. A GraphQL API supports triggering workflow runs, monitoring their status, and reading a task's result location in the used storage. Furthermore, since Dask can be used as an executor, Dask's API for larger-than-memory objects is available for usage as well. Prefect can be either used in its free and open-source version called Prefect server, or as the paid service Prefect Cloud. Prefect Cloud includes further functionalities like user authentication and a secret store. Prefect 2.0 Orion is currently under development, which will include an OpenAPI instead of a GraphQL API.

After the first evaluation, Celery and Prefect seemed to fulfil the requirements better than the other frameworks. To choose between them, the two frameworks were further evaluated in terms of the required service infrastructure to assess their integration in scientific workflows. Figure 1 shows Celery's service structure, while Figure 2 shows a simplified version of Prefect's structure.

Celery itself does not include an API that enables starting workflow runs via HTTP requests. A REST API would have to be implemented using a Python web framework like FastAPI, Flask or Django ("Producer"). This API would then send a task to a task queue system. Celery offers support for RabbitMQ, Redis, Amazon SQS and Zookeeper.

The tasks in the task queue are then scattered across Celery workers running in a cluster. These workers are responsible for executing the tasks ("Consumers") and need all the dependencies required for the task's execution in their environment. The task results are then sent to a central result backend. Out of the box, Celery supports Redis, RabbitMQ and SQLAlchemy as backend.

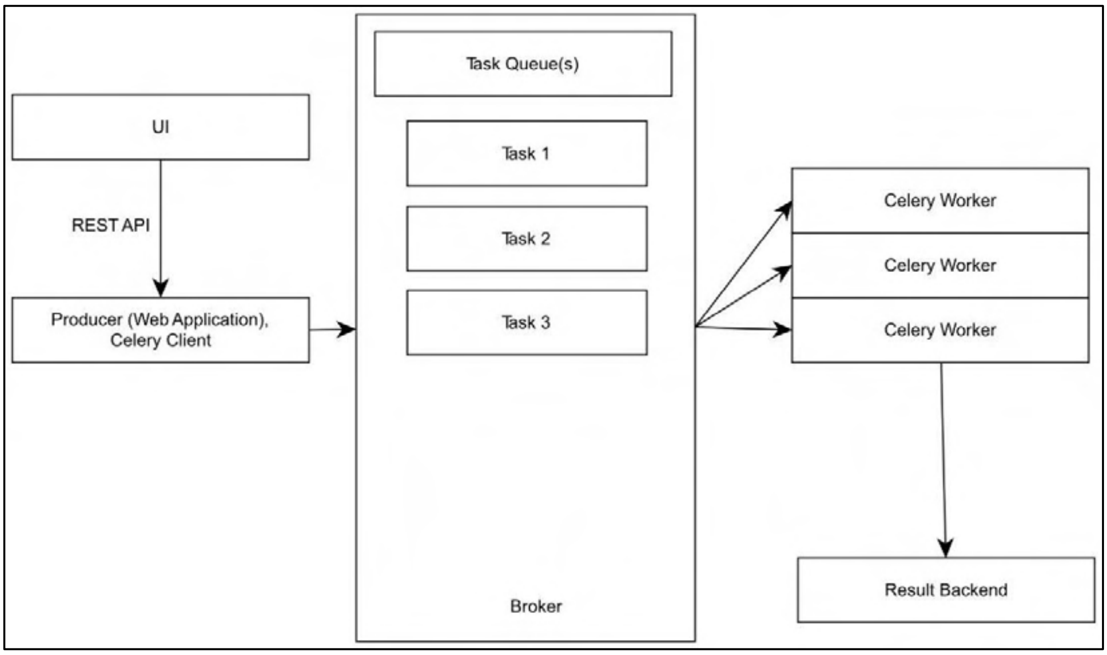


Figure 1. Celery service structure.

Prefect follows a slightly different approach by implementing its own task scheduling system and adding additional layers of abstraction. It also requires more dedicated services to be running. Prefect uses Apollo and Hasura to host a GraphQL API for many functionalities like starting workflow runs and querying workflow states. Prefect 2.0 however will implement an OpenAPI. Using the GraphQL API, users can trigger workflow runs by sending a request to the Prefect Server or Prefect Cloud, depending on whether Prefect's cloud service is being used or if Prefect is being hosted on-premises.

Prefect Agents query Prefect Server/Cloud for any scheduled workflow runs and are responsible for providing an environment that can execute the workflow (i.e., all dependencies and requirements are fulfilled in the environment). There are several different Agents available, that rely on different technologies for providing the environment (e.g., a DockerAgent that starts a workflow run from within docker containers or a KubernetesAgent that runs flows as Kubernetes Jobs).

BY NC SA

How the Agent accesses the workflow definition can be controlled by choosing a suitable storage and including it in the workflow definition. Options range from Python modules accessible within the Agent's environment, to GitStorage that pulls the newest workflow definition via git or DockerStorage which pulls a container image that includes the workflow definition. If a storage solution like DockerStorage or GitStorage is chosen, the Prefect Agent pulls the newest workflow definition before executing it.

How a workflow is then executed, is part of the workflow definition. A LocalExecutor executes the workflow in the Agent's local environment, while a DaskExecutor can execute tasks in a Dask cluster.

Finally, results are stored in a Storage as well and the path to the result can be queried via the GraphQL API.

In addition to the services shown in Figure 2, Prefect also relies on additional services for stopping tasks that no longer communicate with the API, scheduling new tasks and maintenance routines.

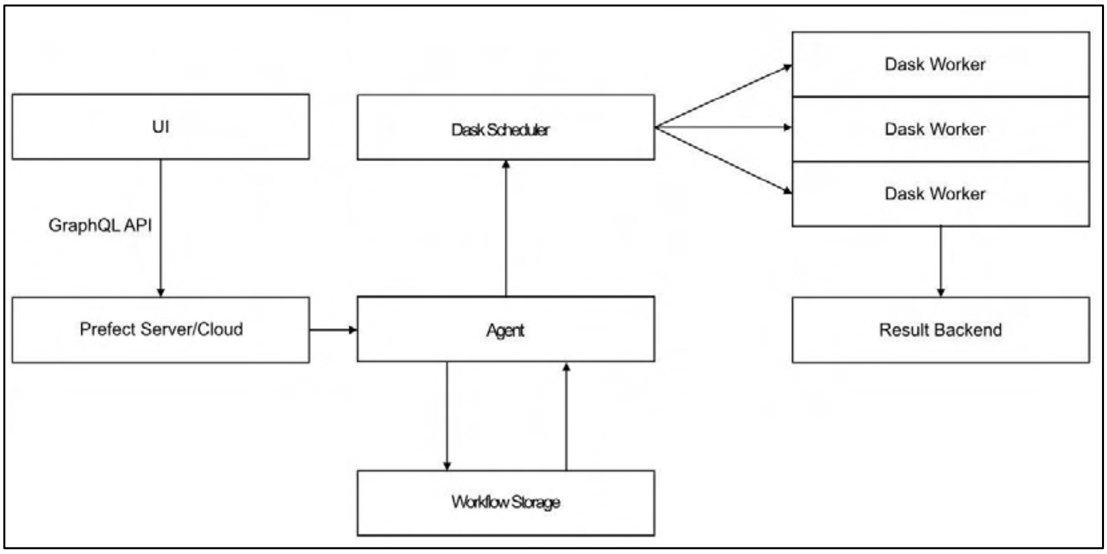


Figure 2. Prefect service structure.

Compared to Celery, Prefect includes more abstractions that can be used to control how artifacts and workflows are stored or how workflows are executed. It also provides a GUI for starting and monitoring workflow runs, whereas Celery does not provide any GUI. Different Agents allow for using different technologies to encapsulate a workflow depending on the available infrastructure and needs.

Native support for Kubernetes and Docker, a good documentation, the easy syntax that uses regular Python constructs for defining task dependencies and the additional functionality provided by Dask in the end lead to Prefect being chosen as basis for COSMOS.

4 COSMOS

COSMOS was created with the intention to provide a platform for executing a variety of different modelling tasks, from simulating simple hydraulic models to more complex analysis workflows if necessary. COSMOS is built upon several existing services, frameworks and standards that allow for an open structure that can be easily enhanced and is focused on reproducibility, easy usage, a central model and result storage and low-latency task execution.

To develop an API for webapps that is based on Python, one of two hydraulic modelling APIs can be used: the Water Network Tool for Resilience (WNTR) [5] and OOPNET [2]. Both provide basic functionalities like parsing EPANET input files, manipulating hydraulic models and simulating hydraulic models with EPANET. OOPNET was chosen as basis for COSMOS, since it has been in use at the Institute for several years and therefore many algorithms have already been implemented using it. This leads to an increased available codebase for future applications.

4.1 Service structure and functionality

COSMOS employs several services to satisfy the requirements specified in section 2:

- Prefect serves as workflow management and scheduling system
- Dask acts as execution backend
- The Amazon Simple Storage Solution (S3) compatible storage MinIO is used as model and simulation result storage
- FastAPI provides a RESTful interface following OpenAPI specifications and authorization
- Elasticsearch for finding models and simulation results in the storage backend

To use COSMOS, users first have to upload their hydraulic models via the provided REST API by sending a POST request to the FastAPI backend. This REST API wraps Prefect's GraphQL API. This was done to add a layer of abstraction so that the execution backend can be easily exchanged in the future, if Prefect 2.0 shows significant advantages after it has reached a stable state. As part of the upload request, users are able to add tags that can later be used for querying stored models, and a description. The hydraulic models are then converted into GeoJSON files, which are more widely compatible with web applications compared to EPANET input files.

GeoJSON is a file format based on JSON (JavaScript Object Notation) but includes geographical features. It implements different "Geometry" types (e.g., Points, LineStrings or Polygons), that include one or more "Positions" which themselves are an array of coordinates. Geometries are then combined with properties to form "Features". A collection of Features can finally be represented as a "FeatureCollection".

In COSMOS, features are equivalent with physical model components. Nodes (junctions, tanks and reservoirs) are modelled as Points and links (pipes, pumps and valves) as LineStrings. The model itself is represented as a FeatureCollection. Curves, patterns, and model settings are converted into regular JSON objects and also added to the GeoJSON model as so called "foreign members". Foreign members extend the GeoJSON specification with additional key-value pairs.

The conversion is handled by a Prefect task that employs pydantic, a package for creating data models based on the native type hints integrated in Python 3. Different validators can be used to for examples validate data types and value ranges in an intuitive manner. Additional custom validators can be added to the models as well (e.g., validating the IDs of model components corresponding to EPANET's ID length requirement or ensuring that a tank's initial tank level is between the tank's minimum and maximum water levels). A basic check of model validity can therefore be optionally run open model upload.

Pydantic can also write model instances into regular JSON files but does not include base models for GeoJSON objects. This functionality is provided by another package: geojson-pydantic. geojson-pydantic provides additional models that follow the GeoJSON specifications for the data types described above (e.g., Points and LineStrings but also Properties and FeatureCollections).

After a model has been uploaded and converted into a FeatureCollection object along with all settings and model components, this new model representation is stored on an S3 storage. The example deployment that comes with COSMOS includes MinIO, a S3 compatible object storage that can be hosted on-premises. Alternatively, Amazon S3 could be used as storage as well. S3 compatible storage was chosen for several reasons:

1. S3 storage is an object storage.

   COSMOS can simply create a GeoJSON object and store it on S3. Any web application can then load the GeoJSON object and parse it. Furthermore, other files can easily be stored on S3 without more complex preparatory steps like writing database models.

2. S3 provides versioning for objects.

   When models or simulation results are updated, the original file is not lost but can still be recovered afterwards.

3. Files on S3 can be encrypted.

   S3 provides the possibility to encrypt data very easily. While this was not a requirement for COSMOS, this might be beneficial for users who want to store sensitive data.

Upon model upload on S3, an event notification pushes the model JSON and any passed metadata to the search engine Elasticsearch. Elasticsearch is a search engine written in Java that stores data as JSONs. Clients can afterwards run queries using a RESTful API to find stored models.

After a model has been stored on S3, it is available for use in COSMOS. Three additional Prefect tasks are available in COSMOS:

- One task is responsible for querying a hydraulic model stored on S3 by model name and tags, loading the corresponding GeoJSON file and converting it back into an OOPNET model object.
- Alternatively, another Prefect task can be used to load a GeoJSON model from S3. This task simply takes the model's path on S3 as argument, loads the model's GeoJSON representation from the provided path from S3 and converts the model back into an EPANET model. This task comes in handy when a model has been uploaded to an S3 storage by another application and stored for use in COSMOS.
- Finally, the third Prefect task handles the simulation of an EPANET model. The task only takes one argument: the hydraulic model to simulate. OOPNET is used for simulating the model and creating a SimulationReport object. This object serves as a container for simulation results like node pressure and pipe flow rates in OOPNET. For further use in other applications, this object is again transformed into a JSON object via pydantic and stored on S3. The path to the result file is automatically stored in Prefect Server/Cloud and can be queried using either Prefect's GraphQL API or the REST API included in COSMOS.

While the tasks are executed, the COSMOS' FastAPI can be used to query their status via a dedicated API endpoint to check if a task has been executed successfully. Finally, a route allows users to query the results for a specific workflow execution.

4.2 Continuous integration

To keep the basis of COSMOS reliable, allow for linking COSMOS versions with simulation results for increased reproducibility and to make releases easier, a continuous integration pipeline has been implemented. The pipeline runs through several stages:

- Testing
- Version number generation
- Python package and Docker image building
- Creating a new release of COSMOS

The testing stage runs several hydraulic simulations (both single period analysis and extended period simulations) in COSMOS and compares the results with pre-calculated results that are included in COSMOS' testing module. This in combination with other unit-tests assures the reliable functioning and reproducibility of results generated with COSMOS. This however also requires an S3 storage available in the testing environment. If any test fails, the CI pipeline itself fails as well and no new COSMOS version is released.

Next, the version of the next COSMOS release is determined. The version numbers follow semantic versioning and is automatically generated based on the git commit messages that haven been submitted since the last release. This

After the new version number has been derived, first a Python package is built and uploaded to a Python package registry. Afterwards, a Docker image based on a Docker image provided by Dask itself (*daskdev/dask*) is built. In addition to the services and packages required for running a Dask worker node, EPANET, OOPNET and COSMOS are added to the Docker image.

In a final step, a new release is created that uses the previously derived version number, the corresponding Python package and Docker image and a changelog based on the Angular git commit messages.

Users are then able to install the framework on their local workstations and develop workflows for usage in COSMOS. If additional dependencies have to be available in a workflow run's execution environment, users can build their own Docker images if needed using the COSMOS image as base image.

4.3 Low-latency execution vs. flexibility during development

For low-latency execution tasks like running a small number of hydraulic simulations, the deployment of Prefect has to be optimized. This includes choosing a suitable Agent and Storage type for the workflow definition.

The Agent type used is an important aspect regarding execution speed on the one hand and flexibility during development on the other. A DockerAgent is able to pull a Docker image to provide the environment necessary for a workflow's execution every time a workflow is executed. While this approach is very flexible since the Agent requires hardly any further setup, it also leads to longer workflow execution times. A LocalAgent however is meant to be running in an environment that already fulfils the requirements and dependencies of the workflow. An easy way for keeping the time from workflow run submission to workflow execution low is running a LocalAgent in a dedicated container that already contains all the dependencies needed. However, this requires users to take care of running an Agent based on the most recent execution environment Docker image. A CI/CD pipeline can be used to simplify this process.

In addition to Agents, different Storage types are available in Prefect as well. Storage types like DockerStorage or GitLabStorage provide users with the possibility to load the most recent workflow definition from a central storage. This is well suited for tasks that don't require a low-latency execution since it also adds overhead to the execution. Pulling the latest workflow version can be skipped if the most recent version of the flow is already available in the Prefect Agent. This can be achieved by packaging all dependencies and workflows in a Docker image and using a ModuleStorage in the workflow definition. A ModuleStorage points to a workflow already available in a local Python module. Similar to the LocalAgent, this means that the most recent workflow definition has to be available in the execution environment image.

4.4 Extending COSMOS

Since COSMOS mainly provides additional features in Prefect, new Prefect tasks can be created and added to the existing workflows using the Prefect syntax. Hydraulic modelling tasks can be implemented using OOPNET's syntax, while simulation results are available as pandas DataFrames. Pandas is a powerful library for data manipulation and analysis which leads to high flexibility regarding result analysis in dedicated analysis tasks.

Due to being based on containerization, any newly added tasks and workflows have to be included in the used container images. This can be done by using a CI pipeline similar to the one used in COSMOS itself, or manually by using the Docker command line interface to build a new container image.

5 CONCLUSIONS

COSMOS is already being actively used while developing a scientific task execution platform. It provides an OpenAPI based RESTful API that can be used in any web application that requires more complex workflows or the execution of several hydraulic models at once. Hydraulic models and simulation results are both stored in conformity with the JSON and GeoJSON standards which makes working with them in a web context easy.

JSON files however tend to be verbose compared to EPANET input files and therefore increase in size rather quickly. This should be mitigated in the future by e.g., using compression or other file size reducing approaches. Also use cases outside of web development might benefit from more concise file formats.

Concerning storage, alternatives to S3 might be added in the future to support storing models and results in a database. Databases like PostGIS would add further usability options to software like the geographic information system QGIS and would enable the modelling of relationships between, for example, measurement data and hydraulic models.

The current version only implements basic functionalities regarding the handling of simulations, hydraulic models and simulation artifacts. Future releases it will be extended to include various algorithms and methods related to hydraulic modelling. Work on migrating already existing algorithms (e.g., roughness calibration and water distribution system sectorization) have already begun. While right now COSMOS is not publicly available yet, it will be hosted on a code sharing platform like GitHub to reach a wider scientific audience and gather a user community that can add new algorithms.

In addition to the platform for running scientific analysis with COSMOS, another useful feature would be a graphical user interface for managing hydraulic models and linked model simulations. There are already plans to implement such an interface for the task execution platform.

Currently, a new version of Prefect is being developed. The new version promises to be easier in usage, has a slimmer structure that requires less services to be running. Furthermore, the integration of Prefect flows and tasks into native Python code is claimed to be improved which would be beneficial for using COSMOS as a general model simulation backend in scripts.

COSMOS' approach could also be used for other types of models such as EPA SWMM using one of the Python to SWMM APIs available. This would shift COSMOS from being focused on water distribution system models to being a more flexible model execution backend.

Since COSMOS was developed while building a scientific platform for modelling and result analysis workflows, it can be used in a wide variety of ways. It supports long running tasks but also rather short running simulation tasks while providing a standard-based interface that allows for the easy creation of clients in many different programming languages.

Exemplary other use cases are more complex online EPANET editors that take factors into account that lead to an increase in necessary model execution runs like different operating conditions or Monte Carlo simulations. Furthermore, COSMOS could be integrated into analysis scripts written in Python to easily parallelize the execution of many hydraulic models at once while also making use of the reproducibility features of Prefect and COSMOS.

6 REFERENCES

[1] L. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, EPANET 2.2 User Manual. Washington, D.C: U.S. Environmental Protection Agency, 2020.

[2] D. B. Steffelbauer and D. Fuchs-Hanusch, “OOPNET: An object-oriented EPANET in Python,” Procedia Engineering, vol. 119, 2015, pp. 710–718.

[3] L. Butler, Á. Bautista, and C. Dzuwa, “EPANET-JS”, https://github.com/modelcreate/epanet-js, 2022.

[4] E. Arandia and B. J. Eck, “An R package for EPANET simulations,” Environmental Modelling & Software, vol. 107, Sep. 2018, pp. 59–63.

[5] K. A. Klisel, R. Murray, and T. Haxton, “An Overview of the Water Network Tool for Resilience (WNTR),” in Proceedings of the 1st International WDSA/CCWI Joint Conference, Kingston, Ontario, Canada, 2018, p. 8.

[6] D. B. Laucelli, L. Berardi, and A. Simone, “A Teaching Experiment Using a Serious Game for WDNs Sizing,” 13th International Conference on Hydroinformatics, vol. 3, 2018, pp. 1104–1112.

[7] D. Savic, M. Morley, and M. Khoury, “Serious Gaming for Water Systems Planning and Management,” Water, vol. 8, no. 10, Art. no. 10, Oct. 2016.

[8] G. Arbesser-Rastburg and D. Fuchs-Hanusch, “Serious Sensor Placement—Optimal Sensor Placement as a Serious Game,” Water, vol. 12, no. 1, Dec. 2019, p. 68.

[9] T. Bayer, D. P. Ames, and T. G. Cleveland, “Design and development of a web-based EPANET model catalogue and execution environment,” Annals of GIS, vol. 27, no. 3, Jul. 2021, pp. 247–260.

[10] Z. Y. Wu and S. M. Elsayed, “Parallelized Hydraulic and Water Quality Simulations for Water Distribution System Analysis,” in World Environmental and Water Resources Congress 2013, Cincinnati, Ohio, May 2013, pp. 892–902.

[11] G. Burger, R. Sitzenfrei, M. Kleidorfer, and W. Rauch, “Quest for a New Solver for EPANET 2,” Journal of Water Resources Planning and Management, vol. 142, no. 3, Mar. 2016, p. 04015065.

[12] O. Piller, M. Le Fichant, and J. E. Van Zyl, “Lessons learned from restructuring a hydraulic solver for parallel computing,” in WDSA 2012: 14th water distribution systems analysis conference, Adelaide, South Africa, Sep. 2012, pp. 398–406.

[13] S. Artina, C. Bragalli, G. Erbacci, A. Marchi, and M. Rivi, “Contribution of parallel NSGA-II in optimal design of water distribution networks,” Journal of Hydroinformatics, vol. 14, no. 2, Apr. 2012, pp. 310–323.

[14] J. M. Alonso et al., “Parallel Computing in Water Network Analysis and Leakage Minimization,” Journal of Water Resources Planning and Management, vol. 126, no. 4, Jul. 2000, pp. 251–260.

[15] Z. Y. Wu and Q. Zhu, “Scalable Parallel Computing Framework for Pump Scheduling Optimization,” in World Environmental and Water Resources Congress 2009, Kansas City, Missouri, United States, May 2009, pp. 1–11.

[16] C. Hu, G. Ren, C. Liu, M. Li, and W. Jie, “A Spark-based genetic algorithm for sensor placement in large scale drinking water distribution systems,” Cluster Computing, vol. 20, no. 2, Jun. 2017, pp. 1089–1099.

[17] B. M. Zaragozí, S. Trilles, and J. T. Navarro-Carrión, “Leveraging Container Technologies in a GIScience Project: A Perspective from Open Reproducible Research,” ISPRS International Journal of Geo-Information, vol. 9, no. 3, Art. no. 3, Mar. 2020.

[18] M. Beauchemin, “Apache Airflow”, https://github.com/apache/airflow, 2014.

[19] Argo Project Authors, “Argo Workflows”, https://argoproj.github.io/argo-workflows/, 2022.

[20] A. Solem, “Celery”, https://docs.celeryproject.org, 2021.

[21] M. Movsisyan, “Flower”, https://flower.readthedocs.io/en/latest/, 2012.

[22] Dask core developers, “Dask”, https://dask.org/, 2019.

[23] Center for High Throughput Computing, University of Wisconsin–Madison, “HTCondor”, https://htcondor.readthedocs.io/en/latest/, 2022.

[24] GitLab Inc, “GitLab”, https://about.gitlab.com/, 2022.

[25] Jenkins, ”Jenkins”, https://jenkins.io, 2018.

[26] Prefect Technologies, Inc., “Prefect”, https://docs.prefect.io/core/, 2022.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14078

WATER CONSUMPTION VARIATION IN LATIN AMERICA DUE TO COVID-19 PANDEMIC

Catalina Ortiz[1], William Clavijo[2], Jürgen Mahlknecht[3], Juan Saldarriaga[4]

[1]Researcher, Water Distribution and Sewerage Systems Research Center, Universidad de los Andes, Bogotá, (Colombia)
[2]Ph.D. Researcher, Centro del Agua para América Latina y el Caribe, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey, (México)
[3]Ph.D. Research Professor, Centro del Agua para América Latina y el Caribe, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey, (México)
[4]Professor, Civil and Environmental Engineering Department, Water Distribution and Sewerage Systems Research Center, Universidad de los Andes, Bogotá, (Colombia)

[1] *pc.ortiz@uniandes.edu.co,* [2] *hwc3@pitt.edu,* [3] *jurgen@tec.mx,* [4] *jsaldarr@uniandes.edu.co*

Abstract

To stop the spread of the COVID-19 pandemic, governments all over the world have applied social distancing measures, which have drastically altered people's lifestyles. Many studies suggest that the water sector, including its demand and supply, has been strongly affected by these regulations. The importance of hygiene practices confers a crucial role to potable water availability as an ally for tackling the spread of the virus, heightening the alteration of water demand patterns during the ongoing pandemic. Therefore, this research aimed to assess the impact of the pandemic on the water consumption patterns in four Latin-American cities and the differences among the type of users. The case studies include two Colombian and two Mexican cities known for their important industrial and touristic features. The outcomes reveal a diminishing effect on water consumption for industrial and commercial customers. Touristic cities were the most affected, even experiencing decreased domestic water demand. Understanding these changes and challenges is essential for keeping and improving the resilience of water systems in different scenarios, especially under fluctuating environmental conditions.

Keywords
Water demand, COVID-19 pandemic, touristic cities, industrial cities.

1 INTRODUCTION

The COVID-19 pandemic was declared by the World Health Organization – WHO on March 11, 2020. The rapid propagation of the COVID-19 disease, caused by the SARS-CoV-2 virus, forced the governments to apply measures to prevent the spread of the virus. Social distancing measures, along with face mask use, constant hand washing, and disinfection of frequently touched surfaces, are standard measures implemented by governments to tackle the spread of the disease. The goal of social distancing measures is to slow down the spread of the virus so that the medical system has enough capacity for treating the sick. Thus, stay-at-home orders have been applied in many countries so that only businesses classified as "essential" could operate in person. These measures abruptly altered the habits of people all over the world, forcing individuals to perform their daily activities from home and, consequently, driving changes in the water consumption patterns of cities, both in daily and in total volumes consumed locally.

Studies have shown that changes in water consumption vary from one city to another because of environmental, socioeconomic, and sociocultural characteristics. A study performed in Hamburg, Germany, revealed an increase of 14.3% in daily water consumption, with a delay of 1-2 hours in the morning peak and a higher peak in the evenings during weekdays [1]. A similar research for five towns in Puglia, Italy, showed a shift of 2-2.5 hours in the morning water use peak for two small towns (Cellemare and Lizzano), while a drop was observed in the peaks and base water demand in the two biggest municipalities (Bari and Molfetta) [2]. This finding relates to that exposed by Bich-Ngoc and Teller [3], who investigated the effect of the lockdown measures and the outbound tourism on water consumption in Liège, Belgium. Liège is a particular case since the historical data shows a lower consumption in the summer months due to people leaving the city. Through a statistical model, the authors found that water demand increased significantly when evaluating restricted trips scenarios [3] as the pandemic spawned. Consistently, the towns that receive a large number of commuters every day, such as Bari and Molfetta, experienced the contrary effect, meaning a decrease in the water consumption because of commuting limitations [2].

Furthermore, the studies mentioned above focused mainly on domestic water consumption; nevertheless, it is evident that the impacts differ according to the different types of water demand. For example, a study conducted in Henderson, Nevada, demonstrated an increase of 11.7-13.1% in residential, a drop of 34.1-35.7% and 55.8-66.2% in commercial and in schools water average daily demand, respectively [4]. Likewise, results obtained from a study in Joinville, Brazil, show a statistically significant decrease of 42%, 53%, and 30% for commercial, industrial, and public water demand, respectively, while a conceivable increase of 11% in residential water demand was recorded [5].

Moreover, the effect of other parameters plays an important role in analysing the sole effect of the pandemic on water consumption. That is why many investigators apply regression models for controlling variables such as seasonality, water price, customer income, weather variations, among others. For instance, researchers demonstrated for residential buildings in Dubai that fasting during Ramadan month modifies water usage patterns [6], which shows how sociocultural aspects are also essential factors to consider. Regarding COVID-19, the same authors observed how daily residential consumption increased (even more during Ramadan month) due to stay-at-home restrictions and augmented cleaning practices [6]. Relatable results from a study with data from several water utilities in California were obtained when applying a multivariate regression model for forecasting residential, commercial, industrial, and institutional (CII) and total water use. The model included variables for seasonality, mandatory and voluntary drought restriction measures, precipitation, temperature, evapotranspiration, water rates, population, and annual inflation [7]. Additionally, the authors demonstrated through an impact index that the influence of the pandemic on water consumption was mildly more substantial than the combined effect of all other factors [7]. This conclusion reflects the relevance of the COVID-19 pandemic in altering water demand patterns.

As expected, all these changes have imposed challenges for water utilities, including effects on financial aspects, like revenues and bills, and operational aspects, such as water quality. Regarding financial features, utilities were faced with suspensions on cut-offs, commercial income reduction, major misconduct in water bill payments, and reduced customer upgrowth [8]. Also, utilities have been impacted monetarily due to augmented domestic and reduced non-domestic water consumption. Concerning operational aspects, as rapid variations in water usage occurred, water infrastructure may have been affected because system operations are designed based on historical demands [9]. Moreover, regular flows maintain drinking water distribution systems absent of leached minerals, corrosion, and microorganisms that could affect tap water quality [9]. Consequently, health risks are present, and water utilities must manage and advertise to customers to avoid potential disease outbreaks. Furthermore, the social distancing measures also affected water utility operations as many workers started to work remotely, and they should operate the systems with reduced skilled staff, carrying off institutional knowledge [8]. Thus, the pandemic obligated numerous water utilities to work in non-designed conditions and with decreased workforce and financial capabilities [10].

This paper aimed to assess the effect of the pandemic on domestic, commercial, industrial, and total water demand in four cities of Latin America. There are several studies of the pandemic impacts on water demand in developed nations, but there are not sufficient records from the Global South. Hence, this study contributes to the progress in water-related research in developing countries. Following this introduction, the four case studies are presented, and the data used for the analysis is described. Then, the following section details the employed methods for estimating the impact of COVID-19 on water demand, which was performed with a regression model and with neural network approaches. Subsequently, the results and discussion of the water demand variation magnitude are presented, and an analysis of the factors influencing these changes and the differences between each city. Finally, conclusions are given on how the results are valuable for water utilities planning and management operations.

2 CASE STUDIES AND DATA

Water demand data for two Colombian and two Mexican cities was obtained from each water utility as the monthly billed water volumes from January 2016 until December 2020. Figure 1 shows the water demand proportion in each city as the average of pre-pandemic data, meaning records before March 2020. Additionally, the number of inhabitants in the metropolitan areas for 2020 is presented as a reference value of the size of the cities.

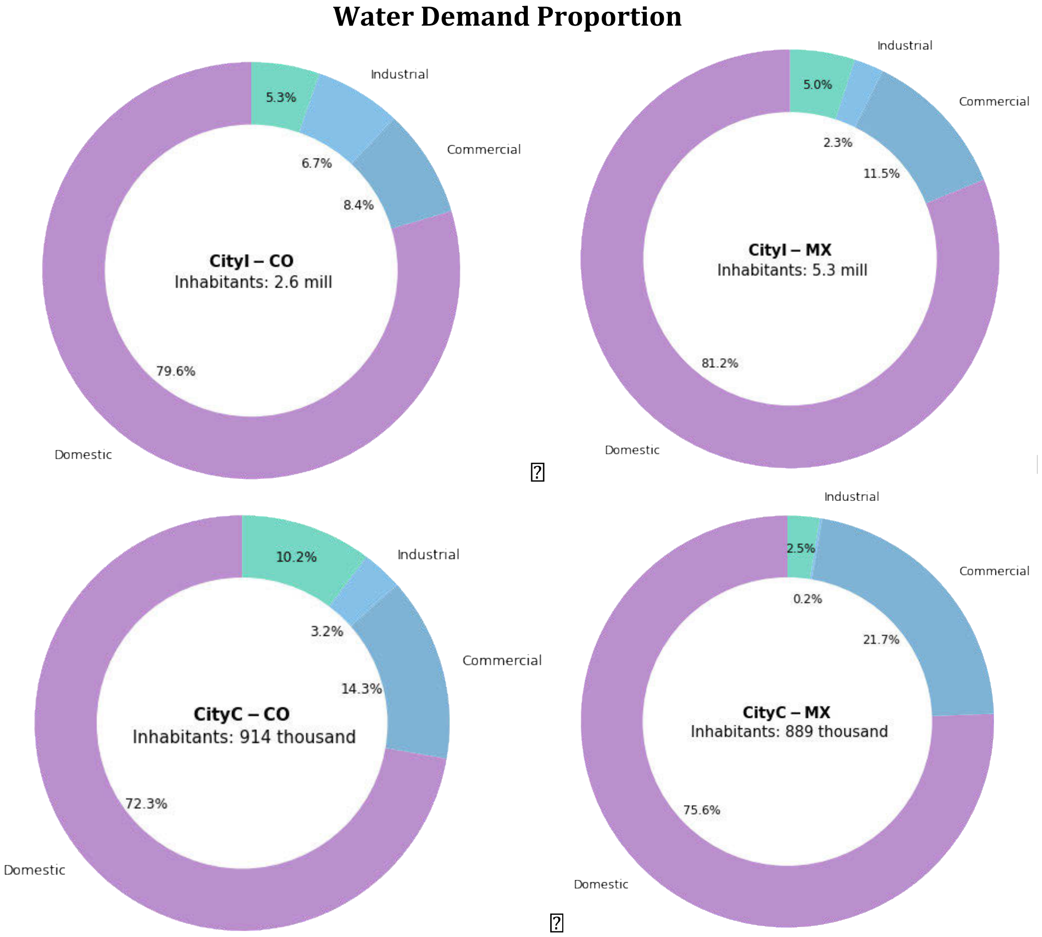


Figure 1. Water demand proportion by domestic, commercial, and industrial categories in each case study, as well as number of inhabitants of the corresponding metropolitan area.

The four cities were carefully chosen for their industrial and commercial characteristics. One in each country performs diverse economic activities (with substantial industrial water demand), and the other is an important touristic city (predominance of commercial demand). The case studies have been named accordingly: City I prefix corresponds to cities with diverse activities and City C to touristic cities. The endings -CO and -MX mean a Colombian or Mexican city, respectively.

As Figure 1 presents, the domestic water demand predominates in all cities, showing the importance of residential users for defining water usage patterns. In contrast, non-domestic water demand represents less than 30% of the total demand in the four case studies. Changes in domestic demands are expected to drastically influence water distribution systems' operations, while only abrupt variations in non-domestic demand would impact the urban water systems.

According to the predominance of commercial and industrial customers in each city, the pandemic might have affected water usage differently. This research focuses on understanding these dissimilarities among the case studies. For instance, an enormous impact on commercial demand can be anticipated in both touristic cities (City C–CO and C–MX) due to travel restrictions. Therefore, the non-domestic water demand changes are more critical in C-cities than in I-cities because of the high commercial proportion out of the total demand.

Moreover, in addition to water demand data, this study also employed temperature information for the analysis. Maximum temperatures for the entire service areas were calculated using monthly data from several meteorological stations distributed throughout the cities. Noteworthy, temperature information was not available for City C-MX, so the analysis was performed without it.

3 METHODS

We analysed the temporal variation of water consumption by comparing the observed water demand during the pandemic with a forecasted non-pandemic scenario. The historic non-pandemic water volumes were used to train and validate water demand models, which were used to forecast the expected demand as if the pandemic did not occur. Different modelling approaches were employed to validate and compare the results: Multivariate Regression Models (MVRM) and Artificial Neural Networks (ANN). Hence, data from January 2016 to February 2020 was used to fit the models to historical water usage patterns. Then, the non-pandemic water demand scenario for March to December 2020 was estimated with those models.

3.1 Multivariate Regression Model

The first approach implemented multivariate ordinary least squares linear regression models using water demand time series. Equation (1) presents the general form of the model. y is a vector representing the volumetric water use per month, X is the matrix containing the values for the n predictor variables in each month, β is the vector with the n related regression coefficients, and E is the vector for the error terms associated with the monthly water use estimations.

$$y = X\beta + E \tag{1}$$

R Software was used to implement the models through the *lm* function [11]. The unbiased estimators are calculated with Equation (2) and are used to produce unbiased least-square estimations of the water volumes $\hat{y}$, with errors $\hat{E}$ following a normal distribution with a mean equal to zero and minimizing the covariance [7].

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{2}$$

It is essential to break down the total water use by sector for obtaining better-modelled water demand results. Therefore, the models were fitted for each category of water demand: domestic, commercial, industrial, and total demand. Regarding the explanatory variables, at first several variables, including precipitation, commuting population, and water tariffs, were evaluated. However, not all the estimators' results showed statistical significance, so fewer variables were considered to avoid noise affecting the results. In fact, Bich-Ngoc and Teller [3] exposed that evaluating meteorological variables, like daily maximum temperature, has been frequently used to forecast short-term variations in the water demand. At the same time, socioeconomic factors, such as income, are more relevant when modelling demand in a long-term period. Hence, since the predictions were only extended to 10 months for this study, 14 explanatory variables were used: 12 binary variables representing seasonality by month and 2 continuous variables for maximum temperature and the number of users over time.

Once the models were fitted, R-squared results and p-values were analysed to assess the fitting results and the influence of the estimator variables on water consumption. Afterward, the matrix X was augmented to include values corresponding to the pandemic, namely, the values from March to December 2020. The augmented X matrix and the regression coefficients vector $\hat{\beta}$ obtained from the historical data were employed to calculate the water volumes in the non-pandemic scenario. These modelled water volumes were used to analyse the differences between the non-pandemic scenario and the experienced pandemic situation.

3.2 Artificial Neural Networks

As water utilities require consumption predictions for operational reasons and updating pricing policies, ANN are one of the principal approaches used to predict water demand [12]. ANN imitate the human brain by performing non-linear calculations in a certain number of neurons, which receive input values that are transformed and transferred to other neurons or the output [13]. The great advantage of ANN is learning based on initial observations and producing equivalent outputs with new input data. Neural networks can work with a single or multiple layers. Several learning algorithms allow the communication between neurons to determine the weights, which are values assigned to every variable from the input layer to the hidden layer [13]. The weights' modification allows the network to adapt to reduce the error between the expected output and the result from the network.

Like the regression model, we used artificial neural networks to model domestic, commercial, industrial, and total water demand. The implementation of ANN in this study was executed using R software by feeding the network and training it with a different number of neurons and layers until the best performance was reached. The *neuralnet* package was used for this aim, using the Resilient Backpropagation algorithm to train the network [14].

Moreover, the data series were normalized before training the networks to adjust the variables to the same scale. The R-squared values from the validation set were used for modifying the number of neurons and layers until the best model fitting was reached. Then, the trained and validated ANN was used to estimate the expected water consumption in a non-pandemic scenario.

Regarding the validation of the models, different approaches were used depending on the city analysed. The entire data sets were randomly divided into training (80%) and validation sets (20%) for the Colombian cities. In contrast, the models were trained with information from 2016 to 2018 and validated with 2019 records for Mexican cities. This approach is due to the seasonality relevance in each city. As can be noticed in the results, the water demand records are clearly influenced by the time of the year in the MX cities. Commercial demand in City C – CO also shows a relation with seasons; however, better water demand estimates were obtained when training the network with the aleatory approach.

BY NC SA

4 RESULTS AND DISCUSSION

4.1 Water Demand Estimation

We used the multivariate regression model results to study the influence of the predictors on the water demand through the computed models. For this aim, Table 1 presents the significance of the explanatory variables measured through the obtained p-values.

Table 1. Explicative variables significance for each computed model (D: Domestic, C: Commercial, I: Industrial, T: Total). Temperature information for City C – MX was not used. p-Value codes: 0 (***), <0.001 (**), <0.01 (*), <0.05 (.)

Explicative variables	**City I – CO**				**City I – MX**				**City C – CO**				**City C – MX**			
	D	C	I	T	D	C	I	T	D	C	I	T	D	C	I	T
Seasonality (12 variables)	***			***	*	*		**			**		*	**	**	**
Temperature					.	*		.			**		-	-	-	-
Number of users	***	***		***	***	***		***	***	***	***	***	***		***	***

The statistical significance of the chosen explicative variables differs for each city according to its socio-economic and environmental conditions. The number of users, which is directly related to water demand, is a relevant variable for most cases. However, industrial demand in I cities does not relate directly to any variable, not even the number of users. These observations showcase how industrial water use is usually unpredictable due to its high variability. In contrast, in the cities where the industrial demand is not predominant (C cities), all the variables are significant, but this is due to the small magnitude of the water volumes.

Moreover, the seasonality condition is prevailing for MX cities. This condition shows the effect of yearly temperature variations on water demand. Hence, the usage of temperature and months as seasonal variables can explain the interannual variations in water consumption. We found for City C-MX that monthly seasonality is enough to represent the tourism-related patterns adequately. In the case of the Colombian cities, the demand is not directly associated with the time of the year. This observation is due to the country's geographic position, affected by tropical weather.

Although not all variables were statistically significant, both approaches employed for water demand forecasting showed a good performance. Different validation sets were studied since the models were trained with different approximations due to the seasonality relevance (see Methods section). Nevertheless, Figure 2 presents the R-squared values for the estimated water volumes between January 2016 and February 2020 to compare both methods consistently. It can be noticed that ANN results for CO cities are much better than the MVRM outcomes. The same conclusion was obtained for industrial demand for MX cities. In contrast, the MVRM approach shows enhanced models' fit for Mexican cities' domestic, commercial, and total water demand.

Nevertheless, all R-squared values show outstanding performance, with values over 0.7 except for some industrial and commercial water use estimations. The worst fitting outcomes relate to industrial demand in City I – CO with MVRM, where the industrial economic sector is remarkably relevant. Similarly, the ANN commercial model fit in City C – MX is not particularly good, where commercial users are the most relevant among non-domestic customers. As previously discussed, forecasting commercial and industrial water demand is problematic due to the unpredictable changes that might occur due to governmental or institutional measures. Therefore, both approaches allow for choosing the best water demand estimations and better approximating what really would have happened if the pandemic had not occurred.

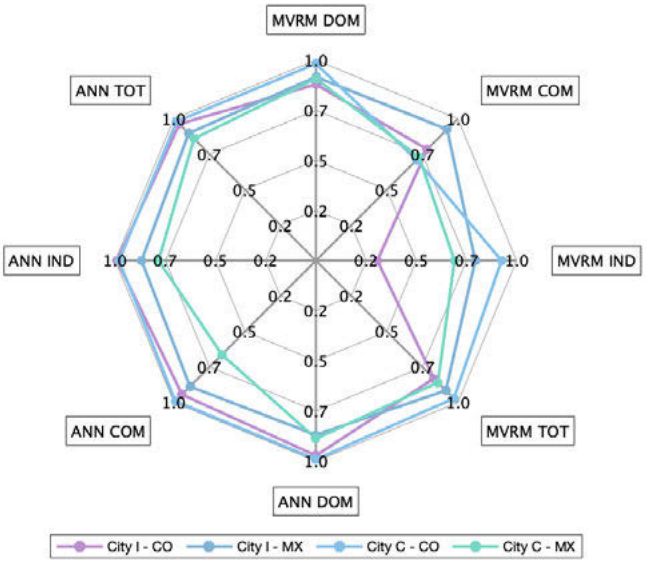


Figure 2. MVRM and ANN models' R-Squared results for domestic, commercial, industrial, and total water demand in the four case studies.

The performance of the models can also be graphically assessed through the time series comparison presented in Figure 3. This graph shows the reported water volumes from January 2016 to December 2020, and the predicted data using MVRM and ANN approaches. All models present a superb fit when comparing the pre-pandemic data. Industrial demand is the most problematic in all cities; however, it gives a good approximation of water demand with enough precision to analyse the pandemic impact.

Moreover, the pandemic effect is clearly evidenced by the data presented from March 2020 forward, highlighted by the red lines. The predicted water volumes show high similarity between the MVRM and ANN methods, giving reliability to the analysis and the estimated pandemic impacts. The only results that have an appreciable inconsistency between both approaches are for industrial demand in the commercial cities. The neural networks tend to underestimate the water volumes, even obtaining similar demands as the pandemic-affected values. In contrast, the MVRM exaggerates the increasing tendency. These anomalies might be related to the small magnitude of the demands as it only accounts for 3.2% and 0.2% of the total demand in City C – CO and – MX, respectively. Hence, it is not possible to state which estimations are better. However, it is preferable to assume an increasing trend since only unexpected events, like the pandemic, might generate such a decrease in industrial demand.

Further, according to both methods of water demand estimation, the domestic component in the CO cities shows a higher trend with time than in the MX cities. It could be related to the economic context of the case studies, despite the constant increase of users billed for each case. In addition, as domestic water use is predominant, with contributions to the total demand of over 70% in all cities, the same patterns are reflected in the total water use.

A similar increasing trend with time is exhibited for commercial use for the four study cases; however, the effect of lockdown breaks the tendency along the pandemic year 2020. In contrast, the industrial component of water demand shows a decreasing and stagnant tendency with time for three cities, excluding City C – MX. This finding could be part of urban development in Latin-American cities, which are implementing industrial hubs out of urban areas.

Finally, the seasonal variation of water demand for Colombian cities is not noticeable because of their geographical location, as previously discussed. Nevertheless, there is some effect of the annual seasonality on domestic and commercial water demand for City C-CO, related to tourism economic activities. On the other hand, City I – MX has a clear and marked water use pattern, showing peaks in the middle of the year. The same observation can be duplicated for City C – MX, but changes occur in a milder magnitude since its population is smaller.

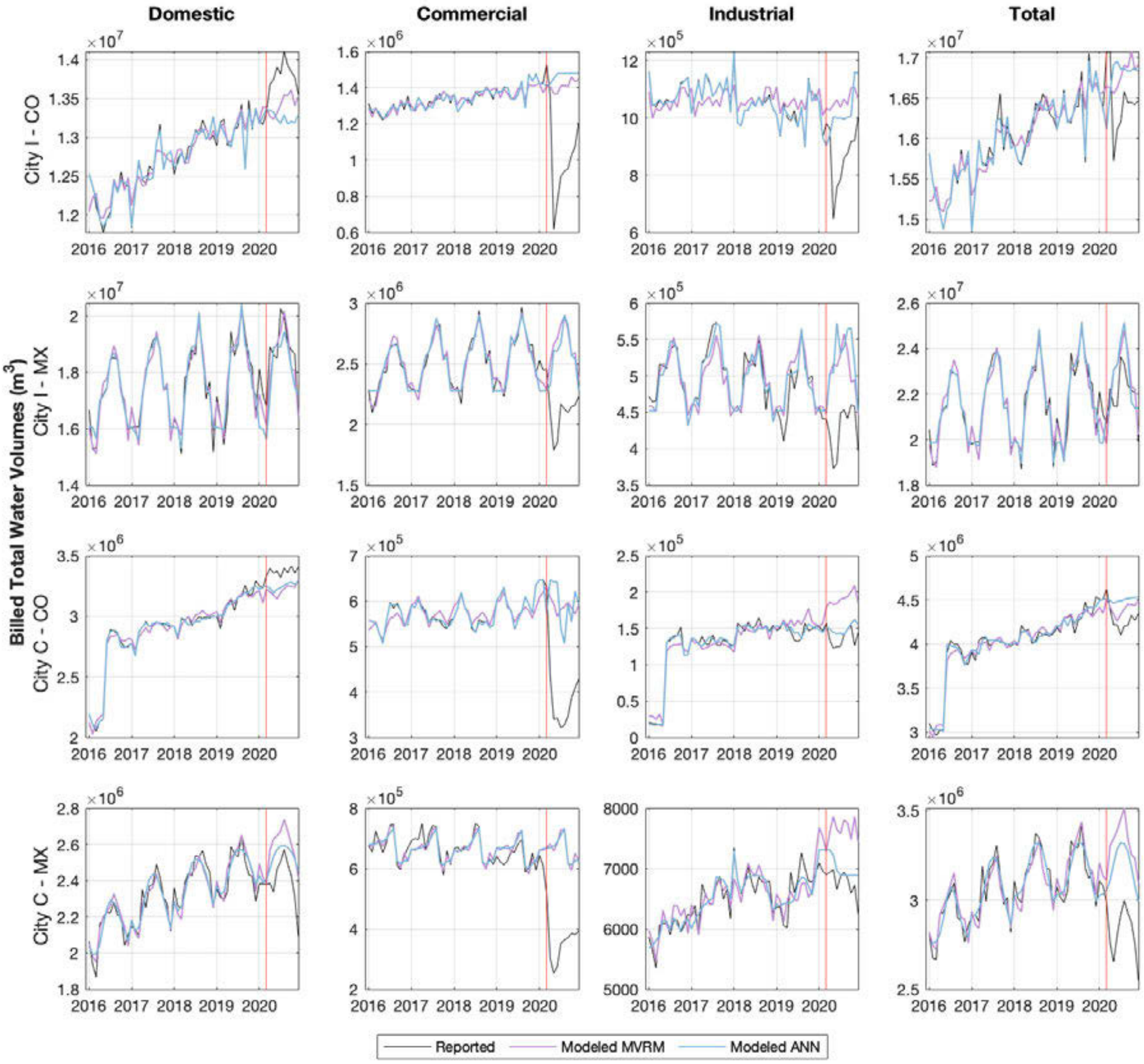


Figure 3. Reported and modelled domestic, commercial, industrial, and total water demand data in the four case studies. The red line marks March 2020, when the pandemic started, and the period for which data was forecasted.

4.2 Variation of water demand during the pandemic

The bibliographic review exposed how the pandemic influenced an upsurge in domestic water demand as people spent more time in their households due to lockdown measures. As Figure 4 presents, the same effect was obtained for industrial cities and City C – CO, with similar increases every month. Considering the models with the best fit results, the domestic component of water demand increased on average by 4% in CO cities and by 1% in City I – MX.

The extent of these increases does not seem influential. However, when contemplating the magnitude order of the residential water use, these surges represent a significant amount of water. Nonetheless, the evaluation of the total water demand changes reveals how the decreases in non-domestic water usage are more influential and balance the increased domestic demand. The total demand presents almost no variation or a moderate decrease. The average diminutions were 2% in City I-CO, 3% in City I-MX, 5% in City C-CO, and 15% in City C-MX.

Further, domestic water demand in City C–MX decreased 7% since the beginning of the pandemic. Lately, it has been more habitual for visitors to arrive at residential dwellings in touristic cities, such as Airbnb's, since it could be more economical than a regular hotel. Therefore, although residential users consume 75.6% of the total demand, this percentage might not relate entirely to the water use of City C – MX's inhabitants. According to 2019 data, the number of monthly tourists reaches the same number of inhabitants for the high seasons [15]. Hence, the decrease in the domestic water demand could be explained by the commuting limitations that the pandemic triggered. The same effect is not present in City C – CO's records because the number of monthly visitors does not exceed 3% of the number of inhabitants, even during the high seasons [16].

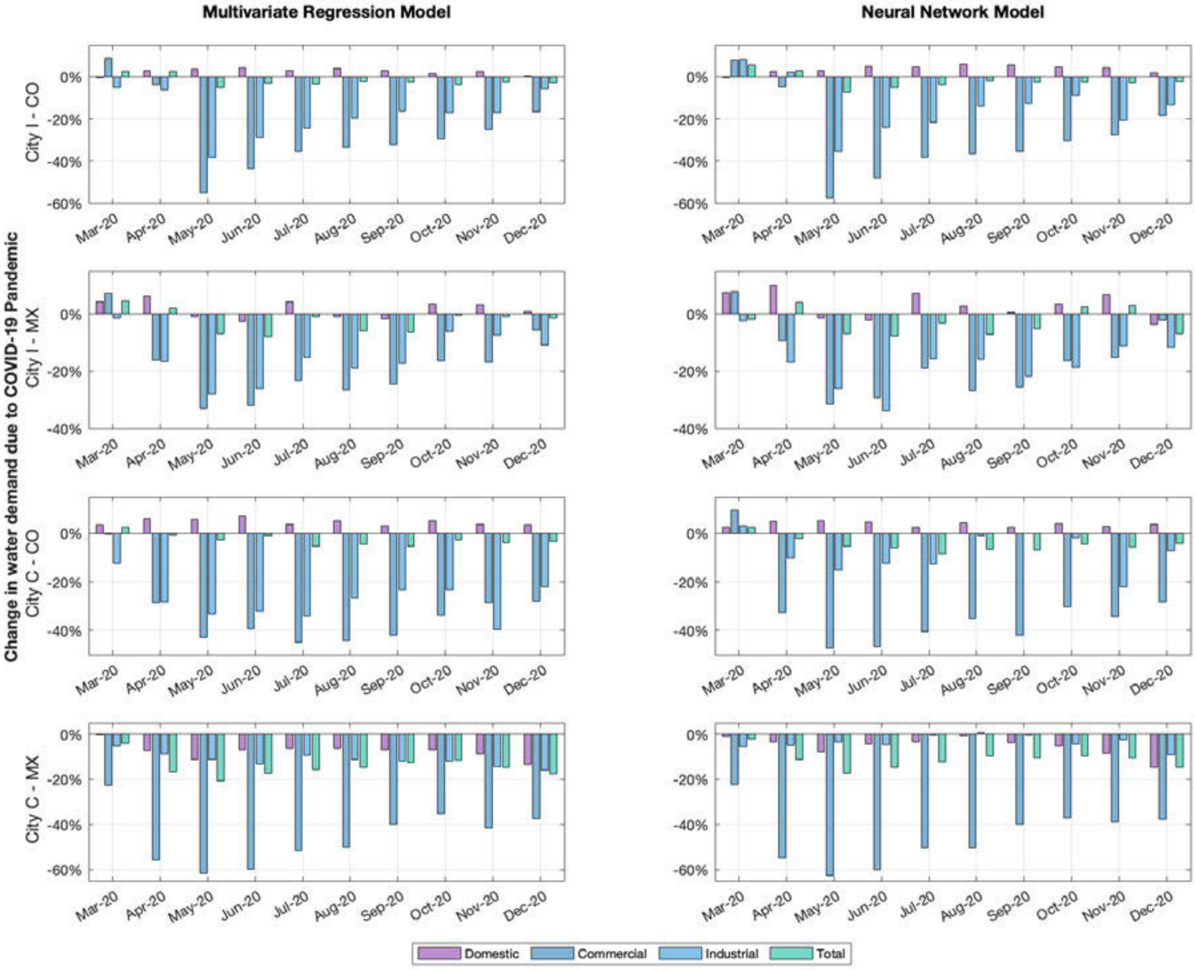


Figure 4. Change in domestic, commercial, industrial, and total water demand due to COVID-19 pandemic with MVRM and ANN estimates in the four case studies.

As it is shown by Irwin, McCoy & McDonough [4], Kalbusch et al. [5], Li et al. [7], among other studies, the non-domestic water demand was the most affected because of the pandemic. Indeed, the data from this research shows a noticeable fall in non-domestic water demand in the following four months after the global pandemic declaration. The variation of commercial water demand due to lockdown measures for industrial cities is about -29% in -CO and -19% in -MX compared to estimates based on the best-validated model. The drop is much more drastic in the commercial cities, with diminutions of 33% and 46% in City C – CO and – MX, respectively. The largest diminution in the C-MX case study compared to C-CO is essentially related to the number of commuters entering the city. More notably, as discussed in previous statements, the importance of tourism in both commercial cities significantly affects water demand since lockdown measures implied the impossibility to travel. Additionally, a considerable fraction of the commercial demand drop is related to ordinary commerce frequented by inhabitants. Hence, the closure of shops and related establishments could explain the decrease in cities I-CO and -MX.

Regarding industrial demand, the data reveals similar down-falling variation in the first three months of the pandemic. Nonetheless, the time series show some recovery in industrial consumption during the remaining months of the year. According to the model estimations, the average decrease in industrial demand was 14% in City I – CO, 18% in City I – MX, 8% in City C – CO, and 3% in City C – MX. The relevance of industrial demand is evidenced when comparing the values for C-cities and I-cities. The industrial decreases are not as abrupt as those obtained for commercial demand. Thus, it is evident how the pandemic affected touristic cities more drastically.

Moreover, the changes in non-domestic water consumption have been consistent with the increase in domestic demand for cities I-CO, I-MX, and C-CO. One might argue that commercial and industrial demand was transferred to households due to work-home activities. However, the cumulative values of non-domestic consumption are higher than those for residential customers. Hence, there is low evidence of volumetric compensation between domestic and non-domestic demand.

5 CONCLUSIONS

This research revealed how the main effect of the pandemic lockdown is the decreased non-domestic water consumption in the commercial and industrial cities analysed. This effect is linked to a slight increase in domestic water demand due to work-at-home activities, another main effect of lockdown restrictions. However, the case of one of the commercial case studies, City C-MX, shows a particular behaviour most probably due to a decline in domestic water demand due to tourism home-ownership renting dropping.

Regarding the employed methods, we can establish that both approaches for water demand estimation can address similar results, which is a relevant sign of a good level of reliability for the demand predictions. Based on the results, we can expect noticeable changes in the consumption patterns due to the effect of home working activities and their consequences on new working schemes in non-domestic sectors. We consider that these findings are related to the expected effect on the post-pandemic water demand behaviour. Hence, these demand variations should be considered in the near future investment and operations master plans for water utilities.

6 ACKNOWLEDGEMENTS

We would like to thank the water utilities of the four case studies for providing access to water use data. Especially to Andrés Fernández and collaborators from Aguas de Cartagena, to Servicio de Agua y Drenaje de Monterrey and Aguakan operators in Cancún. J.M. is thankful to the Chair of Circular Economy of Water FEMSA at Tecnologico de Monterrey for complementary funding.

7 REFERENCES

[1] D. U. Lüdtke, R. Luetkemeier, M. Schneemann, and S. Liehr, “Increase in Daily Household Water Demand during the First Wave of the Covid-19 Pandemic in Germany,” Water, vol. 13, no. 3: 260, Jan. 2021, doi: 10.3390/w13030260

[2] G. Balacco, V. Totaro, V. Iacobellis, A. Manni, M. Spagnoletta, and A. Piccinni, “Influence of COVID-19 Spread on Water Drinking Demand: The Case of Puglia Region (Southern Italy),” Sustainability, vol. 12, no. 15: 5919, Jul. 2020, doi: 10.3390/su12155919

[3] N. Bich-Ngoc, and J. Teller, “Potential Effects of the COVID-19 Pandemic through Changes in Outbound Tourism on Water Demand: The Case of Liège (Belgium),” Water, vol. 12, no. 10: 2820, Oct. 2020, doi: 10.3390/w12102820

[4] N. B. Irwin, S. J. McCoy, and I. K. McDonough, “Water in the time of corona(virus): The effect of stay-at-home orders on water demand in the desert,” Journal of Environmental Economics and Management, vol. 109, no. 102491, Sep. 2021, doi: 10.1016/j.jeem.2021.102491

[5] A. Kalbusch, E. Henning, M. Brikalski, F. Luca, and A. Konrath, “Impact of coronavirus (COVID-19) spread-prevention actions on urban water consumption,” Resources, Conservation and Recycling, vol. 163, no. 105098, Aug. 2020, doi: 10.1016/j.resconrec.2020.105098

[6] S. Rizvi, R. Rustum, M. Deepak, G. B. Wright, and S. Arthur, “Identifying and analyzing residential water demand profile; including the impact of COVID-19 and month of Ramadan, for selected developments in Dubai, United Arab Emirates,” Water Supply, vol. 21, no. 3, May. 2021, pp. 1144-1156, doi: 10.2166/ws.2020.319

[7] D. Li, R. A. Engel, X. Ma, E. Porse, J.D. Kaplan, S. A. Margulis, and D. P. Lettenmaier, “Stay-at-Home Orders during the COVID-19 Pandemic Reduced Urban Water Use,” Environmental Science and Technology Letters, vol. 8, no. 5, Feb. 2021, pp. 431-436, doi: 10.1021/acs.estlett.0c00979

[8] E. Zechman, N. Thelemaque, L. Spearing, K. M. Faust, J. Kaminsky, L. Sela, E. Goharian, A. Abokifa, J. Lee, J. Keck, M. Giacomoni, J. E. van Zyl, Y. Ethan, M. Cunha, A. Ostfeld, and L. Kadinski, “Water and Wastewater Systems and Utilities: Challenges and Opportunities during the COVID-19 Pandemic,” Water Resour. Plann. Manage., vol. 147, no. 5: 02521001, May. 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001373

[9] H. Cooley, P. Gleick, S. Abraham, and W. Cai, “Water and the COVID-19 Pandemic: Impacts on Municipal Water Demand,” https://pacinst.org/publication/coronavirus-impacts-on-municipal-water-demand/, Jul. 2020.

[10] L. A. Spearing, N. Thelemaque, J. A. Kaminsky, L. E. Katz, K. A. Kinney, M. J. Kirisits, L. Sela, and K. M. Faust, “Implications of Social Distancing Policies on Drinking Water Infrastructure: An Overview of the Challenges to and Responses of U.S. Utilities during the COVID-19 Pandemic,” ACS EST Water, vol. 1, no. 4, Dec. 2020, pp. 888–899, doi: 10.1021/acsestwater.0c00229

[11] RDocumentation, “lm: Fitting Linear Models,” https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/lm, Nov. 2021.

[12] E. Farah, A. Abdallah, and I. Shahrour, “Prediction of water consumption using Artificial Neural Networks modelling (ANN),” MATEC Web of Conferences, vol. 295, no. 01004, Oct. 2019, doi: 10.1051/matecconf/201929501004

[13] M. Awad, and M. Zaid-Alkelani, “Prediction of Water Demand Using Artificial Neural Networks Models and Statistical Model,” I.J. Intelligent Systems and Applications, vol. 9, Sep. 2019, pp. 40-55, doi: 10.5815/ijisa.2019.09.05

[14] RDocumentation, “neuralnet: Training of neural networks,” https://www.rdocumentation.org/packages/neuralnet/versions/1.44.2/topics/neuralnet, Nov. 2021.

[15] Mexican Government, “Datatur: Comprehensive Tourism Analysis,” https://www.datatur.sectur.gob.mx/SitePages/Inicio.aspx, Apr. 2021.

[16] Ministry of Foreign Affairs, “Migratory Flows Tableau,” https://www.migracioncolombia.gov.co/planeacion/estadisticas/content/223-tableau, Mar. 2021.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14082

ANALYSIS OF ONLINE PRESSURE FOR RESILIENCE PHASE CHARACTERISATION OF LEAKAGE/BURST EVENTS

Sotudeh Hoseini-Ghafari[1], Jorge Francés-Chust[2], Olivier Piller[3], and David Ayala-Cabrera[4]

[1]PhD student, CWRR-School of Civil Engineering, University College Dublin, Dublin (Ireland)
[2]Business Development, Aguas de Bixquert, S.L., Xàtiva, Valencia (Spain)
[3]Senior Research Scientist, INRAE, UR ETTIS, Cestas (France)
[4]Ad Astra Fellow - Assistant Professor, CWRR-School of Civil Engineering, University College Dublin, Dublin (Ireland)

[1] *Sotudeh.hoseinighafari@ucdconnect.ie*, [2] *Jorge@abxat.com*, [3] *Olivier.piller@inrae.fr*,
[4] *David.ayala-cabrera@ucd.ie*

Abstract

While operating a water distribution network (WDN), it is essential to prepare the system to face with intentional (*e.g.*, cyber-physical attack) or unintentional (*e.g.*, pipe leakage/burst) adverse events or other drivers such as the effects of climate change. Increasing the network's preparedness to deal with anomalous events is an effective manner to improve the system's resilience, reducing the negative impacts of events. In this paper, leakage/burst events, and ordinary network operation, are captured by both sensors and expert knowledge in a WDN in Spain. Event-driven and data-driven approaches are used to characterise the system behaviour, in particular when it is operating under the effects of an anomalous event, based on the resilience phases (*i.e.*, absorptive, adaptive, restorative) for the collected dataset. The relationship of clustering pressure head time series based on their potential state in a particular resilience phase, in three random cases of short-term leakage events, was explored. This paper focuses on capturing the behaviour of the system, through the exploration of the hydraulic parameters of WDNs (in particular the pressure head) before, during, and after a leakage event, by means of a spatial-temporal analysis. It was observed that the network behaviour could be categorised into 1) ordinary operation and 2) during the event, which would allow to characterise the system behaviour when influenced by leakage/burst event and also explore its adaptability to resilience phases. The results show that it is possible to extract relevant patterns (*i.e.*, feature maps) and generate an anomaly indicator from the pressure head heatmaps that facilitate the characterisation of anomalous events for WDNs.

Keywords
Three phases of resilience, Spatial-temporal analysis, Pressure sensors, Water distribution network, Preparedness, Protection of critical infrastructures, Intelligent data analysis, Leakage/burst events.

1 INTRODUCTION

Leakage and bursts are commonly known as minor (less severe) but more frequent than events such as flooding and cyber-physical attacks, widely known to be significant (more powerful) but less frequent events in water distribution networks (WDNs). Apart from these, the former category can intensify the latter's effects and make them more intense. One example is the possibility of a leakage/burst to increase the chance of causing a flood [1]. As a potential consequence of the effects of climate change, water shortage can become more severe if it leaks through the system's pipes. In addition, leakage points are potential sources of pollutant intrusion into the network and can increase the severity of flooding impacts. The balance between input and output water gets is further disturbed by wasted water, known as non-revenue water, from leakage points. Therefore, it is worth making efforts to prevent minor events to avoid more serious outcomes.

As mentioned, WDNs need to be prepared for such events not only to prevent their direct negative impacts on system objectives, but also to prevent significant events from having a wider impact. It should not be overlooked that leakage can cause flooding if they are not dealt with in time [2]. One manner to minimise the potential negative impacts due to the occurrence of these events is to improve the network's resilience through possible interventions (*e.g.*,) that can help restore system performance to near pre-event levels [3].

A WDN can performs at different levels (*e.g.*, ordinary, degraded, or failed operation) when it is under the influence of disruptive conditions, which can lead to a reduction in the network efficiency [3, 4]. In particular, whether the system as a whole or partially is working at degraded or failure performance levels, it can be increased by improving the resilience of the system. System resilience is commonly referred as the ability of the system to withstand disturbances, the ability to adapt more easily to changing conditions, and the ability to recover the system performance at the level prior to the occurrence of the anomalous event [5, 6, 7, 8]. Improving network preparedness for abnormal events is of great interest to water utilities as an effective manner to improve system resilience [9].

In this paper, we propose a strategy to characterise the behaviour of the network (in terms of pressure values) during both ordinary operation and event occurrence taking into account the resilience phases (*i.e.*, absorptive, adaptive, restorative), which will help improve the preparedness of networks.

2 PROPOSED FRAMEWORK

When the anomalous event (in this study leak/burst event) occurs and depending on the relevance of the affected component in the system, the network’s performance may be lead in a decrease. In this sense, the start of the event may not be detected until the changes in the control parameters are representative or until, in the case of leak/burst events, it becomes visible on the surface. Given the uncertainty in the start time of this type of event, the start time of the event is assumed as the time at which the anomaly is detected. The end time of the event is referred to the time in which an acceptable or ordinary/regular level of performance is achieved for the system after the event occurs. The start and end times deploy a box whose area can represents the total resilience among the event. The capacities of system during an event can be divided to absorptive, adaptive and restorative capacities [10, 11]. These capacities are contained in boxed demarcated by no actions (absorptive capacity) or actions (adaptive and restorative) taken to remedy/mitigate the effects of the event (if any). However, some studies refer the adaptive and restorative capacity to an unique capacity (restorative). Due to both the nature of the system's behaviour and the actions to be implemented in each phase, in this work we refer to the three capacities of the system as each of the phases of the event. Resilience function captures the effect of the event in absorptive phase, adapts the system to the new temporarily disrupted conditions in the adaptative phase, and restores the system’s performance in restorative phase if the adaptive capacity is not efficient [10]. Studies showed that one method to increase the resilience of critical infrastructures is to improve the resilience for each phase [12]. Figure 1 is a typical representation of changes in network performance in regular operation and during an event, considering resilience phases.

A potential action deployed in the system can significantly enhance network performance (see Figure 1, green area). For the case of leakage, isolation of the affected area is a temporary (palliative) action to prevent further pressure loss and volume of wasted water. Afterward, repairing the leakage or replacing the damaged component (e.g., pipe) are potential restorative actions. One means of improving preparedness is to capture the behaviour of the network during both ordinary operation and event occurrence. Identifying the network's behaviour to events could facilitate the decision-making process by supporting the implementation of preventive/responsive solutions to keep the network's resilience close to acceptable levels when such events occur.

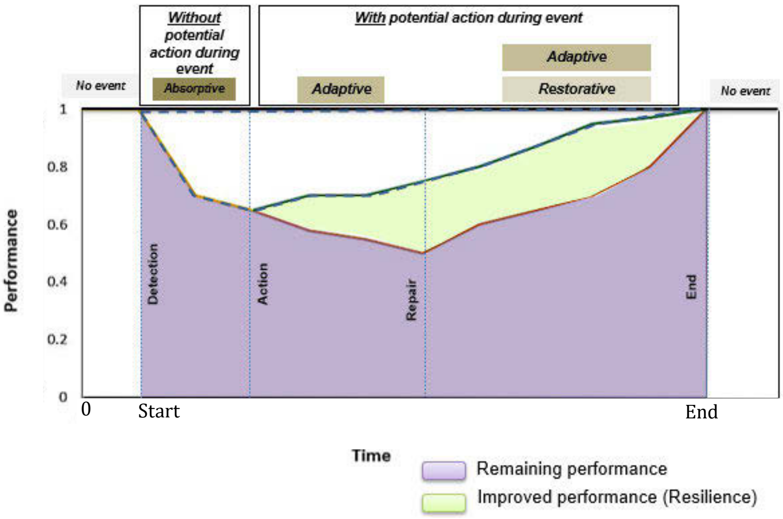


Figure 1. Resilience curve for a disruptive event in WDNs.

The purpose of this paper is to capture the behavioural patterns of the network under normal condition, and during leakage event (pre and post actions) at the pressure heads (which will be called pressure in this paper) through spatial-temporal analysis. It is proposed that extraction of these patterns will allow the network to adapt to each resilience phase. Three random cases of short-term leakage events were selected to implement how effective is the recognition of patterns of pressure values in the whole network.

3 CASE STUDY

In this section, a real WDN, working under abnormal/degraded operating conditions has been selected to apply the proposed methodology. The reason for choosing this network was the availability of data from both sensors and company records by operators, including information on the leak detection and repair process. This medium-size utility network (Figure 2) is located in Spain. The model consists of 146 demand nodes, each node responsible for delivering water to a large number of consumers (mainly houses), 212 links (40 km), two pumping stations, two reservoirs, and four tanks. There were 23 pressure sensors (recording pressure values every 15 minutes) in the network with different working conditions at different times (Table 1).

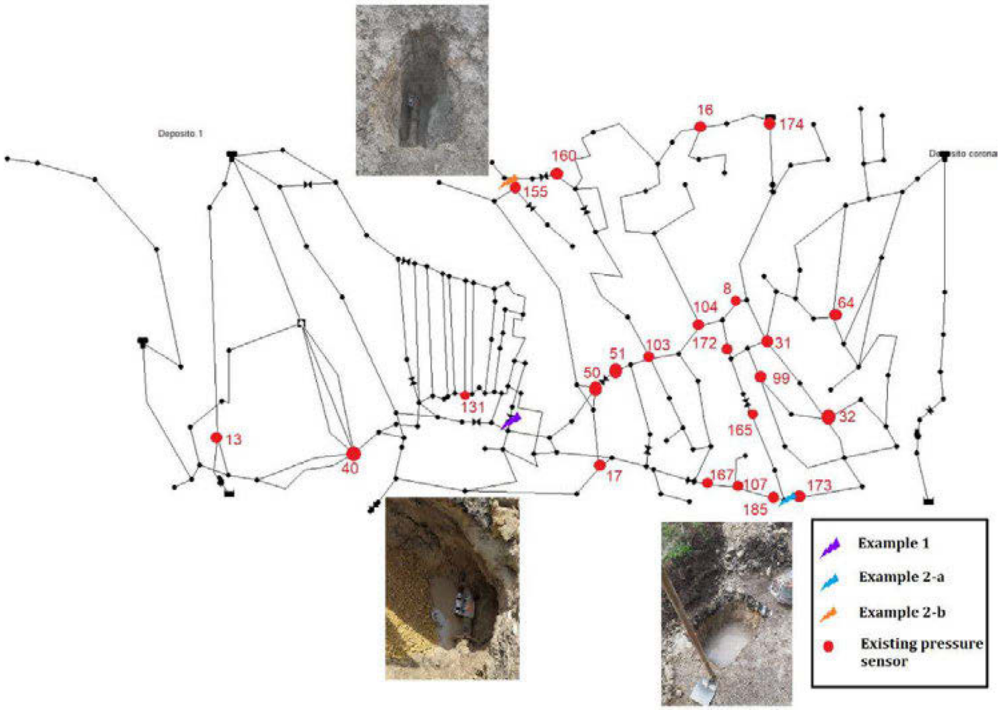

Deposito 1
Deposito corona
16
174
160
155
104
8
64
31
172
103
51
50
99
32
131
165
13
40
17
187
107
173
185
Example 1
Example 2-a
Example 2-b
Existing pressure sensor

Figure 2. Case study network with the location of pressure sensors.

Table 1. Status of pressure sensors. No operative -, and Operative ✓.

Number	ID	Working status		Number	ID	Working status		Number	ID	Working status	
		Example 1	Example 2			Example 1	Example 2			Example 1	Example 2
1	103	-	-	9	165	-	-	17	173	-	-
2	104	✓	✓	10	17	✓	✓	18	167	-	-
3	185	✓	✓	11	155	✓	✓	19	32	-	-
4	16	✓	✓	12	131	✓	✓	20	99	-	-
5	50	✓	✓	13	160	-	✓	21	64	-	-
6	51	✓	✓	14	174	-	-	22	13	-	-
7	172	✓	✓	15	107	✓	✓	23	31	-	-
8	8	✓	✓	16	40	-	✓				

4 DATASET

The dataset includes: 1) historical data from pressure sensors (dataset 1) and 2) utility expert knowledge (dataset 2), including information about the type of leakage, causes, detecting and repairing each leakage, among others. Analysis of leakage was initially conducted through a data-driven method from sensors' time-series data. Then leakages were temporally labelled by an event-driven approach from the records by utility expert knowledge. The removal of outliers in the data was conducted by means of manual inspection and in collaboration with the system operator.

MATLAB's interp2 function was used to construct the pressure matrix, P (of size $m \times n$), for each evaluated time. Where m with $i = 1, \dots, m$ and n with $j = 1, \dots, n$ represent the resolution for the x-axis and y-axis of the spatial coordinates, respectively. The matrix P was built from the nodes with sensors that were operational at the specific time evaluated. A homogeneous mesh was created for both the x and y coordinates. Each of these meshes contained all the nodes of the network. These meshes were used to infer the spatial relationship between the different available sensors presented in P. As a result of multiple iterations, the selected mesh size corresponded to a 100×100 resolution mesh. It should be mentioned that the matrix P can be easily visualised through the use of heatmaps (and this is done as a visual example in some sections below). However, all the calculations are conducted directly in the P matrix for each evaluated time.

A preliminary temporal analysis of the pressures, for the sensors, available in dataset 1 was conducted in order to contrast the information recorded in dataset 2. In a case where the leakage was not observable through Dataset 1, the second dataset source was used to track the location and the information that was not possible to get from sensors, such as time of detection, visit, isolation, repair, and corresponding data. Taking advantage of two data resources, we recognised resilience phases for each of the examples explained in the next section.

5 PRELIMINARY RESULTS

The first example corresponds to a leakage event recorded on 22 August 2021. The location of the leakage point is presented in Figure 2 (violet lightning flash in the bottom middle). Figure 3 shows changes in pressure values recorded by sensors from 20 August to 24 August 2021. Figure 3 shows the effect of the leak event on the pressure values, which was recorded in most of the sensors that were working at that time. This information was confirmed by the utility operator who conducted interviews. The operators' records show that people observation reported the leakage area one day before intervention to solve the problem. As the leak was not repaired on the same day as the detection, further pressure drops occurred and affected almost the entire network at around 7:30-10 am on 23 August (Figure 3). This result shows the crucial role of resilience phases in avoiding severe impacts in time.

The second example includes two overlapping leakage events, which were reported on 24 and 25 October 2021 (locations are shown in Figure 2, blue and orange lightning flashes in the bottom right, and top middle, respectively). Example 2-a was a leakage reported 24 hours before the visit, and the leakage of Example 2-b was reported 6 hours before visiting. Oscillating and high pressure caused leakages in two parts of the network; both were seen, isolated, and repaired on 25 October. This example is a more complex case, compared to Example 1, as the behaviour of the network could be affected by overlapping events, which in some cases might need a more advanced type of analysis. The interesting point about this example is that it is almost impossible to track the first leakage through sensors' records. The pressure curves extracted from most sensors before, during, and after leakage, shown in Figure 4, confirm this.

The pressure drops of sensors 155, 160, and 16 for Example 2-b indicate that the effect of the leak in the surrounding area occurs over a short period. This pressure drop could not have been more severe as the isolation action was conducted quickly and prevented the rest of the network from being affected. But in some cases, isolation might reduce pressure in some areas based on the network's characteristics. It is necessary to mention that the effect of demand on pressure values has been ignored. In both cases observed, the pressure variations due to leakage were much larger than the demand fluctuations.

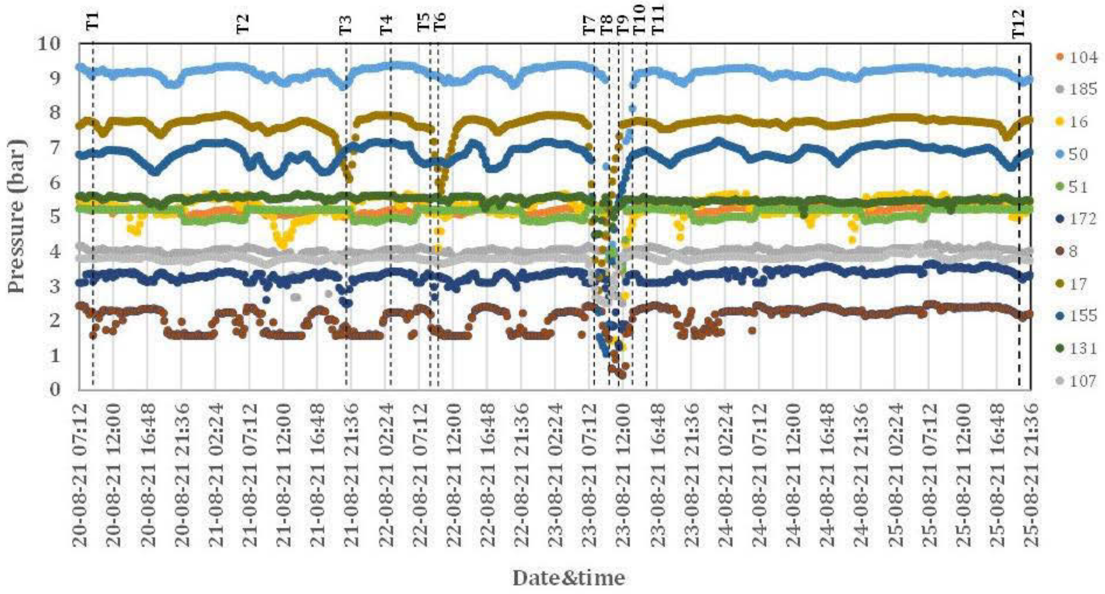


Figure 3. Pressure history from the working sensors for Example 1

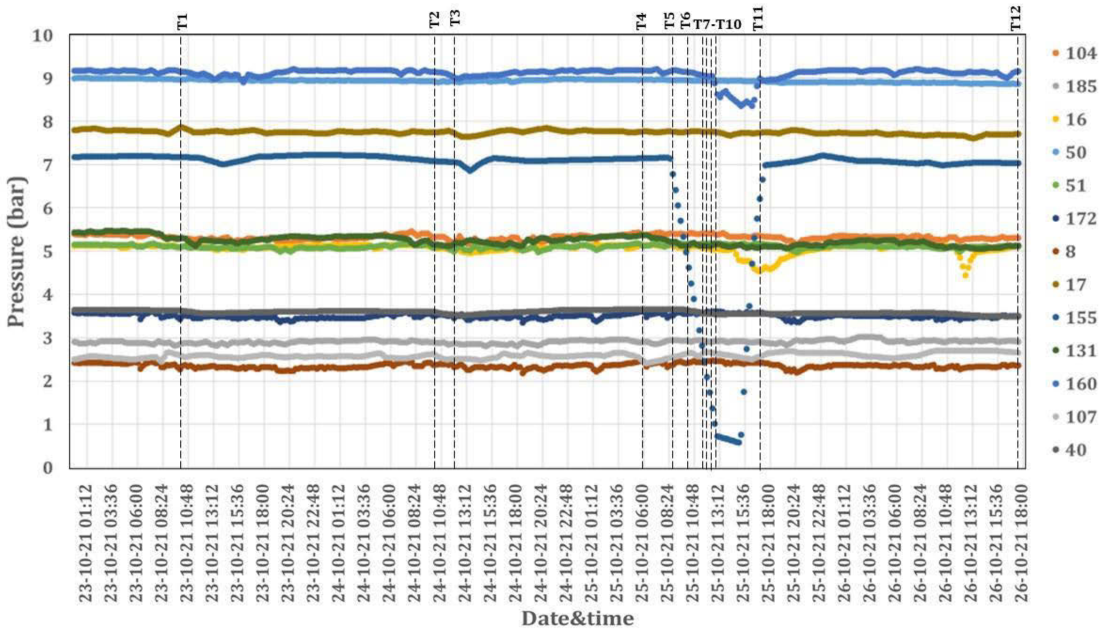


Figure 4. Pressure history from the working sensors for Example 2.

The spatial-temporal distribution of pressure for a selection of 12 timesteps from 20 to 25 August in Example 1 (Figure 5) and from 22 to 26 October in Example 2 (Figure 6) reflects the changes in pressures over time at different points of the network. It should be noted that 12 timesteps were selected to be representative of the leakage behaviour of the network in terms of pressure. In this sense, all the data of every 15-minute record was involved in the calculations. Timesteps were chosen according to historical data records and modified with interviews with the operator. Details are provided in Table 2 and Table 3.

Table 2. Timesteps for Example 1

ID	Date/Time	Description	Resilience phase
T1	20/08/2021 09:00	Time before the start of leakage	Pre-event
T2	21/08/2021 07:00	Random time	Pre-event
T3	21/08/2021 21:15	Pressure drop time	Pre-event
T4	22/08/2021 03:00	Pressure drop time	Pre-event
T5	22/08/2021 09:00	Detection time	Start
T6	22/08/2021 10:15	Random time	Absorptive
T7	23/08/2021 08:00	Visit time	Absorptive
T8	23/08/2021 10:00	Isolation time	Adaptive
T9	23/08/2021 11:00	Repair time	Restorative
T10	23/08/2021 13:00	End of leakage	End
T11	23/08/2021 15:00	A few hours later	Post-event
T12	25/08/2021 20:30	Hours after event	Post-event

Table 3. Timesteps for Example 2

ID	Date/Time	Example 2-a		Example 2-b	
		Description	Resilience phase	Description	Resilience phase
T1	23/10/2021 10:00	Time before the start of leakage	Pre-event	Time before the start of leakage	Pre-event
T2	24/10/2021 10:00	Detection time	Start	Random time	Pre-event
T3	24/10/2021 12:00	Random time	Absorptive	Random time	Pre-event
T4	25/10/2021 06:00	Random time	Absorptive	Detection time	Start
T5	25/10/2021 08:30	Random time	Absorptive	Start of pressure drop time	Absorptive
T6	25/10/2021 10:00	Visit time	Absorptive	Random time	Absorptive
T7	25/10/2021 11:30	Random time	Absorptive	Visit time	Absorptive
T8	25/10/2021 12:00	Isolation time	Adaptive	Isolation time	Adaptive
T9	25/10/2021 12:30	Repair time	Restorative	Random time	Adaptive
T10	25/10/2021 13:00	End of leakage	End	Repair time	Restorative
T11	25/10/2021 17:30	A few hours later	Post-event	End of leakage	End
T12	26/10/2021 17:30	Hours after event	Post-event	Hours after event	Post-event

The changes in the behaviour of the network as a function of the values of pressure (shown in Figure 5 for the first leak example and Figure 6 for the second leak example) can be obtained in the specific parts of the network in the effective times. The pressure maps show that changes in pressure are around the affected area and in the areas that have strong dependencies to this point.

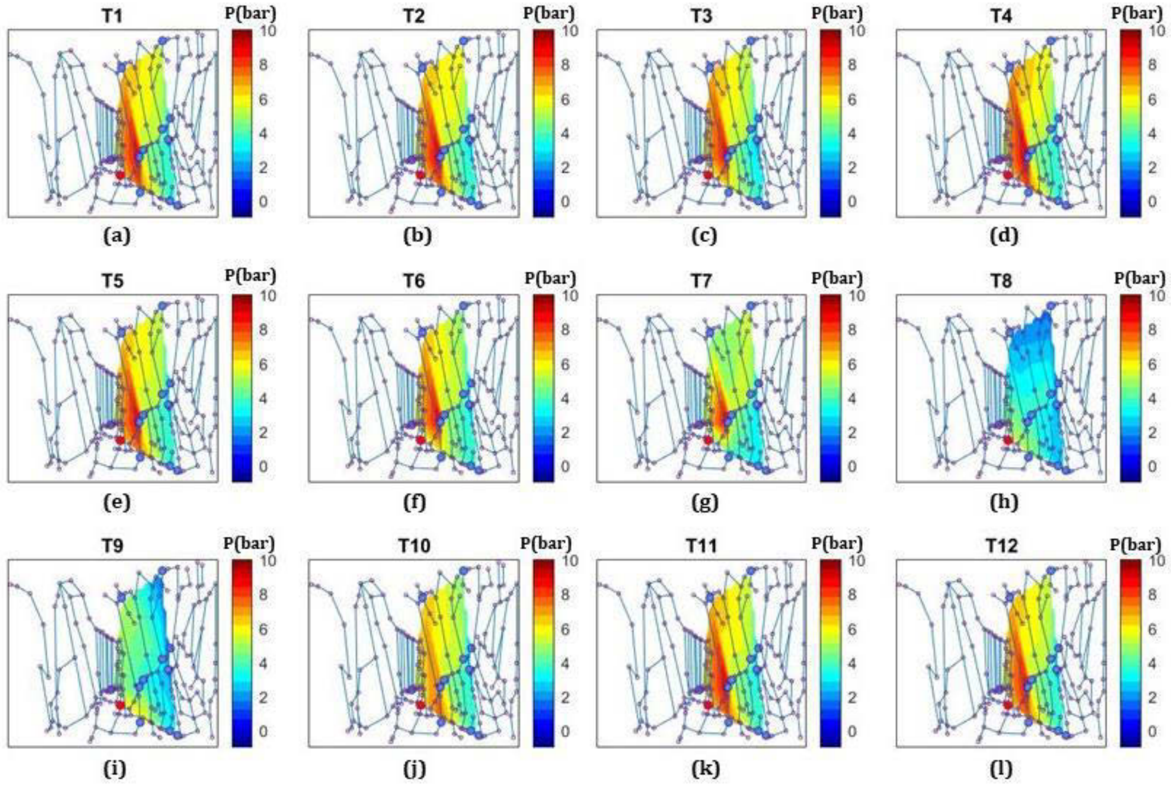


Figure 5. Spatial-temporal distribution of pressure for Example 1. (a)-(d); before, (e)-(j) during, and (k)-(l) after the leakage event.

BY NC SA

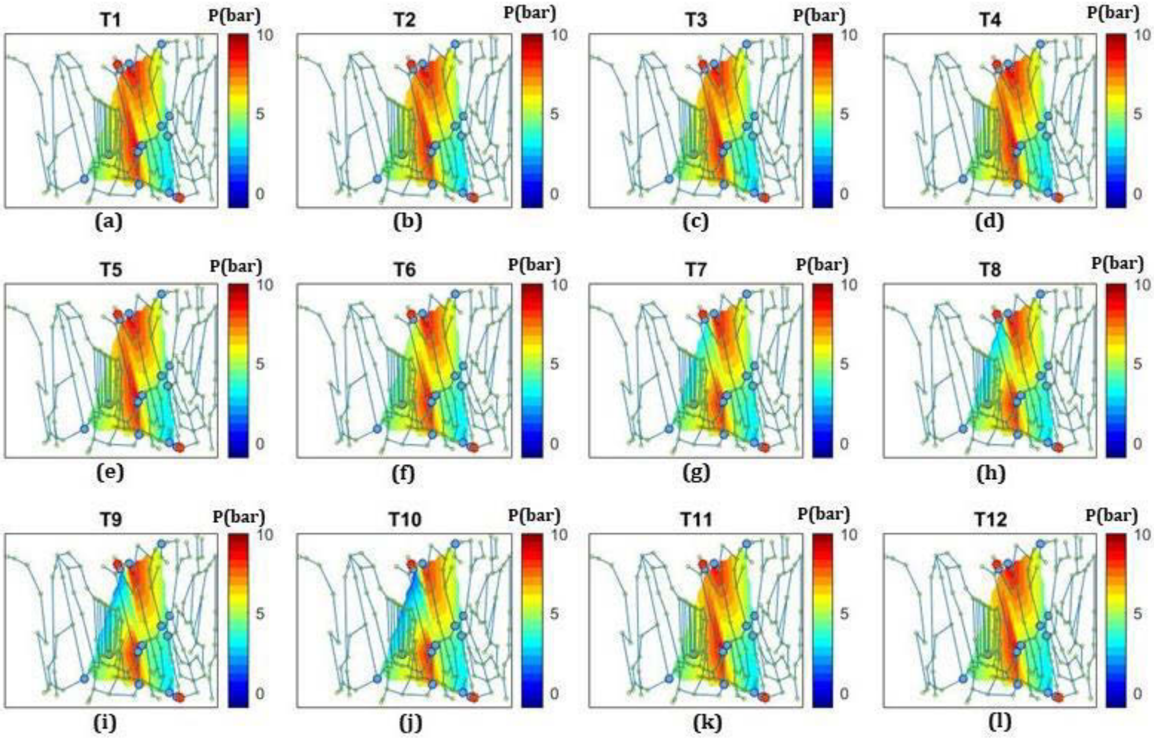


Figure 6. Spatial-temporal distribution of pressure for Example 2-a. (a); before, (b)-(j) during, and (k)-(l) after the leakage, and Example 2-b. (a)-(c); before, (d)-(k) during, and (l) after the leakage

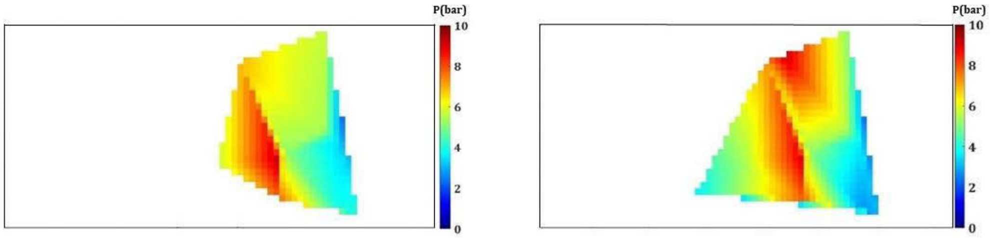
6 ANOMALY INDICATOR

According to the observations, it is possible to construct an anomaly indicator based on pressure to characterise the leakage/burst events. To better understand the pressure response of the network leakage, a matrix of maximum pressure was created for the given period according to equation (1) as a basis for the anomaly indicator. The highest pressure values are assumed to be desired. The maximum pressure map is shown in Figure 7. This map does not correspond to a single timestep, but represents the maximum pressure achieved for each element of P during the study period. Comparing Figure 7a and Figure 7b, we observe that in Example 1, maximum pressure distribution is less abrupt (smoother) than in Example 2. These different trends underline the importance of considering the maximum pressure values as a basis for the analysis of pattern extraction.

$$P_{max_{i,j}} = \max\left(p_{{t_1}_{i,j}}, p_{{t_2}_{i,j}}, \dots, p_{{t_{12}}_{i,j}}\right) \qquad (1)$$

where, $p_{t_{i,j}}$ is the pressure value in the i_{th} row and j_{th} column in the matrix at time t. It should be mentioned that the process includes all the timesteps that were set every 15 minutes.

BY NC SA



(a) (b)

Figure 7. Maximum pressure during the given period for (a) Example 1 and (b) Example 2.

The maps in Figure 7 facilitate the analysis of the behaviour of the network via an anomaly indicator (equation 2) in terms of pressure. The relative pressure resulting from dividing the pressure by the maximum pressure (individually for each element of P) over the whole period was calculated for all the points in the network. Figure 8 and Figure 9 show the spatial-temporal representation of the relative pressure maps for the selected timesteps. These values range between 0 and 1.

$$R_{t_{i,j}} = \frac{P_{t_{i,j}}}{P_{max_{i,j}}} \tag{2}$$

where, $R_{t_{i,j}}$ refers to anomaly indicator in the i-th row and j-th column in the matrix at time t.

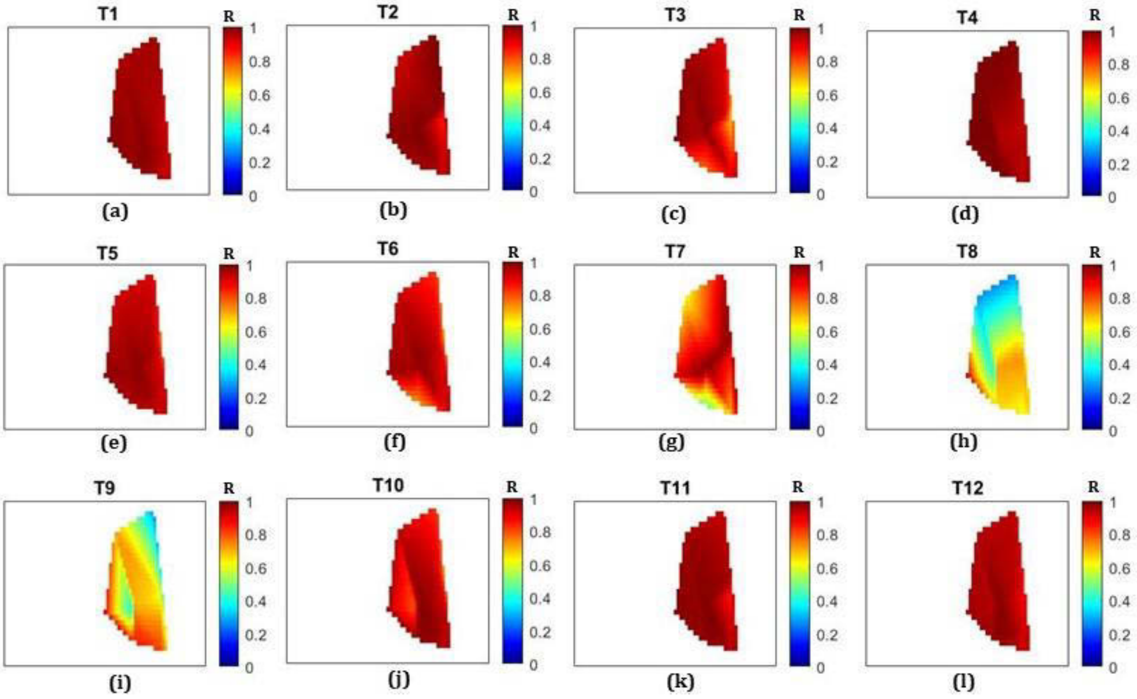


Figure 8. Spatial-temporal distribution of relative pressure for the Example 1. (a)-(d); before, (e)-(j) during, and (k)-(l) after the leakage.

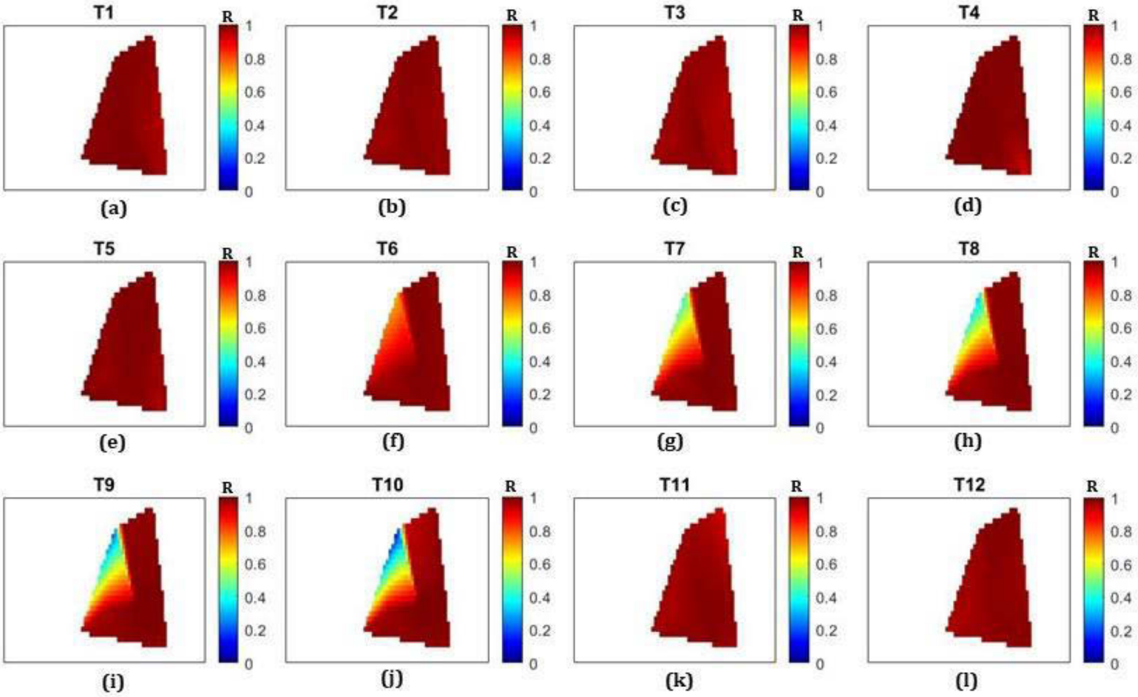


Figure 9. Spatial-temporal distribution of relative pressure for the Example 2.

$$R_t = \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} R_{t\,i,j}}{m \times n} \tag{3}$$

where R_t is the anomaly indicator, representative of pressure behaviour of the entire network (study area) at timestep t.

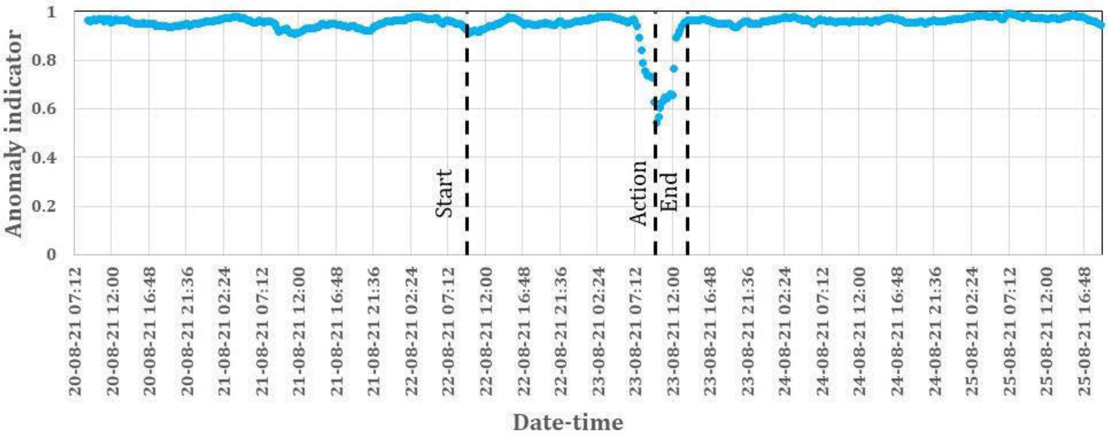


Figure 10. Anomaly indicator curve for the Example 1.

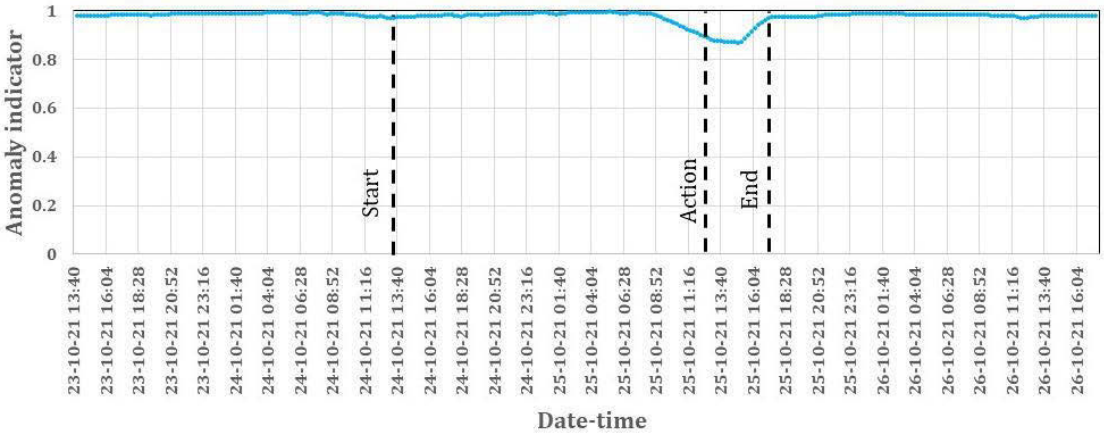


Figure 11. Anomaly indicator curve for Example 2.

Analysis of Example 1. As we can see in Figure 8, the leak becomes detectable at T7 (drops in pressure values), then changed significantly (huge reduction in pressures) in the adaptive and restorative phases, and gradually returns to normal operation (leakage was fully repaired). The classification of this behaviour also represents the phases of resilience. In other words, the behaviour of the network is classified as follow: no leakage, during the leakage with no action (absorptive phase), and with action (adaptive/restorative phases).

Analysis of Example 2. Observing the maps in Figure 8, the first overlapping leakage (Example 2-a) is not considerable, probably due to the more significant impacts of the second overlapping leakage (Example 2-b) in the pressure of the whole network. For this reason, the latter will have the central role in this example. Focusing on Example2-b, the network behaviour started to change significantly (drops in pressure values) at T6 and T7, and pressures gradually drop in the adaptive and restorative phases (T8 to T10), and return to the normal operating condition (T11 and T12). As with Example 1, for Example 2 the behaviour of the network is classified as follow: no leakage, during the leakage with no action (absorptive phase), and with action (adaptive/restorative phases).

The anomality indicator curves for both examples are shown in Figure 10 and Figure 11. These values result from the average relative pressure values of the whole network for each time. This curve is at its highest level during the regular operation (close to or equal to 1) and at its lowest level, when the pressure drop is at its lowest state. These curves can be representative of resilience curves. So, these curves allow the water utility to characterise the network in terms of its resilience phase and to better prepare for leakage by focusing on the points where there are gaps. For instance, in Example 1, the impact of leakage (both before and after action) can be seen in more areas of the network than in Example 2. This analysis shows which parts of the network are most sensitive to leakage according to high-pressure zones and the relevant changes in pressure in some specific regions. Comparing the adaptive/restorative phases for the two examples (shown in Figure 10 and Figure 11), it can be concluded that being prepared to respond to the event as quickly as possible would avoid huge pressure drops.

7 CONCLUSIONS

In this paper, we explored a dataset including two examples of leakage events and the ordinary operation of a real WDN in Spain. The data was provided by sensor records and expert knowledge. The resilience phases (i.e., absorptive, adaptive, restorative) within the collected dataset were rebuilt using event-driven and data-driven approaches. The results illustrated the importance of clustering the pressure head time series according to the phases of resilience. Capturing the behaviour of the pressure head as a determining hydraulic parameter before, during, and after the leakage was achieved by means of a spatial-temporal analysis. The results were promising, recognising the patterns of pressure head values throughout the network. It was observed that the network behaviour could be categorised into 1) ordinary operation and 2) during the event, which would allow to characterise the system behaviour when influenced by leakage/burst event and also explore its adaptability to resilience phases.

The approach was based on the available information from sensors and expert knowledge in a WDN. One of the benefits of this form of analysis is that if any sensor fails or is relocated, it is still possible to identify an abnormal incident in the network by spatial-temporal analysis. It means that the reflection of an event would be independent of only one specific sensor and will be obtained through the extracted patterns. In other words, events can be traced even when there is a lack of either historical data from sensors or records of utility experts. On the other hand, the information given by sensors is useful for checking if the analysis is correct. If the history of an event is missed, temporal-spatial analysis of pressures (and other parameters) can be practical.

The output of this preliminary study would be advantageous to develop research studies in many aspects, such as:

- The ability to extract relevant patterns (i.e., feature maps) from the preliminary results of the pressure head heatmaps allows for appropriate characterisation of these events.
- Increase the capacity of the network to learn from events by anticipating the potential reaction of the network (in this study, pressure head change) to a similar type of event.
- Pressure distribution maps during an event make it possible to recognise the critical areas in the network to a specific parameter before and after action. Many factors can be considered to make the best decisions to improve the preparedness of WDN. For example, a delay in pipe isolation (as an adaptive action) might negatively impact the pressure of the entire network depending on the affected part. Developing this approach will help identify the potential landmarks, the following purpose for our future study.
- This type of analysis with spatial-temporal dimension can be improved by including other hydraulic/non-hydraulic parameters such as flow, weather temperature, and factors causing the leakage (for example, pipe age, human/environmental interventions., etc.).
- The anomaly indicator could serve as a basis for further characterisation processes and support the decision-making process in terms of the implementation/deployment of actions likely to mitigate the effects of the event. In future research, we will investigate how to anticipate future events by increasing the network preparedness, being proactive in preventing the occurrence of an event, and/or responding more quickly to events.
- Intelligent data analysis tools are recommended for a comprehensive study of influencing parameters for this approach, taking into account diversities in the event types, causes of events, network types., etc.

8 REFERENCES

[1] FLOOD RISK MANAGEMENT, “Climate Change Sectoral Adaptation Plan”. Prepared by the Office of Public Works, https://www.gov.ie/en/publication/97984b-climate-change-and-sectoral-adaptation-plan/.

[2] L. Armstrong. “How a Small Water Leak Turns into a Flood,” https://restorationmasterfinder.com/restoration/how-a-small-water-leak-turns-into-a-flood/. (2017).

[3] D. Ayala-Cabrera, O. Piller, J. Deuerlein, M. Herrera, “Key Performance Indicators to Enhance Water Distribution Network Resilience in Three-Stages”. Water Utility Journal, vol.19, 2018, pp. 79–90.

[4] D. Ayala, Olivier Piller, J. Deuerlein, F. Parisini, F. Sedehizade, “ResiWater deliverable report D5.2: Development of Tools for Assessing WDS Vulnerability, Resilience and Robustness and Decision Support for Design”. irstea. 2017, pp.57. hal-02608637.

[5] R. Faturechi, E. Miller-Hooks, "Measuring the Performance of Transportation Infrastructure Systems in Disasters: A Comprehensive Review". J. Infrastruct. Syst. vol. 21, no. 1,. 2015, pp. 1-15.

[6] B. Balaei, S. Wilkinson, R. Potangaroa, N. Hassani, M. Alavi-Shoshtari, "Developing a framework for measuring water supply resilience". Natural Hazards Review. vol. 19,. 2018.

[7] B. Balaei, S. Wilkinson, R. Potangaroa, P. McFarlane, "Investigating the technical dimension of water supply resilience to disasters". Sustainable Cities and Society. vol. 56,. 2020.

[8] E. Zio, "The future of risk assessment". Reliability Engineering & System Safety. vol. 177,. 2018, pp. 176-190.

[9] O. Piller, F. Sedehizade, T. Bernard, M. Braun, N. Cheifetz, J. Deuerlein, et al. "Augmented Resilience of Water Distribution Systems following Severe Abnormal Events," in Proc. CCWI2017: F96. The University of Sheffield. https://doi.org/10.15131/shef.data.5363509.v1. (2017).

[10] D. Ayala-Cabrera, O. Piller, M. Herrera, D. Gilbert, and J. Deuerlein, "Absorptive Resilience Phase Assessment Based on Criticality Performance Indicators for Water Distribution Networks," Journal of Water Resources Planning and Management. vol. 145, no. 9,. 2015, pp. 04019037.

[11] Ouyang, M., L. Due˜nas-Osorio, and X. Min. 2012. "A three-stage resilience analysis framework for urban infrastructure systems." Struct. Saf. 36 (May/Jul): 23–31.

[12] K. Øien, A. Jovanović, L. Bodsberg, A. Øren, A. Choudhary, J. Sanne, L. Bergfors, M. Rahmberg, H. Matschke Ekholm, P. Thörn, R. Molarius, M. Vollmer, G. Walther, Z. Székely, P. Auerkari, P. Klimek, D. Bezrukov, "SmartResilience. Deliverable D3.6: Guideline for assessing, predicting and monitoring resilience of Smart Critical Infrastructures (SCIs) ". 2019. p. 15.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14084

THE IMPACT OF DRINKING WATER NETWORK MODEL SPATIAL AND TEMPORAL SCALE ON HYDRAULIC METRICS INDICATING DISCOLOURATION RISK

Reinar Lokk[1], Mirjam Blokker[2], Joby Boxall[3], Michele Romano[4], Anna Provost[5] and Stewart Husband[6]

[1,3,6]University of Sheffield, Sheffield, South Yorkshire, UK
[2]KWR Water Research Institute, Nieuwegein, Netherlands
[4,5]United Utilities, Warrington, UK

[1] *rlokk1@sheffield.ac.uk*, [2] *Mirjam.Blokker@kwrwater.nl*, [3] *j.b.boxall@sheffield.ac.uk*,
[4]*michele.romano@uuplc.co.uk*, [5]*anna.provost@uuplc.co.uk*, [6] *s.husband@sheffield.ac.uk*

Abstract

Matching model complexity to application and ensuring sufficient complexity to capture the emergent behaviour of interest is a perennial challenge. In this paper, we define a model as the variables, parameters and factors that represent a particular place, time and situation, not the software or algorithms. Specifically, we explore the cross products of spatial and temporal scaling of water demands within extended period 1D network hydraulic model simulations to predict the hydraulic conditions within individual drinking water pipes. High spatial scale hydraulic models investigated include mapping each customer with a unique demand node instead of the current practice of aggregated demand to nodes at the ends of pipe lengths. For demand profiling, we compare top-down DMA inlet patterns at 15-minute resolution with bottom-up stochastic demand patterns down to 1 second timesteps. The value of the resulting increases in resolution of hydraulic model outputs are captured in a range of pipe specific metrics that are likely to be indicative of water quality risk. Results explore different hydraulic metrics some of which may indicate discolouration risk by correlating with consumer reported discolouration events, showing how these change as a function of spatial and temporal resolution. For example, increasing the temporal scale from 15 minutes to 1-second results in more than a 100-fold increase in identifying flow reversal locations that can facilitate settling of network discolouration material and therefore pose a discolouration risk. High temporal scale is shown to capture the on/off nature of customer demands but a significant impact on daily peak velocities can only be observed using a temporal scale of 36s or higher. This work provides an indication of the quantity of added information which can indicate model spatial and temporal resolution required to differentiate pipes according to discolouration risk and hence improve targeting of pro-active maintenance and discolouration management efficiency.

Keywords
Hydraulic modelling, Spatial scale, Temporal resolution, Discolouration metrics, SIMDEUM.

1 INTRODUCTION

1.1 Model spatial scale and temporal resolution

Discolouration of drinking water is a dominant issue, for example in the UK (United Kingdom) the Drinking Water Inspectorate (DWI, independent auditors of UK water companies operating practices) reported that in 2019 1,811,121 customers were affected by significant discolouration events [1]. Several studies have been conducted that estimate discolouration risk via different hydraulic metrics, such as velocity or shear stress [2-4]. The metrics have then been used by drinking water network modellers to design new systems or simulate discolouration risk.

CC BY NC SA

Matching model complexity to application, and ensuring just sufficient complexity to capture the emergent behaviour of interest remains a perennial challenge.

Hydraulic models of drinking water distribution systems (DWDS) have historically been used to model the overall continuity of supply and check water pressures as the principal equations are well-known and conservative. Most of the standard principal equations of the DWDS solvers are based on the Global-Gradient-Algorithm (GGA), which combines energy loss and mass balance equations giving simultaneous solutions for pipe flows and nodal heads [5]. However, to save computing power and increase simulation speed, hydraulic models are often constructed using top-down approach and run at low temporal resolution (industry-standard temporal resolution is 15 minutes). Top-down hydraulic models (standard lumped models) include determining water demand of the whole district metered area (DMA) and then applying it to aggregated demand nodes with a reducing multiplier. This process assumes that all nodes follow identical demand profile thus failing to consider flow variability and stochastic effect [6]. Top down models have been viewed as an efficient way of saving computing power and dealing with highly random-stochastic demands of individual households [7]. However, since the creation of the GGA solver in 1988 computing power has increased millions of times [8] and stochastic demand generators such as SIMDEUM and WUDESIM offer a way of dealing with individual household demands. Both named toolkits can generate domestic water demands on a 1 second basis on a household level which allows averaging and studying the patterns at different temporal resolutions [9, 10]. SIMDEUM has been validated via field measurements and has been shown to result in a realistic representation of the likely network flows. This has the potential to highlight low flow and potential material settling zones as well as the parts of the network that may be considered self-cleaning [11].

It has been shown that the use of high-resolution SIMDEUM patterns has a direct impact on the demand variability, however, the effect of spatial and temporal scale has not been investigated on metrics determining discolouration risk [12]. Most referenced work has used hydraulic models in some form to confirm flow conditions in a network or to evaluate water quality. In this paper, we are taking the definition of a model as the variables, parameters and factors that provide an adequate representation of a particular place, time and situation, not the software or algorithms. Specifically, we explore the cross products of spatial and temporal scale of water demand within extended period 1D network hydraulic model simulations to predict the hydraulic conditions within individual drinking water pipes and the association of this with discolouration risk.

1.2 Discolouration metrics

Discolouration typically occurs when particles that have accumulated inside drinking water pipes become mobilised through hydraulic changes [3]. It is now understood that two primary processes allow material to accumulate generating a discolouration risk: cohesive material layers on all pipe surfaces or sedimentation due to self-weight forces [13, 14]. Research has shown that material accumulation is governed by daily conditioning shear stress [15]. Therefore, as velocity has the most direct impact on the hydraulic conditions (shear stress) in a pipe, for pro-active management it is important to understand the velocity thresholds that affect discolouration. One well-established velocity threshold is the Dutch self-cleaning velocity developed in systems without residual disinfectant. Earlier work on this defined a daily maximum of 0.4 m/s as self-cleaning velocity [16, 17], but more recently a daily peak of 0.2 - 0.25 m/s has been deemed sufficient to achieve self-cleaning conditions [2, 18]. It is important to note that this threshold has been identified as sufficient to keep material accumulation via pipe wall mechanism (cohesive material layers) and sedimentation minimal. Whereas another threshold of 0.05 - 0.06 m/s has been described as a threshold above which suspended particles should not be able to form sediments within the distribution pipes [2, 3, 19]. As described, discolouration is a mobilisation mechanism and thus it is essential to examine the effect of both spatial scale and temporal resolution on the threshold velocities which can be expected to affect discolouration risk.

BY NC SA

Fieldwork has shown that certain locations in drinking water networks can still pose a discolouration risk despite experiencing velocities above the self-cleaning threshold [20]. Therefore, in addition to fluid velocity, it is also necessary to consider other criteria to map discolouration risk across a network. Other metrics affecting discolouration risk do not have as strong and proven correlation with discolouration risk as velocity but can still give an insight into areas of concern. Flow reversals have been theorised to contribute to discolouration risk by effectively creating a zone where incoming suspended material cannot escape, thus allowing sediments to accumulate [14, 20]. Flow reversals alone however might not give a full picture as it does not indicate whether material is truly stuck in any specific area of the network. Instead, flow reversals combined with water age could give a better indication of discolouration risk as water age together with flow routes has been shown to influence water quality [21].

2 METHODOLOGY

To investigate the effect of model spatial scale and temporal resolution a residential study DMA site was selected in the UK. Figure 1 presents the DMA layout in two different formats: standard lumped (top down) and all connections. The standard lumped hydraulic model for the area considers 1358 customers spread over 235 nodes and 257 pipes (average pipe length 28m). To modify the standard lumped model to an all connections model, GIS data was extracted and imported into DNVGL Synergi hydraulic software. Tools within Synergi were then adapted to split existing pipes into smaller pipe lengths (average new pipe length 4m) depending on the location of individual customers from the GIS data (Figure 1). The geospatial accuracy of the new nodes largely depends on the imported GIS data. This process created an all connections model adding 1178 new nodes and splitting the 257 pipes to 1367 pipes. All connection refers to including every customer connection as a node along the pipe lengths.

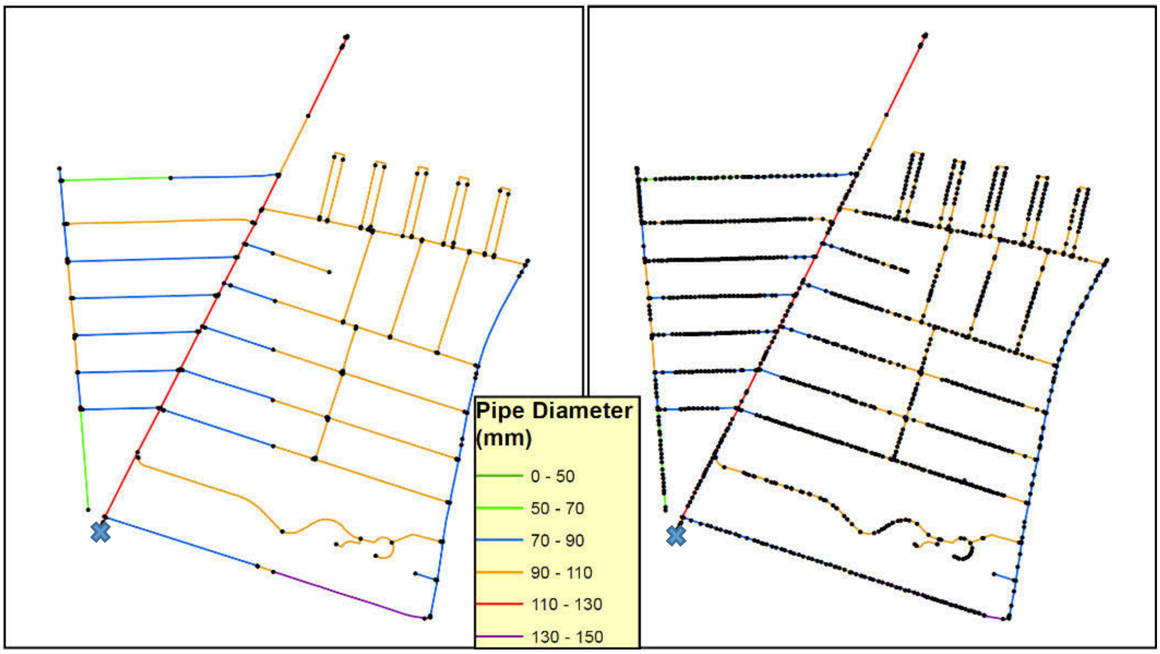


Figure 1. DMA layout in a standard lumped model (left) and all connections model (right). The blue icon represents the location of the DMA inlet.

The next step of the modification process included generating individual stochastic household demands for which SIMDEUM was selected. SIMDEUM patterns were calibrated to measured daily inlet demand by modifying daily consumption to 115.8 Lpppd (litres per person per day) and an average household size set to 2.3. For accurate calibration, social behaviour patterns that inform SIMDEUM were revised based on a recent study conducted in a comparable neighbourhood in The

Hague (NL) [22]. Social behaviour included setting unemployment rates, commuting habits and night use based on measured data. The night use was exaggerated to capture the effects of leakage by considering leakage as part of individual SIMDEUM demand profiles. Good calibration was confirmed by comparing DMA inlet patterns between different configurations, which included averaging SIMDEUM demands over multiple temporal resolutions. The selected time steps were 15 minutes, 9 minutes, 3 minutes, 36 seconds and 1 second (precise fractions of an hour) (Table 1). These individual demands were then allocated to nodes in Synergi so each node has the same profile where only the temporal resolution of the demand changes. Nodes where multiple customers share a single model node (apartments) were identified by checking GIS data. Additional SIMDEUM profiles were generated and assigned to apartment nodes based on the number of people that share a node.

The modifications retain the original lumped demand allocation in nodes and thus this converted all connections model is the base model for the standard lumped model analysis (Configuration 1, Table 2). Configurations 2-6 use the same all connection model with original demand allocation removed and SIMDEUM demands allocated to individual customer nodes. This allows comparable pipe lengths between the standard lumped model and the all connections models. As leakage is considered part of the SIMDEUM profiles, the standard lumped model leakage allocation was removed further highlighting only the effects of spatial scale and temporal resolution.

Table 1. Selected temporal resolutions with the demand profile and spatial allocation. All configurations use the all connections model (Figure 1, right) allowing comparable pipe lengths.

Configuration	**1**	**2**	**3**	**4**	**5**	**6**
Time step	15 minutes	15 minutes	9 minutes	3 minutes	36 seconds	1 second
Profile	Standard	SIMDEUM	SIMDEUM	SIMDEUM	SIMDEUM	SIMDEUM
Allocation	Lumped	Individual	Individual	Individual	Individual	Individual

The metrics investigated include pipe-specific parameters: velocity profiles, peak velocity, flow reversals and more complex metrics combining multiple thresholds. In addition, metrics such as duration at stagnation, above a threshold and below a threshold were aggregated across the model to showcase the effect of spatial and temporal scale. The effect of spatial-temporal scale on water age was investigated as a node property.

3 RESULTS

Figure 2 shows the match of inlet flow between different configurations over 24 hours showing a good calibration between automatically generated SIMDEUM profiles and the standard DMA inlet. There is a slight underestimation of SIMDEUM daytime and night-time demands compared to the standard DMA inlet pattern. The stochastic nature becomes more apparent as the temporal scale increases, which can be seen by sudden increases and decreases in the flow. To highlight the effect of temporal-spatial scale on the hydraulic model, this match between the automatically generated stochastic SIMDEUM demands, and the DMA inlet pattern is considered sufficient for this study.

BY NC SA

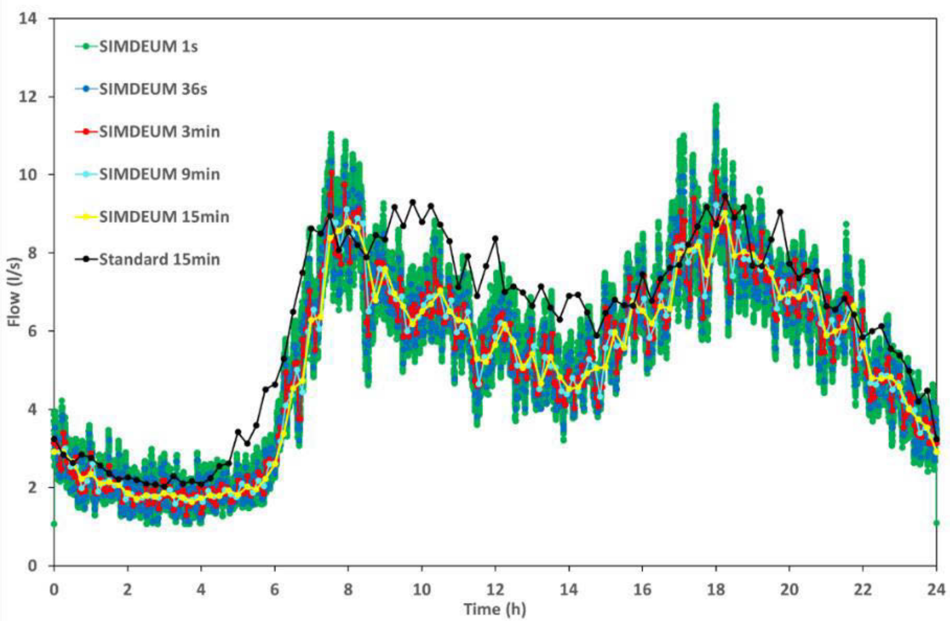


Figure 2. DMA inlet pattern for different configurations.

The effect of spatial-temporal scale on velocities can be viewed as a realistic representation of velocities in pipe sections. Figure 3 presents the velocity profile in a single pipe in the middle of a dead-end and in a pipe in the middle of a looped section. Both pipe sections are not on the main flow path. Regardless of the location of the pipe section, 1s temporal resolution always results in the highest observed velocity. Figure 3 reveals that an increase in temporal resolution has the most significant impact on the observed velocity profile. The change from low spatial scale to high spatial scale and stochastic demands seems to be more significant for pipes in loops compared to pipes at dead-end sections of the network, however, the standard demand fails to capture any stochastic behaviour. It is apparent that the velocity in the looped pipe section is much more stochastic as shown by the multiple sudden velocity changes crossing the x-axis entirely (flow reversal). In addition, configurations using SIMDEUM demands result in a strikingly different velocity profile compared to the configuration using lumped standard demand.

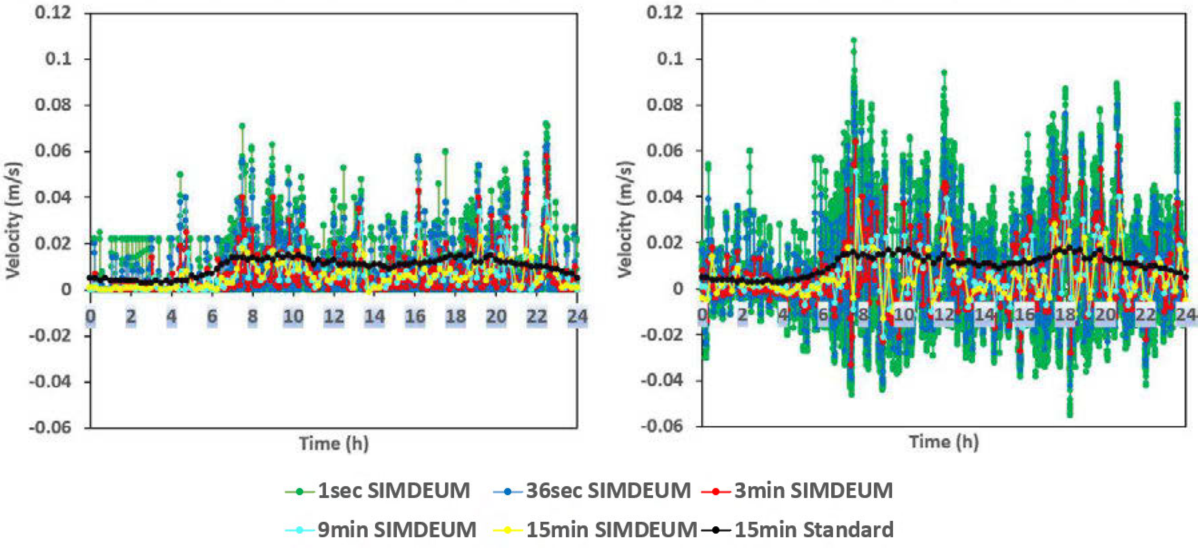


Figure 3. The effect of spatial scale and temporal resolution on velocity in a dead-end pipe (left) and on a pipe in the middle of a looped section (right).

With the potential for material to settle, for example due to low velocities, stagnation may be considered a key factor in estimating discolouration risk. For this, hours of stagnation were calculated and tabulated across different spatial-temporal resolutions, Table 2. Here the increase in spatial and temporal resolution is significant as evidenced by the steeper increase in the total number of hours of stagnation across the whole network. In addition, the total hours spent above or below two different key velocity thresholds (0.06 m/s and 0.25 m/s)[3, 18] were determined. The standard model overestimates the time spent above 0.25 m/s and underestimates time spent below 0.06 m/s compared to the higher spatial and temporal simulations. In addition, the temporal resolution of 1 second results in the highest peak velocity across all pipes. Going from a 15 minute standard lumped model to a 15 minute SIMDEUM all connections model (spatial change) has only a limited impact on peak velocities. Temporal resolution however has an increasing impact as time steps reduce in size.

Table 2. The effect of spatial scale and temporal resolution on metrics aggregated across the model. For reference, the total maximum possible hours across all DMA pipes is 32832 (hours).

	Maximum velocity across the whole model (m/s)	Total hours of stagnation across all pipes (hours)	Total hours all pipes experienced velocity above 0.25 m/s (hours)	Total hours all pipes experienced velocity below 0.06 m/s (hours)
Effect of Spatial Scale and Stochastic Demands				
15min Standard	1.20	1730	1063	24328
15min SIMDEUM	1.16	1900	610	26604
Effect of Temporal Resolution				
15min SIMDEUM	1.16	1900	610	26604
9min SIMDEUM	1.17	2059	752	26644
3min SIMDEUM	1.29	2196	894	26634
36sec SIMDEUM	1.44	2554	954	26643
1sec SIMDEUM	1.53	2754	987	26523

Figure 4 shows the total number of flow reversals in 24 hours. The most notable difference is the standard lumped model shows the total number of flow reversals across the whole network as 0. The move from the lumped to the all connections model with SIMDEUM profiles reveals the location of flow reversals plus further locations occur as the temporal resolution increases. As the temporal scale increases, however, there are more opportunities for flow reversals with 43200 chances for 1 second simulations compared to 48 for 15 minutes, hence contributing to the increase in total flow reversals observed. Configurations using SIMDEUM reveal that only the pipes near the inlet and across the bottom right of the network are mostly unidirectional. Moving away from the main flow paths results in the total number of flow reversals increasing with increasing temporal resolution of the demand patterns.

Figure 5 presents results for maximum average water age for different configurations. Water age analysis was run for 48 hours (sufficient to reach stable repeating daily patterns of water age with the maximum water age being 24hours) at 0.005 hours temporal scale for configurations 1-5, for configuration 6 the water age analysis time step was set to 0.0002778 hours (1 second). Higher propagation timestep and longer total simulation times for water propagation analysis (water age) did not have an impact on water age results. The results in Figure 5 show the maximum average water age between 24 to 48 hours. A significant difference comes when moving from the lumped model to an all connections model. The lumped model appears to have two redundant loops with high water age. Closer inspection revealed that due to the original lumped allocation of demands or top down modelling approach the loops where water age exceeds 18 hours (in 15 min lumped config) have no flow through them. Added spatial scale fixed that error by considering a more realistic spatial representation of water movement down all pipe sections. Temporal resolution, however, seems to have only a limited effect on water age.

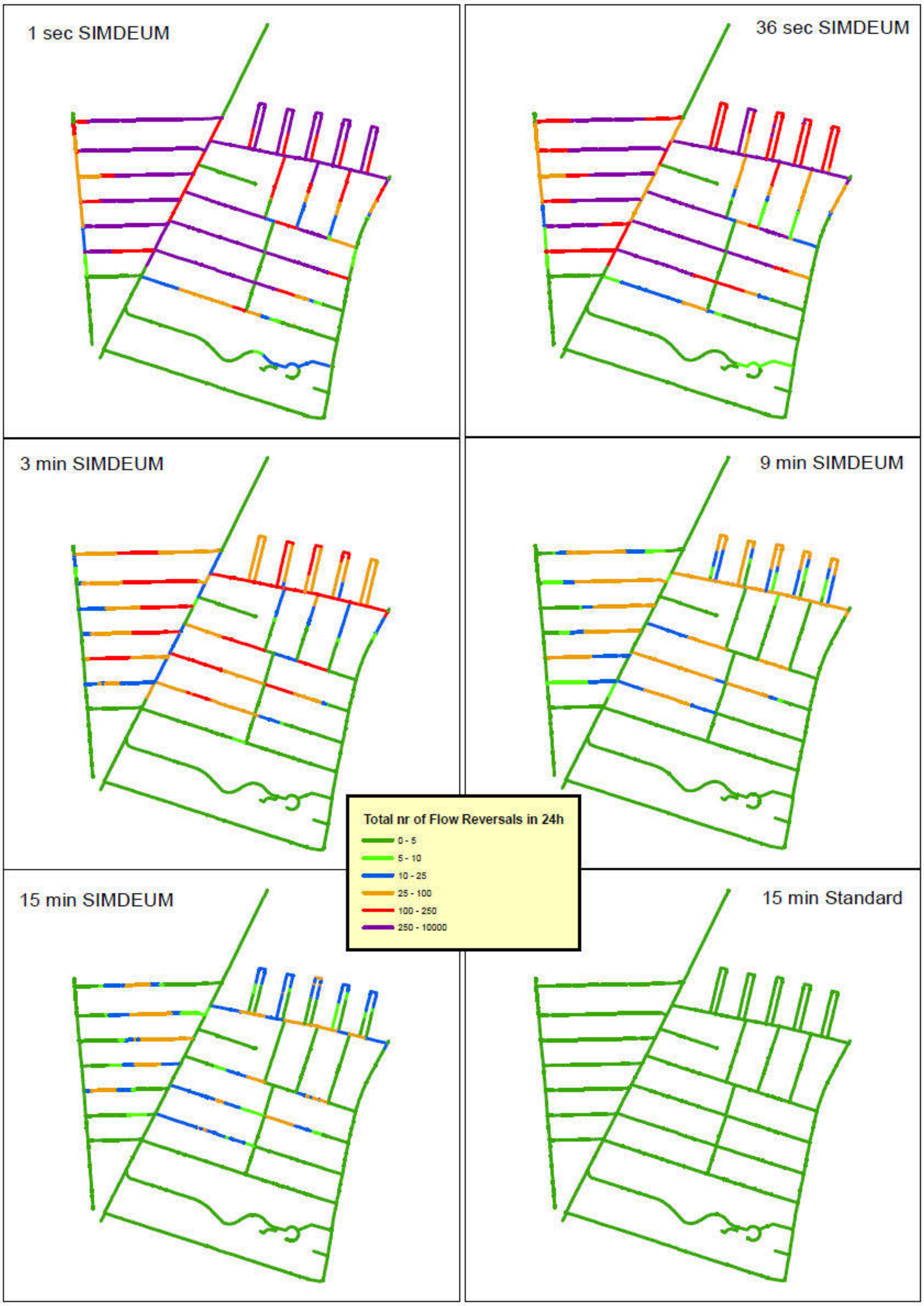

1 sec SIMDEUM
36 sec SIMDEUM
3 min SIMDEUM
9 min SIMDEUM
15 min SIMDEUM
15 min Standard
Total nr of Flow Reversals in 24h
0 - 5
5 - 10
10 - 25
25 - 100
100 - 250
250 - 10000

Figure 4. The total number of flow reversals in each pipe section mapped across the whole network.

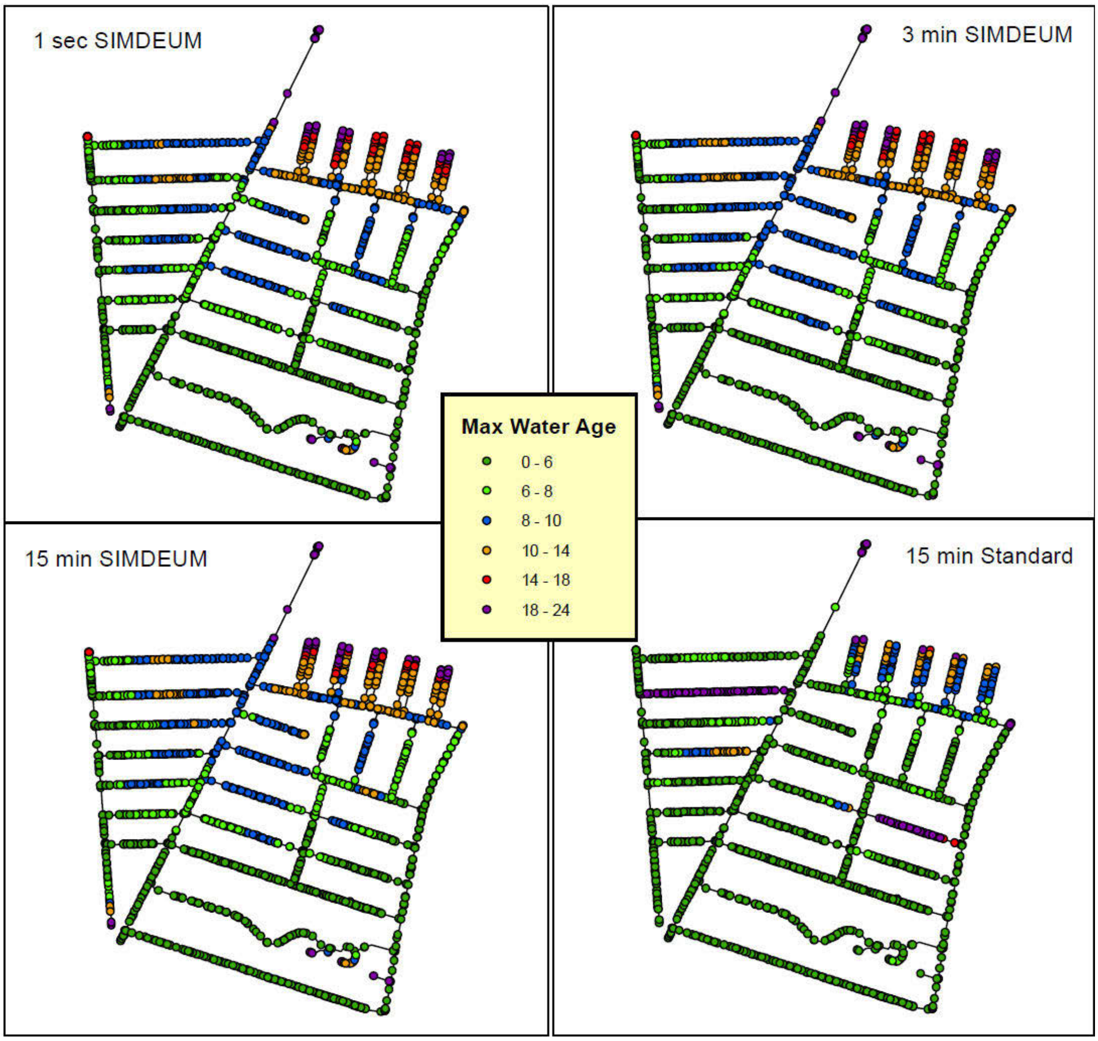


Figure 5. Maximum average water age (hours) in each customer node mapped across the whole network.

Figure 6 investigates which pipes experience the proposed self-cleaning velocities by looking at the maximum velocity in each pipe at different configurations. The figure also shows the apparent random customer discolouration contacts from the past 7 years as black dots. The spatial scale has a limited effect on maximum velocities, agreeing with the overall maximum velocity data in Table 2. However, the increasing temporal resolution does show some differences. For example, comparing 15min SIMDEUM to 1s SIMDEUM reveals that the bottom left dead-end pipe section could experience self-cleaning conditions. Customer contacts are notoriously inconsistent (many factors influence why a customer does or does not make a contact) and the potential correlation between customer contacts and any metrics including self-cleaning velocities can only be considered crude. Results however support a possible correlation with most contacts located towards the top of the DMA where there are no self-cleaning pipes whilst around the inlet with higher, potentially self-cleaning, velocities there are fewer reported discolouration events.

Figure 7 presents an example of a more complex combined metric where pipes are coloured by the number of times experiencing velocity less than 0.06 m/s for more than an hour continuously. This metric considers pipes where the velocity is below a potential sedimentation threshold for long enough to allow the possibility of suspended material to settle. As was seen in Figure 3, the increases in velocity may only be short-lived resulting in a stop-start motion of the flow, rather than quiescent conditions for long enough for the suspended particles to settle out. Pipes with a greater number of occurrences of consistently low velocity for an hour or more, perhaps correlate

CC BY NC SA

even more with the discolouration contacts from the past 7 years. This stop-start stochastic motion in pipes is only picked up by high-end (1s) temporal resolution considering all connections.

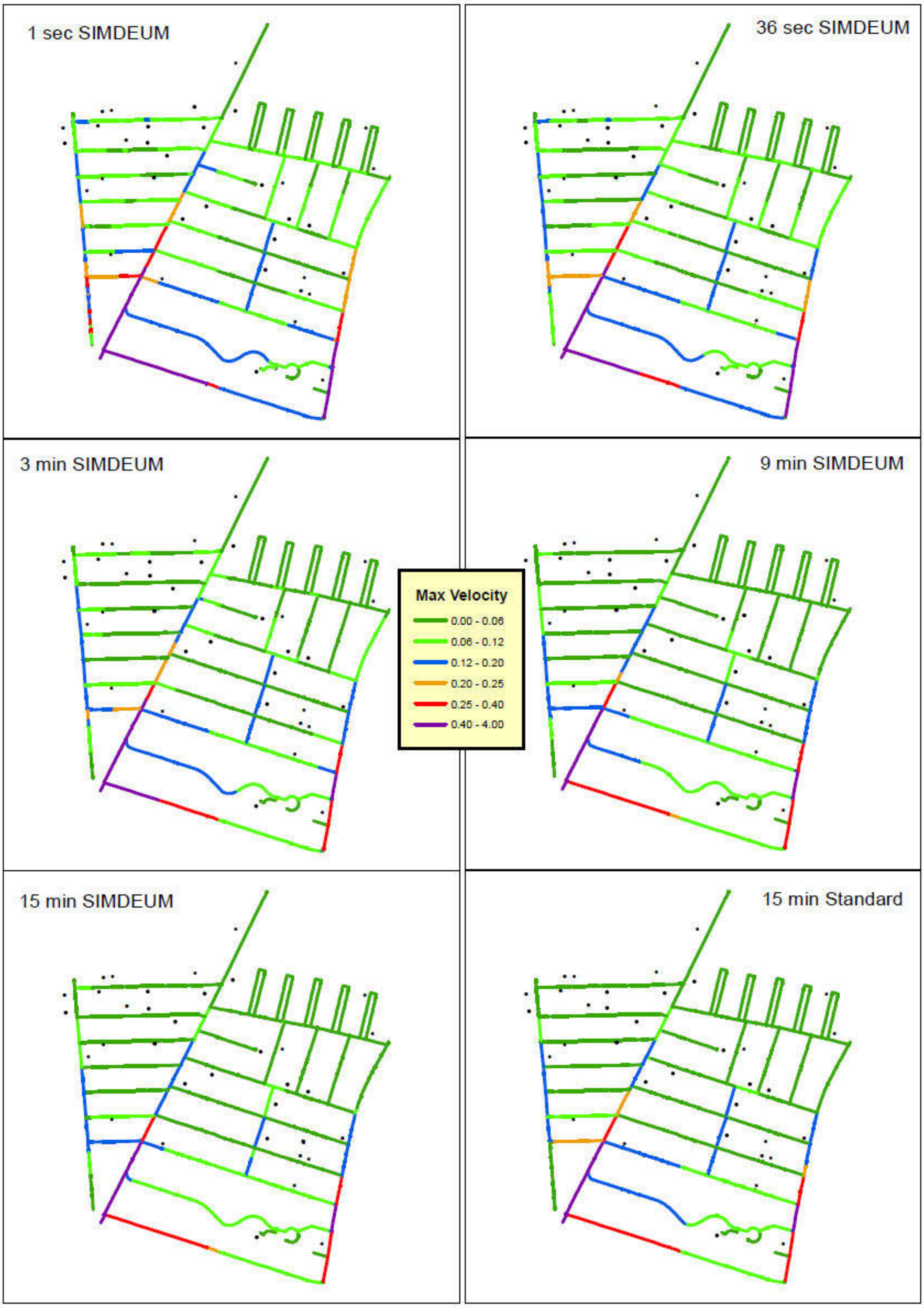

1 sec SIMDEUM
36 sec SIMDEUM
3 min SIMDEUM
9 min SIMDEUM
15 min SIMDEUM
15 min Standard
Max Velocity
0.00 - 0.06
0.06 - 0.12
0.12 - 0.20
0.20 - 0.25
0.25 - 0.40
0.40 - 4.00

Figure 6. The maximum velocity in pipes across the DMA. Black dots mark the customer reported discolouration events from the past 7 years.

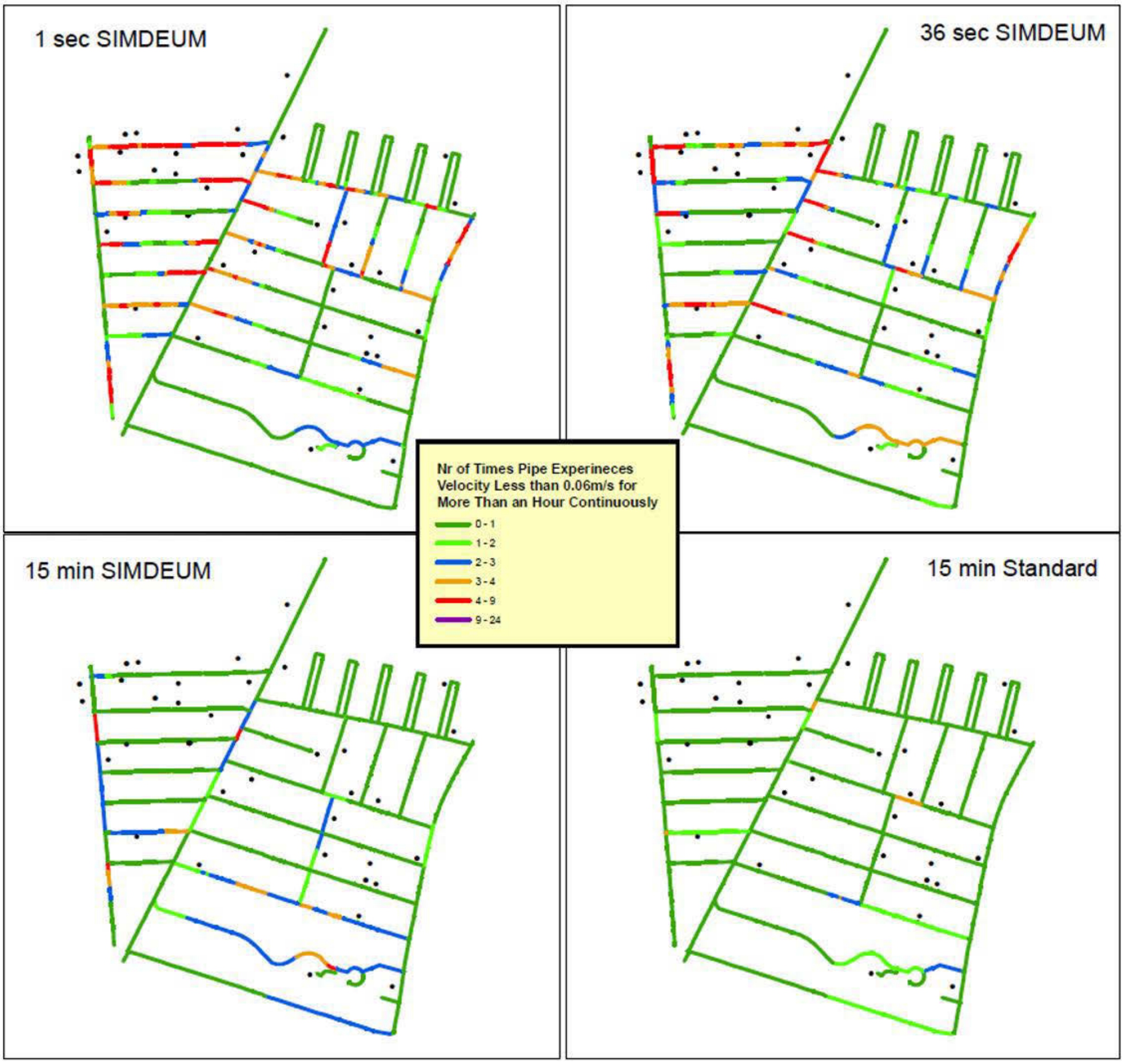

1 sec SIMDEUM
36 sec SIMDEUM
15 min SIMDEUM
15 min Standard
Nr of Times Pipe Experineces Velocity Less than 0.06m/s for More Than an Hour Continuously
0 - 1
1 - 2
2 - 3
3 - 4
4 - 9
9 - 24

Figure 7. Network-wide look at the number of times pipe experiences velocity less than 0.06 m/s for more than an hour continuously. Black dots mark the customer reported discolouration events from the past 7 years.

4 DISCUSSION

Overall, the modification process was mostly automated and did not take much computing power or time. However, simulating results on 1-second temporal resolution was computationally demanding and could take several hours. In total 6 different scenarios were studied (Table 1) each representing a different combination of spatial-temporal scale. The comparison between configurations 1 and 2 highlighted the effect of increasing the model spatial scale and the addition of stochastic demands. Configurations 2-6 showed the effect of the model temporal scale on results.

The results reveal that the biggest effect of spatial resolution can be observed in downstream pipes away from main flow paths. Hydraulic conditions affecting discolouration risk in residential pipes off the main flow paths are not realistically represented by lumping demands to end of pipe nodes. All connection models can, therefore, be viewed as a more realistic representation of hydraulic conditions in residential area pipes, where flows are dominated by individual customer water consumption behaviours. The results are consistent with literature where it has also been shown that low spatial scale and temporal resolution models only suffice modelling trunk mains [23]. Moving from standard lumped models to the all connections model with 15 minute SIMDEUM stochastic demands affected all studied metrics apart from peak velocity. However, there was a limited impact for increasing temporal resolution on metrics that capture cumulative effects of the water moving through a network. Similarly, only spatial scale with stochastic demands seems to matter in determining total time spent under sedimentation threshold (0.06 m/s) [3, 14, 19]. Water quality surrogates such as water age (shown in figure 5) and chlorine [24] aggregate time and pipe effects from source to point of interest thus only capturing cumulative effects of the entire network.

Discolouration has been shown to be mainly a function of the network velocities (shear stress), whilst metrics such as flow reversals may serve as additional criterion to help determine risk zones [4, 13, 20]. In both cases the move to all connections model with stochastic demands seems to have only limited benefit. Instead to capture discolouration risks in detail, increase of temporal scale is needed. The results of this study show that high spatial scale models with stochastic demands at temporal scale of 36s are sufficient to capture more than 90% of peak velocity values, total time at stagnation, map location of most flow reversals and capture total hours above or below a threshold. This is consistent with previous research where it has been shown that temporal resolution of higher than 30s results in only 10% missed variance in instantaneous water demands [12]. Standard lumped models underestimating time at stagnation correlate with Blokker et al. (2008) where 1 minute temporal resolution was shown to be sufficient to determine realistic peak Reynolds number and probability of stagnation [23]. Temporal scale of 36s revealed that there are possibly hundreds of flow reversals in most residential area pipe sections even on or close to main flow paths. Based on these results, the conditions in residential water network pipes have much more start-stop motions and are more multidirectional than previously thought [25]. This could significantly contribute to discolouration risk by trapping material and allowing it to suspend or disturb already accumulated material. Customers either have their water tap on or off for a mostly short amount of time meaning, this high temporal stochastic effect plays a significant role in determining pipe specific hydraulic metrics. Based on the results of this study and literature, metrics affecting discolouration risk have been shown to be primarily a function of the hydraulic conditions in a specific length of pipe, which are influenced by both spatial scale and temporal resolution (Figures 3, 4, 7) [2, 9, 10]. The all connections hydraulic models using stochastic demands with temporal scale of 36s or higher can improve indication of discolouration risk by revealing more pipe specific information.

Identification of discolouration risk may involve a combination of several different metrics and their thresholds. Figures 6 and 7 attempted to correlate customer reported discolouration contacts with self-cleaning pipes and an example of a combined threshold metric. In looped sections of the network, an increase of the temporal scale only resulted in a small increase in number of pipes observed to exhibit proposed self-cleaning velocities which might be due to the extremely looped nature of the selected DMA. This indicates that to achieve self-cleaning velocities and reduce discolouration risk, branched network layouts may offer an improved design [17, 26, 27]. The more complex metric of Figure 7 investigated the stochastic stop-start motion of the velocity in individual pipe sections followed by a period that might facilitate both particle aggregation (increased mass due to flocculation) and settling. This has been reported to take place at velocities below 0.06 m/s [3]. It may then be considered that once settled, factors including particle shielding and potential additional cohesive forces may result in higher shear forces being required to re-mobilise allowing larger deposits to form creating a localised discolouration risk.

It is important to note that investigating potential correlations between discolouration risk and modelled metrics based on customer contacts will remain inconclusive and field evidence is required. This could then evidence both material accumulation zones and self-cleaning pipes facilitating modelling metrics to be developed to help improve targeting of pro-active network maintenance and discolouration management.

5 CONCLUSIONS

This study investigated the effect of spatial and temporal scale on the drinking water distribution system hydraulic model at a DMA level. Key findings from this study are:

- All connections models using stochastic demands capture more detailed hydraulic information in residential area pipes that are considered important in determining discolouration risk.
- Moving from industry-standard 15 minute lumped models to all connections models using 15 minute stochastic profiles (added spatial and stochastic effect) has minimal effect on observed peak velocities, but impacts other possible discolouration metrics including velocity profiles, flow reversals and water age.
- Increasing the temporal scale adds more information on pipe specific properties such as velocity profiles, peak velocity and flow reversals but little additional information on water age (as it is a water quality surrogate aggregating time and pipe effects from source to point of interest) or identifying more potential self-cleaning pipes (daily maximum velocity greater than 0.25 m/s).
- The combined threshold metric of times a pipe experiences velocity less than 0.06 m/s for more than an hour continuously could be of use in identifying discolouration risk areas.

6 ACKNOWLEDGEMENTS

This research was sponsored by United Utilities and CDT WIRe (EPSRC EP/S023666/1). We would like to acknowledge the support of KWR Water Research Institute for providing SIMDEUM Pattern Generator for the purposes of this research. Special thanks to United Utilities for supporting this work. For the purpose of open access, the author has applied a creative commons attribution (CC BY) license to any author accepted manuscript versions arising.

7 REFERENCES

[1] DWI, "The Chief Inspector's Report for Drinking Water in England," 2020.

[2] E. J. Blokker, J. H. Vreeburg, P. G. Schaap, and J. C. van Dijk, "The Self-Cleaning Velocity in Practice," presented at the Water Distribution System Analysis 2010 – WDSA2010, Tucson, USA, September 12-15, 2010.

[3] J. H. Vreeburg and J. B. Boxall, "Discolouration in Potable Water Distribution Systems: A Review," Water Res, vol. 41, no. 3, pp. 519-29, Feb 2007, doi: 10.1016/j.watres.2006.09.028.

[4] J. B. Boxall and N. D. Dewis, "Identification of Discolouration Risk Through Simplified Modelling," presented at the ASCE ERWI World Water and Environmental Water Resources, Anchorage, USA, 15-19 May, 2005.

[5] A. Menapace and D. Avesani, "Global Gradient Algorithm Extension to Distributed Pressure Driven Pipe Demand Model," Water Resources Management, vol. 33, no. 5, pp. 1717-1736, 2019, doi: 10.1007/s11269-018-2174-3.

[6] D. Fiorillo, E. Creaco, F. De Paola, and M. Giugni, "Comparison of Bottom-Up and Top-Down Procedures for Water Demand Reconstruction," Water, vol. 12, no. 3, 2020, doi: 10.3390/w12030922.

[7] F. J. Martínez-Solano, P. L. Iglesias-Rey, D. Mora-Meliá, and V. S. Fuertes-Miquel, "Exact Skeletonization Method in Water Distribution Systems for Hydraulic and Quality Models," Procedia Engineering, vol. 186, pp. 286-293, 2017, doi: 10.1016/j.proeng.2017.03.246.

[8] N. Routley. "Visualizing the Trillion-Fold Increase in Computing Power." Visualcapitalist. https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/ (accessed March, 2022).

[9] E. J. M. Blokker, J. H. G. Vreeburg, H. Beverloo, K. A. M., and J. C. van Dijk, "A Bottom-Up Approach of Stochastic Demand Allocation in Water Quality Modelling," Drinking Water Engineering and Science, vol. 3, pp. 43-51, 2010, doi: www.drink-water-eng-sci.net/3/43/2010/.

[10] A. A. Abokifa, P. Biswas, B. R. Hodges, and L. Sela, "WUDESIM: A Toolkit for Simulating Water Quality in the Dead-End Branches of Drinking Water Distribution Networks," Urban Water Journal, vol. 17, no. 1, pp. 54-64, 2020, doi: 10.1080/1573062x.2020.1734949.

[11] E. Creaco, M. Blokker, and S. Buchberger, "Models for Generating Household Water Demand Pulses: Literature Review and Comparison," Journal of Water Resources Planning and Management, vol. 143, no. 6, 2017, doi: 10.1061/(asce)wr.1943-5452.0000763.

[12] S. Diaz and J. Gonzalez, "Temporal Scale Effect Analysis for Water Supply Systems Monitoring Based on a Microcomponent Stochastic Demand Model," Journal of Water Resources Planning and Management, 2021, doi: 10.1061/(ASCE).

[13] J. Boxall, P. J. Skipworth, and A. J. Saul, "A Novel Approach to Modelling Sediment Movement in Distribution Mains Based on Particle Characteristics," in Computing and Control for the Water Industry, Leicester, 3-5 September 2001 2001.

[14] H. Armand, I. I. Stoianov, and N. J. D. Graham, "A Holistic Assessment of Discolouration Processes in Water Distribution Networks," Urban Water Journal, vol. 14, no. 3, pp. 263-277, 2015, doi: 10.1080/1573062x.2015.1111912.

[15] D. M. Cook and J. B. Boxall, "Discoloration Material Accumulation in Water Distribution Systems," Journal of Pipeline Systems Engineering and Practice, vol. 2, no. 4, pp. 113-122, 2011, doi: 10.1061/(asce)ps.1949-1204.0000083.

[16] M. Boomen, A. Mazijk, and B. R.H.S, "First Evaluation of New Design Concepts for Self Cleaning Distribution Networks," Journal of Water Supply: Research and Technology - AQUA, vol. 53, no. 1, pp. 43-50, 2004.

[17] J. C. van Dijk, P. Horst, E. J. M. Blokker, and J. H. G. Vreeburg, "Velocity-Based Self-Cleaning Residential Drinking Water Distribution Systems," Water Supply, vol. 9, no. 6, pp. 635-641, 2009, doi: 10.2166/ws.2009.689.

[18] J. van Summeren and M. Blokker, "Modelling Particle Transport and Discolouration Risk in Drinking Water Distribution Networks," Drinking Water Engineering and Science, vol. 10, no. 2, pp. 99-107, 2017, doi: 10.5194/dwes-10-99-2017.

[19] G. Ryan et al., "Particles in Water Distribution Systems, Research Report 33," Cooperative Research Centre for Water Quality and Treatment, 2008.

[20] I. W. M. Pothof and E. J. M. Blokker, "Dynamic Hydraulic Models to Study Sedimentation in Drinking Water Networks in Detail," Drinking Water Engineering and Science, vol. 5, no. 1, pp. 87-92, 2012, doi: 10.5194/dwes-5-87-2012.

[21] J. Machell and J. Boxall, "Modeling and Field Work to Investigate the Relationship between Age and Quality of Tap Water," Journal of Water Resources Planning and Management, vol. 140, no. 9, 2014, doi: 10.1061/(asce)wr.1943-5452.0000383.

[22] J. E. van Steen, "Robustness Assessment of Leak Detection and Localization in Water Distribution Networks Under Stochastic Water Demand," Masters Thesis, Delft University of Technology, Delft, 2020.

[23] E. J. Blokker, J. H. Vreeburg, S. G. Buchberger, and J. C. Dijk, "Importance of Demand Modelling in Network Water Quality Models: a Review," Drinking Water Engineering and Science, vol. 1, pp. 27-38, 2008, doi: https://dwes.copernicus.org/articles/1/27/2008/.

[24] M. Blokker, J. Vreeburg, and V. Speight, "Residual Chlorine in the Extremities of the Drinking Water Distribution System: The Influence of Stochastic Water Demands," Procedia Engineering, vol. 70, pp. 172-180, 2014, doi: 10.1016/j.proeng.2014.02.020.

[25] E. J. Blokker, "Stochastic Water Demand Modelling for a Better Understanding of Hydraulics in Water Distribution Networks," PhD, University of Delft, 2010.

[26] E. J. Blokker, J. Vreeburg, P. G. Schaap, and P. Horst, "Self-Cleaning Networks Put to the Test," presented at the World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, 2007.

[27] E. Abraham, M. Blokker, and I. Stoianov, "Decreasing the Discoloration Risk of Drinking Water Distribution Systems through Optimized Topological Changes and Optimal Flow Velocity Control," Journal of Water Resources Planning and Management, vol. 144, no. 2, 2018, doi: 10.1061/(asce)wr.1943-5452.0000878.

BY NC SA

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14085

LEAK DETECTION AND LOCATION IN A REAL WATER DISTRIBUTION NETWORK USING A MODEL-BASED TECHNIQUE

Bruno Ferreira[1], Nelson Carriço[2] and Dídia Covas[3]

[1,2]INCITE, Barreiro School of Technology, Polytechnic Institute of Setúbal, Setúbal, Portugal
[3]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[1] *bruno.s.ferreira@estbarreiro.ips.pt,* [2] *nelson.carrico@estbarreiro.ips.pt,*
[3] *didia.covas@tecnico.ulisboa.pt*

Abstract

This paper presents a practical application of a model-based approach for leak detection and location in a real water distribution network. The methodology is divided into three main steps: 1) identification of the DMA with the highest leakage volume; 2) hydraulic model update and 3) leak location by inverse analysis. The water distribution network is monitored through 14 flowrate and pressure sensors and data were available for 1.5 years. The methodology is applied 15 times, specifically to five weekdays in three different periods with high seasonal variation (summer, spring and winter). Regardless of the analysed period, obtained results point to the presence of a leak with ca. 4 m^3/h in a specific area of the network. This approximate location is a starting point for the application of more precise leak location techniques using acoustic equipment.

Keywords
Leakage identification, Leakage location, Hydraulic simulation, Minimum night flow, Optimization.

1 INTRODUCTION

The occurrence of pipe bursts represent an important source of water losses in water distribution networks (WDN), depending on their frequency and size. Users usually report events to utilities when surface flooding or service disruption occurs. The problem lies in events that do not cause water to come up to the surface, nor cause service disruption. These events can continue unreported for a long time, potentially leading to the loss of large volumes of water and the consumption of energy and water disinfectants, whilst also posing a health risk to the population due to risks of bacteria and pollutant contamination [1].

The first step towards an efficient leakage management is the accurate assessment for the water volume that is lost. To this end, WDN are usually divided in smaller district metered areas (DMA) in which the flowrates are continuously measured in the area's inlets and outlets and the consumed water volume is periodically measured in consumers. Leakage volumes in DMA are usually assessed through the the minimum night flow (MNF) regime [2,3]. During the MNF, consumption from users is usually minimum and the dominant consumption is due to leakage. Such MNF analysis can reduce the search area from the whole WDN to a particular DMA. Nonetheless, additional burst and leak location techniques are necessary to find the approximate location inside the DMA and the exact location at the street level.

Leak location techniques generally involve the use of acoustic equipment (e.g., listening rods and leak correlators) or of non-acoustic methods (e.g. gas tracer injection, ground-penetrating radar, or infrared photography) [4]. Although accurate, these equipment-based inspection techniques are time-consuming and labor-intensive processes, even in relatively small DMA [5]. Thus, a software-based approach is should be used before to reduce the search area to a particular zone in the DMA (e.g., at the street level).

Distinct software-based techniques have been developed in last decades for leak detection and location in WDN. These techniques can be roughly divided in model-based and data-driven methods. Data-driven methods [6–8] use monitoring data that, combined with tools such as data mining or artificial intelligence algorithms, allow the identification of possible leak location zones [9]. Although these methods do not require a deep knowledge regarding the WDN hydraulic characteristics (e.g., pipe characteristics or individual demands), extensive historical data records of monitoring data and precise information on past pipe burst events are necessary. In many water utilities, service work orders are collected by operational technicians who register imprecise and incomplete data [10].

Model-based methods use a hydraulic simulation models and detect and locate leaks based on the comparison of numerical results with measurements from pressure sensors and flowrate meters. Examples of these methods are the inverse analysis [11–13], the network sensitivities' computation and analysis [14,15], the error-domain model falsification [16,17], or based on a classification problem [18,19]. Although promising results have been obtained using these methods (see [13,20]), their application usually requires a robust and well calibrated hydraulic model and a high number of pressure sensors with long time-records. In addition, obtained results are highly sensitive to the uncertainty of model parameters (e.g., water consumption, valve states, and pipe characteristics).

This paper presents the practical application of a model-based approach for leak detection and location in a real WDN located in Porto Metropolitan Area, in Portugal. The network has a total length of 37 km and serves 996 consumers, being divided in seven DMA. The water utility has identified the existence of a leak by analysing the minimum night flow. The leak location is known to be somewhere in an area along the embankment of a river, whose flow affects the noise captured by the acoustic equipment. In this context, the water utility challenged the authors to apply a model-based technique and find the leak location. This approximate location will be the starting point for the application of more precise leak location techniques using acoustic equipment. Data were collected in 14 sensors (seven pressure transducers and seven flowmeters) during the period between 1st of June 2020 and 26th of December 2021. The number and location of pressure and flowrate sensors were established by the water utility to monitor DMA (inlets and outlets) both in terms of water consumption and pressure requirements. The model-based methodology is applied to 15 different days divided into three periods (with different data availability). Genetic Algorithms [21] are used as the optimization method. Obtained results are discussed and the most relevant conclusions are drawn, focusing on the difficulties of implementing model-based techniques to leakage location in real-life contexts.

2 METHODS

The used model-based methodology is divided in three main steps: 1) identification of the DMA with the highest leakage volume; 2) hydraulic simulation model update and 3) leak location by inverse analysis. The first step is aims at identifying the DMA with highest leakage volumes based on the analysis of the MNF and at estimating the burst magnitude. The hydraulic simulation model should be updated with the water consumption and valve settings corresponding to the collected data period. Finally, the leak/burst is located by formulating and solving the inverse problem. The objective is to determine the values of the unknown parameters (leak location and size) in the hydraulic model that minimizes the difference between numerical results and real measurements. These steps are further explained in the following sections.

2.1 Identification of the DMA with the highest leakage volume

A simple yet effective method is proposed herein to identify the DMA with the highest leakage volume, as well as to roughly estimate the respective size. Two metrics may be calculated for this purpose for each DMA by considering a complete day of measurements. The first metric (M1) quantifies the average water consumption during the MNF per service connection (SC) [m^3/(h.SC)] in each DMA. During the MNF, the user water consumption is minimum, the network pressure is high and, thus, the leak volume has the highest contribution to the flowrate. The second metric (M2) quantifies the average daily consumption per service connection [L/(day.SC)]. The water losses can be assessed in each DMA by comparing the metrics M1 and M2 values for DMA with similar characteristics (e.g., similar size or topology) and by assessing if any DMA presents outlier values of these metrics. The first metric aims to assess if unexpected consumption is occurring during the night whilst the second metric aims to assess if unexpected consumption is occurring throughout the day. The occurrence of outlier metric values for a DMA indicate that every service connection of that DMA is presenting (on average) an abnormally high water consumption during either the night (for the first metric) or through the day (for the second metric).

Once the DMA with a high leakage level has been identified, the leak volume can be estimated. To this end, the average of metric M1 values (without the outlier values) should be calculated for DMA with similar characteristics. This value (in m^3/[h.SC]) should be multiplied by the number of SC in the DMA with leakage. This allows estimating the MNF (in m^3/h) in a no-leakage scenario. Finally, the leak size (in m^3/h) can be estimated by subtracting the MNF in the no-leakage scenario from the real MNF.

This analysis can be carried out for distinct periods (e.g., summer or winter, weekdays or weekends) to validate the identified DMA with high leakage levels.

2.2 Hydraulic simulation model update scheme

The application of the leak location technique described in 2.3 requires the use of a hydraulic model of the WDN. To guarantee reliable results, it is essential to minimize the uncertainty in model parameters (which arise predominantly from water consumption [17]). This can be achieved by updating the hydraulic simulation model with the actual water consumption values that occurred (minus the effect of the leak) during the period for which the leak location analysis will be performed. If telemetry at the user level is available, the hydraulic simulation model can be updated with the individual consumption of each user (e.g., through an individual daily measurement). In most cases, the individual consumption of each user is unknown and only the total consumption of each DMA is known. In these situations, a consumption pattern for each DMA can be created (and is equal to the DMA consumption); the base demand of each node (of each DMA) represents the proportion of the consumption of that specific node concerning the overall DMA. This proportion can be obtained, for instance, based on the billing records (i.e., the amount of billed water).

In addition to water consumption, boundary conditions must be properly updated. These include settings for valves, pumps, and tanks. For this purpose, specific time step controls can be developed if measurements are available (e.g., a rule can be created for the setting of a PRV based on the outlet pressure measurements of the PRV).

2.3 Leak location

The leak location can be carried out by solving the inverse problem, that is, by finding the values of the unknown parameters (leak size and location) in the hydraulic model, that minimize the differences between numerical results and real measurements [11–13].

Consider the N vector containing the indexes of all potential leak locations (e.g., all node) in the DMA identified in 2.1. A single-objective optimization problem is formulated and two decision variables are considered, namely, the leakage location (defined as a discrete variable that may assume any value in N) and the leakage size (i.e., the value of the emitter coefficient C, defined as a continuous variable). Two types of constraints are considered, namely, a set of implicit type constraints considering hydraulics systems (which are guaranteed by the hydraulic solver) and an additional constrain which bound the algorithm solution search space by avoiding solutions that may cause negative pressures in the system. The optimization problem can be formulated as follows:

$$\begin{gathered} Search\,for\;X = [L_i, C_i];\; L_i\; \in N \\ Minimize\;F(x) \\ Subject\;to\;P_i > 0 \end{gathered} \tag{1}$$

where L_i and C_i are the node index and emitter coefficient value for leakage in location i, N is the vector containing the indexes of all nodes in the DMA, and P_i is the pressure at location i (for the leak scenario).

Finally, the objective function $F(X)$ is defined as the weighted sum of squared differences between real and simulated values for both pressure and flowrate at the measurement nodes (assuming a leak in location L_i and size C_i) as follows:

$$F(X) = \sum_{k=1}^{T} \left[\sum_{s=1}^{N_p} \frac{\left(\frac{P_{sensor\ s}^{real}(k) - P_{sensor\ s}^{sim}(k)}{P_{pnt}}\right)^2}{N_p} + \sum_{q=1}^{N_q} \frac{\left(\frac{Q_{sensor\ q}^{real}(k) - Q_{sensor\ q}^{sim}(k)}{Q_{pnt}}\right)^2}{N_q} \right] \tag{2}$$

where N_p and N_q are the number of pressure and flowrate sensors, respectively; $P_{sensor\ s}^{real}(k)$ and $P_{sensor\ s}^{sim}(k)$ are, respectively, the real and simulated (with the leak in location L_i and size C_i) pressure values in sensor s with respect to time k; $Q_{sensor\ q}^{real}(k)$ and $Q_{sensor\ q}^{sim}$ are the real and simulated flowrate values in sensor q with respect to time k, respectively, and T is the total number of time steps considered in the analysis.

P_{pnt} is the average pressure measurement error (m) and is based on the average observed pressure values at sensor nodes and the pressure sensor reading percentage error as indicated by the manufacturer. Similarly, Q_{pnt} is the average flowrate measurement error (m^3/h) as is based on the average global system demand and the flow reading percentage error.

3 APPLICATION: RESULTS AND DISCUSSION

3.1 Case study description

The case study is a WDN located in Porto Metropolitan Area, Portugal, with an overall network extension of around 37 km. The WDN is supplied by a single water tank and is divided into seven DMA. Pressure and flowrate measurements were collected through 14 sensors, installed in the inlets and outlets of the DMA, during the period between 1st June 2020 and 26th December 2021 with a 15-minute time step. The WDN is located in a hill, resulting in significant elevation differences between the water source (at 162 m) and the lowest point supply location (at 12 m). As such, pressure-reducing valves (PRV) are installed at different locations in the WDN to reduce excessive pressures. The hydraulic simulation model was developed in EPANET and includes one storage tank, 1,058 node junctions, 344 valves, and 728 pipes. Figure 1 presents a schematic representation of the WDN with the location of measuring points (MP) and the PRV in relation to the DMA inlet and outlets. The number of service connections of each DMA is also presented.

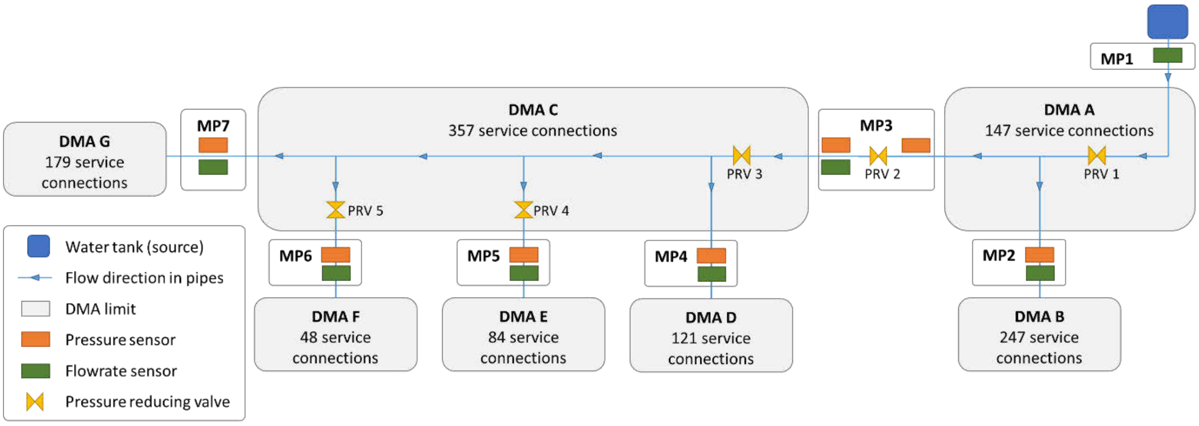

MP1
DMA A
147 service connections
PRV 1
MP2
DMA B
247 service connections
MP3
PRV 2
DMA C
357 service connections
PRV 3
PRV 4
PRV 5
MP4
DMA D
121 service connections
MP5
DMA E
84 service connections
MP6
DMA F
48 service connections
MP7
DMA G
179 service connections
Water tank (source)
Flow direction in pipes
DMA limit
Pressure sensor
Flowrate sensor
Pressure reducing valve

Figure 1. WDN scheme with the location of MP and PRV concerning the DMA inlet and outlets.

Figure 2 depicts the data availability for each of the sensors between 1st June 2020 and 26th December 2021. Although 14 sensors are installed in the network, there is not a single period where data from all the sensors is simultaneously available. The missing data are due to the fact that some sensors had not yet been installed (e.g., MP6 was only installed in September 2021) and due to sensor malfunction (in the remaining sensors with missing data). Based on the data availability, three distinct periods (of five weekdays) are selected for analysis: one week in august 2020 (period P1), one week in March 2021 (P2) and one week in December 2021 (P3).

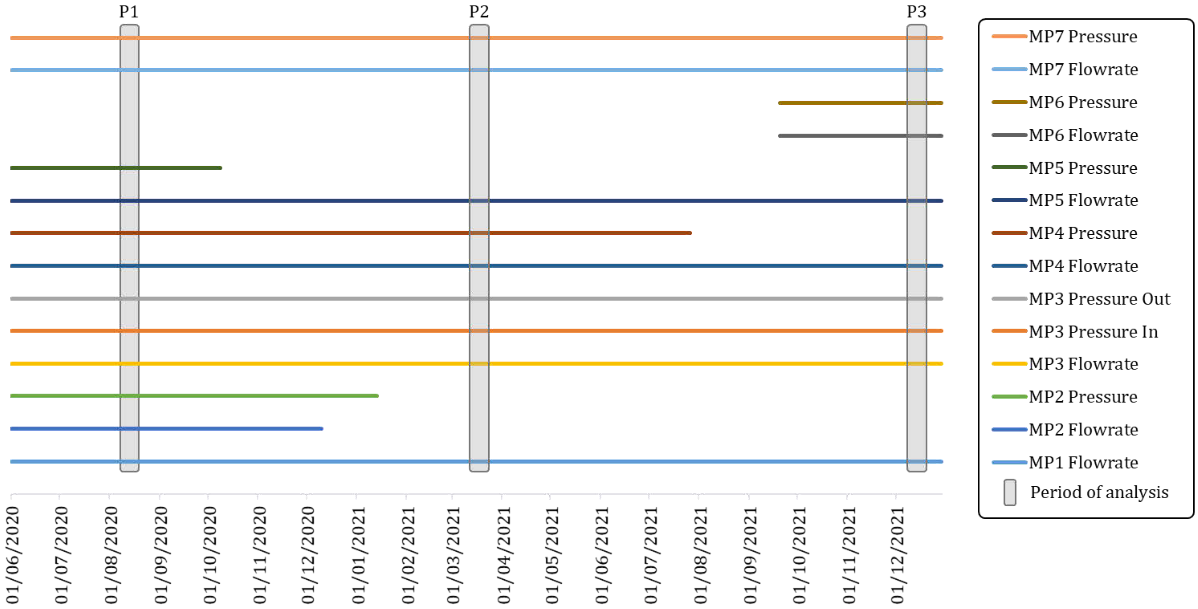


Figure 2. Data availability for the 14 sensors in the period from June 2020 to December 2021 (in horizontal bars) and selected periods for analysis (in vertical bars).

3.2 Identification of the DMA with the highest leakage volume

Leak detection is carried out based on the calculation of metrics M1 and M2 (as explained in 2.1) for each DMA. The metrics are calculated for every weekday of the three periods of analysis and the average metric values are computed in each of the three periods. This calculation requires the consumption values (in each time step) for each DMA to be known. This consumption is directly linked to the inlet flowrate measurements in *downstream* DMA (i.e., DMA B, D, E, F and G). However, in *intermediate* DMA (such as DMA A and C), this consumption is obtained by a water balance between inlets and outlets. For instance, the consumption of DMA C is obtained by a water

CC BY NC SA

balance between the inlet (MP3) and the four outlets (MP4 to MP7). A similar process is applied to DMA A between the inlet (MP1) and the two outlets (MP2 and MP3).

Although the water balance is easily calculated, not all inlet/outlet values were available. This compromises the balance calculation of cascade DMA, since all the unaccounted demand for DMA with missing data is assumed as leakage. In these situations, an estimation of DMA consumption (in each time step) is carried out based on the average consumption per service connection in the remaining DMA.

Table 1 presents the average metric results (for the five weekdays) for each DMA in each period. The highest value of each column is highlighted in bold. DMA A presents no value for the first metric (in either period). The consumption of this area is calculated based on a water balance using data from sensors with different characteristics and measurement errors. Also, overestimation of water consumption for certain DMA might occur. As such, negative values for water consumption in this area are obtained (especially during the night period). In these time steps, the value of zero consumption is assumed, hence the value for the first metric. By assessing the first metric value, it is possible to conclude that leaks are present in several DMA. Nonetheless, DMA C is likely to have a major leak. Note that it is unlikely that, on average, each service connection (of the 357) in DMA C consumes more than double the amount of any service connection of the remaining DMA during MNF, regardless of the period of analysis. This is confirmed by the results of the metric M2, as it is unlikely that the average daily consumption of service connection in DMA C is more than double than in the remaining DMA. A more plausible explanation for the metric values in DMA C might be the existence of a major unknown leak/burst, which results in additional water consumption, both during the night and throughout the day.

Table 1. Average M1 and M2 values for each DMA during each period

DMA	SC	Week in August 2020 (P1)		Week in March 2021 (P2)		Week in December 2021(P3)	
		M1 [m³/(h.SC)]	M2 [L/(day.SC)]	M1 [m³/(h.SC)]	M2 [L/(day.SC)]	M1 [m³/(h.SC)]	M2 [L/(day.SC)]
DMA A	147	0	425	0	220	0	308
DMA B	247	0.0016	273	0.0031	300	0.0045	291
DMA C	357	**0.0157**	**610**	**0.0148**	**595**	**0.0177**	**646**
DMA D	121	0.0006	286	0.0003	232	0.0019	233
DMA E	84	0.0074	413	0.005	353	0.0035	271
DMA F	48	0.0035	334	0.0031	300	0.0068	334
DMA G	179	0.0032	366	0.0034	316	0.005	326

The estimation of leak/burst size in DMA C is carried out as described in 2.1. The leak size is estimated for each day of the three period as follows: 1) the average M1 values for DMAs (with the exception of DMA A and C) is calculated; 2) this value is multiplied by 357 (the number of service connections in DMA C) and a new MNF (without leakage) for DMA C is obtained; 3) the leak size is estimated by subtracting the new MNF (without leakage) from the real MNF.

Table 2 presents the average leak size (for the 5 weekdays) for each DMA in each period. Based on these results, the leak is estimated on 4 m³/h. The two exceptions (in period P2) can be associated with the overestimation of consumption values.

Table 1. Size of leakage in DMA C in distinct periods

	Leakage size $[m^3/h]$				
	Monday	Tuesday	Wednesday	Thursday	Friday
Week in August 2020 (P1)	4.64	4.51	4.54	4.26	4.33
Week in March 2021 (P2)	3.98	4.88	3.08	4.79	4.36
Week in December 2021(P3)	4.73	4.68	4.80	4.69	4.94

3.3 Hydraulic simulation model update

A total of 15 days (divided in 3 periods) are considered for the leak location analysis. As such, a total of 15 hydraulic simulation models (one for each day) are developed. In each model, two major update processes are carried out, namely, the water consumption of DMA and the settings of PRV.

The water consumption of each DMA is updated through a consumption pattern. Note that the base demand of each node (in each DMA) represents the proportion of the consumption of that specific node in relation to the overall DMA. Furthermore, the consumption of DMA C (where the leak is located) should be updated without the leakage effect. This is done by subtracting the leak size (as previously calculated) from the DMA C consumption in each time step. Figure 3 depicts the consumption pattern for each DMA on Monday, 13th December 2021, including the pattern for DMA C with (in dashed red line) and without (in solid red line) the major leak effect.

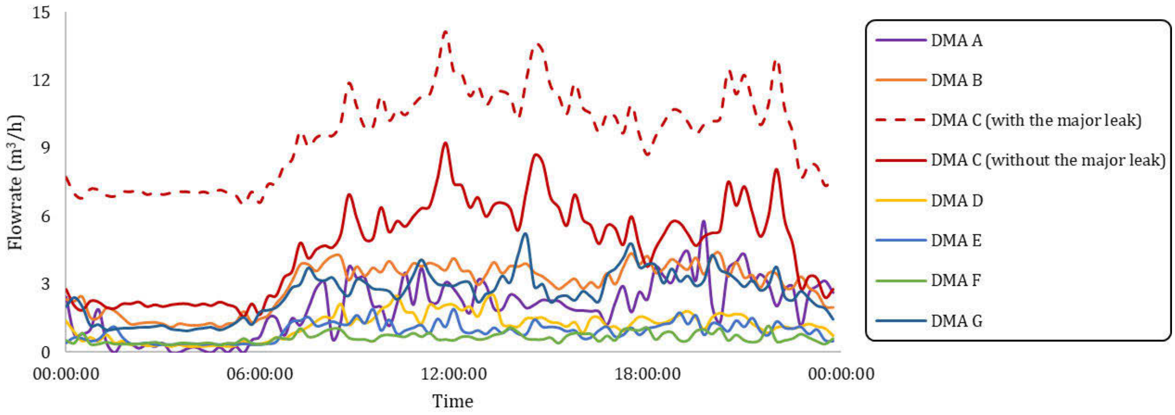


Figure 3. Consumption pattern for each DMA. The pattern for DMA C is depicted both with leakage (in dashed red line) and without leakage (in solid red line).

The second update process focuses on boundary conditions through controls of PRV settings. This is essential as PRV, although having a fixed service pressure value, often fluctuate around that value. Furthermore, time-controlled actions are often associated with such devices (e.g., for reduction of pressure during night hours) which must be accurately translated into the hydraulic simulation model. In this study, controls are created for PRVs 2, 4, and 5 based on the available measurements of PRV outlet pressure. Figure 4 exemplifies the update process of PRV 2, and its effect on MP7 (which is located further downstream). The default setting for PRV 2 is 35 m, as can be observed by the orange line. The value of 35 m may be adequate for the day period, as observed by the proximity of the orange line to the dark blue line. However, it is not adequate for the night period as these lines differ by approximately 10 m. Furthermore, the setup of PRV 2 directly influences the MP7; by considering a fixed set point of 35 m in PRV 2, an almost static pressure value of 110 m is observed in MP7 (in red line). By assigning the measured PRV 2 outlet pressure values (in dark blue line) to controls in the hydraulic simulation model, it is possible to obtain similar valve settings (as shown by the overlap of dark blue and dashed green lines). As a result, simulated and real measurements in MP7 (in dashed grey and green lines) vary just slightly; simulated pressures are in fact higher as a lower overall water consumption (since the leakage is not modelled) is used. Lastly, real and simulated measurements in the inlet of PRV 2 differ greatly. Although modelling error might be present, the real pressure measurements values (around 80 m) are above the maximum possible value of 68 m (considering that the reservoir is located at 162 m of elevation, and the MP3 is located at 94 m). This indicates that a problem might exist with this sensor (either lacking calibration or within the measurement system itself) which, effectively, invalidates its usability.

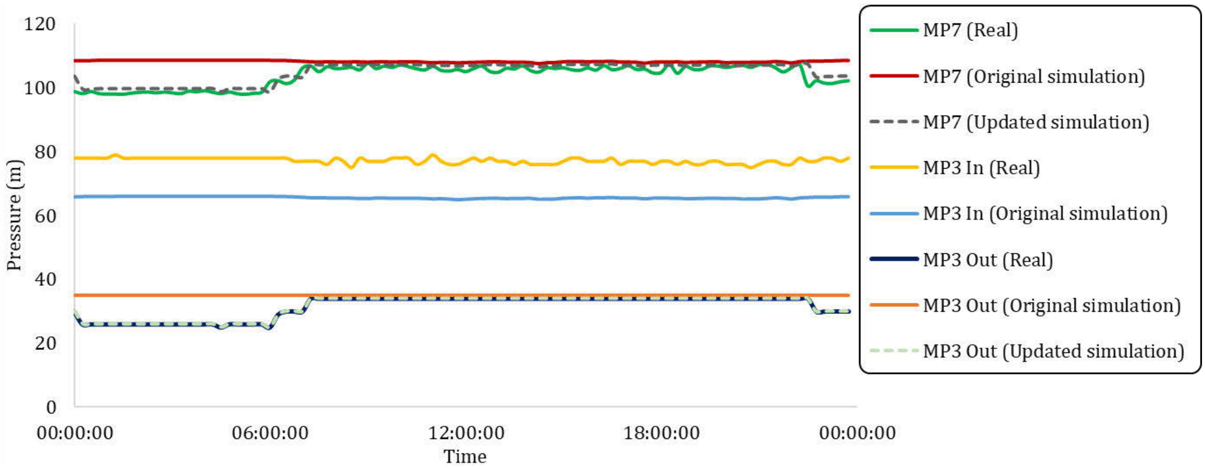


Figure 4. Effect of updating PRV 2 (with values MP3 Out) in the MP7

Although two additional PRV can be found in the network (PRV 1 and 3), they are not easily calibrated through the assignment of controls as no pressure sensor is placed downstream of those valves. As such, the fixed setting is used as provided by the water utility.

3.4 Leak location

Leak location is carried out as described in 2.3 for the night period in each day of each period (with a total number of 15 optimization runs). A specific hydraulic simulation model was prepared for each day (as described in 3.3). Hydraulic simulations are carried out in EPANET [22] by using the Water Network Tool for Resilience (WNTR) Python package [23]. The inverse problem is solved using Genetic Algorithms [21] with Pymoo Python package [24]. In problem formulation, all nodes of DMA C are considered as potential leak locations.

Three pressure sensors (MP2, MP4, and MP7) and two flowmeters (MP1 and MP3) are considered in the analysis for period P1. Due to data availability, two pressure sensors (MP4 and MP7) and two flowmeters (MP1 and MP3) are considered for the analysis in period P2. For period P3, a single pressure sensor (MP7) and two flowmeters (MP1 and MP3) are used. Many of the optimization problems lead to similar locations and, thus, from the total of 15 problems, only 3 distinct locations were obtained. Such locations are depicted in Figure 5 in blue circular markers and are grouped in a specific zone in DMA C (more specifically near the outlet to DMA G).

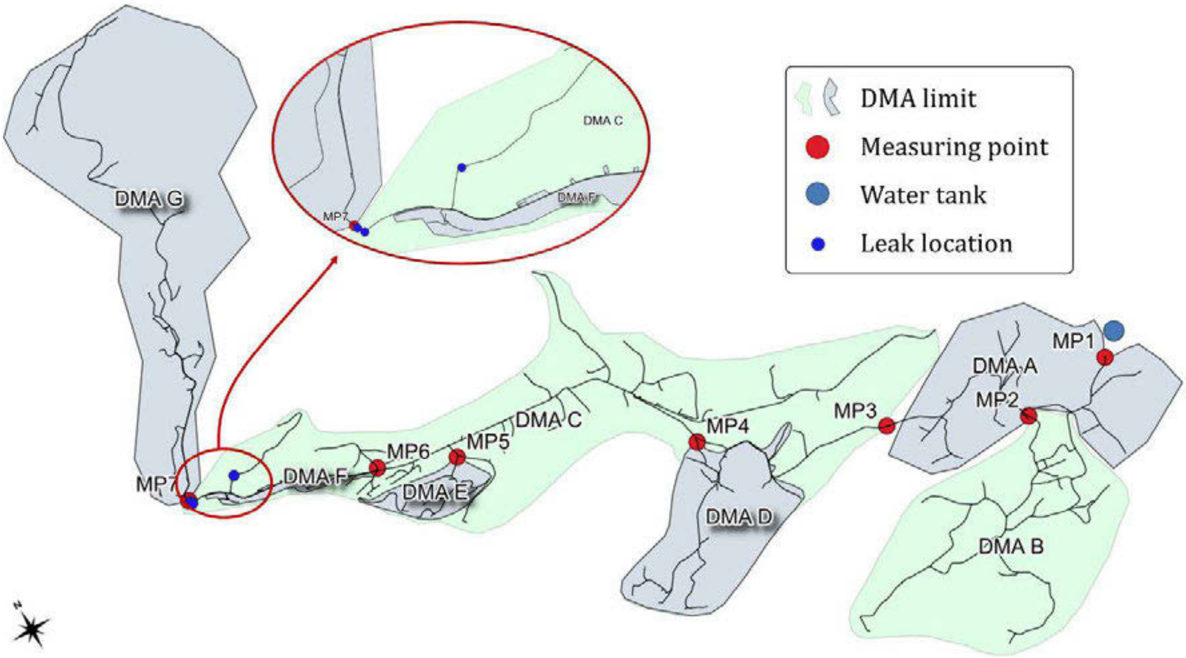

DMA limit
Measuring point
Water tank
Leak location
DMA G
DMA C
MP7
DMA F
MP1
DMA A
MP2
MP3
MP4
DMA C
MP5
MP6
DMA F
DMA E
MP7
DMA D
DMA B
N

Figure 5. Obtained leak locations

Figure 6 depicts the simulated (after optimization) and real measurements for Monday, 13th December 2021. Although the optimization is carried out only for the night period, the simulated and real measurements are presented for the complete day. These results show that, after the optimization, real and simulated flowrate measurements (in MP1 and MP3) overlap almost perfectly. Pressure measurements (in MP4 and MP7) present small differences, most notably in MP7 during the night period. Such differences may be associated with additional head loss effects that are not considered in the analysis, for instance, unknown valve status, which induce local head losses.

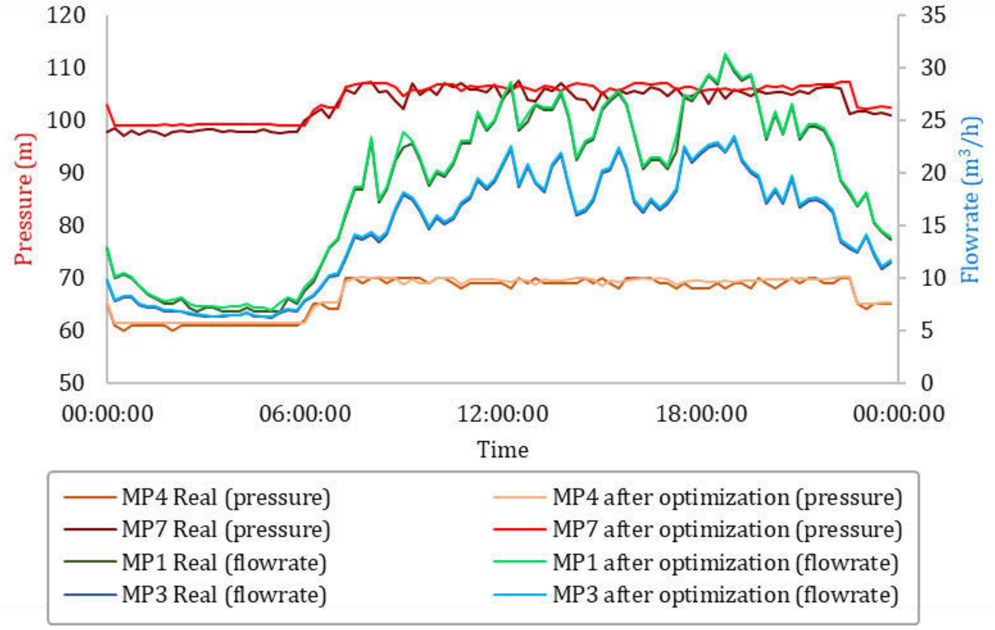


Figure 6. Comparison between simulated and real measurements after leakage location

By assessing the results of the three period, and based on the available sensors and data, it is possible to conclude that the leak may be located at downstream the DMA C. Note that the number and location of pressure and flowrate sensors were established in order to monitor DMAs both in terms of water consumption and pressure requirements and not for leak location. Although such sensors can be used for leak location (as shown), their location is not optimized for that objective. This is a reality in many Portuguese water utilities, as investment in sensors is carried out mainly to monitor water consumption and inlet pressure of DMA. Ideally (and in addition to the already installed sensors), an optimal number of pressure sensors should be established, for instance, based on the analysis of the hypervolume indicator [25], and whose location is optimally defined to carry out leak detection and location [26].

4 CONCLUSIONS

This paper presents the practical application of model-based methodology for leak detection and location in a real WDN. The methodology is based on three main steps, namely: 1) identification of the DMA with the highest leakage volume; 2) hydraulic simulation model update (specifically water consumption and valve settings) and 3) leak location (by formulating and solving an inverse problem).

The WDN is monitored through 14 (flowmeter and pressure) sensors installed in the inlet and outlets of the seven DMA. Data are available over the period between the 1st of June 2020 and the 26th of December 2021. The methodology is applied 15 times, specifically to five weekdays in three periods (August 2020, March 2021, and December 2021). Results from the first step indicate that, regardless of the period being assessed, a leak with 4 m^3/h does exist in DMA C. In the second step, 15 hydraulic simulation models are developed by updating the real water consumptions (occurring during that period) and by creating valve controls (based on the measured PRV outlet pressure). Finally, 15 leakage locations are obtained in the third step by solving 15 optimization problems. Obtained leak locations point to a specific zone in the DMA C and should be treated as the starting point for the application of more precise leak location techniques in situ (e.g., acoustic equipment).

This study also demonstrates the difficulties of implementing model-based techniques for leak location in real-life contexts. Although large capital sums are already invested in network monitoring of water consumption and service pressure by water utilities, the installed sensors are often insufficient and non-optimally located. This is a reality in many Portuguese water utilities, which require further investment in network monitoring so that automatic model-based techniques for leak location can become a common practice in Portuguese water utilities.

5 ACKNOWLEDGMENT

The authors would like to thank the Fundação para a Ciência e Tecnologia (FCT) for funding this research through the WISDom project (grant number DSAIPA/DS/0089/2018) and Ph.D. Research Studentship of Bruno Ferreira (grant number SFRH/BD/149392/2019). The authors also thank the engineers from the water utility for presenting us with such a challenging and rewarding problem.

6 REFERENCES

[1] Fontanazza, C.M., Notaro, V., Puleo, V., Nicolosi, P., and Freni, G. (2015). Contaminant intrusion through leaks in water distribution system: Experimental analysis. Procedia Engineering 119, 426–433. doi:10.1016/j.proeng.2015.08.904.

[2] Farah, E., and Shahrour, I. (2017). Leakage Detection Using Smart Water System: Combination of Water Balance and Automated Minimum Night Flow. Water Resources Management 31, 4821–4833. doi:10.1007/s11269-017-1780-9.

[3] Covas, D.I.C., Cláudia Jacob, A., and Ramos, H.M. (2008). Water losses' assessment in an urban water network. Water Practice and Technology 3, 1–9. doi:10.2166/wpt.2008.061.

[4] Li, R., Huang, H., Xin, K., and Tao, T. (2015). A review of methods for burst/leakage detection and location in water distribution systems. Water Science and Technology: Water Supply 15, 429–441. doi:10.2166/ws.2014.131.

[5] Moors, J., Scholten, L., van der Hoek, J.P., and den Besten, J. (2018). Automated leak localization performance without detailed demand distribution data. Urban Water Journal 15, 116–123. doi:10.1080/1573062X.2017.1414272.

[6] Loureiro, D., Amado, C., Martins, A., Vitorino, D., Mamade, A., and Coelho, S.T. (2016). Water distribution systems flow monitoring and anomalous event detection: A practical approach. Urban Water Journal 13, 242–252. doi:10.1080/1573062X.2014.988733.

[7] Gomes, S.C., Vinga, S., and Henriques, R. (2021). Spatiotemporal Correlation Feature Spaces to Support Anomaly Detection in Water Distribution Networks. Water 13, 2551. doi:10.3390/w13182551.

[8] Romano, M., Kapelan, Z., and Savić, D.A. (2014). Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems. Journal of Water Resources Planning and Management 140, 457–467. doi:10.1061/(ASCE)WR.1943-5452.0000339.

[9] Wu, Y., and Liu, S. (2017). A review of data-driven approaches for burst detection in water distribution systems. Urban Water Journal 14, 972–983. doi:10.1080/1573062X.2017.1279191.

[10] Carriço, N., Ferreira, B., Barreira, R., Antunes, A., Grueau, C., Mendes, A., Covas, D., Monteiro, L., Santos, J., and Brito, I.S. (2020). Data integration for infrastructure asset management in small to medium-sized water utilities. Water Science and Technology 82, 2737–2744. doi:10.2166/wst.2020.377.

[11] Pudar, R.S., and Liggett, J.A. (1992). Leaks in pipe networks. Journal of Hydraulic Engineering 118, 1031–1046. doi:10.1061/(ASCE)0733-9429(1992)118:7(1031).

[12] Wu, Z.Y., Sage, P., and Turtle, D. (2010). Pressure-Dependent Leak Detection Model and Its Application to a District Water System. Journal of Water Resources Planning and Management 136, 116–128. doi:10.1061/(ASCE)0733-9496(2010)136:1(116).

[13] Sophocleous, S., Savić, D., and Kapelan, Z. (2019). Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction. Journal of Water Resources Planning and Management 145, 04019024. doi:10.1061/(ASCE)WR.1943-5452.0001079.

[14] Casillas Ponce, M. V., Garza Castañón, L.E., and Cayuela, V.P. (2014). Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. Journal of Hydroinformatics 16, 649–670. doi:10.2166/hydro.2013.019.

[15] Geng, Z., Hu, X., Han, Y., and Zhong, Y. (2019). A Novel Leakage-Detection Method Based on Sensitivity Matrix of Pipe Flow: Case Study of Water Distribution Systems. Journal of Water Resources Planning and Management 145, 04018094. doi:10.1061/(ASCE)WR.1943-5452.0001025.

[16] Goulet, J.-A., Coutu, S., and Smith, I.F.C. (2013). Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Advanced Engineering Informatics 27, 261–269. doi:10.1016/j.aei.2013.01.001.

[17] Moser, G., Paal, S.G., and Smith, I.F.C. (2018). Leak Detection of Water Supply Networks Using Error-Domain Model Falsification. Journal of Computing in Civil Engineering 32, 04017077. doi:10.1061/(ASCE)CP.1943-5487.0000729.

[18] Capelo, M., Brentan, B., Monteiro, L., and Covas, D. (2021). Near–Real Time Burst Location and Sizing in Water Distribution Systems Using Artificial Neural Networks. Water 13, 1841. doi:10.3390/w13131841.

[19] Zhang, Q., Wu, Z.Y., Zhao, M., Qi, J., Huang, Y., and Zhao, H. (2016). Leakage Zone Identification in Large-Scale Water Distribution Systems Using Multiclass Support Vector Machines. Journal of Water Resources Planning and Management 142, 04016042. doi:10.1061/(ASCE)WR.1943-5452.0000661.

[20] Perez, R., Sanz, G., Puig, V., Quevedo, J., Cuguero Escofet, M.A., Nejjari, F., Meseguer, J., Cembrano, G., Mirats Tur, J.M., Sarrate, R., et al. (2014). Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in barcelona [applications of control]. IEEE Control Systems 34, 24–36. doi:10.1109/MCS.2014.2320336.

[21] Goldberg, D. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning (New York, New York, USA: Addison Wesley).

[22] Rossman, L. (2000). EPANET 2 user's manual (United States Environmental Protection Agency).

[23] Klise, K.A., Bynum, M., Moriarty, D., and Murray, R. (2017). A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. Environmental Modelling & Software 95, 420–431. doi:10.1016/j.envsoft.2017.06.022.

[24] Blank, J., and Deb, K. (2020). Pymoo: Multi-Objective Optimization in Python. IEEE Access 8, 89497–89509. doi:10.1109/ACCESS.2020.2990567.

[25] Ferreira, B., Carriço, N., and Covas, D. (2021). Optimal Number of Pressure Sensors for Real-Time Monitoring of Distribution Networks by Using the Hypervolume Indicator. Water 13, 2235. doi:10.3390/w13162235.

[26] Ferreira, B., Antunes, A., Carriço, N., and Covas, D. (2022). Multi-objective optimization of pressure sensor location for burst detection and network calibration. Computers & Chemical Engineering 162, 107826. doi:10.1016/j.compchemeng.2022.107826.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14091

DUAL ESTIMATION OF IRON OXIDE DEPOSITION ON DRINKING WATER PVC PIPES USING CALIBRATED TURBIDITY DATA AND BRIGHTFIELD MICROSCOPY IN A FULL-SCALE LABORATORY SYSTEM

Artur Sass Braga [1] and Yves Filion [2]

[1,2] Queen's University
Kingston, Ontario, (Canada)

[1] *16asb2@queensu.ca*, [2] *yves.filion@queensu.ca*

Abstract

The assessment of accumulated sediments inside drinking water pipes is an important step for determining the risk of water quality deterioration for a sector of a distribution network and for scheduling the required maintenance activities that minimize this risk. Water utilities and researchers have traditionally used turbidity data collected during flushing operations to quantify the discolouration potential in isolated pipe lengths. Flushing has an elevated cost of specialized personnel, consumes large quantities of drinking water, and offers poor information about sediment conditions prior to mobilization (e.g. structure, position on the pipes). The last problem must be overcome by gaining a better understanding of the processes driving material accumulation, which might help in the development of strategies to prevent sediment deposits. In addition, a complex relationship between turbidity and SSC also makes it difficult to accurately translate turbidity units (NTU) into physical units of concentration (e.g. mg/L). This paper aims to consolidate the macroscopic estimations of sediment deposits in drinking water pipes using turbidity data and to propose a microscopic complement that provides richer data about sediment deposits at the pipe wall. The research was developed through a controlled experiment using a full-scale PVC pipe system that mimics the operational conditions of drinking water distribution systems. In the experiments, the drinking water was amended with iron oxide particles that progressively adhered to the pipe walls during 30 days of steady flow conditioning. After the conditioning period, the pipes were flushed to mobilize the sediment deposits. The SSC of water samples collected during the experiments were used to produce translation factors for the online turbidity data. Macroscopic sediment loads were estimated based on the difference between suspended sediments at the inlet and outlet of the pipe loop, while microscopic loads were estimated through the direct observation of particles on pipe wall samples using automated brightfield microscopy. Physical metrics were proposed to adequately represent the sediment load data. Results from the turbidity data analysis produced insights about the impacts of experimental conditions on the SSC translation factors, while microscopy images allowed a detailed assessment of particles deposited on the pipe walls including information about their particle size distribution and dispersion.

Keywords
Discolouration, Iron Oxide Particles, Suspended Sediment Concentration, Brightfield Microscopy.

1 INTRODUCTION

Pipes in drinking water distribution systems (DWDS) commonly experience a chronic material loading [1] that contributes to the accumulation of several contaminants on their pipe walls, including inorganic particles, metals and viable biofilms [2]. This increases the risk of a drinking water contamination event, where a considerable fraction of accumulated materials is quickly mobilized by the bulk flow [3, 4]. This can lead to interactions between the bulk water and the pipe walls that can gradually degrade drinking water quality [5]. The case of material mobilization has historically been the focus of research due to severe water discolouration events that often

CC BY NC SA

surpass established guidelines for potable water turbidity and regulated metals. However, health-related impacts of chronic drinking water degradation are gaining interest because of the prevalence of key contaminants in material deposits (e.g., anti-microbial resistant genes [6, 7]), that have a direct path to the population through drinking water. In either case, water utilities typically have the goal of reducing accumulated material in pipes so as to reduce the risk of contamination in their drinking water systems [4, 8-10].

The first step toward effective material accumulation management is the establishment of a reliable method to quantify the material load, which is already a challenge due to several factors, including: 1) material accumulation is often localized in DWDSs [8, 10]; 2) material composition varies across networks and can include a large range of inorganic and organic elements [3, 4, 11, 12]; 3) accumulation and mobilization phenomena occur over short time scales [11, 13]; 4) the majority of pipes are buried, and a commercialized solution to easily access material deposits is not currently available. Due to these factors, previous approaches to quantify material load in DWDSs have been restricted to the analysis of specific drinking water elements (e.g., turbidity, iron, viable microorganisms), and not fully interchangeable among systems. Historically, the investigation of material accumulation is connected to events of rapid material mobilization commonly known as *discolouration events* [13], where the increase of pipe flow rate provokes the mobilization of materials from the pipe wall. The increase of particulate material in suspension produces the “discoloured” aspect of the water, as a consequence of a decrease in the translucid quality of the water rather than a change in colour. Therefore, the measurement of water turbidity – a common water quality parameter that measures the level of light scatter of fine particles in suspension in *Nephelometric Turbidity Units* (NTU) relative to an established standard in a reference solution [14] – has emerged as the main parameter to quantify material accumulation loads in DWDSs.

Previous research has been focused on turbidity data from flushing operations that were intentionally produced to mobilize materials and clean pipes, from field and laboratory DWDSs [11, 15-18]. These included approaches that considered a pseudo-mass balance of turbidity to estimate material accumulation load in selected systems [19], and the development of discolouration models based on the conservation of turbidity for a wide range of systems [20-22]. Because turbidity is an aesthetic aspect of drinking water, it is also strongly related to customer perception of drinking water quality—something that has motivated its adoption as a key parameter in previous research focused on improving customer satisfaction [15, 16]. Turbidity-based approaches in discolouration research are mostly founded on the argument that the correlation between turbidity and suspended sediment concentration (SSC) remains stable with a consistent particle size distribution (PSD) and particle composition [22, 23]—two parameters that determine their interaction with light. By comparison, other researchers have shown that these conditions are not commonly met in DWDSs, and that the relationship between SSC and turbidity can change substantially among systems and during abrupt changes in operational conditions [11, 24, 25]. Recent research has shown that drinking water PSD is directly connected to the fluid velocity in the pipes and might rapidly change due to the selection of particles that remain in suspension for individual operational conditions—a factor that may drastically alter the relationship between turbidity and SSC [25].

Researchers have also argued that turbidity rather than material load and SSC is a better indicator to manage material accumulation because it directly translates to customer satisfaction [1, 15]. However, drinking water turbidity does not necessarily correspond to water quality, since contaminants might still be present in dissolved form (originating from accumulated materials), or as particles that only make a small contribution to turbidity. Larger particles in the range of 50 μm – 500 μm produce little turbidity due to their low abundance in drinking water [22], but they have an equivalent mass of thousands of fine particles (1 μm – 10 μm). These large particles might still be mobilized during elevated flow events, and it is reasonable to assume that they have

BY NC SA

a greater potential to harbour higher contaminant loads than do fine particles while still being small enough to be undetected by customers. In this context, the customer's perception of drinking water quality is biased and insufficient to fully evaluate the risks of contamination events from pipe wall material deposits. Meanwhile, several aspects of the material accumulation problem are still poorly understood, and these are still under-represented by current drinking water quality assessment practices. Nonetheless, there is no proof of a conservation of water turbidity since the changes of PSD are independent of changes of SSC [26]—something that may affect the reliability of long-term models that only consider water turbidity to forecast material accumulation.

The lack of direct assessment of pipe wall conditions and the indirect nature of data acquired through flushing and material mobilization also creates additional challenges to fully understand and quantify material accumulation on drinking water pipes. Turbidity approaches use a short mobilization period (e.g., minutes) at the end of long accumulation periods (e.g., months) to detect a final time-average of material load. In addition, during flushing activities the period of material mobilization from pipe walls is proportional to the flow acceleration period [27] while the water is travelling downstream in the pipes. This produces a mixing of released material from long pipe sections into the same control volume. This tends to decrease the accuracy of material deposits localized through turbidity data and prevents the investigation of distinct attachment along the pipe circumference. By comparison, the processes that account for material accumulation occur over long-time scales (e.g., years) and occur at a microscopic scale at the pipe wall. For this reason, the investigation of the detailed mechanics of material accumulation requires methods of observation of similar time and space scales. Figure 1 contrasts the time and space scales of the processes of material accumulation and mobilization.

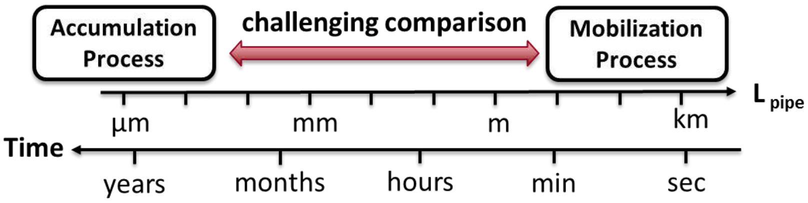


Figure 1 – Comparison of length and time scales of material accumulation and mobilization processes in drinking water pipes.

In this context, previous research has proposed direct methods for assessing biofilms in DWDSs [28, 29] supported by established methods in microbiology that are traditionally used to investigate microorganisms. However, only recent approaches have allowed the direct quantification of inorganic particles on the pipe walls through acquisition of pipe wall samples and direct observation using brightfield microscopy [30]. The method was specifically developed to identify iron oxide particles (the dominant particle that causes in discolouration events [3, 12]) for PVC pipes in a controlled full-scale water distribution laboratory. Using this setup, this paper is focused on comparing two independent pipe wall material quantification approaches based on turbidity and microscopy observations. A controlled experiment was developed in a full-scale PVC laboratory system that partially mimics the conditions of operational DWDSs. The main simplifications of the study were the induced absence of biofilms and the use of stable and homogeneous iron oxide particles in suspension. These simplifications were adopted to facilitate the estimation of particulate deposits in physical mass units.

2 METHODS

2.1 Laboratory experiments

The experiments were realized in the Drinking Water Distribution Laboratory (DWDL) at Queen's University. The DWDL can mimic the hydraulic conditions of real distribution systems in a controlled environment. The laboratory has two independent pipe loops, that are comprised of a water tank with a volume of 3.6 m^3, two variable-speed centrifugal pumps (low-flow and high-flow) and 11 coils of IPEX Blue Brute PVC pipe Class 235 (DR18) with an internal diameter 108 mm, and a total length of 193 m. Prior to each experiment, the tanks were cleaned, and the entire system (tanks and loops) was disinfected using sodium hypochlorite at a free chlorine concentration of 20 mg/L with a contact time of 24 h. A unidirectional flushing flow rate of 16 L/s was used to mobilize any remaining material deposits accumulated in the pipes, and the flushed water was discarded and replaced by local drinking water from the City of Kingston. Several water samples were collected at the tap during the experiments and tests showed an average SSC of 60 µg/L with a corresponding turbidity of 0.08 NTU.

The experiment consisted of a 28-day conditioning phase where a fixed volume of water was recirculated at a constant flow rate between the pipe loop and the tank, and a flushing phase where the pipe loop was flushed in 3 sequential steps [31-33]. The hydraulic conditions for each experimental stage are presented in Table 1. A steady conditioning flow rate of 0.6 $L\ s^{-1}$ was chosen based on the average flow rate measured in distribution water mains of Canadian DWDSs. The steady flow was used to simplify the interpretation of the particle attachment phenomenon on the pipe walls. The use of a closed system volume (no water renewal) was chosen to make it possible to perform a mass balance on the material entering and leaving the pipe loop. In doing so, it was possible to quantify the amount of particulate mass accumulated in the pipe loop and the amount of mass mobilized in each flushing stage. The closed volume also assisted in preventing biofilm formation due the limited source of nutrients in the initial drinking water. In addition, a free chlorine concentration of approximately 0.2 mg/L was maintained through weekly addition of sodium hypochlorite to the system based on weekly monitoring and an empirical decay curve.

The experiment was completed in duplicate using the North and South pipe loops, where the water was inoculated with 1 $mg\ L^{-1}$ of particulate iron oxides to encourage the formation of iron oxide deposits on the pipe wall. The SSC value is substantially higher than common values for drinking water but chosen to amplify the material attachment phenomena and facilitate its detection. During the experiments, both pipe loops were subject to the same hydraulic, environmental, and physico-chemical conditions to obtain a duplicate set of data. A chemical grade of red iron oxide powder from Alpha Chemicals, with a composition of 82% of Iron(III) oxide – Fe_2O_3 was chosen to be the source of particles for the experiments due to its stable particulate form and insolubility in water. The powder particle size distribution analysis through laser diffraction showed a predominance of small sizes, with 10% of particles smaller than 0.8 µm, a median particle diameter (D_{50}) of 6.3 µm, and 20% of particles larger than 10.8 µm. A stable water conductivity and pH measured through the experiments suggested that little changes in the water chemistry occurred during all the experiments.

To start the experiment, the iron oxide particles were added as a single load into the tanks. This method was preferred to a progressive addition of particles mostly due to the practical challenges of maintaining a consistent concentration in the system due to the lack of control of particles mixing in the tanks. After the inoculation, the pipe loops were operated at a steady, conditioning flow rate (Table 1) for 28 days to allow particles to adhere to the pipe walls. Following the conditioning phase and prior to the flushing phase, the tank and the pipes on the suction and discharge sides of the pumps were pre-flushed to remove whatever iron oxide particles may have accumulated in these sections and to prevent reintroducing materials from these components into the pipe loops during the flushing steps of the experiment. In the flushing stage, three unidirectional flushing steps were performed in sequence in each pipe loop to mobilize material deposits from the pipe walls at different WSS levels (Table 1). During the flushing steps, the pipe loops were operated in a non-recirculatory manner, where the water was discarded at the outlet of the pipe loop instead of re-entering the tank.

During the experiments, the flow rate and turbidity were continuously monitored at 1 Hz in each pipe loop with two Sierra InnovaSonic® 205i Ultrasonic Flow Meters (±0.5%) and two online Hach TU5300sc turbidimeters (±0.01 NTU). The turbidimeters were configured to monitor the water at the pipe loop inlet and at the pipe loop outlet. The TU5300sc turbidimeters were specifically chosen for use in the DWDL due to their small flow cell of only 10 mL. However, because the sensor uses a "flow-through approach" and required the extraction of the water from the system, a special installation scheme was needed to achieve a reliable operation and prevent the depletion of the finite water of the system (Figure 2a). First, unique plastic sampling ports that collect water from the centreline of the pipes was developed using 3D printing technology (Figure 2b). The aim of this bespoke sampling device was to achieve an approximate isokinetic sampling condition – same fluid velocity entrance at the collection tubing as the local flow velocity – to reduce sampling errors associated with the inertia of the suspended particles. Second, a modification of the sensor operation was realized by the addition of a small plastic centrifugal pump in the tubing downstream of the sensor location that pumped the water back into the pipe loop (Figure 2a). These modifications guaranteed that the required flow rate for the sensor operation was consistent and maintained regardless of the hydraulic conditions in the pipes. In addition, the installation of the small centrifugal pump downstream of the sensor prevented possible impacts of this pump to the sensor readings. Several preliminary tests using multiple flow configurations and controlled concentrations of similar iron oxide particles demonstrated that turbidity readings were very sensitive to hydraulic conditions of the sensor tubing. These justified the substantial efforts placed to guarantee a consistent operation of the turbidity sensors during the experiments.

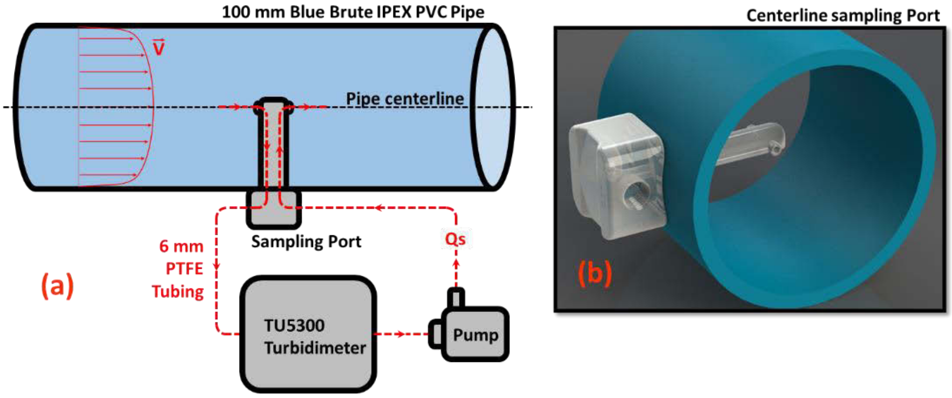


Figure 2 – a) Special scheme for the installation of turbidimeters in the pipe loop; b) rendered model of the 3D printed sampling ports to collect/return water at the centreline of the pipes.

A similar approach was used to collect grab water samples from the centreline of pipes during the experiments for measuring the bulk water SSC. 3D printed sampling ports designed to maintain isokinetic conditions were produced for low- and high-sampling flow rates. Low-flow ports were installed at the inlet and outlet positions of the pipe loop, while a single high-flow sampling port was installed at the outlet only. The low-flow ports were mainly used to collect samples during the conditioning phase of the experiments under stable conditions. By contrast, the high-flow sampling ports were used to collect large sample volumes during short periods of the flushing phases under potentially variable SSC conditions.

A unique pipe wall coupon sampling system [30] was used to directly monitor material deposits on the internal surface of the PVC pipes. The coupon sampling system was developed by cutting out circular PVC coupons from the native PVC pipe and mounting them on a 3D-printed coupon support base. The circular PVC coupon and support base were designed to be inserted in custom, pre-drilled holes and affixed with a metal clamp to prevent leakage when the pipe loop was pressurized. The coupon system allows for a precise surface alignment of the internal surface of the coupon with the surrounding pipe wall surface that is in the range of ± 0.1 mm. This guarantees that the internal surface of the coupon will experience the same flow conditions as does the surrounding pipe wall without disturbing the local velocity profile in the viscous sub-layer region (VSL). Only coupons installed in the invert positions were used for the material quantification, since no particles were found in the obvert and springline positions [34]. During the conditioning phase of the experiments, pipe wall coupons in the invert pipe position were acquired every 7 days to assess the pipe wall conditions. During the flushing phase, invert pipe wall coupon samples were acquired between the flushing steps to assess the pipe wall conditions.

2.2 Material estimation based on turbidity data

The continuous, high-frequency monitoring of turbidity at the inlet and the outlet of the pipe loops was used to estimate changes in the SSC of the bulk water that was sequentially used to calculate the transport of iron oxide particles from the bulk water to the system pipe wall and tank. For this, turbidity data was transformed to continuous SSC data through calibration coefficients (κ (mg L^{-1} NTU^{-1})) obtained from valid SSC samples that were also complemented by previous experimental data under similar conditions [34]. Turbidity values for the calibration were obtained directly from the pipe loop – from the correspondent control volume from where the SSC samples were collected – rather than values from portable turbidimeters. For the flushing stage of the experiment, SSC sample data was regressed against turbidity data to estimate a constant calibration coefficient. However, for the conditioning stage of the experiment, a decrease of the calibration coefficient was observed over time. To cope with this, a *calibration coefficient curve* (κ (t)) that varied as function of time was used to transform the turbidity data to SSC values. The calibration coefficients were used to transform the turbidity data series (T(NTU)) to SSC (C(mg L^{-1})) using Equation 1.

$$C = \kappa \times T \tag{1}$$

Following this, the flux of suspended sediments (Q_{SS} (mg s^{-1})) through the turbidity monitoring sections was calculated using Equation 2 as the product of the measured flow rate in the pipes (Q (L s^{-1})) and the SSC (C(mg L^{-1})). The total accumulated load of iron oxide particles attached to the pipe wall (M_{PW}(mg)) in each experimental stage was estimated by integrating the difference in suspended material flux at the inlet and outlet of the pipe loop length between the turbidimeters over time (Equation 3).

$$Q_{SS} = Q \times C \tag{2}$$

$$M_{PW} = \int_{t} (Q_{SS,\,in} - Q_{SS,\,out})dt \tag{3}$$

where t = period of time over which suspended materials are deposited on pipe wall (seconds); $Q_{SS,\,in}$, $Q_{SS,\,out}$ = flux of suspended sediments estimated from turbidity data measured at the inlet and outlet of the pipe loop. The duration of the inoculation stage was 180 minutes (corresponded to 3 complete turn-overs of the pipe loop volume), and the conditioning stage lasted for 30 days. The period that corresponds to the passage of 3 pipe-loop volumes was used for each flushing step.

The material accumulated in the tanks (M_{Tank}(mg)) was calculated using Equation 4 as the difference between the total mass of iron oxide particles initially added (M_{Fe} (mg)), the material load estimation of particles attached to the pipe wall (M_{PW}(mg)), and the material mass that remained in suspension (M_{SSC} (mg)). The load of material in suspension (M_{SSC} (mg)) was estimated with Equation 5 by multiplying the instantaneous SSC (C (mg L^{-1})) at the inlet position by the total volume of the water (V_T (L)).

$$M_{Tank} = M_{Fe} - M_{PW} - M_{SSC} \tag{4}$$

$$M_{SSC} = C \times V_T \tag{5}$$

Equations 1 – 5 were used to estimate the accumulation of iron oxide particles on the internal surface of the pipe loops during the inoculation and conditioning phases (positive material load), and the mobilization of iron oxide particles from the pipe surface after each flushing step (negative material load). Before the pipe flushing stage, the tank and pump side of the system was pre-flushed at the maximum flow rate, and all particles that could be mobilized were discarded with the renewal of the tank water. Because of that, and due to the non-recirculating strategy adopted during the pipe flushing stage, all material load entering in the pipe loop during the flushing stage was assumed to be negligible.

2.3 Material estimation based on microscopy data

Pipe wall samples were retrieved from a middle section of the pipe loop during the conditioning and flushing phases of the experiment and were processed according to the monitoring scheme developed by the authors [35]. Pipe wall samples collected from the invert pipe positions were used in this study. All samples were collected within a pipe length of 20 m located approximately 100 m from the inlet of the pipe loop and assumed to be invariant with the pipe length. The pipe wall samples were carefully unmounted from their 3D-printed support base and air dried at 30°C for 24 h before they were imaged with a brightfield microscope. Four samples and up to 44 fields of view (FOV) of each coupon sample were captured per sampling cycle using an automated upright microscope Nikon Eclipse Ni-E in brightfield mode and a CFI60 Super Plan Fluor ELWD 40x objective lens. The final images were used to characterize the attachment of iron oxide particles on the pipe wall and their adherence to the PVC pipe substrate by means of a MATLAB image-processing script and a quantitative particle analysis. The image-processing script was used to detect individual iron oxide particles by differentiating them from the PVC pipe substrate and calculating their individual coverage area (μm^2). The multiple FOVs captured from each coupon sample were used to calculate the particle coverage area mean and standard deviation.

The microscopic material load M_{FOV} (mg) per field of view area was calculated by following a number of steps. First, the total volume of iron oxide particles detected in each FOV (V_P (μm^3)) was estimated using Equation 6 by multiplying the mean particle cross section area (A_P (μm^2)) by the average particle diameter (d_P (μm)) that was determined to be 1.5 μm from the analysis of a previously generated dataset [34]. For the total volume calculation, it was further assumed that the particles were spherical and therefore a volume reduction factor of 0.52 (ratio between the volume of a sphere imprint into a cube) was applied (Equation 6). In comparison to Braga and Filion [34], the new method proposed here aims to improve the previous calculation of particle volume through its equivalent diameter. In addition, the new approach is also supported by data from recent experiments [25], which suggests that only small particles (< 5μm) should be able to reach the coupon sections under the hydraulic conditions tested. An estimate of the average iron oxide particle mass per FOV (M_{FOV} (mg)) was calculated by multiplying a constant iron oxide density (ρ) of 5.24 mg cm^{-3} by the total volume (Equation 7), and the average iron oxide area density per FOV (S_{FOV} (mg m^{-2})) was calculated by dividing M_{FOV} (mg) by the microscope image FOV area (A_{FOV}) of 0.071×10^{-6} m^2 (Equation 8).

$$V_P = 0.52 \times A_P \times d_P \tag{6}$$

$$M_{FOV} = \rho \cdot V_P \tag{7}$$

$$S_{FOV} = \frac{M_{FOV}}{A_{\text{FOV}}} \tag{8}$$

In order to calculate the total material load on the pipe loop using the microscopy data was required to extrapolate the average iron oxide area density per FOV to the total area covered by particles, while experimental evidence from the experiments suggest that particle attachment occurred only in the invert region of the pipe circumference [34]. A corrected surface area (A_{Fe} (m^2)) corresponding to the area occupied by particles, was estimated from the total pipe area (A_T (m^2)) using Equation 9. The correction factor α corresponds to the fraction of the pipe circumference where materials were deposited. The observation of inner pipe walls after preliminary experiments suggested that most of the material deposits had accumulated on the invert region of the pipes. Therefore, in this study it was assumed $\alpha = 1/16$ (angle of 11.25° for each side from the pipe invert). The average iron oxide area density per FOV (S_{FOV} (mg m^{-2})) was also assumed to represent the average area density in the invert region of the whole pipe loop length.

$$A_{Fe} = \alpha\, A_T \tag{9}$$

Finally, the estimation of the total material load accumulated on the pipe wall based on the microscopy data (M_{PW_mic} (mg)) was calculated using Equation 10 by multiplying the average iron oxide area density per FOV (S_{FOV} (mg m^{-2})) and the corrected surface area (A_{Fe} (m^2)).

$$M_{PW\_mic} = S_{FOV} \times A_{Fe} \tag{10}$$

3 RESULTS & DISCUSSION

Figure 3 presents the relationship between SSC and turbidity used for the estimation of particulate material transport in the experiments. Figure 3a shows the decay of the linear transformation coefficient κ along the conditioning stage period of the experiment, and the empirical equation used to interpolate the data. After approximately 20 days of conditioning, there was no more measurable fractions of the SSC in the bulk water, but there was still a detectable level of turbidity caused by the iron oxide particles that resulted in a coefficient $\kappa = 0$. The decrease in the coefficient over time was likely caused by the more rapid attachment of larger particle fractions to the pipe wall during the experiment— a phenomenon that was also observed in Braga and Filion [25]. But in this case the effect was hastened by the passage of the recirculating water through the tank region where the flow is laminar and insufficient to maintain particles in suspension and causes the particles to settle more rapidly. The disproportional reduction in larger particle fractions relative to the fine fractions likely changed the PSD of the particles in suspension which in turn had an impact on the relationship between turbidity and SSC [14].

By comparison, Figure 3b shows that an adjusted constant coefficient $\kappa = 0.35$ mg L^{-1} NTU^{-1} fit the sample data well during the flushing stage of the experiments. The goodness-of-fit adjustment of the calibration coefficient indicate that particle suspension from all flushing steps had a similar PSD. This suggests that all flushing operations mobilized particles of similar sizes. This observation agrees with observations from previous experiments that showed that a narrow range of fine particle sizes are selected to populate the downstream sections of the pipe loop, while larger fractions are rapidly lost in the initial pipe lengths [25]. The SSC samples collected during the flushing stage of the present experiment corresponded to the middle section of the pipe loop (same as the coupon sections) and did not incorporate material from the initial pipe sections.

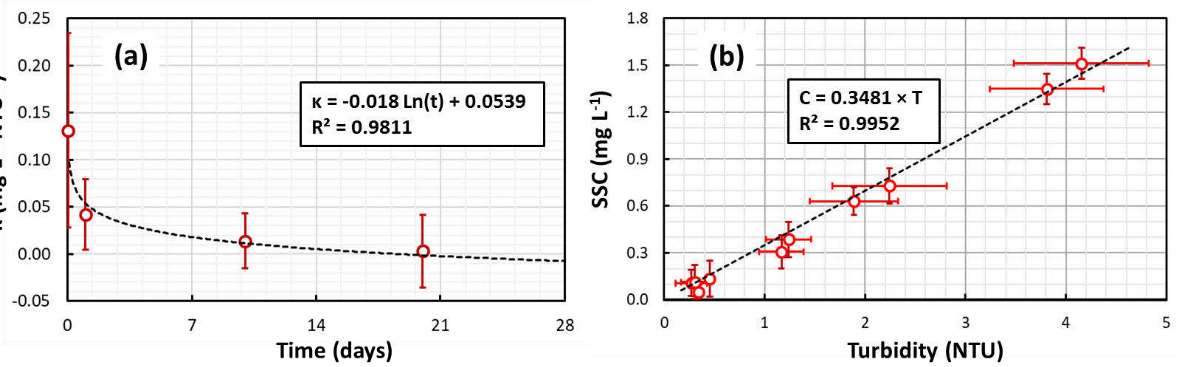


Figure 3 – a) Decay of the calibration coefficient κ along the conditioning phase of the experiment; b) linear SSC vs. turbidity relationship for the flushing phase of the experiments.

The results of the SSC and estimations of iron oxide particle accumulation on the pipe wall using the turbidity data for the conditioning phase of the experiment are presented in Figure 4. In particular, Figure 4a shows the SSC and the iron oxide particle accumulation on the pipe wall in the first 3 hours following the particle inoculation. The data here suggests that a rapid accumulation occurred in the first hour during the first passage of the particle plume through the pipes. This observation is likely explained by the entrainment of large particles (>6 µm) from the tank and into the loop; these large particles were likely rapidly deposited in the initial sections of the pipe loop [25].

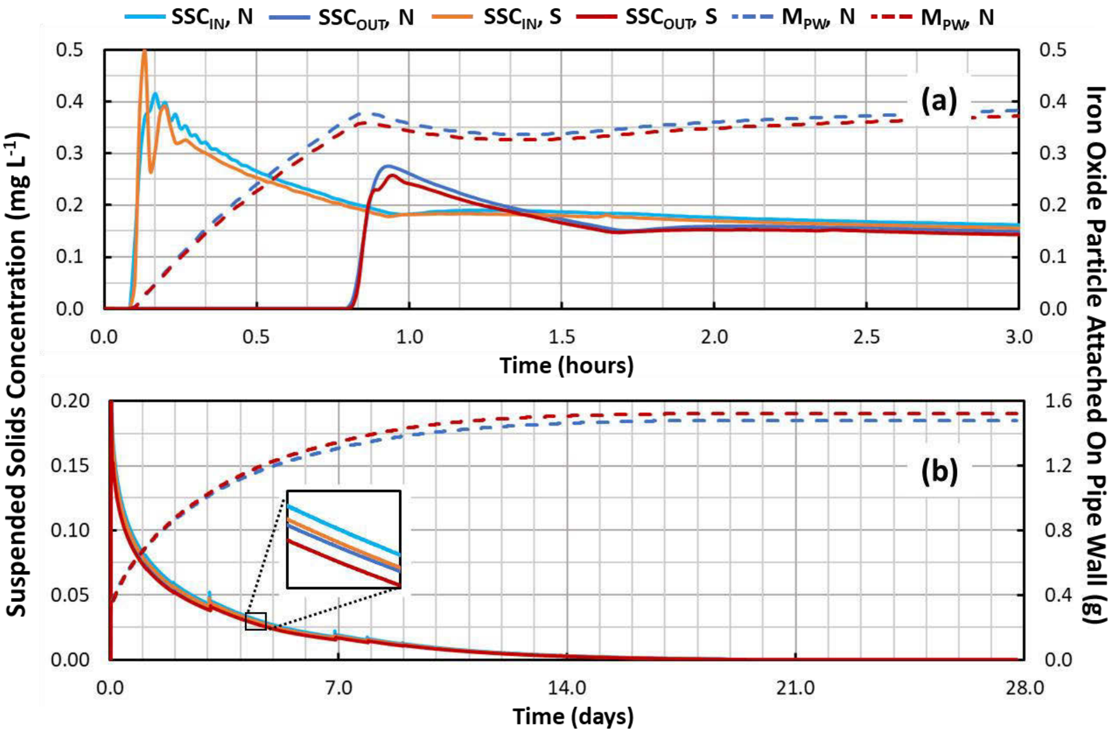


Figure 4 - SSC and iron oxide particle attached on pipe wall for the the conditioning phase of the experiment after a) 3 hours, and b) 28 days.

Once the larger particles settled to the bottom of the tanks and onto the internal surface of the pipes, the concentration of particles held in suspension decreased at a lower rate (Figure 4b). Particle accumulation on the pipe wall progressively increased based on the small difference between the inlet SSC and the outlet SSC as shown in Figure 4b. Despite this, the accumulation rate decreased over time and reached a value of zero after 20 days of conditioning when the SSC levels measured in the bulk water fell to zero. Approximately 1.5 g of iron oxide particles were deposited in the pipe loop at the end of the conditioning phase.

Results for the SSC and particle mobilization during the flushing phase of the experiments are presented in Figure 5. The first flushing step (FS1– Figure 5a) mobilized the larger particles with a mobilization of 0.8 g for both pipe loops (72% of total mobilized load). The second flushing step (FS2 – Figure 5b) also mobilized a substantial amount of material (21%). Also, the particles that remained attached to the pipe wall had a shear strength higher than the self-cleaning level of 1.2 Pa (WSS of the FS1) previously proposed [17]. Lastly, the third flushing step (FS3 – Figure 5c) showed that a small fraction of particles (7%) was able to resist a shear stress of 3.0 Pa (WSS of the FS2); these particles however were mobilized at 5.0 Pa. However, it is noted that very low turbidity levels likely resulted in higher proportional errors in the estimation of the turbidity profile during the third flushing step FS3.

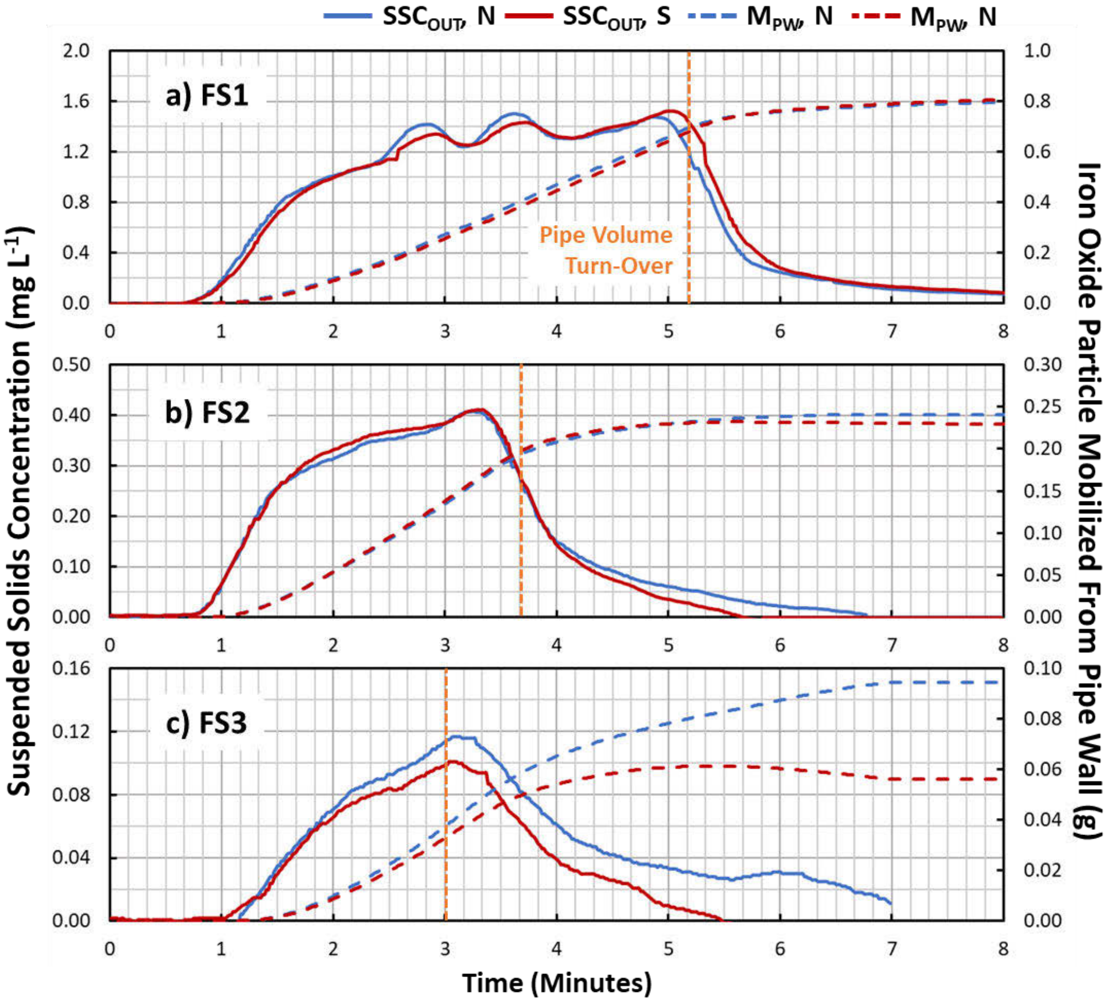


Figure 5 - SSC and iron oxide particles mobilized from the pipe wall for the the three flushing steps (FS1, FS2 and FS3) of the flushing phase of the experiment.

The SSC profiles of all flushing steps were similar and showed a slight tendency to reach peak turbidity at the single pipe volume turn-over point (Figure 5). This suggests that particle shear strength was independent of the pipe length, but that a higher accumulation load occurred in the initial sections of the pipe and progressively decreased along the pipe length. Approximately 1.1 g of iron oxide particles were mobilized from the pipe wall during the flushing phase.

The location of the iron oxide particles in all phases of the experiment was estimated by way of a mass balance analysis and the results are shown in Figure 6. In the conditioning phase, from an initial total mass of 4.4 g of iron oxide particles added to the system, only 19% remained in the bulk water 3 h after the inoculation, while 73% were estimated to be deposited in the tanks and 9% were attached to the pipe loop. The results suggest that the bulk fraction was mostly transferred to the pipe loop rather than to the tanks, while the material accumulated in the tank was reduced from 73% to 67% during the conditioning phase. This error is likely caused by an underestimation of the initial bulk water SSC, which was calibrated mostly through SSC samples collected from the turbidity positions at the pipe loop – a position that most large particles never reached. However, the error affected the early estimations of the material deposited in the tanks, since the fraction of particles accumulated in the pipe wall was estimated based on a SSC difference. In this case, the error in the initial bulk water SSC was resolved by the SSC difference operation. The only error remaining in the pipe wall load estimation arose from a longer period for the coefficient κ to reach zero, which in this case had a negligible contribution to the final load.

In the case of the flushing phase (right side of Figure 6), an initial load of 1.5 g of iron oxide particles accumulated on the pipe wall was progressively mobilized through the three flushing steps (FS1, FS2 and FS3). The flushing step FS1 mobilized 54% of the estimated attached particles, the FS2 step mobilized 16% and the FS3 step mobilized 5%, while an estimated fraction of 25% of the initial particles remained attached to the pipe walls. These correspond to 9% of the initial fraction of particles inoculated in the system.

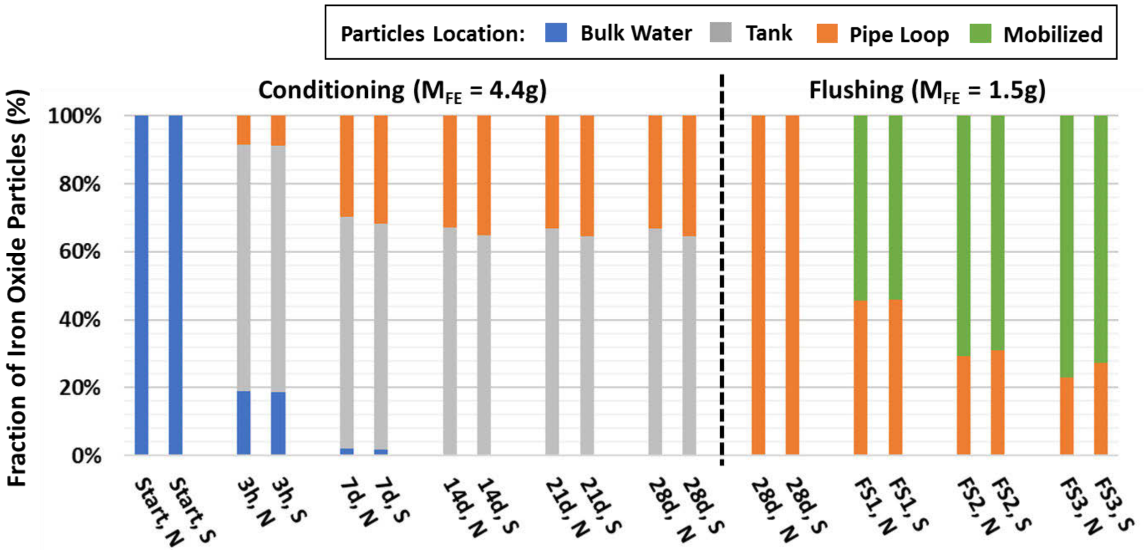


Figure 6 - Location of the iron oxide particles introduced in the pipe loop at various stages of the experiment (estimated through a mass balance analysis). "Tank" denotes the particles that settled to the bottom of the tank, "Bulk Water" denotes the particles held in suspension in the pipe loop, "Pipe Loop" denotes the particles attached to the internal surface of the pipe wall, and "Mobilized" denotes the particles mobilized during the flushing steps. It is noted that at the beginning of the flushing phase of the experiment (after 28 d), 1.5g of particles were adhered to the pipe wall and this mass is represented as 100% in this figure.

The results of the comparison between the turbidity and microscopy approaches to estimate the mass of particles attached to the pipe wall are shown in Figure 7 for each stage of the experiment where pipe wall samples were collected. During the conditioning phase, the microscopy estimations were substantially smaller than the turbidity-based method. The discrepancy may be explained by results from previous experiments that suggested that particle accumulation occurs to a greater extent in the upstream sections of the pipe loop near the inlet [25] and to a lesser extent in the downstream sections near the outlet. Because pipe wall coupons were only available in the middle section of the pipes, their data extrapolation to the whole pipe loop largely underestimates deposits from initial pipe sections. By comparison, results for the flushing phase shows good agreement between both methods, confirming the decrease in particles on the pipe loop along the three flushing steps and a remaining fraction on the pipe wall even after the third flushing step FS3.

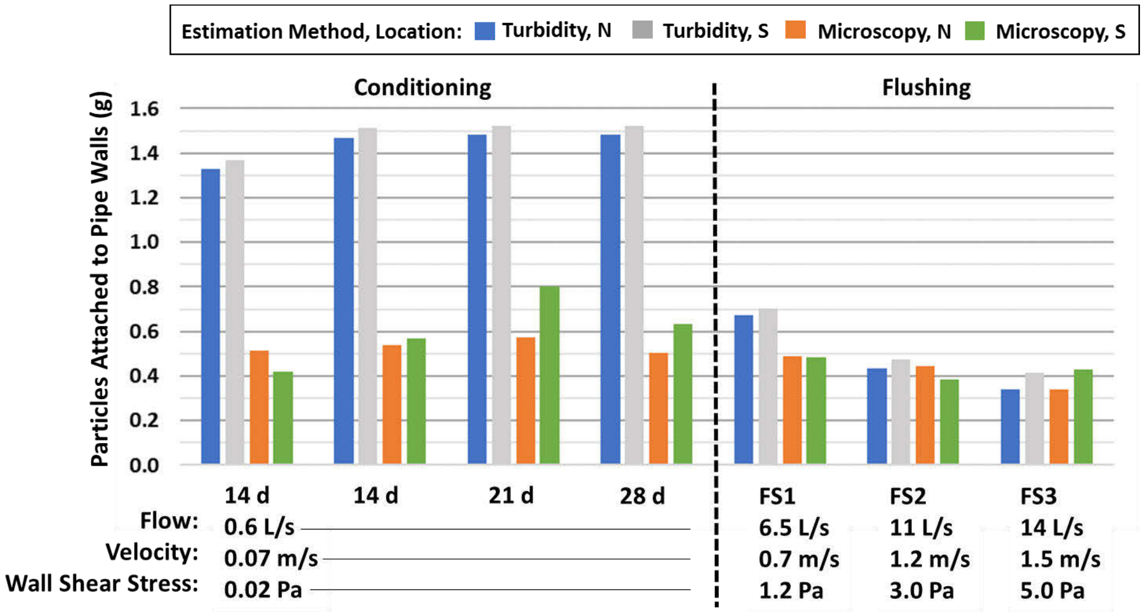


Figure 7 – Comparison of turbidity and microscopy approaches to estimate the mass of particles attached to the pipe wall for the conditioning and flushing stages of the experiment.

Lastly, the results of the microscopy approach to estimate the mass of particles attached to the pipe wall during the conditioning phase are shown in Figure 8. Particle accumulation followed a linear trend over time and suggested that particles accumulated on the pipe wall at the rate of 6.3 mg day^{-1} during the conditioning phase, while an approximate load of 0.46 g was initially deposited during the first passage of the particle plume. These estimations correspond only to particle fractions observed on the coupon wall samples, but they support the hypothesis that a small attachment rate of fine particles does prevail in suspension for longer periods [25]. It is worth noting that the error bars only considered the variation of material estimation across multiple FOVs of microscopic images, and do not include the extrapolation error of the data for the entire pipe area.

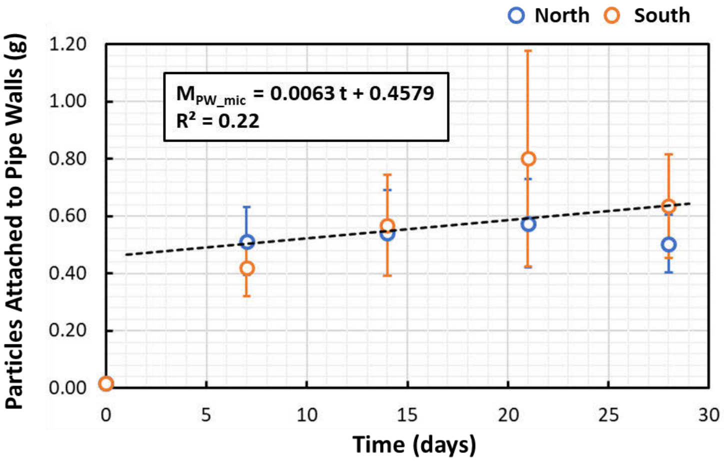


Figure 8 – Estimations of particle attachment with the microscopy approach for the conditioning phase of the experiment.

cc BY NC SA

4 CONCLUSION

This paper presented a turbidity approach and a direct microscopy approach to quantify the accumulation of iron oxide particles in a full-scale PVC pipe loop facility that mimics an operational DWDS. Results showed that particle accumulation on the pipe walls ended after 20 days of conditioning. At this time, approximately 30% of the material introduced into the loop had accumulated on the pipe wall and the rest accumulated in the system tank. Microscopy results showed that the accumulation of particles on coupon samples was 3 times smaller during the conditioning phase due to the fact that the bulk of particles were deposited in the initial sections of the pipes, away from pipe wall samples. However, material estimations after the three flushing steps showed good agreement between the turbidity and microscopy data, suggesting the larger fractions causing earlier discrepancy were mobilized in the first flushing step. Approximately 8% of the particles introduced into the pipe loop remained in the pipes after the last flushing step with a WSS of 5 Pa.

The efforts made to estimate material loads in physical units were fundamental to the mass balance approach used in the experiments, and for the direct comparison of the turbidity and microscopy quantification methods. The assumption of a constant relationship between SSC and turbidity for all stages of the experiment would have created a strong discrepancy in the results. Specifically, a large fraction of particles added to the system that did not contribute to turbidity would not have been quantified. The proposed methods still have considerable sources of errors, but these are likely smaller than errors incurred by using uncalibrated turbidity data or porting coefficients from systems with different or unknown particles characteristics.

Future research will aim to better assess drinking water SSC and PSD. This will have the potential to reduce errors of turbidity data usage for material accumulation estimations, while the evolution of the microscopic assessment methods has also potential to contribute to complementary data. New studies are also needed to examine the occurrence of biofilms and possible precipitation of particles as sources of material accumulation at the pipe wall, as well as additional pipe wall materials and other operational conditions such as variable flow.

5 REFERENCES

[1] I. Sunny, P. S. Husband, and J. B. Boxall, "Impact of hydraulic interventions on chronic and acute material loading and discolouration risk in drinking water distribution systems," Water Research, vol. 169, p. 115224, 2020/02/01/ 2020, doi: https://doi.org/10.1016/j.watres.2019.115224.

[2] G. Liu et al., "Hotspots for selected metal elements and microbes accumulation and the corresponding water quality deterioration potential in an unchlorinated drinking water distribution system," Water Res, vol. 124, pp. 435-445, Nov 1 2017, doi: 10.1016/j.watres.2017.08.002.

[3] A. Seth, R. Bachmann, J. Boxall, A. Saul, and R. Edyvean, "Characterisation of materials causing discolouration in potable water systems," Water Science and Technology, vol. 49, no. 2, pp. 27-32, 2004, doi: 10.2166/wst.2004.0080.

[4] A. Carrière, V. Gauthier, R. Desjardins, and B. Barbeau, "Evaluation of loose deposits in distribution systems through: unidirectional flushing," Journal - American Water Works Association, vol. 97, no. 9, pp. 82-92, 2005, doi: 10.1002/j.1551-8833.2005.tb07474.x.

[5] S. Triantafyllidou, D. A. Lytle, A. S. C. Chen, L. Wang, C. Muhlen, and T. J. Sorg, "Patterns of arsenic release in drinking water distribution systems," AWWA Water Science, vol. 1, no. 4, p. e1149, 2019, doi: https://doi.org/10.1002/aws2.1149.

[6] V. Rilstone, L. Vignale, J. Craddock, A. Cushing, Y. Filion, and P. Champagne, "The role of antibiotics and heavy metals on the development, promotion, and dissemination of antimicrobial resistance in drinking water biofilms," Chemosphere, vol. 282, p. 131048, Nov 2021, doi: 10.1016/j.chemosphere.2021.131048.

[7] L. K. Kimbell et al., "Cast iron drinking water pipe biofilms support diverse microbial communities containing antibiotic resistance genes, metal resistance genes, and class 1 integrons," Environmental Science:

Water Research & Technology, 10.1039/D0EW01059F vol. 7, no. 3, pp. 584-598, 2021, doi: 10.1039/D0EW01059F.

[8] B. Barbeau, K. Julienne, A. Carriere, and V. Gauthier, "Dead-end flushing of a distribution system: Short and long-term effects on water quality," Journal of Water Supply: Research and Technology-Aqua, vol. 54, no. 6, pp. 371-383, 2005, doi: 10.2166/aqua.2005.0035.

[9] M. J. Lehtola, T. K. Nissinen, I. T. Miettinen, P. J. Martikainen, and T. Vartiainen, "Removal of soft deposits from the distribution system improves the drinking water quality," Water Res, vol. 38, no. 3, pp. 601-10, Feb 2004, doi: 10.1016/j.watres.2003.10.054.

[10] O. M. Zacheus, M. J. Lehtola, L. K. Korhonen, and P. J. Martikainen, "Soft deposits, the key site for microbial growth in drinking water distribution networks," Water Research, vol. 35, no. 7, pp. 1757-1765, 2001/05/01/ 2001, doi: https://doi.org/10.1016/S0043-1354(00)00431-0.

[11] J. H. Vreeburg, D. Schippers, J. Q. Verberk, and J. C. van Dijk, "Impact of particles on sediment accumulation in a drinking water distribution system," Water Res, vol. 42, no. 16, pp. 4233-42, Oct 2008, doi: 10.1016/j.watres.2008.05.024.

[12] A. Mussared, R. Fabris, J. Vreeburg, J. Jelbart, and M. Drikas, "The origin and risks associated with loose deposits in a drinking water distribution system," Water Supply, vol. 19, no. 1, pp. 291-302, 2019, doi: 10.2166/ws.2018.073.

[13] J. H. Vreeburg and J. B. Boxall, "Discolouration in potable water distribution systems: a review," Water Res, vol. 41, no. 3, pp. 519-29, Feb 2007, doi: 10.1016/j.watres.2006.09.028.

[14] B. G. B. Kitchener, J. Wainwright, and A. J. Parsons, "A review of the principles of turbidity measurement," Progress in Physical Geography: Earth and Environment, vol. 41, no. 5, pp. 620-642, 2017, doi: 10.1177/0309133317726540.

[15] H. Armand, I. I. Stoianov, and N. J. D. Graham, "A holistic assessment of discolouration processes in water distribution networks," Urban Water Journal, vol. 14, no. 3, pp. 263-277, 2015, doi: 10.1080/1573062x.2015.1111912.

[16] D. M. Cook and J. B. Boxall, "Discoloration Material Accumulation in Water Distribution Systems," Journal of Pipeline Systems Engineering and Practice, vol. 2, no. 4, pp. 113-122, 2011, doi: 10.1061/(asce)ps.1949-1204.0000083.

[17] P. S. Husband and J. B. Boxall, "Asset deterioration and discolouration in water distribution systems," Water Res, vol. 45, no. 1, pp. 113-24, Jan 2011, doi: 10.1016/j.watres.2010.08.021.

[18] S. Husband and J. B. Boxall, "Field Studies of Discoloration in Water Distribution Systems: Model Verification and Practical Implications," Journal of Environmental Engineering, vol. 136, no. 1, pp. 86-94, 2010, doi: 10.1061/(asce)ee.1943-7870.0000115.

[19] E. J. M. Blokker and P. G. Schaap, "Particle Accumulation Rate of Drinking Water Distribution Systems Determined by Incoming Turbidity," Procedia Engineering, vol. 119, pp. 290-298, 2015, doi: 10.1016/j.proeng.2015.08.888.

[20] J. van Summeren and M. Blokker, "Modeling particle transport and discoloration risk in drinking water distribution networks," Drinking Water Engineering and Science, vol. 10, no. 2, pp. 99-107, 2017, doi: 10.5194/dwes-10-99-2017.

[21] W. R. Furnass, R. P. Collins, P. S. Husband, R. L. Sharpe, S. R. Mounce, and J. B. Boxall, "Modelling both the continual erosion and regeneration of discolouration material in drinking water distribution systems," Water Supply, vol. 14, no. 1, pp. 81-90, 2014, doi: 10.2166/ws.2013.176.

[22] J. B. Boxall and A. J. Saul, "Modeling Discoloration in Potable Water Distribution Systems," Journal of Environmental Engineering, vol. 131, no. 5, pp. 716-725, 2005, doi: doi:10.1061/(ASCE)0733-9372(2005)131:5(716).

[23] J. B. Boxall, A. J. Saul, J. D. Gunstead, and N. Dewis, "Regeneration of Discolouration in Distribution Systems," in World Water &amp; Environmental Resources Congress 2003, 2003, pp. 1-9.

[24] F. Pourcel, S. Duchesne, and M. Ouellet, "Evolution of the relationship between total suspended solids concentration and turbidity during flushing sequences of water pipes," Journal of Water Supply: Research and Technology-Aqua, vol. 69, no. 4, pp. 376-386, 2020, doi: 10.2166/aqua.2020.127.

[25] A. S. Braga and Y. Filion, "The interplay of suspended sediment concentration, particle size and fluid velocity on the rapid deposition of suspended iron oxide particles in PVC drinking water pipes," Water Research X, vol. 15, p. 100143, 2022/05/01/ 2022, doi: https://doi.org/10.1016/j.wroa.2022.100143.

[26] T. Walski, K. Minnich, C. Sherman, L. Strause, and B. Whitman, "Can There be a Law of Conservation of Turbidity," Procedia Engineering, vol. 186, pp. 372-379, 2017, doi: 10.1016/j.proeng.2017.03.233.

[27] A. S. Braga, R. Saulnier, Y. Filion, and A. Cushing, "Dynamics of material detachment from drinking water pipes under flushing conditions in a full-scale drinking water laboratory system," Urban Water Journal, vol. 17, no. 8, pp. 745-753, 2020, doi: 10.1080/1573062x.2020.1800759.

[28] K. E. Fish et al., "Characterisation of the physical composition and microbial community structure of biofilms within a model full-scale drinking water distribution system," PLoS One, vol. 10, no. 2, p. e0115824, 2015, doi: 10.1371/journal.pone.0115824.

[29] P. Deines, R. Sekar, P. S. Husband, J. B. Boxall, A. M. Osborn, and C. A. Biggs, "A new coupon design for simultaneous analysis of in situ microbial biofilm formation and community structure in drinking water distribution systems," Appl Microbiol Biotechnol, vol. 87, no. 2, pp. 749-56, Jun 2010, doi: 10.1007/s00253-010-2510-x.

[30] A. S. Braga and Y. Filion, "A novel monitoring scheme to detect iron oxide particle deposits on the internal surface of PVC drinking water pipes," Environmental Science: Water Research & Technology, 10.1039/D1EW00614B vol. 7, no. 11, pp. 2116-2128, 2021, doi: 10.1039/D1EW00614B.

[31] W. Furnass, I. Douterelo, R. Collins, S. Mounce, and J. Boxall, "Controlled, Realistic-scale, Experimental Study of How the Quantity and Erodibility of Discolouration Material Varies with Shear Strength," Procedia Engineering, vol. 89, pp. 135-142, 2014, doi: 10.1016/j.proeng.2014.11.169.

[32] R. L. Sharpe, C. J. Smith, J. B. Boxall, and C. A. Biggs, "Pilot Scale Laboratory Investigations into the Impact of Steady State Conditioning Flow on Potable Water Discolouration," in Water Distribution Systems Analysis 2010, 2011, pp. 494-506.

[33] P. S. Husband, J. B. Boxall, and A. J. Saul, "Laboratory studies investigating the processes leading to discolouration in water distribution networks," Water Res, vol. 42, no. 16, pp. 4309-18, Oct 2008, doi: 10.1016/j.watres.2008.07.026.

[34] A. S. Braga and Y. Filion, "Initial stages of particulate iron oxide attachment on drinking water PVC pipes characterized by turbidity data and brightfield microscopy from a full-scale laboratory," Environmental Science: Water Research & Technology, 10.1039/D2EW00010E 2022, doi: 10.1039/D2EW00010E.

[35] A. S. Braga and Y. Filion, "A novel monitoring scheme to detect iron oxide particle deposits on the internal surface of PVC drinking water pipes," Environmental Science: Water Research & Technology, 2021, doi: 10.1039/d1ew00614b.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14093

MODEL FAST DEPOSITION OF FINE IRON OXIDE PARTICLES ON PVC PIPE MAINS DURING THE PASSAGE OF A SUSPENDED PARTICLE PLUME IN A FULL-SCALE LABORATORY SYSTEM

Artur Sass Braga [1] and Yves Filion [2]

[1,2] Queen's University
Kingston, Ontario, (Canada)

[1] 16asb2@queensu.ca, [2] yves.filion@queensu.ca

Abstract

This paper aims to investigate the interplay of suspended sediment concentration (SSC) and the fluid velocity that are both responsible for the rapid formation of sediments deposits in a full-scale laboratory system with PVC pipes. The specific objectives of the paper were to: (1) estimate the average rate of iron oxide particle deposition on PVC pipes immediately after the rapid release, and during the passage of, a highly-concentrated plume of iron oxide particles that typifies discolouration events in real systems; (2) determine whether previously deposited particles on the pipe wall from the passage of previous iron oxide particle plumes affect the average rate of iron oxide deposition and attachment to the pipe wall. Experiments were realized in a test facility that mimics the operation of drinking water systems using a special design method to accurately inoculate and detect the passage of iron oxide particles through the pipes under steady flow conditions commonly found in water networks. The experiments consisted of the passage of 3 sequential waves of particles, tested for three different SSC and three different flow velocities.

The results showed that increasing the inlet water quality conditions and the concentration of the particulate plume tended to increase the total deposited load while fluid velocity had a negligeable impact on the total deposited load. The results also showed that both an increase of concentration and fluid velocity produced an increase in the average deposition rate, and that this was mainly caused by the shorter conditioning period of the experiments at higher velocities. Further, the experimental results did not show any change in average particle deposition rate and attachment across the three plumes. This suggests that the presence of particle deposits from previous plumes had a negligible impact on particle deposition and attachment.

Keywords
Drinking water discolouration, particle deposition, wall shear stress.

1 INTRODUCTION

It is well established that important changes in the water quality conditions often occurs between the discharge works of a water treatment facility and the consumer's tap. Indeed, as treated water is conveyed through the distribution system it is subject to several processes (e.g., internal corrosion, resuspension of fine and coarse particles from the pipe wall) that can cause its quality to change dramatically before it reaches the consumer's tap [1]. For this reason, it is not surprising that red water problems and customer complaints are still a major concern for water managers responsible for the operation and maintenance of distribution systems in North American cities.

Heavy metals such as iron and manganese are of particular concern for utilities since they are often responsible for the creation of discolouration events and for the deterioration of water aesthetic that can often lead to customer complaints [2]. Iron and manganese particles can be imported from upstream metallic pipes that have undergone some level of corrosion, as these metal products are resuspended into the bulk flow and transported to downstream pipes. Despite this, it is still unclear to researchers and practitioners alike what are the key processes that cause particles to migrate to the pipe wall from the bulk flow and to adhere to pipe wall under normal flow conditions. Moreover, water utilities often lack the tools and methods to adequately assess the build-up of material deposits in operational networks.

In this context, this paper aims to examine the extent of acute particulate attachment on pipe walls following the rapid release and travel of a concentrated plume of iron oxide particles immediately following a simulated discolouration event in controlled experiments completed in a full-scale PVC pipe loop laboratory. The specific objectives of the experiments were to: 1) estimate the load of iron oxide particle deposition on PVC pipes immediately after the rapid release, and during the passage of, a highly-concentrated plume of iron oxide particles that typifies discolouration events in real systems; 2) determine whether previously deposited particles on the pipe wall from the passage of previous iron oxide particle plumes affect the average rate of iron oxide deposition and attachment to the pipe wall.

2 METHODS

The experiments were performed in the Drinking Water Distribution Laboratory (DWDL) at Queen's University (Figure 1), using a full-scale pipe loop rig that simulates the operation of drinking water mains. The pipe laboratory system is comprised of a water tank with a volume of 3.6 m^3, two variable speed centrifugal pumps and 11 loops of IPEX Blue Brute PVC pipe Class 235 (DR18) with an internal diameter of 108 mm, and a total length of 193 m. During the experiments, the pipe loop system was operated in a non-recirculatory manner, where drinking water from the City of Kington was continuously added to the tank, pumped through the pipe system, and discarded at the end. The water from the City of Kingston has a pH of 8.1, a hardness of 123 mg L 1-CaCO3 and an alkalinity of 92 mg L 1 CaCO3, and has approximately 1.7 mg L 1 of dissolved carbon and 1.0 mg L 1 of total nitrogen [3]. During the experiments the water temperature ranged between 12°C and 14°C.

In the experiments, drinking water was amended with iron oxide particles of known particle size distribution at the inlet to the pipe loop to induce the formation of material deposits on the pipe walls. A chemical grade of red iron oxide powder from Alpha Chemicals, with a composition of 82% of Iron (III) oxide – Fe_2O_3, was chosen as the source of particles for the experiments due their representativeness of iron oxide particles found in DWDSs and their stable particulate form and insolubility in water. The powder was sent to a commercial laboratory to split it into sub-samples of 50 g with identical particle size distributions (PSD) using a mini rotary splitter from Retsch company model PT100. The PSD of two sub-samples picked at random was analyzed with a Malvern Mastersizer 3000 particle size analyzer (laser diffraction). The insolubility of the

BY NC SA

particles and negligible amounts of dissolved iron in the local drinking water guaranteed that particles remained stable during the experiments.

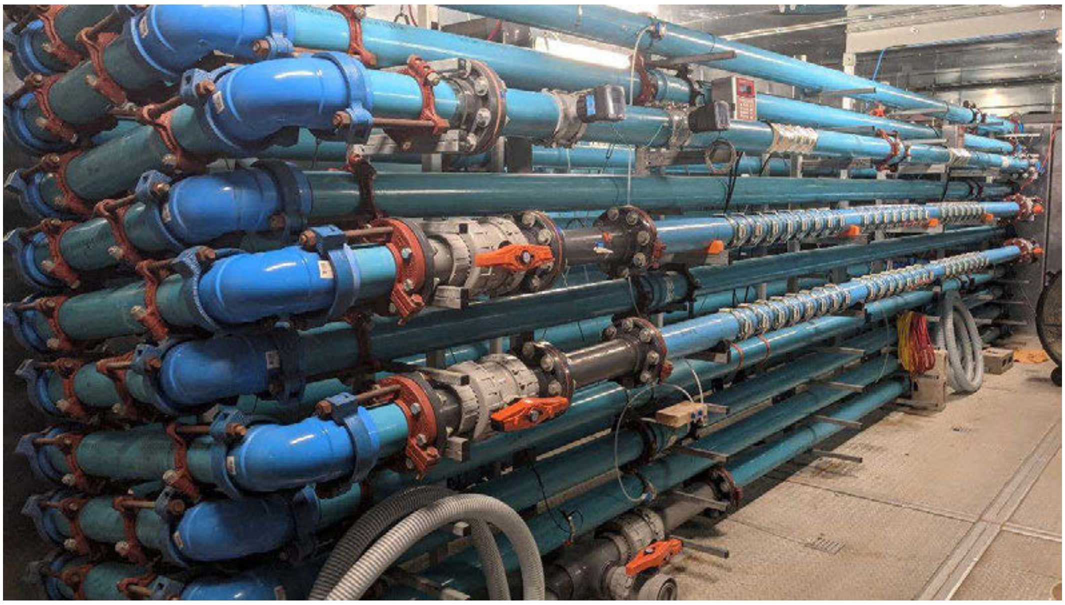


Figure 1. Drinking Water Distribution Laboratory (DWDL) at Queen's University.

A total of five experiments were performed with a variable concentration of suspended particles (F1C1, F2C2 and F1C3 in Table 1) and a variable conditioning flow rate (F1C1, F2C1 and F3C1 in Table 1). During all the experiments, the pipe flow rate and turbidity were continuously monitored at a sampling frequency of 1 Hz using two Sierra InnovaSonic® 205i Ultrasonic Flow Meters (±0.5%) and two online Hach TU5300sc turbidimeters (±0.01 NTU). The turbidimeters were configured to monitor turbidity at the inlet (7% of the pipe length) and outlet (98% of the pipe length). Turbidity was measured from water continuously sampled from the centre of the pipe at a constant sampling flow rate. Only steady flow conditions were tested, but due to the short duration of the experiments, these flows can be interpreted as the highest daily peak flows of operational systems commonly used to determine the conditioning WSS in systems with variable demands [4, 5].

***Table 1**. Hydraulic and water quality conditions for the experiments.*

Experiment	***SSC*** [a] ***(mg*** L^{-1}***)***	***Flow (L*** s^{-1}***)***	***Velocity (m*** s^{-1}***)***	***Reynolds Number***	***WSS*** [b] ***(Pa)***	***Pipe loop RT*** [c] ***(min)***
F1C1	*20.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F1C2	*40.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F1C3	*60.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F2C1	*20.0*	*1.80*	*0.20*	*21 000*	*0.11*	*17.0*
F3C1	*20.0*	*2.75*	*0.30*	*32 000*	*0.24*	*11.1*

a. Suspended solids concentration.
b. Wall shear stress based on the Darcy-Weisbach equation.
c. Pipe loop residence time (RT) calculated by dividing the pipe loop volume of 1.7 m^3 *by the flow rate.*

Each experiment was divided in three independent sub-stages (P1, P2 and P3), where plumes of suspended particles were produced at the inlet of the pipe loop and completely transported through the pipe loop outlet before the introduction of the next plume in the next sub-stage. Each plume was created by injecting a known volume of a concentrated solution of iron oxide particles

(CSIOP) at the inlet of the pipe loop with a known SSC over an injection period equal to the time required for the water to travel 60 m downstream of the injection point given the prevailing flow rate in the pipe loop. The plume length of 60 m was mainly defined by a limitation of the maximum injection volume under stable mixing conditions in the CSIOP. A new CSIOP was prepared for each plume, by adding a known mass of iron oxide particles (IOP) into a known drinking water volume in a mixing baffled tank operated at a fixed mixing intensity. After the IOP addition in the tank, a period of 5 minutes was used to allow the particle suspension to stabilize in the mixing tank. A diaphragm pump was used to pump water from the bottom of the CSIOP tank (approximately 10 cm away from the tank walls) from a position of intense mixing between two of the four vertical baffles and inject it at a constant flow at the inlet of the pipe loop, at the centreline cross-sectional position of the pipe. At the end of each injection period, the steady flow conditions were maintained in the pipe loop for an additional period required to refresh the equivalent of 1.5 pipe loop volumes to guarantee that all suspended particles from the sediment plume had exited the pipe loop. Between each of the different experimental conditions tested, the pipe system was cleaned at the highest flowrate (16 L s^{-1}) to re-establish identical initial condition for all the experiments.

Following each injection, grab samples were collected at the inlet and outlet sampling port locations to determine SSC at those locations and to assess changes in the particles in suspension after the passage of each plume. The collection of grab samples was synchronized with the fluid velocity in the pipe loop to collect 4 L of water from the centre of the passing sediment plume. The SSC of the grab samples was determined using a dry-weight method, by filtering the sample volume with pre-weighed 0.45 µm glass microfiber filters, drying the filters at 105°C for 1 hour, and weighing them again after drying with a precision scale. The full details of the experimental design can be found in Braga and Filion [6].

3 RESULTS & DISCUSSION

The impact of source concentration and velocity on the average deposition rate were examined. Figure 2 plots the fraction of the inoculated particles that was deposited in the pipe wall after the passage of each plume and the average deposition rate for different inlet concentrations and fluid velocities. Note that the deposited load is calculated by taking the difference between the total material load at the injection point and the total material load at the outlet. The results in Figure 2 shows that a stable fraction of 74% of particle was deposited in the walls for the experiments with same velocity and three different concentrations, while the same fraction reduced for 69% for the two experiments with higher velocities and similar concentration of the experiment F1C1. This result suggests that the conditioning velocity of 0.07 m s^{-1} were not sufficient to maintain a large fraction of the particles in suspension (74%), and a velocity increase above 0.2 m s^{-1} was only sufficient to keep an extra 5% of the particles in suspension, preventing their deposit in the pipes. Figure 2 also shows that both an increase of concentration and fluid velocity produced an increase in the average deposition rate (secondary y-axis in Figure 2). Such increase was mainly caused by the shorter conditioning period of the experiments at higher velocities. Is hypothesized that most of particle deposition occurred right after their injection into the pipe loop. Therefore, the instantaneous deposition rate at the beginning of the experiment must have exceeded several times the average deposition rate estimated here, which was reduced by the longer period required for the plume passage through the pipes.

At the outset of the study, it was hypothesized that the particle accumulation rate of the 2nd and 3rd plumes might be affected by the presence of particles deposited on the pipe wall after the first plume. The basis for this hypothesis is that the particles from the first plume might make the pipe substrate more favourable to particle attachment. The results however did not show a change in total deposited load or average deposition rate across the three plumes introduced into the pipe loop. This suggests that the presence of particle deposits from previous plumes had a negligible impact on particle deposition and attachment.

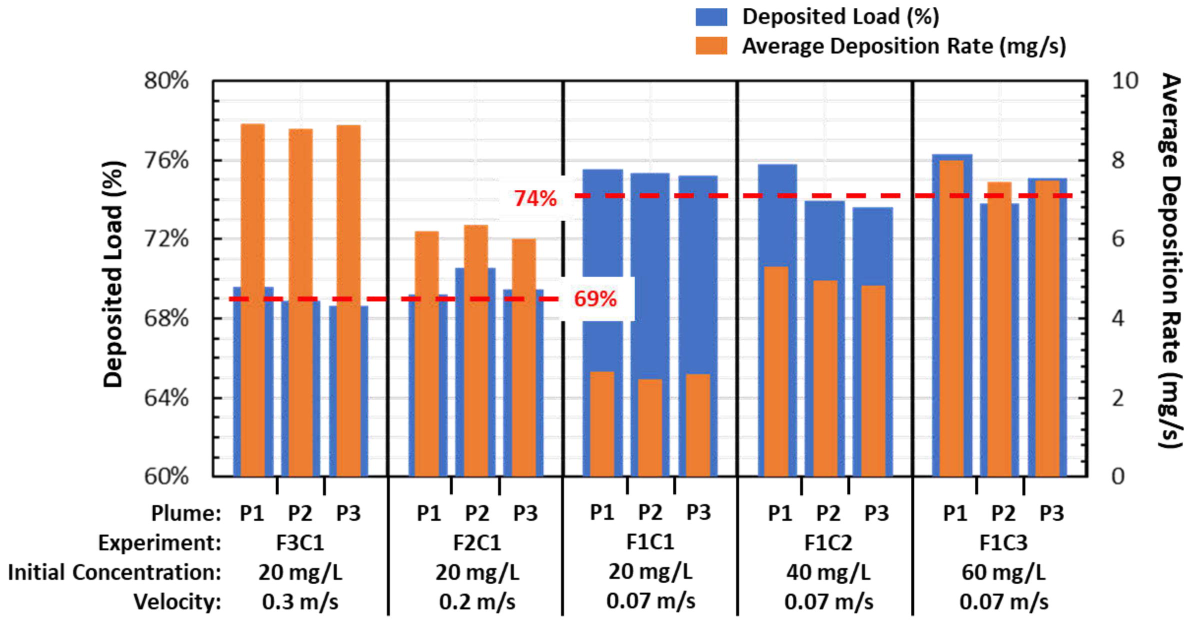


Figure 2. Deposited load and average deposition rate for each plume of each experiment performed at three different velocities and similar concentration (F3C1, F2C1 and F1C1), and at three different initial concentrations and similar velocity (F1C1, F1C2 and F1C3).

4 CONCLUSION

The aim of the paper was to examine the extent of particulate attachment following the passage of a concentrated plume of iron oxide particles that may arise in systems following a discolouration event. A set of controlled experiments were performed in a full-scale water distribution laboratory to 1) estimate the average rate of iron oxide particle deposition on PVC pipes immediately after the passage of the plumes, and 2) determine whether previously deposited particles on the pipe wall from the passage of previous iron oxide particle plumes affect the average rate of iron oxide deposition and attachment to the pipe wall.

The results showed that increasing the inlet water quality conditions and the concentration of the particulate plume tended to increase the total deposited load, but its proportion to the load of added particles remained constant for all three concentrations at the same fluid velocity. The increase of fluid velocity above 0.2 m s^{-1} produced a slight decrease of the proportional deposited load of particles of 5%. The results also showed that both an increase of concentration and fluid velocity produced an increase in the average deposition rate, and that this was mainly caused by the shorter conditioning period of the experiments at higher velocities.

At the outset of the study, it was hypothesized that the particle accumulation rate of the 2nd and 3rd plumes might be affected by the presence of particles deposited on the pipe wall after the first plume. The experimental results did not show any change in average particle deposition rate and attachment across the three plumes. This suggests that the presence of particle deposits from previous plumes had a negligible impact on particle deposition and attachment.

5 REFERENCES

[1] H. Armand, I. I. Stoianov, and N. J. D. Graham, "A holistic assessment of discolouration processes in water distribution networks," Urban Water Journal, vol. 14, no. 3, pp. 263-277, 2015, doi: 10.1080/1573062x.2015.1111912.

[2] A. Seth, R. Bachmann, J. Boxall, A. Saul, and R. Edyvean, "Characterisation of materials causing discolouration in potable water systems," Water Science and Technology, vol. 49, no. 2, pp. 27-32, 2004, doi: 10.2166/wst.2004.0080.

[3] Utilities Kingston, "2017 ANNUAL REPORT ON DRINKING WATER QUALITY - KING STREET WATER TREATMENT PLANT," Utilities Kingston, 2017. Accessed: 30 August 2018. [Online]. Available: https://utilitieskingston.com/Water/Quality/Reports

[4] P. S. Husband and J. B. Boxall, "Asset deterioration and discolouration in water distribution systems," Water Res, vol. 45, no. 1, pp. 113-24, Jan 2011, doi: 10.1016/j.watres.2010.08.021.

[5] P. S. Husband, J. B. Boxall, and A. J. Saul, "Laboratory studies investigating the processes leading to discolouration in water distribution networks," Water Res, vol. 42, no. 16, pp. 4309-18, Oct 2008, doi: 10.1016/j.watres.2008.07.026.

[6] A. S. Braga and Y. Filion, "The interplay of suspended sediment concentration, particle size and fluid velocity on the rapid deposition of suspended iron oxide particles in PVC drinking water pipes," Water Research X, vol. 15, p. 100143, 2022/05/01/ 2022, doi: https://doi.org/10.1016/j.wroa.2022.100143.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14094

THE IDENTIFICATION OF VARIABLE SHEAR STRENGTH OF PARTICLES DEPOSITED ON DRINKING WATER PVC PIPES AFTER THE PASSAGE OF A SUSPENDED PARTICLE PLUME IN A FULL-SCALE LABORATORY SYSTEM

Artur Sass Braga [1] and Yves Filion [2]

[1,2] Queen's University
Kingston, Ontario, (Canada)

[1] 16asb2@queensu.ca, [2] yves.filion@queensu.ca

Abstract

Understanding the adhesion properties of sediments that accumulate on the wall of drinking water pipes is an important step in the development of mitigation strategies to prevent the formation of deposits and protect drinking water. Research based on flushing of operational pipe mains that mobilized particulate sediments from isolated pipe sections has shown that fine iron oxide particles are a recurrent major component of these deposits. In addition, it has been established that adhesion forces proportional to the flow wall shear stress (WSS) develop between the pipe wall and accumulated particles, which prevents the washing-off of particles during common conditioning flows and provokes a rapid resuspension during high-flow events that cause water discolouration. Discolouration models have also showed that sediments have a variable shear strength, and, therefore, a fraction of material may resist a first increase in WSS but then be mobilized after a second increase in WSS. To explain the variable shear strength of layers, researchers have hypothesized that sediments accumulate as cohesive layers, which might be explained by the growth of biofilm among particulate material. Although current models have successfully explained sediment mobilization during flushing, the prediction of material accumulation and its shear strength is more challenging due to the lack of a comprehensive understanding about the accumulation process.

The aim of this paper is to examine how particulate iron oxide that are rapidly deposited on PVC pipes develop variable shear strength under common hydraulic conditions found in drinking water distribution networks. A set of experiments were performed in a full-scale laboratory facility, where selected iron oxide particles were controlled and used to amend the feed water at the entrance of a 200 m pipe loop during a short period of time to create a suspended sediment plume with constant concentration. Experiments were realized at three different concentrations and three different velocities. In each set, three sequential plumes were used to accumulate particles on the pipe walls, followed by three sequential flushing steps used to mobilize the particulate material. The deposition of iron oxides in the PVC pipes were assessed indirectly through suspended sediment concentration (SSC) and turbidity data. Results showed that iron oxide particles predominantly accumulated in the first sections of the pipe wall. Most sediments were found to have weak shear strength and were easily mobilized with the first flushing step. However, the mobilized load from the second and third flushing steps revealed a consistent mobilization of sediment with higher shear strengths. These shear strengths were higher in the experiments with a higher inoculation concentration, and they were lower in experiments performed with a higher conditioning fluid velocity. The results suggest that variable shear strength can develop without biofilm. Additional long-term experiments are still required to evaluate the evolution of sediment shear strength which possibly can increase with time.

Keywords
Discolouration, Pipe Flushing, Iron Oxide Particles, Shear Strength, Wall Shear Stress.

1 INTRODUCTION

Pipe flushing is an established method in the drinking water industry to recover accumulated materials on the internal surface of pipe walls before they can have a negative impact on drinking water quality and cause customers complaints [1-3]. However, the practice of flushing operations by water utilities is not yet standardized and rely mostly on local empirical experience, where results from previous attempts are generally used as a guidance to determine key control variables for new pipe flushing. Research shows that the fluid velocity and wall shear stress (WSS) are the most important parameters that have an impact on material mobilization [4]. But it is well known that material accumulation is dominant in old cast iron pipe sections due to corrosion processes [5], and prolonged periods of time without flushing that produce larger material loads [6, 7]. In this context, a water utility's challenges consist of optimizing flushing operations in terms of intensity (e.g., flushing duration, fluid velocity/WSS) and recurrence, to reduce their elevated cost and water consumption.

During flushing, the increase of the WSS over the pipe walls promotes the mobilization of material accumulation. This implies that there is an adhesion force that binds the accumulated materials have an to the pipe wall that prevents their mobilization under regular flow demands. This concept is named *material shear strength* and it was first quantified by the *prediction of discolouration events in distribution system* (PODDS) model [8]. PODDS hypothesized that accumulated material have a finite shear strength, and material can be mobilized if the WSS exceeds the shear strength of that material. In addition, the fraction of material within each shear strength layer was pre-defined based on the assumption of a linear decrease in material load according to their shear strength. However, later research showed that the highest WSS levels achieved on flushing of DWDSs are often below the ultimate shear strength of accumulated materials [7, 9], and that material strengths at all levels are developed simultaneously without conforming to the constraints imposed by the PODDS framework [10]. In practice, these findings revealed that pipe flushing has a limited capacity to remove material from pipe walls, and that the effectiveness of such operations is strongly determined by the material shear strength.

Inspired by the PODDS concept, research has focused on determining the shear strength of accumulated materials to better estimate optimal flushing levels [4, 6, 7, 10]. Researchers have proposed that material shear strength is conditioned by the daily peak flow, although substantial differences between the shear strength behaviour for different pipe materials are also reported and yet not fully explained [7, 9]. In this context, the known contributions of biofilm for material accumulation in drinking water pipes [11] motivated the explanation of material shear strength based on cohesiveness and high-adhesion properties of extracellular polymeric substances (EPS) that form biofilms [12, 13]. Biofilm may explain the occurrence of materials with high shear strength that are able to resist elevated WSS during flushing, and these are hypothesized to assist in the deposition of iron oxide particles – commonly released during discolouration events [14] – by enhancing pipe wall adhesion properties [15]. By comparison, recent research realized under controlled laboratory conditions that prevented biofilm formation observed iron oxide deposits on PVC pipes with shear strength that was able to resist a WSS of 5 Pa [16]. Microscopy observations of the particles also revealed that the particles with the highest shear strength were small (<2 µm) and were uniformly distributed, rather than aggregated, along the pipe wall. The findings suggest that cohesion was not responsible for the variable shear strength, but rather that shear strength originates from the placement of the particles in microscopic 'valleys' of the pipe wall roughness that shield them from the wall shear stress.

Despite the current advances in understanding material accumulation shear strength, several questions remain unanswered. Specifically, differences in the mechanisms of particulate material accumulation of low and high shear strength are not yet explained. The aim of this paper is to examine how particulate iron oxide that are rapidly deposited on PVC pipes develop variable shear strength under common hydraulic conditions found in drinking water distribution networks. The research was completed by means of a series of experiments where iron oxide was introduced into a full-scale laboratory system to produce controlled plumes of particles that were conveyed through the system to simulate the occurrence of a discolouration event in a DWDS. Five experiments independently tested three different suspended sediment concentrations (SSC) and three different conditioning fluid velocities. The shear strength of particle deposits was assessed through their mobilization using three incremental flushing stages realized after the pipes were conditioned.

2 METHODS

The experiments were realized in the Drinking Water Distribution Laboratory (DWDL) at Queen's University, using a full-scale pipe loop rig that simulates the operation of drinking water mains. The pipe laboratory system is comprised of a water tank with a volume of 3.6 m^3, two variable speed centrifugal pumps and 11 loops of IPEX Blue Brute PVC pipe Class 235 (DR18) with an internal diameter of 108 mm, and a total length of 193 m. During the experiments, the pipe loop system was operated in a non-recirculatory manner, where drinking water from the City of Kington was continuously added to the tank, pumped through the pipe system, and discarded at the outlet of the loop. Previous assessment of the local drinking water suggests a negligible amount of SSC (<0.1mg L^{-1}).

For the experiments, the incoming drinking water that was constantly fed into the inlet of the pipe loop was amended with iron oxide particles. This feed location was located after a straight pipe section of approximate 2 m reserved for the stabilization of the flow velocity profile. A chemical grade of red iron oxide powder from Alpha Chemicals, with a composition of 82% of Iron (III) oxide – Fe_2O_3, was chosen as the source of particles for the experiments due to their representativeness of iron oxide particles found in DWDSs and their stable particulate form and insolubility in water. To guarantee the repeatability of the powder particle size distribution (PSD), the full iron oxide sample was sent to a commercial laboratory to split it into sub-samples of 50 g using a mini rotary splitter from Retsch company model PT100. The PSD of two sub-samples picked at random was analysed with a Malvern Mastersizer 3000 particle size analyser (laser diffraction).

A total of five experiments were performed with a variable concentration of suspended particles (F1C1, F2C2 and F1C3 in Table 1) and a variable conditioning flow rate (F1C1, F2C1 and F3C1 in Table 1). Each experiment was divided into three independent sub-stages (P1, P2 and P3), where plumes of suspended particles were produced at the inlet of the pipe loop and transported to the pipe loop outlet before the introduction of the next plume in the next sub-stage. Each plume was created by injecting a known volume of a concentrated solution of iron oxide particles (CSIOP) at the inlet of the pipe loop with a known SSC over an injection period equal to the time required for the water to travel 60 m downstream of the injection point given the prevailing flow rate in the pipe loop. A diaphragm pump was used to pump water from the bottom of the CSIOP tank (approximately 10 cm away from the tank walls) from a position of intense mixing between two of the four vertical baffles. This water was injected at a constant flow rate at the inlet of the pipe loop at the centreline cross-sectional position of the pipe. At the end of each injection period, the steady flow conditions were maintained in the pipe loop for an additional period required to refresh the equivalent of 1.5 pipe loop volumes to guarantee that all suspended particles from the sediment plume had exited the pipe loop.

At the end of the conditioning stage of each experiment (passage of the three sequential plumes P1, P2 and P3), the system was flushed with three independent incremental steps at different intensities (FS1, FS2 and FS3 at Table 1) to assess the mobilization of particle deposits from the pipe walls. The first flushing step had a WSS of 1.2 Pa, which was previously suggested as an ultimate shear strength for materials accumulated in plastic pipes [7]. By comparison, the last flushing step produced an elevated WSS of 5.0 Pa, that is not commonly achieved in DWDSs even under fire fighting/flushing conditions [4]. After the end of the flushing phase, and before starting a new experiment with a specific set of conditions, the pipe system was flushed and cleaned at the highest achievable flow rate of 16 L s^{-1} (slightly above the flow of the FS3) for 5 minutes to establish identical initial condition for all the experiments.

***Table 1**. Hydraulic and water quality conditions for the experiments.*

Experiment	*SSC [a] (mg L^{-1})*	*Flow (L s^{-1})*	*Velocity (m s^{-1})*	*Reynolds Number*	*WSS [b] (Pa)*	*Pipe loop RT [c] (min)*
F1C1	*20.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F1C2	*40.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F1C3	*60.0*	*0.60*	*0.07*	*7 000*	*0.02*	*50.9*
F2C1	*20.0*	*1.80*	*0.20*	*21 000*	*0.11*	*17.0*
F3C1	*20.0*	*2.75*	*0.30*	*32 000*	*0.24*	*11.1*
FS1	-	*6.5*	*0.71*	*76 600*	*1.20*	*4.7*
FS2	-	*11.0*	*1.20*	*129 700*	*3.09*	*2.8*
FS3	-	*14.3*	*1.56*	*168 600*	*4.96*	*2.1*

a. Suspended solids concentration.

b. Wall shear stress based on the Darcy-Weisbach equation.

c. Pipe loop residence time (RT) calculated by dividing the pipe loop volume of 1.7 m^3 by the flow rate.

During all the experiments, the pipe flow rate and turbidity were continuously monitored at a sampling frequency of 1 Hz using two Sierra InnovaSonic® 205i Ultrasonic Flow Meters (±0.5%) and two online Hach TU5300sc turbidimeters (±0.01 NTU). The turbidimeters were configured to monitor turbidity at the inlet (7% of the pipe length) and outlet (98% of the pipe length). Turbidity was measured from water continuously sampled from the centre of the pipe at a constant sampling flow rate. Following each injection, grab samples were collected at the inlet and outlet sampling port locations to determine SSC at those locations and to assess changes in the particles in suspension after the passage of each plume through the pipe loop. The collection of grab samples was synchronized with the fluid velocity in the pipe loop to collect 4 L of water from the centre of the passing sediment plume. During the flushing phase, SSC grab samples were synchronized to collect water from the centre region of the pipe loop that corresponded to a region between ¼ and ¾ of the total pipe loop length. The SSC of the grab samples was determined using a dry-weight method, by filtering the sample volume with pre-weighed 0.45 μm glass microfiber filters, drying the filters at 105°C for 1 hour, and weighing them again after drying with a precision scale. Additional details of the experimental design can be found in Braga and Filion [17].

The continuous turbidity data of the conditioning and flushing phases of the experiments was offset by the background turbidity of the local drinking water and converted to SSC data through the calibration of linear conversion coefficients κ (mg L^{-1} NTU^{-1}). For the conditioning phase, the constant source of particles and steady flow resulted in very stable relationships between SSC and the turbidity increase caused by the iron oxide particles addition [17]. It is noted here that the calibration of two different coefficients for the inlet and outlet positions was required due to changes in the PSD during the passage of the plume through in the pipes. A coefficient $\kappa = 0.24$ mg L^{-1} NTU^{-1} was obtained for the inlet, and $\kappa = 0.15$ mg L^{-1} NTU^{-1} was obtained for the outlet. In the case of the flushing stage of the experiment, the PSD of the particles mobilized from the walls was not stable and varied across each flushing step, experimental condition, and position of the pipe where particles were mobilized. To a reliably convert turbidity data into SSC data, it was assumed that each flushing step produced a similar PSD, and coefficients κ were calculated through the average of individual SSC grab sample coefficients from all experiments for each flushing step. A $\kappa = 2.5$ mg L^{-1} NTU^{-1}, $\kappa = 6.0$ mg L^{-1} NTU^{-1}, and $\kappa = 5.2$ mg L^{-1} NTU^{-1} were obtained for the FS1, FS2 and FS3, respectively. The average error of the coefficients is approximately 30%.

Following the conversion of the continuous turbidity data (TURB) to SSC data using the constant coefficients κ for each conditions highlighted above, a continuous particle mass flux Q_{SS} (mg s^{-1}) was obtained by the product of the instantaneous flowrate Q (L s^{-1}) and the SSC (mg L^{-1}) (Equation 1). A mass balance analysis of the continuous particle mass flux at the inlet and outlet of the pipe loop made is possible to estimate the mass of iron oxide particles that remained attached to the pipe walls during the conditioning phase and the mass of iron oxide particles that were mobilized during the flushing phase.

$$Q_{SS} = Q \times SSC = Q \times (\kappa \times TURB) \tag{1}$$

3 RESULTS AND DISCUSSION

The results of the iron oxide particle mass flux produced by the passage of the plumes during the conditioning phase of each experiment is presented in Figure 1 and highlights the differences between the plumes at the inlet and outlet positions of the pipe loop. Figure 1a shows the three experiments at different SSC, where an evident decrease of the particle mass flux occurred between the inlet and outlet. The results among the three experiments were consistent and they point to a decrease of 46% of particles between the two turbidimeter positions (22% of the injected particle load). However, it was previously estimated that 55% of the particles remained in the short length of pipe between the particle injection point and the inlet turbidimeter for these 3 experiments [17]. The combined total fraction of the particles injected in the pipe loop that accumulated in the pipes for the F1C1, F1C2 and F1C3 was estimated to be 75%.

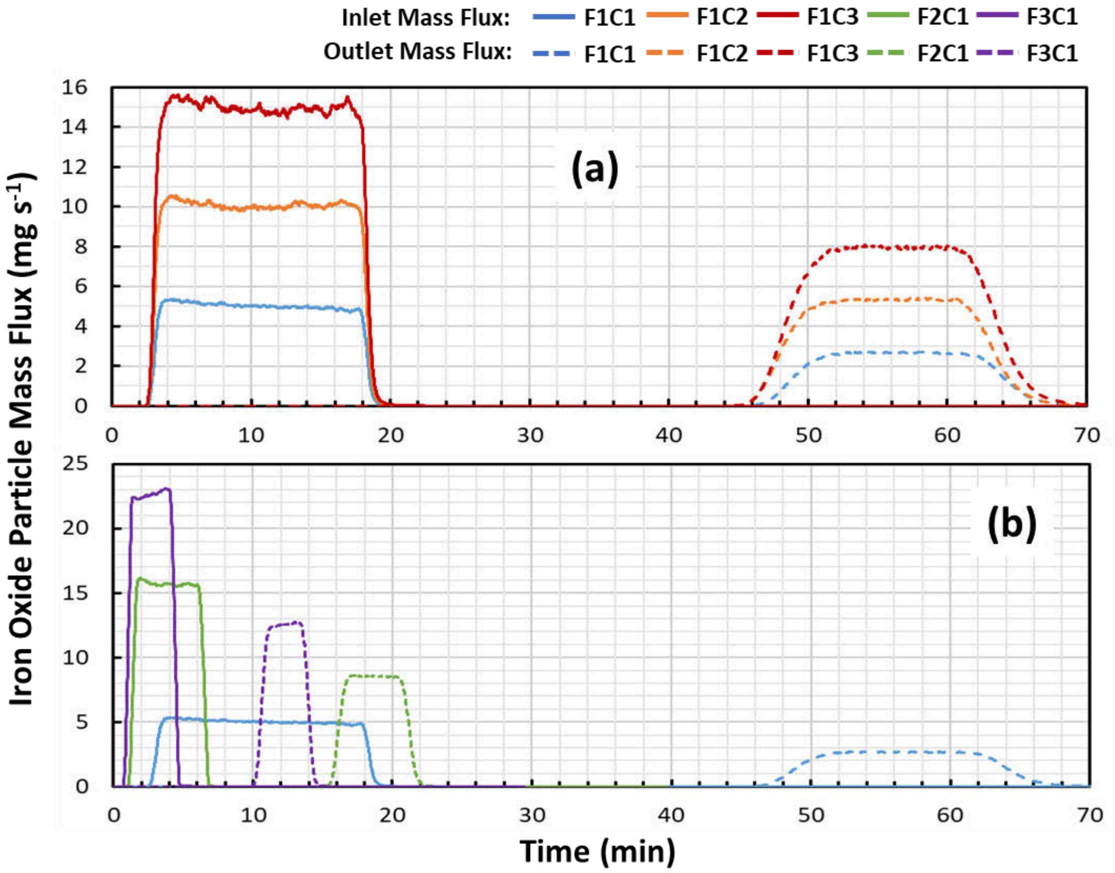


Figure 1 – Iron oxide particle mass flux of the suspended sediment plumes produced during the conditioning phase of the experiments at the inlet and outlet positions during the a) three experiments at different SSC and constant velocity (F1C1, F1C2, F1C3); and b) three experiments at different velocities and constant SSC (F1C1, F2C1, F3C1).

In the case of the experiments with different fluid velocities and constant SSC (Figure 1b), the decrease in the peak of the particle plume mass flux is even more evident in experiments F2C1 and F3C1 which were notably different than F1C1. Note that even if the SSC between the experiments is similar, the particle mass flux of the experiments at higher fluid velocities is substantially higher due to their higher flow rates. In terms of the mass balance for F2C1 and F3C1, a similar constant reduction fraction of 46% was obtained between the turbidimeters locations (in this case representing 25% of the injected particle load). However, a slight decrease of the particle fraction in the initial pipe length to 44% caused a reduction of the total load accumulated in the pipes to 69%. These results suggest that the main difference in the particle deposition load between the experiments at higher velocities occurred in the initial pipe sections.

Further, the results for the particle mass flux at the pipe outlet during the flushing phase of the experiments is presented in Figure 2. Substantial particle loads were mobilized during the three sequential flushing steps. The highest load was observed during flushing step FS1 (Figure 2a), and smaller loads were observed during the subsequent flushing steps FS2 (Figure 2b) and FS3 (Figure 2c). The results provide evidence that: i) the particles accumulated in the pipes had different shear strengths, and ii) these material deposits were produced after the rapid passage of particle plumes and not as a result of a continuous particle load as reported in the literature [16, 18]. However, in comparison to previous experiments that tested for chronical loading of particles, the results of FS1 in Figure 2a shows unusual peaks in the particle mass flux around 5 min, which corresponded to the initial pipe sections where most of the accumulation occurred.

The results suggest that this fraction of accumulated particles likely settled onto the invert of the pipes once the conditioning velocity was not able to maintain their suspension and were held to the pipe wall by weak adhesion forces [16]. In addition, the peak in particle mass flux in flushing step FS1 decreased in the experiments with higher velocities and this agrees with the observed reduction of particle fractions that accumulated on the initial pipe lengths in these experiments.

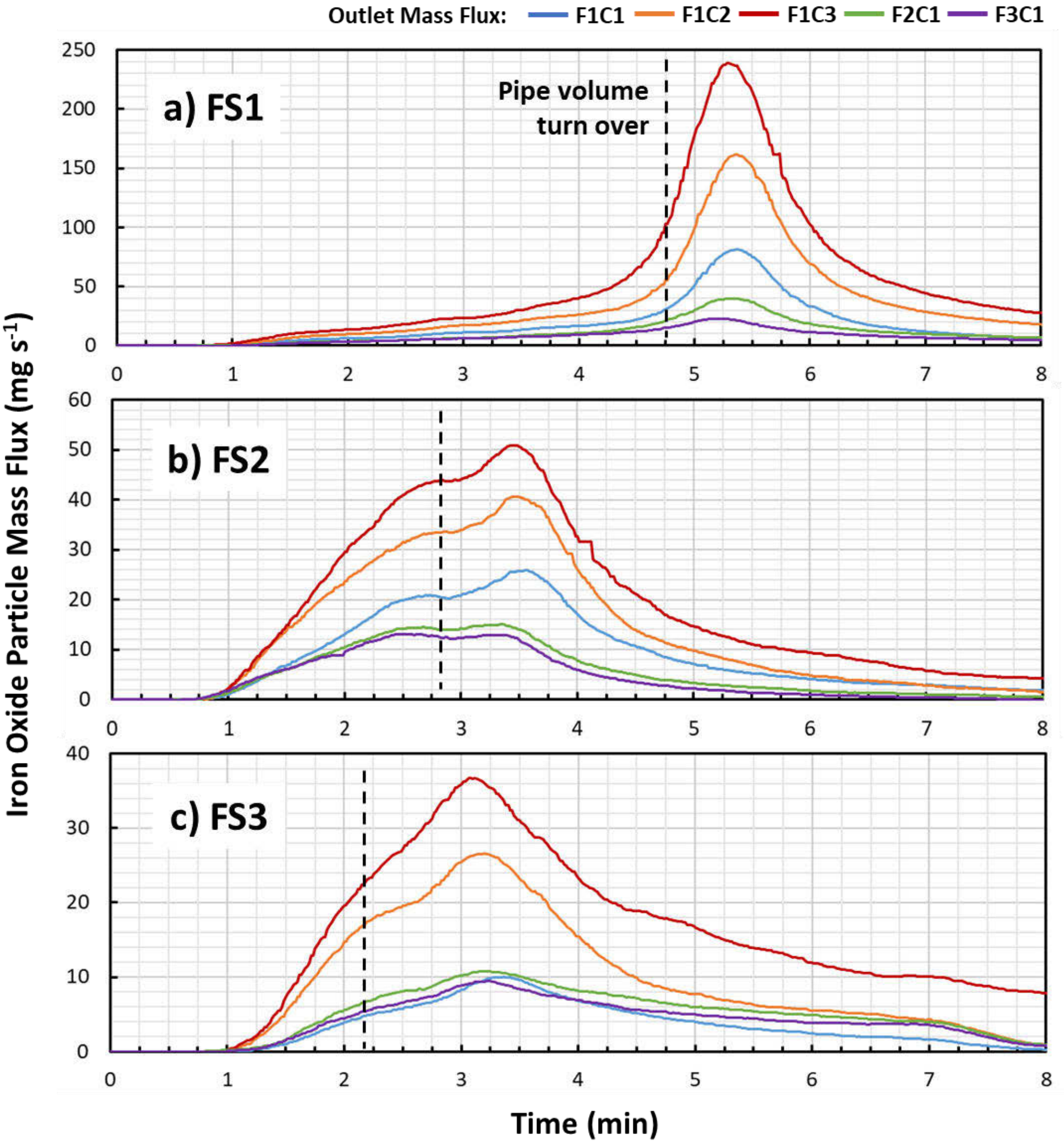


Figure 2 – Iron oxide particle mass flux observed at the outlet turbidimeter location of the pipe loop during the three flushing steps (FS1, FS2 and FS3) for all experiments.

Higher loads of mobilized particles for all flushing steps were observed for the experiments at higher concentrations (F1C2 and F1C3) that produced higher loads of deposited particles on the pipe walls. By comparison, differences between the three experiments at similar SSC (F1C1, F2C1 and F3C1) were larger for the FS1 (Figure 2a), where the F1C1 had the larger mobilized load of particles. However, these differences tapered out for the FS2 and FS3. The particle mass flux peaks of the flushing steps FS2 and FS3 (with corresponding to the initial pipe sections) was also substantially reduced in comparison to the flushing step FS1 (Figure 2bc). This demonstrates that particle attachment with higher shear strength occurred for longer lengths of the pipe loop rather

BY NC SA

than predominantly at the first sections. However, the particle mass flux profiles of the flushing steps FS2 and FS3 also shows a slight increase towards their peaks (slightly after the pipe loop volume turn over) rather than stable level. This result suggests that the accumulated particle load decreased along the pipe loop length. Previous analysis of the data proposed that only a finite fraction of particles was suitable for attachment on the pipes in each experimental condition [17] For this reason, the depletion of the particle fraction along the plume passage explains the decrease in mobilized particles in downstream sections of the pipe loop (firstly measured by the outlet turbidimeter between the times of 1 and 3 minutes in Figure 2bc).

Lastly, the difference between the load of accumulated and mobilized particles in Figure 2 revealed that a consistent fraction of 31% of the accumulated particles were not mobilized during the flushing in the F1C1, F1C2 and F1C3 experiments. The non-mobilized fraction of particles increased to 39% and 52% in the F2C1 and F3C1 experiments, respectively. Although previous research suggested that the last flushing step at a WSS of 5.0 Pa was insufficient to mobilize all particles from the pipe loop [16, 17], it is unlikely that the remaining particle load in the pipes totalled 30% of the inoculated load. Is also unlikely that the non-mobilized fraction of particles increased for experiments with higher velocities. In this case, the discrepancies are better explained by errors in the particle load estimations of the flushing stage based on turbidity data. Since the collection of only one grab sample per flushing step was feasible to analyse SSC due to the requirements of a large sample volume. This sample corresponded to the middle section of the pipe loop, and it is likely that the SSC peak (corresponding to the initial pipe sections) was underestimated. A better estimation of the particle load during flushing would require a higher resolution of SSC data, specifically during the passage of the particles mobilized from the initial pipe lengths where a variable PSD is expected. Alternatively, constant turbidity to SSC calibration factors may be obtained inversely from the assumption that all attached particles during the conditioning phase were mobilized. But this approach would average the differences of PSD along the mobilization profile, and still be inaccurate to investigate the rapid dynamics of SSC during the flushing operations.

4 CONCLUSION

The aim of this paper is to examine how particulate iron oxide that are rapidly deposited on PVC pipes develop variable shear strength under common hydraulic conditions found in drinking water distribution networks. A series of controlled laboratory experiments in a full-scale system were completed in which selected particles were inoculated in the system under steady conditions and rapidly deposited on the pipe walls of the system through the controlled passage of a sediment plume that simulated a discolouration event. Three different SSC and three different fluid velocities were tested. After the pipe wall conditioning through the sediment plume passage, the pipes were flushed at three incremental steps with WSS of 1.2 Pa, 3.1 Pa and 5.0 Pa to mobilize accumulated particles to test their shear strength resistance to the imposed WSS.

The results showed that large fractions of the inoculated particles (69% – 75%) remained in the pipe loops after the passage of the concentrated sediment plume. The predominant fraction of the accumulated particles occurred in the initial pipe sections of the system, and experiments with higher velocities (>0.2 m s^{-1}) produced only a slight decrease in particle accumulation. Data obtained from the flushing stage of the experiments confirmed that most of particles were accumulated in the initial pipe sections of the loop and revealed that a considerable fraction of the accumulated particles developed a higher shear strength. Particle mobilization profiles from the second and third flushing steps showed that particles with higher shear strength accumulated through longer pipe sections, but progressively decreased with the pipe length. This behaviour agrees with a previous hypothesis that a finite pool of particles amenable to pipe wall attachment is determined by the conditioning fluid velocity.

The experimental results presented here were obtained from conditions that only partially represent operational DWDSs. However, the results contributed to substantial improvements in the understanding of particulate deposit formation and have the potential to also contribute to the better management of material accumulation in operational systems. Further research is required to examine the current results against more complex conditions of DWDSs and connect them to concurrent material accumulation processes.

5 REFERENCES

[1] A. Carrière, V. Gauthier, R. Desjardins, and B. Barbeau, "Evaluation of loose deposits in distribution systems through: unidirectional flushing," Journal - American Water Works Association, vol. 97, no. 9, pp. 82-92, 2005, doi: 10.1002/j.1551-8833.2005.tb07474.x.

[2] A. Mussared, R. Fabris, J. Vreeburg, J. Jelbart, and M. Drikas, "The origin and risks associated with loose deposits in a drinking water distribution system," Water Supply, vol. 19, no. 1, pp. 291-302, 2019, doi: 10.2166/ws.2018.073.

[3] J. H. Vreeburg and J. B. Boxall, "Discolouration in potable water distribution systems: a review," Water Res, vol. 41, no. 3, pp. 519-29, Feb 2007, doi: 10.1016/j.watres.2006.09.028.

[4] S. Husband and J. B. Boxall, "Field Studies of Discoloration in Water Distribution Systems: Model Verification and Practical Implications," Journal of Environmental Engineering, vol. 136, no. 1, pp. 86-94, 2010, doi: 10.1061/(asce)ee.1943-7870.0000115.

[5] A. S. Benson, A. M. Dietrich, and D. L. Gallagher, "Evaluation of Iron Release Models for Water Distribution Systems," Critical Reviews in Environmental Science and Technology, vol. 42, no. 1, pp. 44-97, 2012, doi: 10.1080/10643389.2010.498753.

[6] D. M. Cook and J. B. Boxall, "Discoloration Material Accumulation in Water Distribution Systems," Journal of Pipeline Systems Engineering and Practice, vol. 2, no. 4, pp. 113-122, 2011, doi: 10.1061/(asce)ps.1949-1204.0000083.

[7] P. S. Husband and J. B. Boxall, "Asset deterioration and discolouration in water distribution systems," Water Res, vol. 45, no. 1, pp. 113-24, Jan 2011, doi: 10.1016/j.watres.2010.08.021.

[8] J. B. Boxall and A. J. Saul, "Modeling Discoloration in Potable Water Distribution Systems," Journal of Environmental Engineering, vol. 131, no. 5, pp. 716-725, 2005, doi: doi:10.1061/(ASCE)0733-9372(2005)131:5(716).

[9] R. L. Sharpe, C. J. Smith, J. B. Boxall, and C. A. Biggs, "Pilot Scale Laboratory Investigations into the Impact of Steady State Conditioning Flow on Potable Water Discolouration," in Water Distribution Systems Analysis 2010, 2011, pp. 494-506.

[10] W. R. Furnass, R. P. Collins, P. S. Husband, R. L. Sharpe, S. R. Mounce, and J. B. Boxall, "Modelling both the continual erosion and regeneration of discolouration material in drinking water distribution systems," Water Supply, vol. 14, no. 1, pp. 81-90, 2014, doi: 10.2166/ws.2013.176.

[11] S. Liu, C. Gunawan, N. Barraud, S. A. Rice, E. J. Harry, and R. Amal, "Understanding, Monitoring, and Controlling Biofilm Growth in Drinking Water Distribution Systems," Environ Sci Technol, vol. 50, no. 17, pp. 8954-76, Sep 6 2016, doi: 10.1021/acs.est.6b00835.

[12] K. E. Fish, A. M. Osborn, and J. Boxall, "Characterising and understanding the impact of microbial biofilms and the extracellular polymeric substance (EPS) matrix in drinking water distribution systems," Environmental Science: Water Research & Technology, vol. 2, no. 4, pp. 614-630, 2016, doi: 10.1039/c6ew00039h.

[13] J. Boxall, I. Douterelo, K. E. Fish, and S. Husband, "Linking discolouration modelling and biofilm behaviour within drinking water distribution systems," Water Supply, vol. 16, no. 4, pp. 942-950, 2016, doi: 10.2166/ws.2016.045.

[14] A. Seth, R. Bachmann, J. Boxall, A. Saul, and R. Edyvean, "Characterisation of materials causing discolouration in potable water systems," Water Science and Technology, vol. 49, no. 2, pp. 27-32, 2004, doi: 10.2166/wst.2004.0080.

[15] S. Husband, K. E. Fish, I. Douterelo, and J. Boxall, "Linking discolouration modelling and biofilm behaviour within drinking water distribution systems," Water Supply, vol. 16, no. 4, pp. 942-950, 2016, doi: 10.2166/ws.2016.045.

[16] A. S. Braga and Y. Filion, "Initial stages of particulate iron oxide attachment on drinking water PVC pipes characterized by turbidity data and brightfield microscopy from a full-scale laboratory," Environmental Science: Water Research & Technology, 10.1039/D2EW00010E 2022, doi: 10.1039/D2EW00010E.

[17] A. S. Braga and Y. Filion, "The interplay of suspended sediment concentration, particle size and fluid velocity on the rapid deposition of suspended iron oxide particles in PVC drinking water pipes," Water Research X, vol. 15, p. 100143, 2022/05/01/ 2022, doi: https://doi.org/10.1016/j.wroa.2022.100143.

[18] I. Sunny, P. S. Husband, and J. B. Boxall, "Impact of hydraulic interventions on chronic and acute material loading and discolouration risk in drinking water distribution systems," Water Research, vol. 169, p. 115224, 2020/02/01/ 2020, doi: https://doi.org/10.1016/j.watres.2019.115224.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14095

W-NET4.0 – INTEGRATED PLATFORM FOR SMALL AND MEDIUM SIZED WATER SUPPLY UTILITIES

Jochen W. Deuerlein[1], Thomas Bernard[2], Naga Mamatha Gonuguntla[3], Jorge Thomas[4], Armin Canzler[5], Heiko Keifenheim[6], Susanne Wiese[7], Rüdiger Höche[8], Joachim Rapp[9], Salomé Parra[10]

[1, 10] 3S Consult GmbH, Garbsen, Germany
[2, 3, 4] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany
[5, 6, 7] COS Geoinformatik GmbH & Co. KG, Ettlingen, Germany
[8] Stadtwerke Buehl GmbH, Buehl, Germany
[9] Schwarzwaldwasser GmbH, Buehl, Germany)

[1] deuerlein@3sconsult.de, [2] thomas.bernard@iosb.fraunhofer.de,
[3]naga.mamatha.gonuguntla@iosb.fraunhofer.de, [4] jorge.thomas@iosb.fraunhofer.de,
[5]armin.canzler@cosgeo.de, [6]heiko.keifenheim@cosgeo.de,
[7]susanne.wiese@cosgeo.de, 8Ruediger.Hoeche@stadtwerke-buehl.de,
[9]joachim.rapp@sw-wasser.de, [10]parra@3sconsult.de

Abstract

More than 5,800 companies ensure the supply of drinking water in Germany. The vast majority of these are small and medium-sized water supply companies. So far, they are using information and automation technology to a very limited extent, measurement data is usually not collected systematically. In addition, the companies often have neither a sufficiently well-maintained database of the water network nor geoinformation systems (GIS), simulation software or data analysis tools that can be used to plan and optimise interventions in the drinking water system (e.g. expansion of the network). Large water suppliers are often better equipped technically, but the data is often not used in an efficient manner.

The objective of the W-Net 4.0 research project (www.wnet40.de) was to develop a modular and scalable platform that combines GIS, simulation software and data analysis tools and meets high IT security standards. Combined with novel service concepts, value-added networks and training concepts, also small and medium-sized water utilities will be enabled to use these technologies. The high degree of usability and accessibility of the platform supports the daily work of planning engineers, and network operators. The paper includes a detailed description of the platform and use cases. It also emphasizes the training program tailored for small and medium sized utilities.

Keywords
Integration, platform, GIS, simulation, digital twin, monitoring, dashboard, training, e-learning.

1 INTRODUCTION

Successful management of drinking water supply systems depends highly on the availability and quality of data describing the system and its operation. During the last decades sophisticated software tools for hydraulic network simulation, planning and management have been developed and are widely used by researchers and practitioners. However, the successful application of these models to solve real world systems requires sufficient information and data from the real system, which is often not available. The creation of a hydraulic simulation model is very often a tedious and time-consuming task with several remaining open questions at the end of the process. In this paper a holistic approach will be presented that places an emphasis on data acquisition, data management and sharing of data between different processes and that is tailored to the specific needs of small and medium sized water supply utilities.

One of the main problems in data collection and processing, especially for small water utilities, is that they do not know the actual condition of their water distribution systems [1,2]. The lack of data collection or the inadequate documentation are the result of the following deficiencies:

- The documentation within a pipeline network cadastre (GIS) and a regular data updating often do not take place. In the case of small water utilities, digital information about the pipe system (location, material, year of installation) is often missing and network maps, if available, only exist as analogous print outs from the time the network was built.
- Pipeline damages and maintenance measures are often not documented, so that failure statistics (necessary for forecasting rehabilitation measures as part of a modern asset management) cannot be kept.
- The amount and development of water losses are often not known, so that a water balance cannot be drawn up.
- After a significant change in the water consumption (e.g. after the expansion of the network or the connection of new large-scale consumers), the hydraulic capacity of the water distribution system is usually not verified.
- If available, operational data recorded using SCADA systems is usually not visualized and analysed adequately for achieving a better understanding of the real system behaviour.

Another decisive factor for a successful water system management are the human resources. Especially small and medium sized water suppliers are facing difficulties in acquiring adequate staff. As an example, the analysis of the small water utilities in Bavaria, Germany, revealed that the main challenge for the water utilities rely mainly on the lack of qualified personnel [3].

For improving the situation, the SchwarzwaldWASSER e.V. Cooperation Association was founded in 2008 in the southwest of Germany, the region of the Black Forest. It has roughly 60 members and serves as a strategic partner for small and medium sized communal water supply utilities. It bundles different kind of technical and commercial services for its members including lab testing for ensuring drinking water quality and engineering consultancy. SchwarzwaldWASSER e.V. also provides a platform for exchange of experience and knowledge, for example by regular meetings. In recent years, one technological focus has been on the digitization in water supply.

In this context, the R&D project W-Net 4.0 was initiated with the objective of making digitization and industry 4.0 technologies available also to small and medium sized water supply utilities. The idea is to provide a common central platform to the members so that they can share not only their knowledge but also hardware and specialized software tools that are tailored to their specific needs. Using the same digital platform instead of implementing individual solutions reduces cost and supports a lively exchange between the members. The project consortium of the R&D project W-Net 4.0 consists, besides SchwarzwaldWASSER, of the two SMEs (COS Geoinformatik: GIS software and services and 3S Consult: hydraulic simulation software and services), the Fraunhofer Research Institute ISOB and four end-users as associate partners.

Within the project, an integrated digital platform for improving the documentation and operational processes in water utilities was developed and implemented. It combines system documentation and mobile maintenance apps with hydraulic simulation and analysis tools for measurement data [4]. It can be understood as a first step towards the implementation of a digital twin of the water distribution system keeping in mind that there exist numerous definitions for a digital twin. In the context of W-Net 4.0, the platform contains a virtual model of the real physical system with all its properties that are necessary for network maintenance and operation. So far, the data flow is mostly one way from the physical system to the model, so that also the term Digital

Shadow may be used. However, in future development steps also interactions from the virtual to the real world are foreseen. That is why the term Digital Twin is used. For a detailed discussion about the differences between the two concepts see for example [5, 6].

The software is tailored to small and medium sized utilities, that usually do not have specialized academic staff for asset management and hydraulic system analysis. The objective of the project is to provide a tool that is simple in its application and can assist the network operators on their daily work on site.

The W-Net4.0 platform includes:

- a compact and user-friendly User Interface (UI)
- a complete definition of the standard processes
- efficient interfaces for model integration (GIS, hydraulic simulation, and sensor data analysis).

The project aims at increasing the attractivity of the jobs in water supply utilities by introducing modern digital tools that simplify the work of the operators and continuously improve the data base and documentation capabilities of the utility. Examples include the mobile apps for documentation and for the implementation of digital failure statistics.

The W-Net4.0 approach also includes a comprehensive training for the operators based on a self-learning program, which includes theoretical background, the application of technical rules (provided by the DVGW – “Deutsche Verein des Gas- und Wasserfaches e.V”, in English: “German Technical and Scientific Association for Gas and Water”) and the usage of the integrated software platform. For the training, a digital water distribution system model was created using data of a real water network. The model servers as training simulator and is available online.

In the following chapter the central platform is introduced, and its three main components are briefly summarized: 1) documentation and GIS; 2) GIS-integrated hydraulic simulation, 3) data analysis, with a clear focus on 2). Then, in the next chapter, the use cases defined by the end users of the project are briefly described followed by the final chapter on the developed training concept and services.

2 W-NET 4.0 CENTRAL PLATFORM

2.1 Overview

The W-Net 4.0 platform consists of the three main applications GIS, data analysis and hydraulic simulation. The data is stored in a central database that relates to the GIS construction tools and the web-based information system. The latter presents the most recent secured state of the data model. The web service component integrates the API (Application Programming Interface) of the hydraulic simulation engine and the data analysis tools for network monitoring and dashboard visualization of sensor data (see Figure 1).

In the following subsection the GIS is briefly outlined followed by a more detailed description of the GIS-integrated hydraulic simulation which is accessible by the web-based information system through the browser running on a personal computer or mobile devices such as tablets or smart phones. Also accessible via the web is the data analysis section of the platform that is presented in the subsequent subsection.

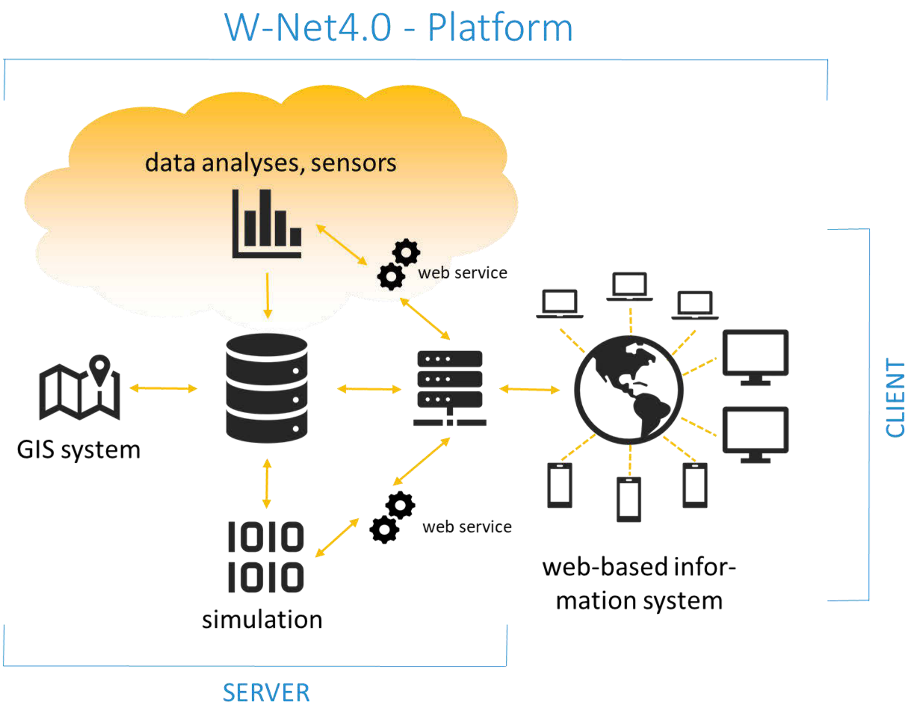


Figure 1. Components of the W-Net 4.0 digital platform

2.2 Geographic Information System

The GIS system that is used in the project provides numerous specialist modules, e.g., for water, gas, electricity, district heating and telecommunication networks. The data is stored in a central database server (Oracle). As an open and flexibly configurable system it enables to dynamically connect data from third-party databases, for example the consumption data or real time data from the SCADA system.

The presentation, visualization and evaluation of the entire database takes place in a web-based information system (see Figure 2). Here, authorized users can evaluate also the effects of operations by playing through scenarios that are interlinked with the hydraulic simulation module; for example, how pressures and flow velocities change when closing certain shut-off valves. Central rights management regulates the access of the various user groups. Data can be sent on site by smartphone directly via the web to the data server including for example the documentation of maintenance measures, valve manipulations etc.

Data collection via app with mobile devices is integrated into the data flow of the W-Net4.0 platform, from the water system on site to the simulation. The information flows automatically into the databases of the W-Net 4.0 platform, where it is processed further. The water utility's staff can handle a variety of tasks with these easy-to-use apps, such as recording maintenance measures for the plants and transmitting newly detected leaks, damages, valve position changes that are important for the simulation and much more.

As part of the W-Net 4.0 project, the integration of the GIS software COSVega with the simulation platform SIR 3S® and the data analysis tools developed by Fraunhofer IOSB has been realized.

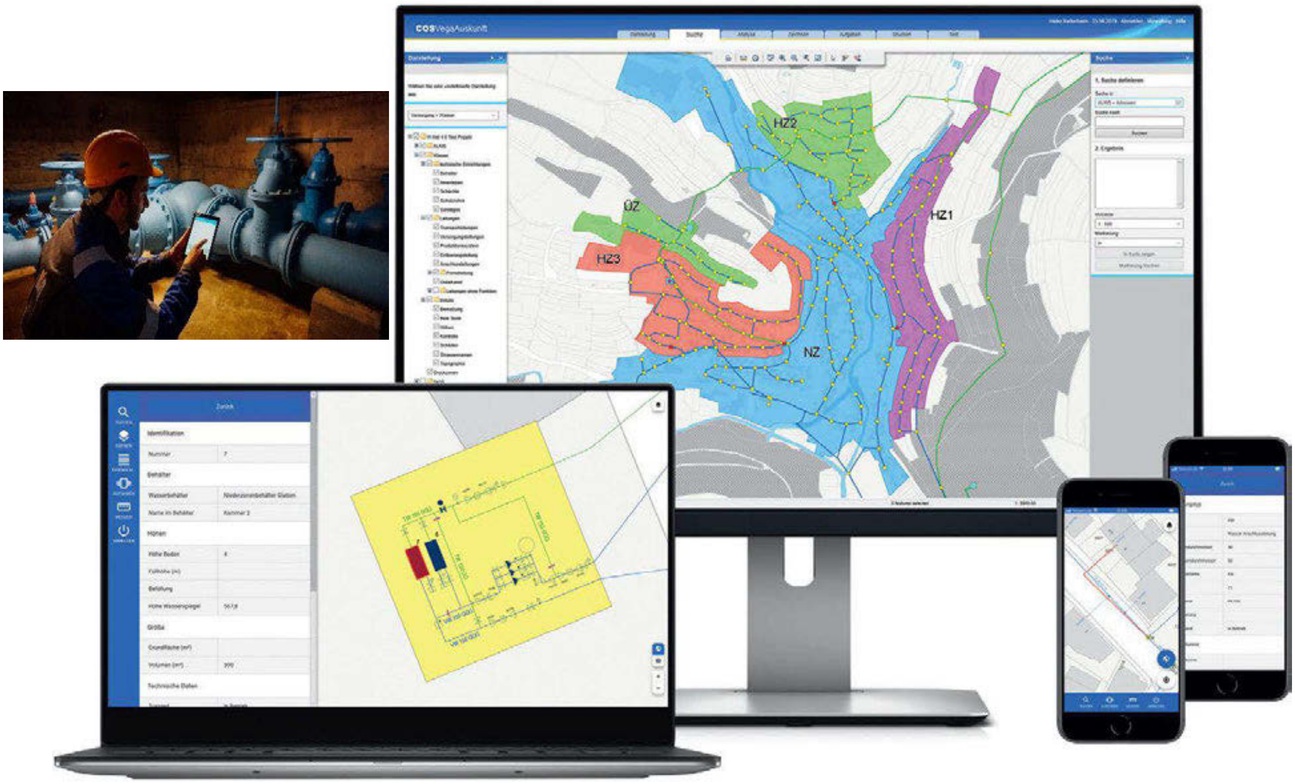


Figure 2: Web-based GIS platform with mobile maintenance apps and integrated hydraulic simulation

2.3 GIS-Integrated Hydraulic Simulation (HS)

The web-platform also provides modules for the hydraulic simulation of the water distribution system. The traditional way of creating a hydraulic simulation model is to collect data from different sources (GIS, Consumption billing, Digital Elevation Model, …) and to compile it into a consistent model. To enhance this process different interfaces have been developed by the provider of hydraulic simulation software. With their help data can be automatically imported in the simulation model. However, this process is one way in nature, without any automatic feedback to the GIS data source about the quality of the data. In most cases the data that is imported is insufficient for hydraulic network calculations because of incompleteness, topological errors, data errors, missing data, wrong data types etc.

In W-Net 4.0 the import of data is replaced by so-called model data integration of GIS and hydraulic simulation. There is a continuous connection between the two data models while each of them maintaining its individual structure. This allows data exchange in two directions. From both sides, GIS and HS, an agreement was made on the properties of an exchange database that includes the most important properties of the network and its operation. The exchange database is the minimal compromise of data that are relevant for hydraulic modelling. For implementation of the two-directional data integration with the exchange database both sides, GIS and HS, had to develop their individual interpreters (see Figure 3).

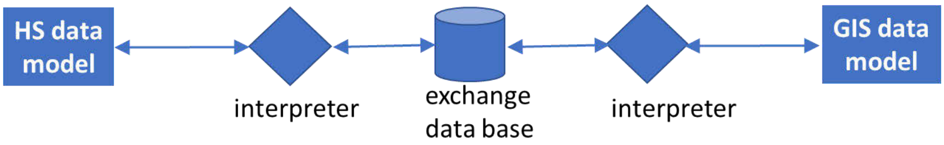


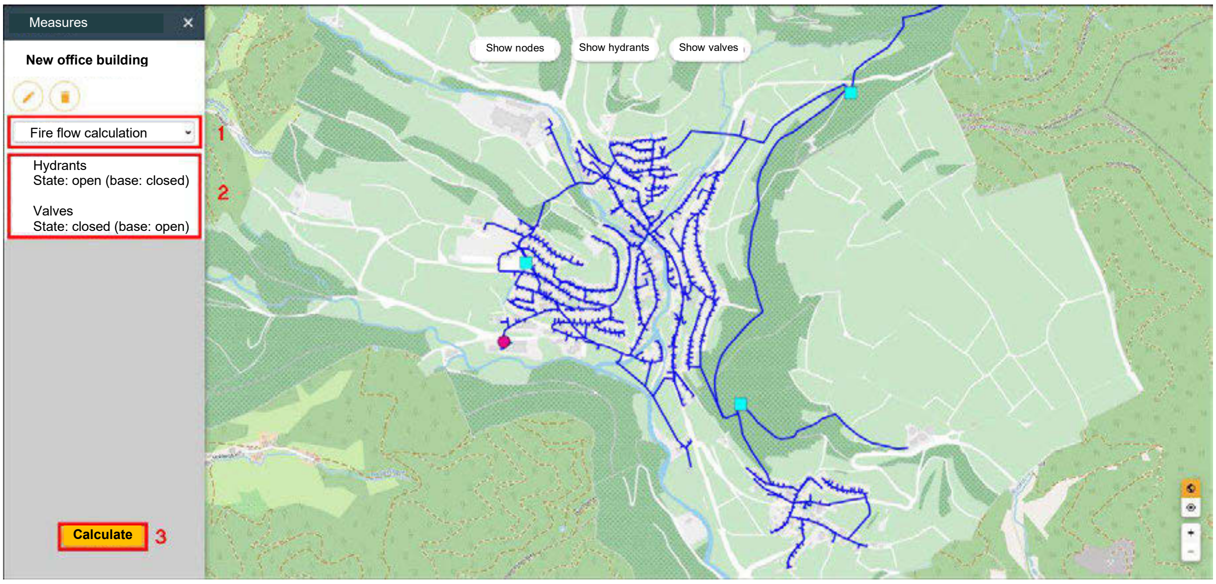
Figure 3: Model data integration of GIS data with Hydraulic Simulation data

CC BY NC SA

The GIS data model includes all the information required for running steady-state calculations. Since pipe systems of public water supply utilities are dynamic systems, the properties of their components are continuously changing. The changes may be caused by construction work, new consumer connections, pipe removal or rehabilitation, valve operations etc. To maintain the integrity of the model a defined state is frozen and disconnected from the construction platform. When the constructions are closed and have reached another defined state, the frozen model is replaced by an updated version. This continuous process guarantees, on the one hand, that always a calculable model is available in the web information system. On the other hand, the intervals for model updates are short enough to represent the physical system adequately enough in the hydraulic simulation model.

The connection between the GIS information system and the hydraulic simulation software is realized by the integration of a newly developed API. In addition to the interpreter for model data integration, the API includes a model edit interface for inserting, updating, and deleting single network components. The model edit interface is used for scenario calculations. A scenario consists of the base model (current published frozen state of the GIS model) and a collection of selected data changes. The advantage of this approach is that the base model must be loaded only once at the beginning of the session. The scenario calculations require minimal data manipulations that are executed using the mode edit interface.

The management of the scenarios is part of the web information system. Any number of scenarios can be created by the user. Each scenario includes a list of actions that distinguish it from the base model. For the hydraulic calculation the scenarios can be combined with four predefined load cases: peak hour demand, fire flow calculation, average demand, and stagnation. Figure 4 shows a fire flow calculation scenario with the corresponding load case selected (ComboBox 1). The text box 2 includes a brief description of the measures (changes compared to the base model): hydrant opened, and valve closed. By pushing button 3 the hydraulic calculation is launched.



New office building

Fire flow calculation

Hydrants
State: open (base: closed)

Valves
State: closed (base: open)

Figure 4: Web-based GIS platform with fire flow calculation scenario

The underlying GIS model distinguishes from other GIS in the fact that all the interior of pumping stations, storage tanks, valve chambers etc are also included in the model. Conventional GIS normally include the pipe system, end at the wall of buildings and exclude the interior of operational stations. In the W-Net 4.0 approach the GIS includes all data that is necessary for calcualting the hydraulic steatdy-state of the base model. No further adaptation is required in the hydraulic simulation engine. Figure 5 shows as an example the model of a storage tank with two chambers (1). The tank is supplied by the transport pipe on top (2). The inflow is controlled by a pressure sustaining valve (3) in order to maintain sufficient pressure in the upstream network. From the storage tank the water flows into the low pressure zone (4) and part of it is pumped to a high pressure zone (5). The pumping station consists of three parallel branches of series of valve (6), pump (7), check valve (8) and valve (9).

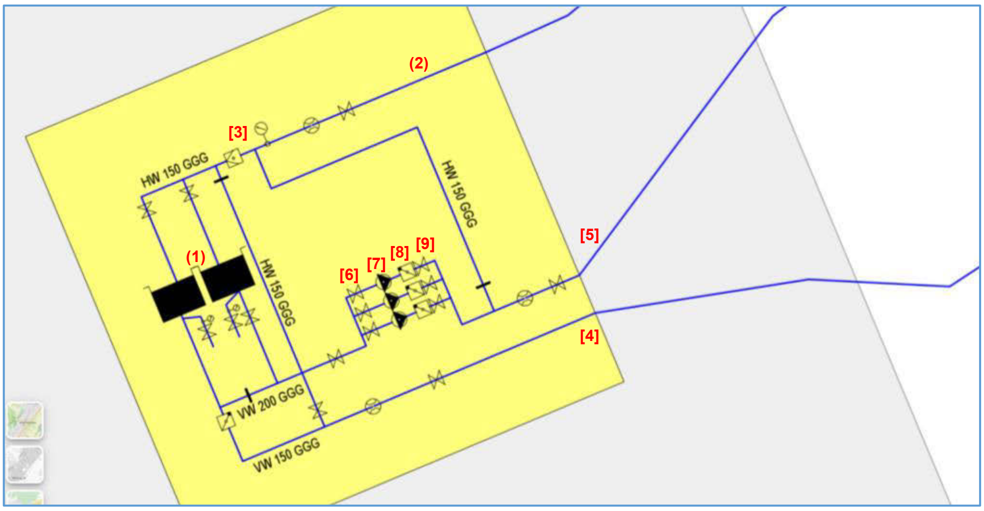
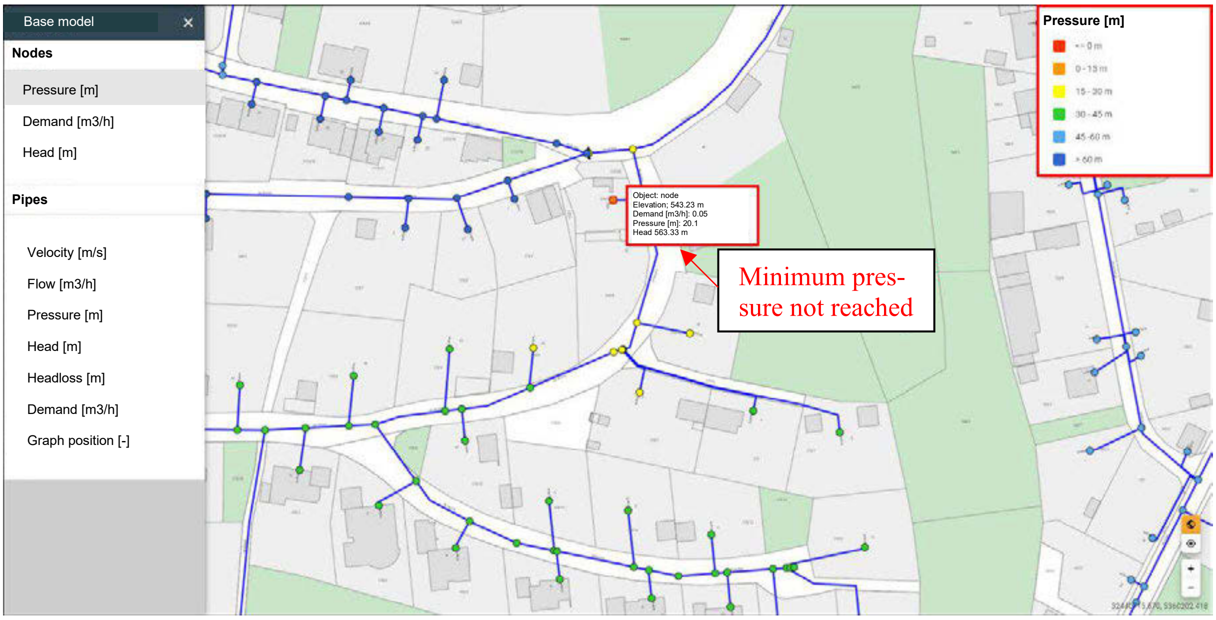

(2)
[3]
HW 150 GGG
(1)
HW 150 GGG
HW 150 GGG
[6]
[7]
[8]
[9]
[5]
[4]
VW 200 GGG
VW 150 GGG

Figure 5: Interior of storage tank with two chambers, pressure sustaining valve and pumping station

After the hydraulic calculation the results are provided by the API of the simulation module through a memory data set that is accessible by the web information system. The results are presented by individual tool tips for the network components (nodes, pipes, valves, etc.) and by coloration of nodes and pipes (see Figure 6).


Base model
Nodes
Pressure [m]
Demand [m3/h]
Head [m]
Pipes
Velocity [m/s]
Flow [m3/h]
Pressure [m]
Head [m]
Headloss [m]
Demand [m3/h]
Graph position [-]
Pressure [m]
<= 0 m
0 - 15 m
15 - 30 m
30 - 45 m
45 -60 m
> 60 m
Minimum pressure not reached

Object: node
Elevation; 543.23 m
Demand [m3/h]: 0.05
Pressure [m]: 20.1
Head 563.33 m

Figure 6: Calculation results for nodal pressures of fire flow calculation with violation of minimum pressure requirement of 1.5 bar at elevated demand node (DVGW W 405 [7])

For getting realistic simulation results that reflect the real system's behaviour, it is very important that the topology of the network is modelled correctly. In the simulation mode also the results of the decomposition of the network graph into looped blocks, bridges and forest [8] can be visualized. In Figure 7 the pipes at position 1 and 2 visually seem to be part of the looped network. Although they are signed as forest links. A more careful investigation shows that the pipes are disconnected (see Figure 8). The reason for the disconnection should be clarified: is it intentional and does it reflect the real system or is it a possible data error?

CC BY NC SA

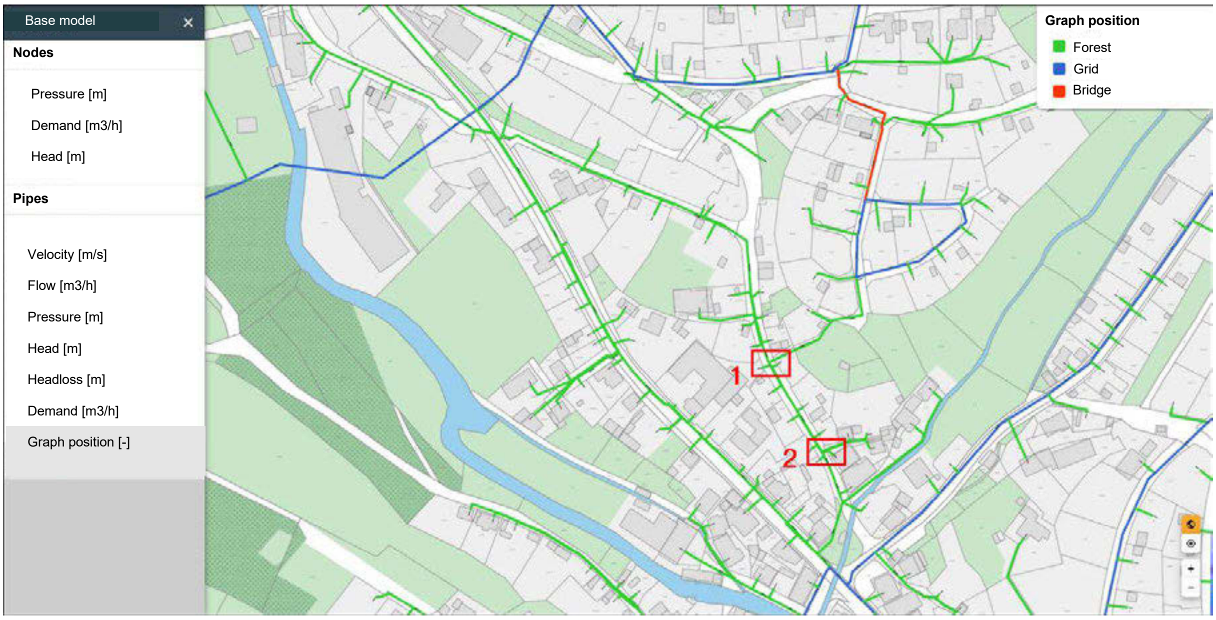

Base model
Nodes
Pressure [m]
Demand [m3/h]
Head [m]
Pipes
Velocity [m/s]
Flow [m3/h]
Pressure [m]
Head [m]
Headloss [m]
Demand [m3/h]
Graph position [-]
Graph position
🟩 Forest
🟦 Grid
🟥 Bridge
1
2

Figure 7: Graph decomposition with looped blocks (blue), bridges (red), forest (green)

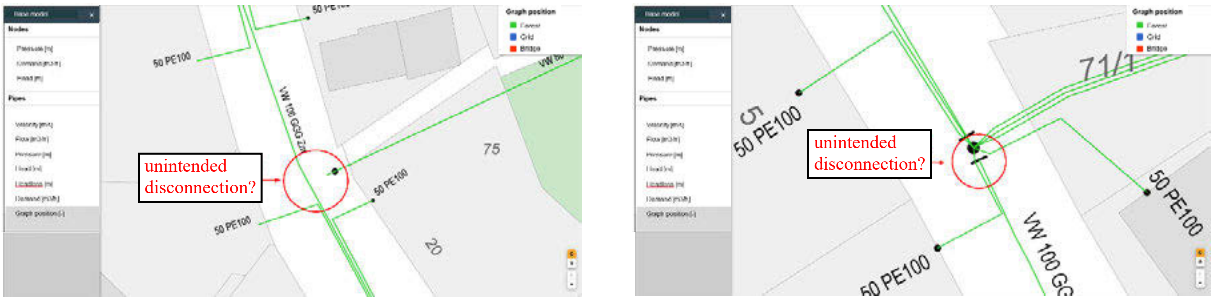

50 PE100
W 100 GGG
75
20
unintended disconnection?
50 PE100
71/1
unintended disconnection?
W 100 GG
Graph position
Forest
Grid
Bridge

Figure 8: Possible sources of disconnection: distant node (left), valve closure (right)

2.4 Data analysis tools

High-quality data analysis tools have been offered for several years, both as commercial and as freely available software. However, small and medium-sized water utilities use them only to a very limited extent due to insufficient data collection. At large utilities, data platforms with sufficient measurement data are often available. However, there is a lack of specialized personnel to use analysis tools. Therefore, easy-to-use data analysis tools were developed in W-Net 4.0. The structure of the data analysis platform is shown in Figure 9. The individual areas are described briefly below (according to Figure 9 from left to right):

a) Data transmission of the sensors/meters: The sensors/meters installed in the network (or at the waterworks) usually transmit their measured data to a control system. Direct connection to a control system is often difficult. Therefore, the export of this data is often done by means of a cyclic filing (e.g., every 10 minutes) in a CSV (Column Separated Value) file on an FTPS (FTP over TLS or File Transfer Protocol over Transport Layer Security) sever or a cloud-based file repository. Data from retrofitted sensors is often stored in cloud solutions. From there, they can be retrieved via REST (Representational State Transfer) or MQTT (Message Queuing Telemetry Transport).

b) Storage of all sensor/meter data in a common database: The data is stored in a common time series database. In the project, InfluxDB has proven to be a very powerful time series database. Important here is the introduction of a suitable data model so that all essential

cc BY NC SA

properties of the measuring point are mapped (e. g. suitable location designation, measuring point name, physical quantity, physical unit).

c) Algorithm toolbox: By means of an algorithm toolbox, the data are suitably pre-processed (e. g. removal of outliers, resampling). Suitable key figures are calculated (e. g. minimum night flow value). Furthermore, based on the historical data, prognosis modules are learned, which allow a prediction of the flow in the network. An alarm tool allows to define simple threshold alarms as well as more complex alarms (e. g. linking of several measuring points; comparison of current sensor values with values from the past).

d) Transmission of results to the users: (1) Alarms: If an alarm occurs, the user is informed by e-mail. A meaningful plot of the measurement data and alarm threshold is sent along. The alarms are stored in an alarm database so that they can be subsequently evaluated at longer intervals. (2) Dashboards, Reports, Prediction Tools: The time series of the sensors/meters, the calculated metrics and predicted flows are visualized in clear browser-based dashboards. Monthly and annual reports are generated, which contain an overview of the most important sensors, key figures and alarms. (3) Coupling to GIS: The data analysis platform was linked to the GIS via a REST interface. The interface is used, for example, to transmit the alarms that have occurred to the GIS.

The realized data analysis platform was implemented and successfully tested in the project at three smaller water utilities.

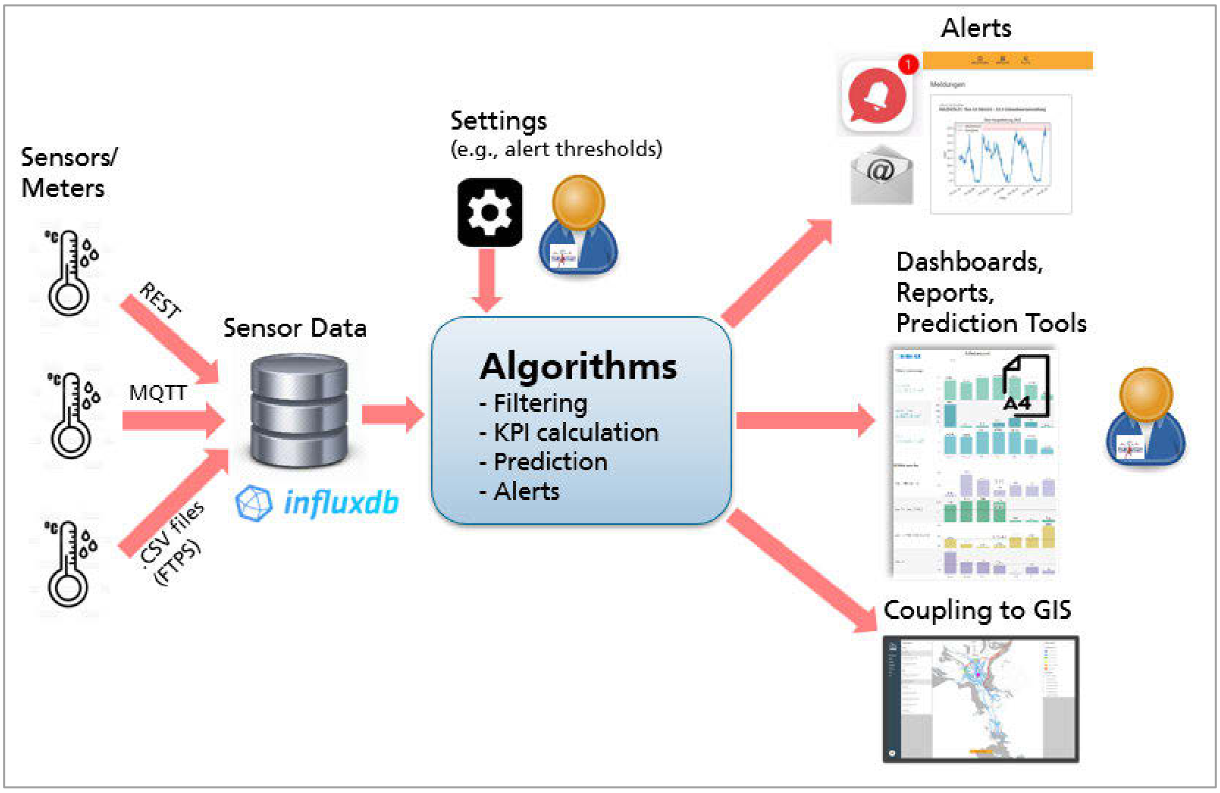

Sensors/
Meters
REST
MQTT
CSV files
(FTPS)
Sensor Data
influxdb
Settings
(e.g., alert thresholds)
Algorithms
- Filtering
- KPI calculation
- Prediction
- Alerts
Alerts
Dashboards,
Reports,
Prediction Tools
A4
Coupling to GIS

Figure 9: Structure of the data analysis platform

3 USE CASE APPLICATIONS

Five use cases have been defined by the end users, which are supposed to be the most relevant for small and medium sized water supply utilities. The integrated GIS and simulation platform was tailored to the use cases and specialized engineering services have been developed for assisting the utilities. As a general concept, simple calculations can be done by the operational staff using the information system's web services. The calculations can be carried out also in the field on tablet computers or mobile phones for supporting ad-hoc decisions. The solution of more complicated problems, comprehensive analysis and investigations are offered as engineering services. In contrast to conventional approaches where a simulation model must be built or at least updated in a time-consuming process, here, an up-to-date model always exists in the GIS, which is ready for instant use in the simulation software. The model that runs the steady-state calculations in the information system is the same that runs in the specialist software.

The benefits of the dual use approach are:

- no additional import or export of data is required for creation of the simulation model
- the GIS-web-application is not overloaded and tailored to the carefully identified needs of practitioners. It doesn't have to show up all the details and parameters that are required for in-depth investigations.
- Only few steps necessary for preparing the model for extended period, slow transient or water hammer calculations.
- Additional apps are available as plugins for the solution of use cases (fire flow app, diameter optimization app, supply reliability app). The plugin concept is designed for any additional extensions.

The five use cases defined in the project are summarized in the following table:

Table 1. W-Net 4.0 use cases and solutions provided by the new development

Use Case	**Web service (online modus)**	**Hydraulic Simulation (expert modus)**
1. Introduction to general network calculations	steady-state calculation	Steady-state and transient calculations
2. provision of firefighting water	capacity calculation for individual hydrants	App for automatized fire flow calculation of entire system
3. network operations	Steady-state behaviour of pumps, valves, control devices	Dynamic behaviour of pumps, valves, control devices
4. Expansion of the settlement area and new development of commercial areas	Manipulation of demands, what-if calculations	App for mathematical optimization of pipe diameters and pumps, reliability calculation, asset strategy
5. Water losses and leakage	Continuous monitoring of pressure and flows	Dual model: prototype for leak detection and isolation [9]

4 TRAINING AND SERVICES

The software development in the project is accompanied by a comprehensive training package and engineering services.

4.1 E-Learning Module and Training courses

The training is focussed on theoretical background, application of German technical rules and usage of the software. The content is guided by the five use cases of the previous chapter. There will be training courses for the end users as well as E-Learning modules that are designed for individual learning. The structure is the same for all modules and includes the sub-chapters:

- Introduction (why is the topic important for the target group of the training?).
- What do the rules and regulations say?
- Thematic content depending on the module
- Consolidation
- Learning test (2-3 questions per module)
- Literature, rules and regulations

The in-depth part of the module “network calculation” also includes the possibility of interactive pressure loss calculation for individual pipes. A JavaScript module for the calculation of frictional pressure losses in pipes based on the Darcy-Weisbach equation was implemented. The pressure loss calculation is a central component of the pipe network calculation. Often, however, only a rough estimation of the pressure loss in an isolated pipe is required.

The second block includes an introduction on how to work with the web platform, especially the simulation subsection. For this purpose, an anonymized training simulator model that is based on a real system is introduced. All training exercises are based on this model. The handling of the simulation platform is explained in detail followed by the interpretation of calculation results.

The third block contains exercises, both for working with the web platform and for self-studying and practical application of the theoretical basics. The exercises contain detailed solutions, which are initially invisible and can be displayed on demand.

4.2 Technical services

The holistic concept of W-Net 4.0 includes a close collaboration of the technical staff of the water supply utilities and external service engineers. The technical responsible at the utility can use the web information system for solving daily problems arising with network operation and management. If the problem is getting more complicated or time-consuming outsourcing is more efficient. SchwarzwaldWASSER in collaboration with the GIS, HS and IOSB partners provides a service package that covers all steps from data acquisition and creation of the initial GIS model over model calibration and field measurements up to specialized problems such as monitoring and data analyses, rehabilitation planning and restructuring, reliability calculations and water hammer analysis.

5 CONCLUSION

The aim of the project W-Net 4.0 is to enable small and medium sized water supply utilities to benefit from industry 4.0 technologies and digitization. A centralized digital platform has been developed whose core is a GIS database that is connected to offline applications for construction and model development and an online information system realized as web services. The GIS is interlinked with hydraulic simulation and data monitoring tools, each of them remaining individual applications. For the end-user it feels like one single application and all tools being integrated in the central platform. Keeping the tools separate has the big advantage of higher flexibility. For the provision of engineering services, the integrated models can also be loaded in specialist software.

The platform has been implemented and tested for the four associate partners of the project, each of them with different focus and prerequisites in terms of data availability, topographical characteristics of the supply area and size of the network. The intermediate results presentation to the utilities already created great interest and consent. It is planned that the results are also presented to a bigger audience in workshops and meetings of SchwarzwaldWASSER e.V. and beyond.

As next logical step of development the simulation model of the platform shall be extended to cyclic, automatic online calculations. The central platform already integrates all the necessary data: an updated GIS that includes all information for hydraulic calculation. And the data analysis tools that provide time series for all kind of process data coming from the SCADA system and from the sensors in the field.

More research has to be done for the definition of the best suited time intervals for updating the GIS data for hydraulic simulation. If constructions take place in the field, at least a state update of the isolation valves that are closed for the separation of the pipes under construction should be effectuated.

6 ACKNOWLEDGEMENT

This project W-Net 4.0 received funding from the German Ministry for Education and Research (BMBF Project W-Net 4.0 02WIK1477C).

7 REFERENCES

[1] Alegre, H., Is strategic asset management applicable to small and medium utilities?, Water science and technology: a journal of the International Association on Water Pollution Research, vol. 62, no. 9, 2010, pp. 2051–2058. http://dx.doi.org/10.2166/wst.2010.509. – DOI 10.2166/wst.2010.509. – ISSN 0273–1223

[2] Scholten, L.; Scheidegger, A.; Reichert, P.; Mauer, M., Lienert, J., Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis, Water research, vol. 49, 2014, pp. 124–143. http://dx.doi.org/10.1016/j.watres.2013.11.017. – DOI 10.1016/j.watres.2013.11.017. – ISSN 1879–2448

[3] Platschek, C; Krause, S.; Günthert, W., Situationsanalyse der kleinräumig strukturierten Wasserversorgung in Süddeutschland am Beispiel Bayerns, energie wasser praxis, vol. 65, no. 11, 2014.

[4] Bernard, T.: W-Net 4.0: Platform for optimizing the operation of water systems. https://www.iosb.fraunhofer.de/en/projects-and-products/platform-operation-optimization-water-systems.html (accessed 04 April 2022)

[5] Bergs, T.; Gierlings, S.; Auerbach, T.; Klink, A.; Schraknepper, D.; Augspurger, T.: The Concept of Digital Twin and Digital Shadow in Manufacturing, Procedia CIRP, Volume 101, 2021, Pages 81-84,ISSN 2212-8271, https://doi.org/10.1016/j.procir.2021.02.010.

[6] Deuerlein, J. (2020): Der Digitale Zwilling: Betriebsnahe hydraulische Modelle – Anforderungen und Einsatz. Aqua und Gas, Nummer 3. SVGW, Zürich. 2020.

[7] DVGW (2006):   DVGW W 405-B1:2016-06: „Supply of Fire Water via the Public Drinking Water Supply - Supplement 1: Prevention of Detriments to the Drinking Water und to the Network During the Abstraction of Fire Water“; DVGW – German Technical and Scientific Association for Gas and Water, 2016.

[8] Deuerlein, J. W.: Decomposition model of a general water supply network graph. ASCE Journal of Hydraulic Engineering, 134, 6, 2008, 822 - 832.

[9] Steffelbauer, D. B.; Deuerlein, J. W.; Gilbert, D.; Abraham, E.; Piller, O: “Pressure-Leak Duality for Leak Detection and Localization in Water Distribution Systems”, Journal of water resources planning and management, Vol.148 (3), 2022-03-0, 2022. doi: 10.1061/(ASCE)WR.1943-5452.0001515

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14098

PUMPING STATION DESIGN WITH AN ANALYSIS OF VARIABILITY OF DEMAND AND CONSIDERING TECHNO-ECONOMIC AND ENVIRONMENTAL CRITERIA THROUGH THE AHP METHOD

Christian X. Briceño-León[1], Pedro L. Iglesias-Rey[2], F. Javier Martínez-Solano[3], Enrico Creaco[4]

[1]PhD Student. Department of Hydraulic and Environmental Engineering. Universitat Politécnica de València. Camino de Vera s/n. 46022 Valencia (Spain).
[2]Associate Professor. Department of Hydraulic and Environmental Engineering. Universitat Politécnica de València. Camino de Vera s/n. 46022 Valencia (Spain).
[3]Associate Professor. Department of Hydraulic and Environmental Engineering. Universitat Politécnica de València. Camino de Vera s/n. 46022 Valencia (Spain).
[4]Associate Professor. Civil Engineering and Architecture Department-Hydraulic and Environmental Engineering Section. Pavia University, Via Ferrata 3, 27100, Pavia, Italy.

[1] *chbri1@alumni.upv.es*, [2] *piglesia@upv.es*, [3] *jmsolan@upv.es*, [4]*creaco@unipv.it*

Abstract

Pumping Station costs including capital and operational costs are some of the highest costs in urban water distribution systems. A proper pumping station design could be defined as the solution with the minimum life cycle cost and satisfying extreme scenarios in water distribution system. These costs are associated with investment, operational, and maintenance costs. However, there are some important aspects to consider in a pumping station design, such as the feasibility of infrastructure construction, the size of the infrastructure, and the complexity of operation in the pumping station. These aspects are associated with technical criteria. In a classical pumping station design, the number of pumps is determined in arbitrary form according to the criteria of the engineer, and the pump model is selected according to the maximum requirements of flow and pressure of the network. In summary, these variables in a pumping station design are not usually analyzed deeply. In addition, global warming acceleration in the last decades has gained momentum to be considered in engineering problems to mitigate the environmental impact. Hence, it is imperative to consider environmental aspects, such as greenhouse emissions, energetic efficiency of the pump in modern pumping systems of water distribution networks. Finally, the most suitable solution is determined only by analyzing economic aspects. Therefore, this work proposed a methodology to design pumping stations in urban water networks considering technical, environmental, and economic criteria and link them together through the Analytic Hierarchy Process (AHP) method. This method proposes to determine the importance priority of these aspects to assess the possible solutions and determine the most suitable solution in the pumping station design. In addition, this work considers the variability of demand pattern. This work analyses several scenarios of demand patterns from the minimum possible demand to the maximum possible demand in a water distribution network and the respective probabilities of non-exceedance. It allows the pumping station design be more robust. This methodology has been applied in different case studies to analyze how affects to determine the most suitable solution when the characteristics of the network change.

Keywords
Pumping station, AHP, technical criteria, environmental criteria, and economic criteria.

1. INTRODUCTION

Water demand has increased constantly around a rate of 1% per year because of the development of urban settlements. It has led to an increase in water stress in the last century. In addition, water distribution systems (WDS)consume a great amount of energy. Approximately 95% of this energy consumption is due to pump station (PS) operation [1]. Hence, climate change issues, such as greenhouse gas (GHG) emissions have been increasing in the last decades. All these problems have been of concern to the authorities in the world. The United Nations (UN) established the Sustainable Development Goals (SDG) of the 2030 agenda, SDG 6 dedicated to water and sanitation, and SDG 7 dedicated to affordable and non-polluting energy. Therefore, water management companies have made efforts to focus on reducing energy consumption and improving the operation and service in WDS.

Urban supply systems are one of the infrastructures of the most vulnerable to climate change; therefore, it is necessary that their projection is made considering energy efficiency, and responding to the variability of demands, without neglecting the optimization of the costs of investment and operation. Life Cycle Costs (LCC) supposes to be a major component in the analysis of a pumping station in closed networks. These costs are mainly composed of operational costs associated with energy consumption, maintenance costs, and investment costs. In fact, the reduction of energy consumption and maintenance costs are the most common efforts to improve the operation and water service for WDS [2].

The main element of the annual operating budget in a WDS is the energy consumption costs of the PS. Therefore, the most common objective in the design of closed networks is to optimize energy consumption. Chang Y. et al. [3] developed a methodology to save energy costs for water networks by transferring the water demand at storage systems when the unit price of energy is high and the amount of water demand is increased when the unit price of energy is low. On the other hand, Lipiwattanakarn S. et al. [4] created a theoretical estimation of assessing the energy efficiency of water distribution systems based on energy balance. These components of energy were: outgoing energy through water loss, friction energy loss, and energy associated with water loss. Besides, Giudicianni et al. [5] developed a methodology to improve the management and monitoring of water distribution systems based on regrouping the original network into dynamic district metered areas. The idea of this proposed framework is to locate determined energy recovery devices and reduce water leakage in a water network.

In addition, several works deepened the operation of PS. For example, Walsky and Creaco[6] evaluated different pumping configurations for closed networks combining a different number of Fixed Speed pumps (FSP) with different sizes and adding a Variable Speed pump (VSP) to select the most suitable configuration for different scenarios of flow and required head. Then, Leon Celi et al. [7] optimized the allocation of flow and the energy consumption in water networks with multiple PSs determining the optimal set-point curve in every PS. This term is referred to the head required of PSs to satisfy demand requirements in the critical node maintaining the minimum service pressure throughout the time. In a similar way, Briceño et al. [14] create a new methodology of control system for PSs to determine the optimal number of pumps and decrease energy consumption using the set-point curve concept.

There are other improvements of pumping systems, such as multiobjective optimizations including, energy costs, maintenance costs, and treatment costs [8], [9] or optimizing energy costs and maximizing the reliability of the systems [10], [11]. In addition, Mahar and Singh [12] developed a methodology to optimize capital and operational costs for PSs. Similarly, Nault and Papa (2015) improved the operational costs of PSs considering environmental aspects, such as GHG emissions associated to pump operation. Then, Candilejo et al. [13] optimized the construction cost and energy cost in a pressurized water network with variable flow demands based in an equivalent flow rate and equivalent volume.

In the last decades, several research have proved that projects related to serving WDS including PS design could be potentially harmful to the climate change [13] [14]. However, there are few works that faced the environmental impact. These works are related with reducing GHG emissions [15] and leakage in pumping systems [16].

In general, most of the previous works in pumping systems aimed to assess the solution from an economical point of view, such as the minimization of operational and construction costs. However, there are hidden important aspects were not considered in the design and could be hardly determined in economic terms. For example, the viability of required size in the construction stage and the flexibility of operation are usually neglected in PS design. These aspects are closely related to the number of pumps, which is arbitrary defined according to the designer's judgement or experience. Another important aspect that is difficult to convert in economic terms is the complexity of the operation of the pumping system. The optimization of the pump operation needs sophisticated devices, especially Programmer Logic Control (PLC). However, it supposed that PS operation be more complex. For example, In real-time control operation of water networks that use Supervisory Control and Data Acquisition (SCADA), it is too complex to schedule the pumps with the PLCs [17], [18]. Another problem of the previous works lies in the fact that they omitted the yearly demand variability of the WDS. In fact, they typically consider a single daily demand pattern, and it could make that the operation of the system be less feasible.

Therefore, the idea of this work is to integrate economic aspects with other important aspects that are usually neglected in PS, such as the flexibility of operation, the size of the station, the complexity of operation, and also consider environmental aspects. Some of these aspects have conflict of interest, such as operational costs with the complexity of operation. Throughout the multicriteria analysis can evaluate different criteria and stablish an alternative or group of alternatives that meet all criteria. One of the methods of multicriteria analysis developed in last years is the AHP method that is an important method for complex management decision problems [19].

AHP is a method developed by Saaty [20]. It allows the resolution of complex problems involving multiple criteria. The AHP process requires the decision-maker to do so through subjective assessments regarding a relative importance of each of the criteria and to specify their preference with respect to each of the alternatives for each of the criteria. In fact, in the last decades the AHP method has been widely applied in the hydraulic engineering field. For example, in WDS and sewer system rehabilitation to determine the priorities of maintenance or substitution of the elements of those systems [21] [22], or water management sustainability [23]. Finally, Briceño-León et al. [24] developed a approach of PS design integrating techno-economic factors.

In this way, the objective of this work is to develop a comprehensive methodology for PS design through the multicriteria analysis (the AHP method). One contribution of this work is to determine the priorities of the aspects considered in the design (Technical, Environmental and Economic Factors). Then, this methodology evaluates the potential alternatives integrating the considered aspects for the PS and select the most suitable solution. The alternatives will be gotten from a pump database that will have different number of pumps and different control system alternatives. The criteria considered in this work are technical factors that considered the size of the pumping station, the complexity of regulation mode and the flexibility of operation, economic factors include investment, operational and maintenance costs and environment factors are associated with minimum energy efficiency (MEI), CO2 emissions and perfomance regulation. In addition, another contribution of this work is to consider demand variability and their respective probability of occurrence of every demand scenario. The consideration of different demand scenarios will make the design be more feasibly and robust.

2. METHODOLOGY

2.1 Pumping Station Statement

The design of a PS contemplates three stages. The first stage includes the definition of the set-point curve of the network. It means the required flow (Q) and head (H) of the pump to satisfy the requirements of the network, and the maximum flow (Q_{max}) and the maximum required head (H_c) at the critical node. The second stage is about the selection of the pump model and determine the required number of pumps to satisfy the operation of the system. The third stage includes the selection of type of operation control system for the PS according to the necessities of the system.

Traditionally, the design of PS for closed WDS is made up of centrifugal pumps and their respective number of pumps, usually installed in parallel. The traditional method of selecting the pump model is searching for a pump model that provides the maximum demand flow (Q_{max}) and maximum head (H_{max}) of the network. Once is defined the pump model, the required number of pumps is obtained by the relation of the maximum required flow (Q_{max}) and the flow of a single pump of the model selected (Q_{b1}) associated with the maximum required head (H_{max}). Nevertheless, in some cases, firstly the number of pumps is fixed. Then, the model is selected according to the relation of the maximum flow (Q_{max}) and the number of pumps and also considering the maximum required head (H_{max}). Then, the designer established different pumping configuration modes, including the number of FSP or VSP and the type of control to use: Pressure Control (PC) or Flow Control (FC).

This methodology considers five different control systems. The first (1.- No control system). In this configuration the pumps operate all the time without restrictions. The second configuration (2. _ FSP with PC) operates only FSPs and their operation are associated with fixed switch on/off pressures and the set point curve. The third configuration (3. _ FSP with FC) operates only FSPs and their operation are associated with intersection flows of the pumping curves and the set-point curve. The fourth configuration (4. _ FSP and VSP with PC) is a combination of FSP and VSP and their operation consist of the operational points of the pump (Q, H) follow a fixed pressure. Finally, the fifth configuration (5. _ FSP and VSP with FC) is the combination of FSP and VSP and their operation consist of the operational points of the pumps (Q, H) follow the set-point curve.

In brief, the operational conditions of the system, the characteristic of the pump models, the number of installed pumps, and different control system configurations are the restrictions in a PS design. Hence, these restrictions drives that the design be arbitrary and subjective according to the criteria of the designer. Therefore, the obtained solutions do not guarantee be the optimal in technical, economic, and environmental aspects at the same time. In addition, it drives that the selection of the ultimate solution be not generic.

Therefore, this work aims to diversify the traditional design process of pumping stations, in which the methodological design development is proposed through a multi- criteria perspective based on technical, environ-mental, and economic criteria that respond to the current dynamics of infrastructure projection.

2.2 Hypothesis and Required Data

One assumption of this methodology is that the WDS is a closed system, and the PS injects directly the flow to the consumption nodes. In addition, they hydraulic characteristics of the system (set-point curve, demand pattern) are assumed as known data. In this way, the required data to develop the methodology are the following:

- **Setpoint curve**: it is the definition of the hydraulic requirements. It represents the minimum head necessary leaving of the pumping station to guarantee the demands with the minimum pressure conditions required

$$H_c = \Delta H + R \cdot Q^C \tag{1}$$

The term ΔH is the static head of the PS including the minimum pressure service of the network, R represents energy losses produced in the system, and c is an exponent that depends on the characteristic of the system.

- **Schematic Model of a PS**: Iglesias Rey et al. [25] proposed a parameterized model of a PS. The basic scheme includes a back-up pump to guarantee the reliability of the PS. The scheme is represented in Figure 1. In this figure, the parameters *N1*, *N2* and *N3* are the characteristic lengths of the PS, which are considered proportional to the nominal diameter (ND) of the pipe:

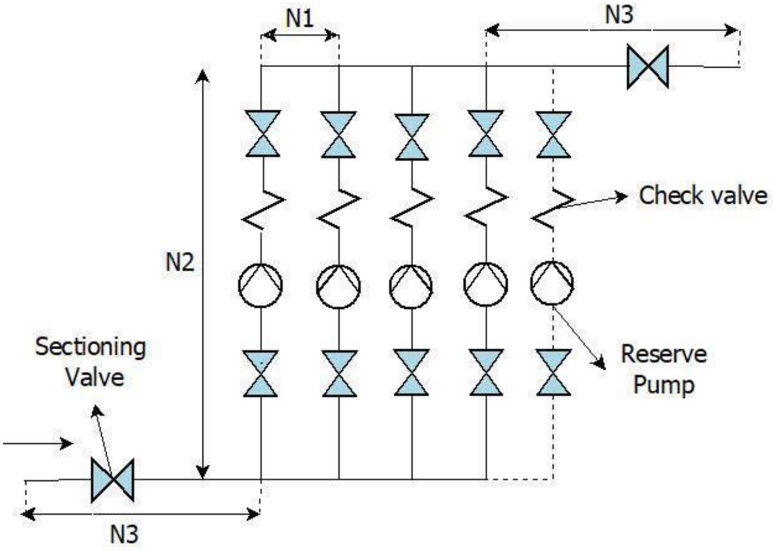


Figure 1. Basic Scheme of a PS.

- **Pump Model Database**: Every pump model in the commercial catalogue is defined by the Best Efficient Point (BEP). The BEP includes the nominal Head (H_0), the nominal flow (Q_0), the nominal efficiency (η_0), and the nominal rotational speed (N_0). These variables determine the pumping curve (H-Q) and the efficiency curve (η-Q). The characteristic of the pumping curve is defined by fixed parameters: A, H_1, and B (Equation 2), and the characteristic of the efficiency curve is defined by fixed parameters: E and F (Equation 3). In addition, both curves are defined by the variables: the ratio of the current rotational speed and the nominal rotational speed ($\alpha = N/N_0$), the flow (Q), and the number of installed pumps (b).

$$H = H_1 \alpha^2 - \alpha^{(2-B)} A \cdot \left(\frac{Q}{b}\right)^B \tag{2}$$

$$\eta = E \cdot \frac{Q}{\alpha \cdot b} - F \cdot \left(\frac{Q}{\alpha \cdot b}\right)^2 \tag{3}$$

In addition, the parameters of a PS includes the correction of the pump efficiency of the affinity laws developed by Sarbu and Borza [26] (Equation 4). The term η_c is the pump efficiency correction and η is the efficiency of the affinity laws. On the other hand, Briceño et al. [27] developed an expression to estimate the efficiency of the frequency drive (Equation 5). The terms k_1, k_2, k_3 are constant parameters of the equation of the frequency drive efficiency, $\eta_{v,0}$ is the maximum frequency drive efficiency. Finally, P_s in equation (6)

BY NC SA

is the consumed power of the PS, and Q_{FSP} and Q_{VSP} are the flow delivered by FSP and VSP respectively.

$$\eta_c = 1 - (1-\alpha)^3 \cdot \eta \tag{4}$$

$$\eta_v = \eta_{v,0} \cdot (\beta_v^{k_1} - k_2 \cdot (1-\alpha)^{k_3}) \tag{5}$$

$$P_s = \frac{\gamma \cdot Q_{FSP} \cdot H}{\eta_c} + \frac{\gamma \cdot Q_{VSP} \cdot H}{\eta_c \cdot \eta_v} \tag{6}$$

- **Demand Patterns**: This methodology incorporates the analysis of variability of demand in the PS design. The different scenarios are defined by the probability of non-exceedance of the demand from 0 to 1. Every scenario has its probability of occurrence to determine the number of days of occurrence of every demand scenario.
- **Electricity rates**: correspond to electricity rates that change hourly depending on the type of power contracting that the supply system has.
- **MEI**: Is a dimensionless index that defines the ratio minimum efficiency of the pump between the operating point of 75% of the BEP and the overload of 110% of the BEP. This index is calculated according to EU Regulation 547/2012.
- **CO2 Emission**: This factor is obtained from a local energy maker and it is used to calculate the CO2 Emission by the PS.
- **Economic Factor**: This factor is associated with the annual interest rate to amortize the investment cost in yearly costs.

2.3 Evaluation of the Pump Models

Every pump model in the database is evaluated the feasibility with the hydraulic characteristics of the WDS. The maximum head (H_1) of every pump model must be higher than the maximum required head of the system (H_{max}). The infeasible pump models are discarded, and the feasible pump models could be considered as potential solution. In addition, in every feasible pump model is determined the minimum number of pumps (b_{min}) and checked if b_{min} is not greater than the maximum allowed number of pumps (b_{max})

In every feasible pump model are five different configuration of control system obtaining a maximum potential solution of 5xN_{viable}. In control system configurations 4 and 5, have the combination of m number of FSP and n number of VSP ($m+n=b_{min}$). In these configurations are optimized the number of pumps ($b=m+n$) in every time slot and in every scenario of demand. The optimization process consists of adding a unit number of VSP in every combination of ($b = m + n$) until is not possible to improve the consumed power of the PS. In this context, the optimal number of pumps ($b=m+n$) could be greater than $b_{min.}$ In summary, the optimal configuration of control systems 4 and 5 search the optimal number m FSP and n VSP in operation, and the rotational speed to minimize the consumed power ($P_{T,OP}$). The following figure 2, describes the optimization process of the control system configurations.

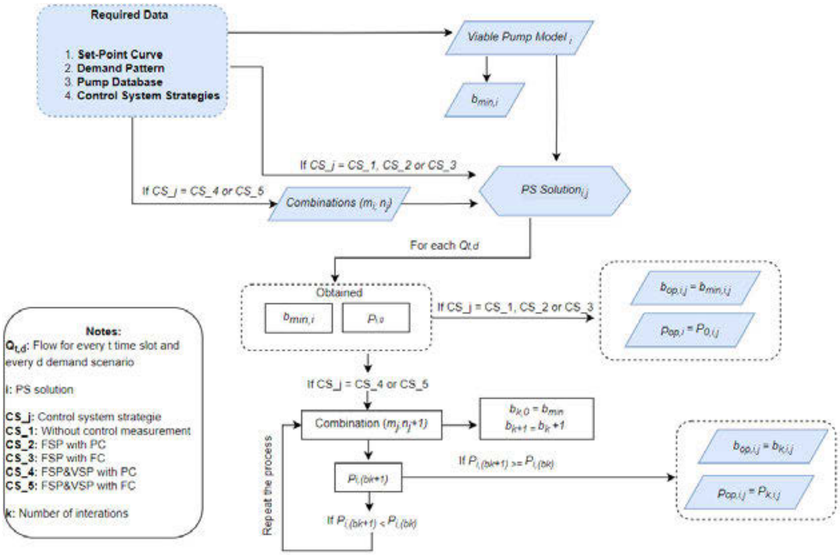

Required Data
1. Set-Point Curve
2. Demand Pattern
3. Pump Database
4. Control System Strategies
Viable Pump Model $_i$
$b_{min,i}$
If CS_j = CS_1, CS_2 or CS_3
If CS_j = CS_4 or CS_5
Combinations (m_j, n_j)
PS Solution$_{i,j}$
For each $Q_{t,d}$
Obtained
$b_{min,i}$
$P_{i,0}$
If CS_j = CS_1, CS_2 or CS_3
$b_{op,i,j} = b_{min,i,j}$
$P_{op,i} = P_{0,i,j}$
Notes:
$Q_{t,d}$: Flow for every t time slot and every d demand scenario
i: PS solution
CS_j: Control system strategie
CS_1: Without control measurement
CS_2: FSP with PC
CS_3: FSP with FC
CS_4: FSP&VSP with PC
CS_5: FSP&VSP with FC
k: Number of interations
If CS_j = CS_4 or CS_5
Combination (m_j,n_j+1)
$b_{k,0} = b_{min}$
$b_{k+1} = b_k + 1$
$P_{i,(bk+1)}$
If $P_{i,(bk+1)} >= P_{i,(bk)}$
$b_{op,i,j} = b_{k,i,j}$
$P_{op,i,j} = P_{k,i,j}$
If $P_{i,(bk+1)} < P_{i,(bk)}$
Repeat the process

Figure 2. Optimization Process of Control System Configurations

The potential solutions are evaluated in every criterion of technical, environmental, and economic criteria. The obtained results of the potential solutions in every criterion are ranked in a numeric scale of 0 to 1, where 0 is the worst and 1 is the best solution.

The criteria considered in this methodology is described in detail below.

Technical Criteria

1. Size: The size of the PS is in function of the number of pumps installed and the length of the pipelines in the station. A higher score is assigned to this sub-criterion if the installation area is small. In this way, the highest size is assigned a score of 0 and the smallest size is assigned a score of 1.

2. Flexibility: The flexibility of the PS is associated with the number of pumps installed, i.e. as higher is the number of pumps installed, the flexibility is larger. In fact, a greater number of pumps in the PS allows that the perfomance of the system increase. A higher score is assigned to this sub-criterion if the number of pumps installed is large. The potential solution with the highest number of pumps (b) is assigned a score of 1 and the solution with the smallest number of pumps is assigned a score of 0.

3. Complexity of control: This sub-criterion is associated to the number of elements needed in every control system strategy. The control system is considered less complex if the number of control elements in the system is small. Hence, as smaller is the number of control elements installed, the score assigned is higher. Every control system configuration is assigned a numeric score from 0 to 1 (Table 1), where 1 is the least complexity of the control systems and 0 the highest complexity of the control systems. The scores are obtained from pairwise comparisons of the different control system configurations applying the AHP method.

BY NC SA

Table 1. Numeric Score of the Control System Configurations

Control System Configuration	Complexity Level	Numeric Score
1. Without CS	1	1.00
2. FSP with PC	2	0.57
3. FSP with FC	3	0.32
4. FSP-VSP with PC	4	0.15
5. FSP-VSP with FC	5	0.07

Economic Criteria

4. Investment cost: It includes the costs of supplying and installing pipes, fittings and control elements, as well as the cost of the pumps. Additionally, it includes the costs of supplying and installing accessories and tubing for the reserve pump and its value. The equations to determine the purchase and installation costs of the accessories were developed by Briceño-León et al. [24]. The investment cost is annualized considering the life cycle of the elements, and the annual interest rate. A higher score is assigned to this sub-criterion if the investment cost is small. In this way, the solution with the lowest investment cost is assigned a score of 1 and 0 to the solution with the highest investment cost.

5. Operational cost: This sub-criterion is associated to the yearly cost of consumption energy (€) for the PS, and it is calculated by the following equation.

$$C_{E,year} = 365 \times \sum_{d=1}^{N_d} Pr_{DP,d} \left(\sum_{t=1}^{N_t} P_{T,t,d} \times \Delta t \times TE_t \right) \tag{7}$$

The term N_d is the number of demand scenarios, the sub-term d corresponds to each demand scenario, the duration of the time slot is represented by Δt, the sub-term correspond to every time slot, and TE is the electric tariff. The number 365 is used to obtain the number days of occurrence of every demand scenario. The lower the operational cost, the higher the score to be assigned to this sub-criterion. Hence, the lowest operational cost is assigned a score of 1 to the solution and the highest operational cost is assigned a score of 0 to the solution.

6. Maintenance cost: It represents the cost of maintenance activities to implement in the PS to keep it under good conditions. The frequency of maintenance activities for the elements of the PS and their costs are obtained by a database to determine the annual maintenance costs. A higher score is assigned to this sub-criterion if the maintenance cost is small. In this way, the solution with the lowest maintenance cost has a score of 1 and 0 the solution with the highest maintenance cost.

Environmental Criteria

7. MEI: The EU regulation 547/2012 developed the calculation of the MEI index. According to this regulation, a MEI value of 0.7 is excellent, whereas a MEI below 0.4 is not acceptable. This sub-criterion is evaluated in a numeric score, where a high score is assigned if the MEI index is high. These scores are detailed in the following table 2:

Table 2. Numeric Scale for MEI Index values

MEI Index	Numeric Score
0.1	0.05
0.2	0.07
0.3	0.12
0.4	0.27
0.5	0.40
0.6	0.61
0.7	1.00

8. CO2 Emission: It represents the amount of CO2 produced by the PS when it is in operation. CO2 emission is obtained by the multiplication of energy consumed by the PS with an emission factor *EF*. This sub-criterion is evaluated in terms of Kg of C02 in a year.

$$GHG_{e,year} = EF \times C_{E,year} \tag{8}$$

A high score is assigned to this sub-criterion if the CO2 emission is low. Therefore, the solution with the lowest CO2 emission is assigned a score of 1 and 0 the solution with the highest CO2 emission.

9. Performance of regulation: The performance of the regulation system (η_{RS}) relates to the ratio of the head of the set-point curve (H_c), to the head of the PS (H) obtained as a result of the application of the control strategy in every time slot (t) (Equation 9). The constraint is that $H \geq H_c$. A high value of this ratio means that the PS is working close to the set-point curve, resulting in an improvement of energy wastes. The overall performance of the regulation system is obtained as the flow of PS-weighted average of η_{RS} in all time slots and in all demand scenarios.

$$\eta_{RS,t} = \frac{H_{,t}}{H_{c,t}} \tag{9}$$

A high score is assigned to this sub-criterion if the performance of regulation is high. Hence, the highest perfomance regulation of a solution is assigned a score of 1 and 0 with the lowest perfomance regulation of a solution.

2.4 Adaptation of the AHP method in PS Design

The propose to apply the AHP method is to determine the importance weight or the priority of every criterion and sub-criterion in a PS Design. These priorities are obtained by the judgment of group of experts in PS. Saaty established a numeric scale to compare how important is a criterion over another in pairwise comparison and organized in a quadratic matrix. The scale is established by the following values: 1. The value corresponds to equal importance between one criterion and another; 3. Moderate importance of one criterion over another; 5. Strong importance of one criterion over another; 7. Very strong or proven importance of one criterion over another; and 9. Extreme importance of one criterion over another.

In this way, technical, environmental, and economic criteria are compared among themselves and organized in the comparison matrix to obtain an eigenvector of the importance weight of every criterion. In the same way, the sub-criteria of every criterion are compared among themselves in the comparison matrix to obtain a local eigenvector of the importance weight of the sub-criteria with respect to the criterion it belongs. Finally, the product of the importance weight of the criteria with the local importance weight of the sub-criteria determines the global importance weight of the sub-criteria.

Another element of the AHP is the "Consistency ratio" (CR), which corresponds a tool that allows controlling the consistency of paired comparisons. Being subjective value judgments, consistency is not absolute in the comparison procedure. Saaty [20] defined that the CR should not be higher than 0.1 regardless of the nature of the problem. Consistency does not imply a "good" final selection, it only guarantees that there are no conflicts in the comparisons. This methodology proposes to weight the global importance weight of the criteria with the obtained CR to decrease the subjectivity of judgments by the group of experts.

On the other hand, the dominant and dominated solutions of the assessment in every criterion are identified by the Pareto Front. In this way, the dominant solutions are considered as potential solutions and continues in the process of the methodology, whereas the dominant solutions are discarded.

Then, the dominant solutions are assessed weighing the score of every criterion with their importance weight. The global assessments of the solutions are transformed in a numeric score from 0 to 1, where 1 is assigned to the solution with the highest value of the assessment and 0 to the solution with the lowest value of the assessment. Finally, the solution with the best score is considered as the ultimate solution in the PS.

The following flowchart (Figure 3) describes the different stages of the proposed methodology: the required data, determine the feasibility of the pump models to the system, and the assessment of the potential solution and the selection of the ultimate solution.

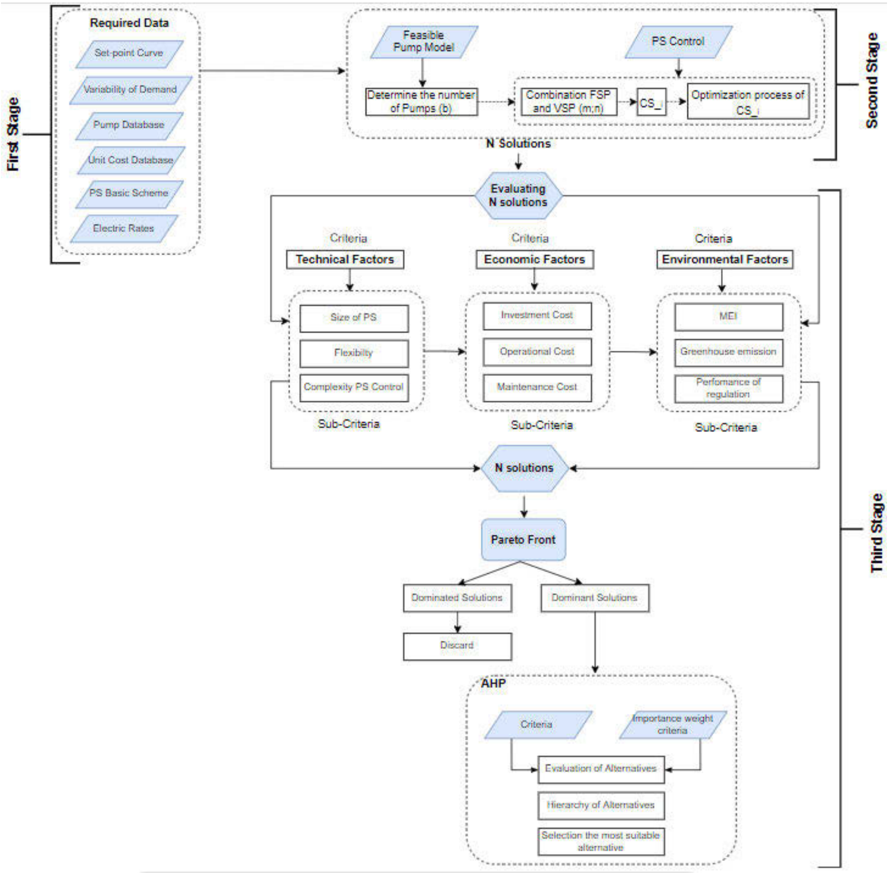

First Stage
Required Data
Set-point Curve
Variability of Demand
Pump Database
Unit Cost Database
PS Basic Scheme
Electric Rates
Feasible Pump Model
PS Control
Determine the number of Pumps (b)
Combination FSP and VSP (m;n)
CS_i
Optimization process of CS_i
Second Stage
N Solutions
Evaluating N solutions
Criteria
Technical Factors
Criteria
Economic Factors
Criteria
Environmental Factors
Size of PS
Flexibility
Complexity PS Control
Investment Cost
Operational Cost
Maintenance Cost
MEI
Greenhouse emission
Perfomance of regulation
Sub-Criteria
Sub-Criteria
Sub-Criteria
N solutions
Pareto Front
Dominated Solutions
Dominant Solutions
Discard
AHP
Criteria
Importance weight criteria
Evaluation of Alternatives
Hierarchy of Alternatives
Selection the most suitable alternative
Third Stage

Figure 3. Flowchart of the proposed methodology

3. CASE STUDY

This work considered a closed WDS namely BN for PS design. This WDS feature a yearly average demand of 25.0 L/s. The hydraulic characteristic of the system is represented by the set-point curve (See table 3).

Table 3. Set Point Curve of BN-WDS

Data	ΔH (m)	R	c
	32.50	0.0312	1.75
$H_c = 32.50 + 0.0312 \times Q^{1.73}$			

The database of the pump models used in this work is conformed by 67 different pump models with their respective BEP, the parameters of the head pumping curve and efficiency curve. In addition, every pump model has its purchase cost and its cycle life.

The velocity design considered for the design of the PS scheme is ($V=2.0\ m/s$). The parameters for the length of the pipes in the PS are $N_1=20$, $N_2=40$ and $N_3=20$. In addition, the maximum number of installed pumps (b_{max}) allowed in the design is 10 pumps.

This case study considered 21 different scenarios of non-exceedance probabilities (P_c) of BN-WDS demand from 0 to 1 with and interval of 0.05. In this way, the different P_c of demand are featured by (0; 0.05; 0.10; 0.15….. 1.0). These data were obtained from Alvisi and Franchini Work [28]. The probability of occurrence ($P_{r,DP}$) of the maximum and minimum P_c of demand is 2.5 % and the $P_{r,DP}$ of the other P_c of demand is 5.0 % (See Figure 4)

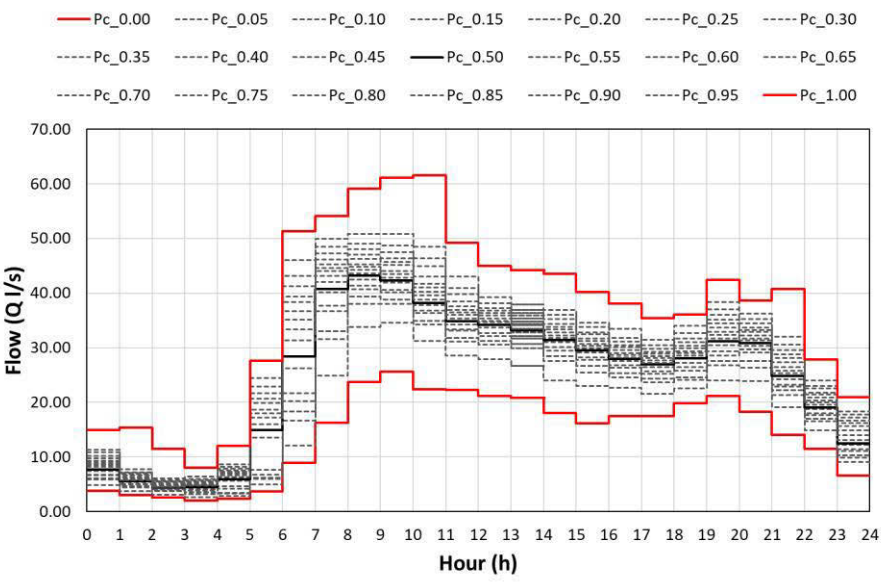


Figure 4. Variability of Demand of BN-WDS

The BN-WDS uses a single Electric Tariff with three kinds of hours: off-peak, peak and plain hours. The costs of the tariff of every kind of hour are specified in the following table 4.

BY NC SA

Table 4. Electric Tariff of BN-WDS

Type of hours	Electric Tariff TE (€/kWh)	Initial hour	Final hour
Off-peak hours	0.069	0	8
Peak hours	0.095	11	15
Plain hours	0.088	8 16	10 23

4. RESULTS

The obtained priorities of the criteria and the obtained local and global priorities of the sub-criteria of every criterion from the judgment of the group of experts in the AHP method are detailed in the Table 5. The priorities are expressed in a numeric scale from 0 to 1. Table 5 shows that the most important criteria in the PS design are Economic and Technical criteria with a score of 0.44 and 0.41, respectively, whereas the environmental criteria have less importance with a score of 0.15. The most important sub-criteria are Complexity of PS and Operational Cost with a scores of 0.18 and 0.16, respectively. Other sub-criteria, such as: Size of the PS, flexibility of the PS, investment cost and maintenance cost have a moderate priority with scores over 0.11. On the other hand, the priority of environmental sub-criteria: MEI, CO2 emission and perfomance of the regulation system are considerably less priority than the others with scores lower of 0.05.

Table 5. Obtained Priorities of the Criteria and Sub-Criteria for the PS Design

Criteria	Priority	Sub-Criteria		Local Priority	Global Priority
Technical Criteria	0.41	C1	Size of the PS	0.26	0.11
		C2	Flexibility of the PS	0.31	0.13
		C3	Complexity of the PS	0.43	0.18
Environmental Criteria	0.15	C4	MEI	0.35	0.05
		C5	CO2 Emission	0.22	0.03
		C6	Perfomance of the regulation system	0.43	0.06
Economic Criteria	0.44	C7	Investment Cost	0.30	0.13
		C8	Operational Cost	0.37	0.16
		C9	Maintenance Cost	0.33	0.14

In this WDS, two different methods of PS were performed. The method 1 is the typical methodology, where it is only considered economic factors. The selection of the ultimate solution is based on the minimization of LCC. The method 2 is the proposed methodology, where it is considered technical, environmental, and economic criteria basing on the AHP method. The solution with the highest overall score is selected as ultimate solution. The objective of this design framework is to analyse the effects of including Technical and Environmental aspects with respect to the classical method considering only economic aspects.

Table 6 shows the ultimate solutions of method 1 and method 2 for the PS design in BN-WDS. In every solution is detailed the BEP of the pump model (Q_0, H_0, η_0), the values and the numeric scores in every of the 9 sub-criteria, and the LCC.

Table 6. The Characteristics of the Ultimate Solutions of Method 1 and Method 2 for BN-WDS

			Method 1 (LCC Minimization)		**Method 2 (AHP with Tech. Env. And Eco. Criteria)**	
Pump Characteristics		**Pump Model Number**	33		45	
		Q_0	24.32 l/s		9.06 l/s	
		H_0	78.73 m		78.19 m	
		η_0	63%		77%	
			Values	**Score**	**Values**	**Score**
Technical Aspects	**C1**	**Size PS**	151.20 m^2	1.00	158.40 m^2	0.80
	C2	**Flexibility (*b*) *m* FSP; *n* VSP**	0 FSP- 3 VSP	0.01	7 FSP- 0 VSP	0.67
	C3	**Control System**	5	0.07	3	0.32
Environmental Aspects	**C4**	**MEI**	0.11	0.07	0.70	1.00
	C5	**GH Emission**	68,954.94 KgCO2	0.93	74,695.35 KgCO2	0.91
	C6	**Perfomance Regulation System**	100%	1.00	79%	0.49
Economic Aspects	**C7**	**Investment Cost**	6,604.70 €/year	0.92	12,347.87 €/year	0.56
	C8	**Operational Cost**	15,877.26 €/year	0.93	17,182.22 €/year	0.91
	C9	**Maintenance Cost**	1,193.21 € /year	0.90	2,105.30 €/year	0.37
Overall Score			0.96		1.00	
Life Cycle Cost			23,675.17 €/year		31,724.93 €/year	

The following radial chart (Figure 5) shows a comparison of the obtained scores for every sub-criterion of the ultimate solutions of method 1 and method 2. As it can see in this figure, the scores in the sub-criteria of the ultimate solution in Method 2 are over 0.30, whereas the solution in Method 1 has low scores in sub-criteria C2, C3 and C4. Hence, the ultimate solution in Method 2 is more equilibrate with the 9 sub-criteria than the solution in Method 1.

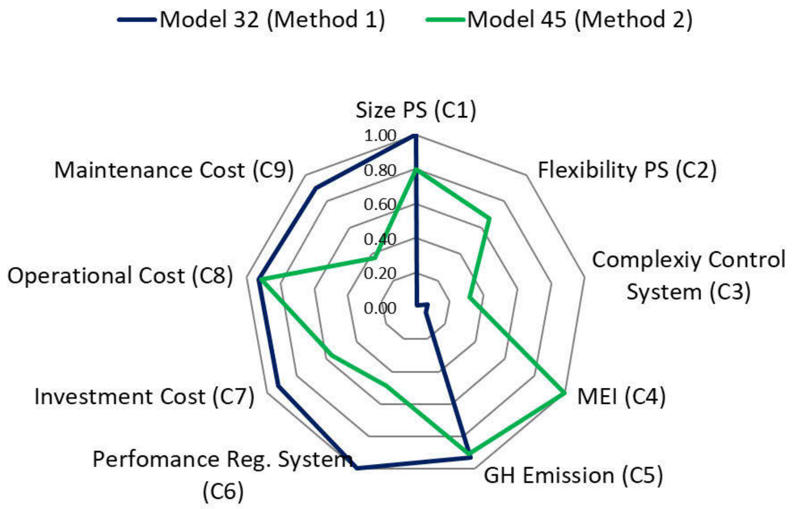


Figure 5. Radial chart of the scores for every sub-criterion of Method 1 and Method 2

The figures 6 and 7 show the number of pumps in operation (b) and the consumed power of the PS (P_T) in every time slot and for every scenario probability of non-exceedance of the demand (P_c). The scenarios considered in these figures are P_c_0.00, P_c_0.25, P_c_0.50, P_c_0.75, and P_c_1.00. The objective to display these figures is to feature how is the operating behaviour of the ultimate solution in the method 2. As it can see in these figures, the number of pumps in operation (b) and the Consumed Power (P_T) of the PS increase as the demand is higher.

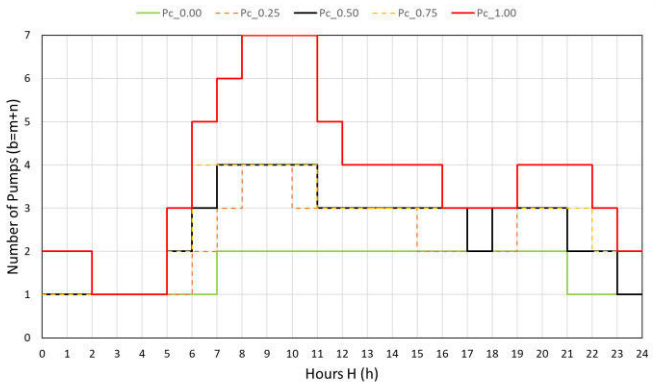


Figure 6. Number of pumps (b) in operation for every time slot and in every demand scenario

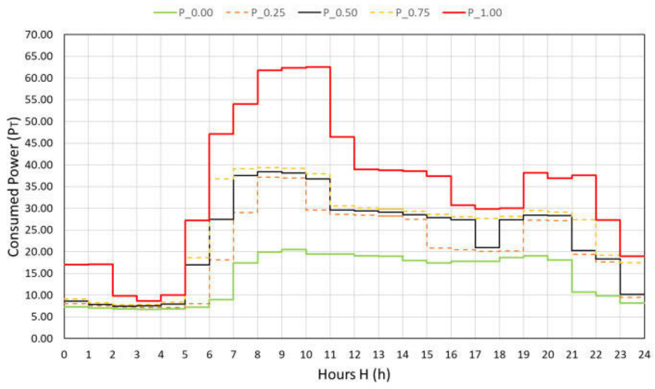


Figure 7. Consumed Power (P_T) for every time slot and in every demand scenario

5. DISCUSSION

The obtained solutions show interesting insights with different methodologies of a PS design, including the differences and effects of considering technical, environmental, and economic criteria based on the AHP method in contrast with classical method based on minimization the LCC.

The ultimate solution of Method 1 has a configuration of 0 FSP- 3 VSP with a FC system. It yields low investment cost, operational cost, and maintenance cost with scores over 0.90. This method is based on minimization LCC, so this solution has the lowest LCC of all the potential solutions considered. In addition, this solution yields low Kg of CO2 consumption and an excellent perfomance of regulation (100%) because are closely related with operational cost.

On the other hand, the ultimate solution of Method 2 has a configuration of 7 FSP- 0 VSP with a FC system. This method besides considering economic criteria considers technical and environmental criteria. Hence, this solution yields good qualities with the criteria: size of the PS and flexibility of operation with scores over 0.67. While the complexity of control system of this solution has a score of 0.32, but less complex than the control system of the solution of Method 1. In addition, the solution of Method 2 has excellent qualities of the environmental criteria: MEI and CO2 consumption with scores over 0.92, while the score of the sub-criterion perfomance of regulation is only 0.49. The environmental criteria have low importance weight with an overall weight of 0.15 in the PS design. The solution of Method 2 yields low operational costs with a score of 0.91, whereas the scores for investment costs and maintenance costs are 0.56 and 0.37, respectively.

In summary, the main effects of considering technical, economic, and environmental criteria in a PS design in contrast with only considering the minimization of LCC lie in that the ultimate solution of Method 2 uses greater number of pumps than in the Method 1, but the size of the pump model in terms of flow in Method 2 is lower than in the Method 1. In addition, the configuration of the control system in the Method 2 is less complex than in the Method 1. These obtained results could be justified because the sub-criterion complexity of control system has the highest importance weight in the AHP method. In addition, the importance weight of technical criteria obtained in the AHP method are high with an overall value of 0.41.

6. CONCLUSIONS

This work developed a methodology for a PS design that consider together technical, environmental, and economic criteria in the design. The process of this methodology is based in the AHP method. This methodology proposed a quantitative assessment of the potential solutions in every one of the criteria. In this way, this work has achieved that the methodology be standardized and could be applied in any kind of PS. In addition, this methodology, has solved the limitations of the classical PS design including technical aspects in the design process.

This methodology has introduced an optimization process in the control systems strategies searching the optimal number of pumps in operation and the current rotational speed to minimize energy consumption in the PS. This contribution has achieved to mitigate polluted energy produced in the PS and be friendly to the environment.

The principal effect of design a PS with the proposed methodology (Method 2) in contrast to the classical methodology (Method 1) is in the configuration of the PS. The ultimate solution based on Method 2 tends to use a greater number of pumps than in Method 1, but with a smaller flow in the pumps. The control system of the solution in Method 2 is less complex than the solution in Method 1. In summary, the principal effects of Method 2 in contrast with Method 1 are visualized the characteristics in technical aspects because the importance weight of these criteria obtained in the AHP method is high with a weight of 0.41

The inclusion of environmental aspects in a PS design (Method 2) allow that the ultimate solution could have better characteristic in environmental criteria, especially in the MEI, though the importance weight of this criteria is low in comparison with other technical or economic criteria.

7. REFERENCES

[1] United Nations Water, “Development Report 2020: Water and Climate Change,” París, 2020.

[2] S. M. Bunn and L. Reynolds, “The energy-efficiency benefits of pumpscheduling optimization for potable water supplies,” IBM J. Res. Dev., vol. 53, no. 3, pp. 1–13, 2009, doi: 10.1147/JRD.2009.5429018.

[3] Y. Chang, G. Choi, J. Kim, and S. Byeon, “Energy cost optimization for water distribution networks using demand pattern and storage facilities,” Sustain., vol. 10, no. 4, 2018, doi: 10.3390/su10041118.

[4] S. Lipiwattanakarn, S. Kaewsang, N. Charuwimolkul, J. Changklom, and A. Pornprommin, “Theoretical Estimation of Energy Balance Components in Water Networks for Top-Down Approach,” Water, vol. 13, no. 8, p. 1011, 2021, doi: 10.3390/w13081011.

[5] C. Giudicianni, M. Herrera, A. di Nardo, A. Carravetta, H. M. Ramos, and K. Adeyeye, “Zero-net energy management for the monitoring and control of dynamically-partitioned smart water systems,” J. Clean. Prod., vol. 252, 2020, doi: 10.1016/j.jclepro.2019.119745.

[6] T. Walski and E. Creaco, “Selection of Pumping Configuration for Closed Water Distribution Systems,” J. Water Resour. Plan. Manag., vol. 142, no. 6, p. 04016009, Jun. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000635.

[7] C. León-Celi, P. L. Iglesias-Rey, F. J. Martínez-Solano, and D. Savic, “Minimum energy and pumping cost in looped networks with multiple pumping systems and reservoir tanks through the setpoint curve concept,” J. Water Resour. Plan. Manag., 2018.

[8] M. Abdallah and Z. Kapelan, “Iterative Extended Lexicographic Goal Programming Method for Fast and Optimal Pump Scheduling in Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 143, no. 11, pp. 04017066 (1–10), 2017, doi: 10.1061/(ASCE)WR.1943-5452.0000843.

[9] M. Abdallah and Z. Kapelan, “Fast Pump Scheduling Method for Optimum Energy Cost and Water Quality in Water Distribution Networks with Fixed and Variable Speed Pumps,” J. Water Resour. Plan. Manag., vol. 145, no. 12, p. 04019055, 2019, doi: 10.1061/(asce)wr.1943-5452.0001123.

[10] N. Mehzad and M. Tabesh, “Optimum Reliable Operation of Water Distribution Network Considering Pumping Station and Tank,” Iran. J. Sci. Technol. Trans. Civ. Eng., vol. 43, no. s1, pp. 413–427, 2019, doi: 10.1007/s40996-018-0174-4.

[11] S. Beygi, M. Tabesh, and S. Liu, “Multi-Objective Optimization Model for Design and Operation of Water Transmission Systems Using a Power Resilience Index for Assessing Hydraulic Reliability,” Water Resour. Manag., vol. 33, no. 10, pp. 3433–3447, 2019, doi: 10.1007/s11269-019-02311-x.

[12] P. S. Mahar and R. P. Singh, “Optimal Design of Pumping Mains Considering Pump Characteristics,” J. Pipeline Syst. Eng. Pract., vol. 5, pp. 1–6, 2014, doi: 10.1061/(ASCE)PS.1949-1204.

[13] L. J. Blinco, A. R. Simpson, M. Asce, M. F. Lambert, M. Asce, and A. Marchi, “Comparison of Pumping Regimes for Water Distribution Systems to Minimize Cost and Greenhouse Gases,” Water Resour. Manag., vol. 142, no. 6, 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000633.

[14] M. Hajibabaei, S. Hesarkazzazi, M. Lima, and F. Gschösser, “Environmental assessment of construction and renovation of water distribution networks considering uncertainty analysis,” Urban Water J., vol. 17, no. 8, pp. 723–734, 2020, doi: 10.1080/1573062X.2020.1783326.

[15] D. Torregrossa and F. Capitanescu, “Optimization models to save energy and enlarge the operational life of water pumping systems,” J. Clean. Prod., vol. 213, pp. 89–98, 2019, doi: 10.1016/j.jclepro.2018.12.124.

[16] E. Creaco, E. Lanfranchi, C. Chiesa, M. Fantozzi, C. A. Carrettini, and M. Franchini, “Optimisation of leakage and energy in the Abbiategrasso district,” Civ. Eng. Environ. Syst., vol. 33, no. 1, pp. 22–34, 2016, doi: 10.1080/10286608.2015.1135136.

[17] E. Salomons, M. Housh, and M. Asce, “A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems,” J. Water Resour. Plan. Manag., vol. 146, no. 4, pp. 1–12, 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001188.

[18] M. Manteigas, A. Andrade-Campos, A. André, and B. Coelho, “Cost-Efficient Algorithms for a Pump Horizon Control in Water Supply System,” Water Resour. Manag., vol. 148, no. 1, pp. 1–15, 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001491.

[19] W. Ossadnik, S. Schinke, and R. H. Kaspar, “Group Aggregation Techniques for Analytic Hierarchy Process and Analytic Network Process: A Comparative Analysis,” Gr. Decis. Negot., vol. 25, no. 2, pp. 421–457, 2016, doi: 10.1007/s10726-015-9448-4.

[20] T. L. Saaty, "Decision making with the analytic hierarchy process," Int. J. Serv. Sci., vol. 1, no. 1, pp. 83–98, 2008.

[21] L. K. Raminelli and D. Costa, "Hierarchy of hydraulic and energy conservation actions at water supply systems Hierarchy of hydraulic and energy conservation actions at water supply systems," Urban Water J., vol. 00, no. 00, pp. 1–11, 2020, doi: 10.1080/1573062X.2020.1729386.

[22] F. Taillandier, S. M. Elachachi, and A. Bennabi, "A decision-support framework to manage a sewer system considering uncertainties," Urban Water J., vol. 17, no. 4, pp. 344–355, 2020, doi: 10.1080/1573062X.2020.1781908.

[23] M. Ward, C. Poleacovschi, and M. Perez, "Using AHP and Spatial Analysis to Determine Water Surface Storage Suitability in Cambodia," Water (Switzerland), pp. 1–18, 2021.

[24] C. X. Briceño-León, D. S. Sanchez-Ferrer, P. L. Iglesias-Rey, F. J. Martinez-Solano, and D. Mora-Meliá, "Methodology for pumping station design based on analytic hierarchy process (AHP)," Water (Switzerland), vol. 13, no. 20, p. 2886, Oct. 2021, doi: 10.3390/w13202886.

[25] P. L. Iglesias Rey, F. J. Martínez Solano, F. Arango Gil, and J. Lozano cortés, "Methodology for the selection of pumping stations considering its mode of operation," IPWE, p. 13, 2018.

[26] I. Sarbu, Ioan; Borza, "Energetic Optimization Of Water Pumping in Distribution Sytems," Period. Polytech. Mec. Eng., vol. 42, no. 2, pp. 141–152, 1998.

[27] C. X. Briceño-León, P. L. Iglesias-Rey, F. J. Martinez-Solano, D. Mora-Meliá, and V. S. Fuertes-Miquel, "Use of fixed and variable speed pumps in water distribution networks with different control strategies," Water (Switzerland), vol. 13, no. 4, p. 479, Feb. 2021, doi: 10.3390/w13040479.

[28] S. Alvisi and M. Franchini, "A robust approach based on time variable trigger levels for pump control," J. Hydroinformatics, vol. 19, no. 6, pp. 811–822, 2017, doi: 10.2166/hydro.2017.141.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14103

AN IMPROVED CONTROL STRATEGY FOR HIGH-PRESSURE PUMPING IRRIGATION SYSTEMS

Ye Wang[1], **Qi Zhao**[2], **Erik Weyer**[3], **Wenyan Wu**[4], **Angus R. Simpson**[5],
Ailsa Willis[6]

[1,3] Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia
[2,4] Department of Infrastructure Engineering, The University of Melbourne, Victoria 3010, Australia
[5] School of Civil, Environmental & Mining Engineering, The University of Adelaide, South Australia 5005, Australia
[6] Lower Murray Water, Mildura, Victoria 3500, Australia

[1] *ye.wang1@unimelb.edu.au*, [2] *qzhao4@student.unimelb.edu.au*,
[3] *ewey@unimelb.edu.au*,[4] *wenyan.wu@unimelb.edu.au*,
[5] *angus.simpson@adelaide.edu.au*, [6] *ailsa.willis@lmw.vic.gov.au*

Abstract

Pumped irrigation systems are critical infrastructure that supply water from water sources to rural users through interconnected pressurized water pipes. To supply enough pressure head to all the irrigation outlets located in the irrigation network, it is required to have an appropriately high pressure head supplied by the pump station. Therefore, the energy consumption related to pumping could be correspondingly high, which may lead to significant operating costs. With the aid of real-time flow data available from each irrigation outlet, the proposed improved pump head setpoint selection algorithm can locate the most critical outlet in terms of the smallest downstream pressure head and therefore guarantee to deliver the minimum pressure required for all active outlets in the network. By using the proposed algorithm, it has been demonstrated with a two-day simulation that a 4.74% savings in pumping energy cost as well as a reduction in the associated greenhouse gas emissions can be achieved.

Keywords
Pump pressure setpoint, Energy cost savings, Greenhouse gas reduction, High-pressure pumping irrigation systems.

1 INTRODUCTION

Pressurized irrigation water networks supply water to municipal (industrial, commercial and residential) and rural (irrigation and residential) users through pipes [1], [2]. The growth in water demand has led to increased pumping and correspondingly larger energy costs, as well as increased greenhouse gas (GHG) emissions (when electricity is sourced from fossil fuel sources). The water sector in Victoria, Australia has committed to reducing its emissions by 42% by 2025 and to net-zero emissions by 2050, under Victoria's water plan, Water for Victoria [3]. For Lower Murray Water (LMW) - a local water authority in Victoria, this objective translates to a total reduction of 15,535 tonnes of GHG emissions by 2025 [4], [5].

Pumping energy cost often makes up a large portion of the operating cost in a pressurized irrigation water network. Within the LMW's four irrigation systems, the Robinvale High-Pressure System (RVHPS) accounts for approximately 60% of the total electricity cost. During the peak demand season, the RVHPS pump station will often run at close to full capacity for 24 hours a day. As a result, the energy cost and associated GHG emissions from this system are very large. The energy cost savings and GHG reduction can be achieved by improving the pump head setpoint selection strategy and applying better pump operations [6]. Currently, the pump station of the

RVHPS is operated using a proportional-integral (PI) controller with a pressure setpoint found from a pressure-setpoint curve. However, depending on the demand and the location of the active irrigation outlets at times this curve may give a setpoint value that is too high, leading to unnecessary energy costs; while at other times the setpoint may be too low, leading to a pressure below the service requirement at some irrigation outlets.

In this paper, we investigate an improved real-time control strategy to find pump pressure setpoints. This new strategy takes advantage of the flow measurement information from all the irrigation outlets and a well-calibrated hydraulic simulation model [7] to identify the most critical irrigation outlet in terms of pressure. Then, a new pump pressure setpoint can be determined based on this critical outlet. The improved control strategy can potentially lead to energy cost savings, GHG emissions reduction and an improved level of service to irrigators.

With a case study of the Robinvale irrigation network, we have demonstrated that the proposed strategy can obtain 4.74% savings in pumping energy cost as well as a reduction in the associated greenhouse gas emissions. The economic and environmental benefits that can be achieved using the new strategy for selecting pump setpoints are demonstrated using historical data from 28th and 29th Dec 2019. The current pump setpoints are for most of the time higher than the setpoints produced using the proposed selection strategy. The average reduction in pump setpoint values is 4.50 m. This shows that the new strategy can lead to lower energy consumption and lower associated GHG emissions while delivering the minimum service pressure head downstream to all the active irrigation outlets.

2 CONTROL SYSTEM DEVELOPMENT

2.1 Current Control Strategy

A proportional-integral (PI) controller is currently employed for the operational management of the pumps for the high-pressure irrigation system. A control input can be generated using the PI controller so that the output pressure of the pump station can track a desired pressure setpoint $r(t)$ at each sampling time t. In general, a PI controller can be formulated below as shown in equation (1):

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau, \quad (1)$$

where $e(t)$ is an error signal obtained by $e(t) = r(t) - p(t)$ with the measured pressure head at the exit of the pump station. Furthermore, K_p and K_i are proportional and integral gains, which are tuning parameters. In equation (1), the proportional term $K_p e(t)$ contributes to a response adjusted by multiplying the error signal by a constant gain K_p, while the integral term $K_i \int_0^t e(\tau) d\tau$ contributes to an adjusted accumulated error offset over time that can accelerate the setpoint tracking and eliminate steady state setpoint errors.

The pump pressure setpoint $r(t)$ has been normally selected based on a pressure-flow curve. For instance, as shown in Figure 1, a pump pressure setpoint can be selected by using the measured system total flow at the exit of the pump station. Considering that the system flow is in the range between 0 L/s and 3500 L/s, the corresponding pressure setpoint varies from 81.9 m to 102.3 m. Since irrigation outlets are located in each branch of the network, some outlets far away from the pump station (e.g., critical outlets at the end of a certain branch as shown in Figure 2) may suffer low pressure heads but the outlets near the pump station may have exceeded minimum required pressure heads. To meet the minimum required pressure head (i.e. 35 m) at all active irrigation outlets, the pressure setpoint selected from the pressure-flow curve in Figure 1 is usually conservative and even in this case, there is no guarantee that the minimum required pressure can always be satisfied for all the irrigation outlets.

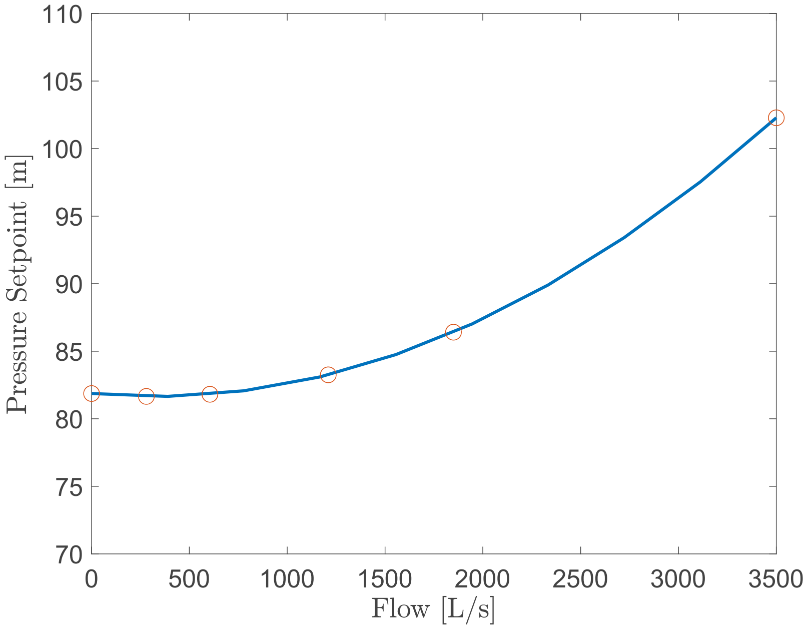


Figure 1. The Pump Pressure Setpoint Curve used by LMW.

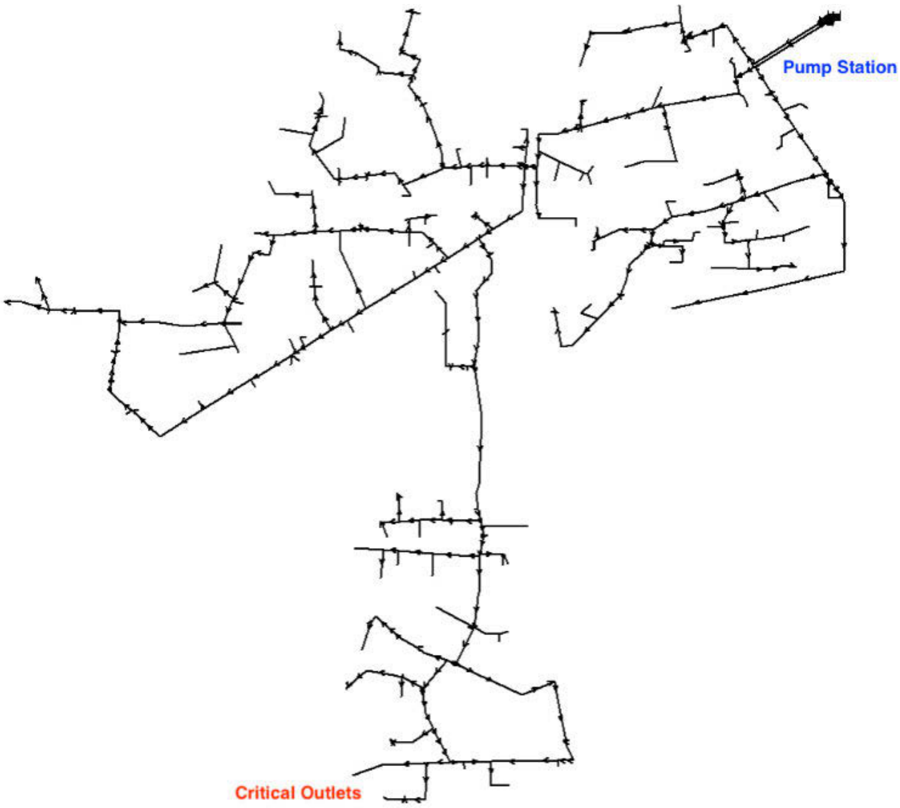

Pump Station
Critical Outlets

Figure 2. Robinvale High-Pressure Irrigation Network Layout.

BY NC SA

2.2 Improved Pump Pressure Setpoint Selection

The new pump setpoints for the proportional-integral controller are adjusted considering actual irrigation deliveries. In principle, all irrigators who have placed an irrigation water order will receive at least the agreed minimum level of service (35 m of pressure head) regardless of what other irrigators do. Meanwhile, no specific attempt will be made to provide sufficiently high pressure to irrigators who are taking irrigation water but who have not ordered water.

Based on these principles, the steps for finding the pump setpoints based on a new method can be summarized as follows:

1. Initialize the pump setpoint with an arbitrarily selected value (e.g. 100 m of pressure head).
2. At every 15 minutes, for each outlet
    a. Obtain the ordered demand and measured irrigation delivery flows at active outlets (the outlets actually ordered and takes water from the network); and
    b. If both flows are greater than zero (indicating an active irrigation outlet), then
        i. Obtain the pressure head at the upstream side of each irrigation outlet from the EPANET hydraulic model simulation;
        ii. Compute the head loss across the irrigation outlet based on the minimum of the ordered demand and actual measured flow (also taking into account different diameters of irrigation outlets); and
        iii. Obtain the outlet downstream pressure head by subtracting the outlet head loss from the upstream pressure head obtained from EPANET.
3. Find the most critical active irrigation outlet by finding the minimum downstream pressure head.
4. Obtain the difference in pressure head by subtracting the downstream pressure head at the most critical irrigation outlet from the required minimum pressure head.
5. Adjust the pump setpoint by adding (or subtracting) the computed difference in pressure head to the pump setpoint 15 minutes ago.
6. Repeat Step 2.

3 CASE STUDY AND RESULTS

3.1 System Description

The Robinvale area had a population of 3,313 at the 2016 census. The Robinvale population includes people living in the Robinvale township and the surrounding rural area, some within the Robinvale Irrigation District. As shown in Figure 2, the RVHPS consists of one pump station and pipelines to supply water from the River Murray to 244 irrigation outlets across a 2700 ha area.

The pump station capacity is about 3,700 L/s. A total of 3,000 L/s is available for irrigation water orders and 300 L/s is reserved for domestic and stock (D&S) use. The remaining capacity of about 400 L/s accommodates irrigators using flow rate in excess of orders, turning an outlet off late or starting an order early. Better compliance with irrigation orders would allow some of the 400 L/s to be made available for irrigation purposes.

3.2 Economic and Environmental Outcomes

The new pump setpoint selection algorithm has been evaluated in simulation and compared with old setpoints using historical data from 28th and 29th Dec 2019. The comparison result with old and new pump setpoints is shown in Table 1. From this table, it can be seen that an average reduction in pressure setpoints is 4.50 m. In addition, operations with the new pump setpoints over the two days investigated resulted in a 7.08 MWh savings in pumping energy, an $800 savings in pumping energy cost, and a reduction of 7.72 tonnes in GHG emissions. The operations with the new pump setpoints can lead to 4.74% savings in operation related energy consumption and GHG emissions compared to old operations during the chosen summer period (peak demand).

Table 1. Comparison of Economic and Environmental Outcomes (over 2-day Simulation Period).

	With Old Setpoints	With New Setpoints
Average Pump Setpoints [m]	94.0	89.5
Total Energy [MWh]	149.3	142.2
Total Energy Cost [$]	16,300	15,500
Volume Pumped [ML]	464	464
Unit Energy Usage [kWh/ML]	322	306
GHG Emissions [tonne]	163	155

3.3 Comparison: Mismatch between Total Ordered and Delivered Irrigation Flows

Before an irrigator takes water, an order should be placed via a computerised water ordering system. Actual measured irrigation deliveries at some outlets may not match the orders. As shown in Figure 3, total ordered demands and sum of measured irrigation flows are different over the two-day simulation period (28th and 29th Dec 2019). In order to understand the impact of the mismatch between total ordered demands and sum of actual delivered irrigation flows on the operation of the RVHPS, its energy consumption and on GHG emissions, the above comparison and analysis are carried out again for two different operation periods when ordered and delivered flows differ. In operation period 1 (within the 2-day simulation period), the sum of flow rates through irrigation outlets exceeds the sum of irrigation orders; and in operation period 2, the sum of ordered flow rates exceeds the sum of flow rates being used through irrigation outlets. A comparison of the results for the two operation periods is presented in the following two sub-sections.

While the sum of actual flow rates through irrigation outlets may exceed the sum of irrigation orders during a period, the actual flow rate may not have necessarily exceeded the ordered flow rate at every irrigation outlet, and vice versa. Compliance with the ordered flow rate at an irrigation outlet may vary during a period, and an outlet may vary between taking more and taking less than ordered.

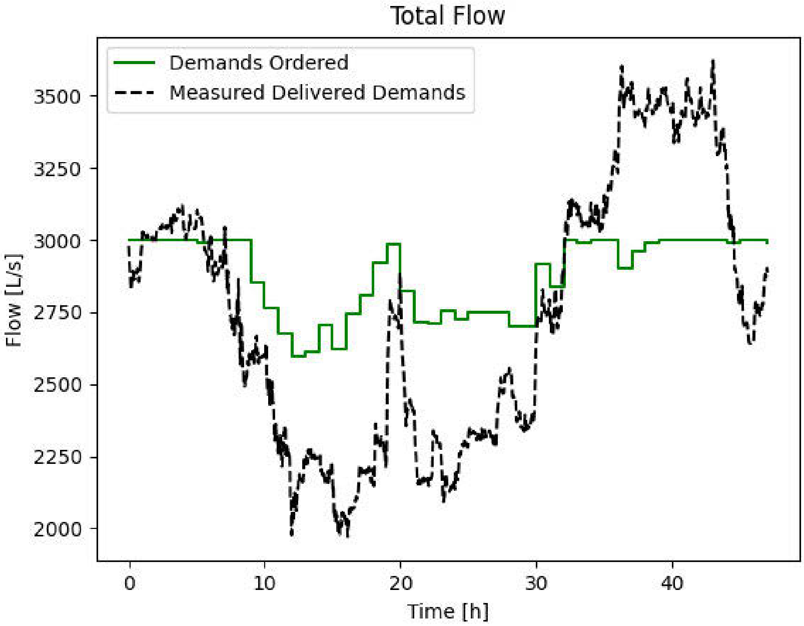


Figure 3. Comparison of Total Pumping Flows with Ordered versus Measured Irrigation Deliveries.

3.4 Potential Savings for the Period of Ordered Flows Less Than Flows Delivered

Within the two-day simulation, this operation period occurred from 32:05 hours to 44:31 hours (Period 1). The sum of actual flow rates through irrigation outlets exceeded the total ordered flow rates. The comparison of the calculation results of energy, energy cost and GHG emissions is reported in Table 2. As can be seen in Table 2, operations with the new pump setpoints over this period can lead to approximately 2.7 MWh savings in pumping energy, $300 savings in pumping energy cost, and a reduction of 2.9 tonnes in GHG emissions. In summary, the improved operation can lead to 5.21% in operation related energy consumption and GHG emissions, when ordered flows are less than delivered irrigation flows.

Table 2. Comparison of Results for Economic and Environmental Outcomes (Period 1).

	With Old Setpoints	With New Setpoints
Average Pump Setpoints [m]	100.1	95.0
Total Energy [MWh]	50.8	48.1
Total Energy Cost [$]	5,500	5,200
Volume Pumped with Non-compliance [ML]	149	149
Unit Energy Usage [kWh/ML]	340	323
GHG Emissions [tonne]	55.3	52.4

Non-compliance with irrigation orders can result in actual delivered flows being higher than ordered demands and hence larger head losses in pipes. To overcome this, the pressure setpoint at the pump station must be increased in order to guarantee minimum required pressure heads at active irrigation outlets. It is also worth mentioning that non-compliance behavior is not always satisfied. The minimum required pressure head is taken into account the ordered flow if farmer takes more water than ordered. During the operation period from 32:05 hours to 44:31 hours in the two days that were investigated, some irrigators were taking more water than they ordered. The total ordered volume was 134 ML (Table 2) and the actual pumped irrigation volume was 149 ML (Table 3). The compliant behavior considered here is that all the irrigators take water strictly in accordance with orders, that is, staying within the start and stop times and adhering to the magnitude of the ordered flow rate. As can be seen in Table 3 the non-compliant behavior can lead to approximately a $1.50/ML increase in unit energy cost compared with the old pump station setpoint operation. In a similar way, for non-compliant behavior, an approximately $1.70/ML increase in unit energy cost occurs with the new pump station setpoint operation. In summary, non-compliance with irrigation orders can cause a 4.21% and 5.07% increase in unit energy cost taking into account old and new pump operations, respectively.

Table 3. Comparison of Results for Energy Cost with Compliant and Non-compliant Water Use (Period 1).

	With Old Setpoints	With New Setpoints
Average Pump Setpoints [m]	96.1	90.5
Volume Pumped with Compliance [ML]	134	134
Unit Energy Cost with Compliance [$/ML]	35.6	33.5
Unit Energy Cost with Non-compliance [$/ML]	37.1	35.2
Percentage of Cost Increase for Non-Compliance	4.21%	5.07%

3.5 Potential Savings for the Period of Ordered Flows Greater Than Flows Delivered

Within the two-day simulation, this operation period occurred from 05:31 hours to 31:00 hours (Period 2). The sum of actual delivered flow rates through irrigation outlets was less than the total ordered flow rates. The comparison of the calculated energy use, energy cost and GHG emissions is reported in Table 4. As can be seen in the table, operations with the new pump setpoints over this period can lead to 3.4 MWh savings in pumping energy, $300 savings in pumping energy cost, and a reduction of 3.7 tonnes in GHG emissions. In summary, the improved operation can lead to a 5.10% reduction in operation related energy consumption and GHG emissions, when the sum of ordered flows is greater than the sum of actual flows being used through irrigation outlets.

Furthermore, over this operation period, the average total pumping flow with ordered demands is 2,779 L/s compared to the one with measured irrigation deliveries of 2,384 L/s. The difference between the two average total pumping flows is 394 L/s, which means the network can accommodate delivery to an extra 7.5 irrigators on average (with an assumed ordering capacity of 52 L/s per irrigator) during the same time period of 26 hours.

Table 4. Comparison of Results for Economic and Environmental Outcomes (Period 2).

	With Old Setpoints	With New Setpoints
Average Pump Setpoints [m]	90.6	85.9
Total Energy [MWh]	66.4	63.0
Total Energy Cost [$]	7,200	6,900
Volume Pumped with Non-compliance [ML]	216	216
Unit Energy Usage [kWh/ML]	308	292
GHG Emissions [tonne]	72.4	68.7

3.6 Comparison of Level of Service Delivered by the Old and New Pump Setpoints

A comparison of the level of service has been assessed with irrigators using the ordered demands, as shown in Figure 4. This simulates the case where all irrigators comply with their water orders. In this case, operation with the new pump setpoints ensures that the required minimum pressure head (35 m) is delivered at all the irrigation outlets every 15 minutes when the pump setpoint is recalculated. As a comparison, the operations with the old pump setpoints did not always ensure all the irrigation outlets have enough pressure. As shown in Figure 4, one irrigation outlet has the largest unsatisfied pressure head magnitude was 3.51 m for a duration of approximately 3 hours.

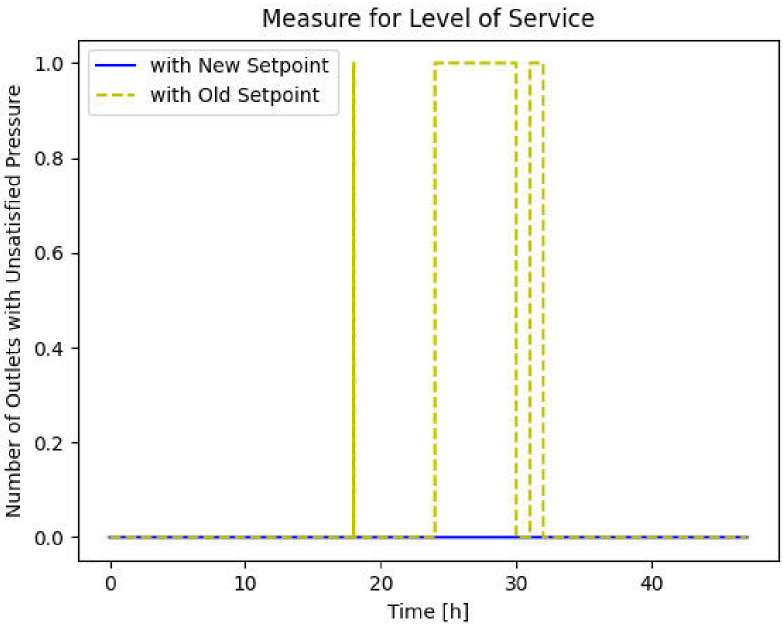


Figure 4. Comparison of the level of service.

4 CONCLUSIONS

In this paper, an improved pump pressure setpoint has been developed for a high-pressure pumping irrigation system. The proposed algorithm can select improved pump pressure setpoint based on real-time measured flows sent from each irrigation outlet to identify the most critical outlet currently active outlets with minimum downstream pressure. From the simulation results, it can be seen that the PI controller with the improved pump pressure setpoint can gain 4.74% savings in pumping energy cost as well as the reduction in the associated greenhouse gas emissions.

Non-compliance with irrigation orders can have a significant impact on pumping energy needed to meet the minimum pressure requirement at all outlets with irrigation orders. The pump setpoint selection can be further incorporated with the water ordering system. Moreover, potential water storage can be designed into the RVHPS network, which gives more degree of freedom to further investigate better control operations leading to more energy savings and GHG emission reduction.

5 ACKNOWLEDGEMENT

We would like to thank the LMW for providing real data of the RVHPS, and also thank the Faculty of Engineering and Information Technology (FEIT), the University of Melbourne for funding through the 2020 interdisciplinary grant program. Ye Wang and Wenyan Wu acknowledge support from the Australian Research Council via the Discovery Early Career Researcher Awards (DE220100609 and DE210100117), respectively.

6 REFERENCES

[1] M. Cantoni, E. Weyer, Y. Li, S. Ki Ooi, I. Mareels, and M. Ryan, "Control of large-scale irrigation networks," Proceedings of the IEEE, vol. 95, no. 1, 2007, pp. 75-91.

[2] Y. Kim, R. G. Evans, and W. M. Iversen, "Remote sensing and control of an irrigation system using a distributed wireless sensor network,". IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 7, 2008, pp. 1379-1387.

[3] Department of Environment Land Water and Planning. "Statement of Obligation (Emissions Reduction)," https://www.water.vic.gov.au/climate-change/reduced-emissions-in-the-water-sector/net-zero-emissions-by-2050, 2018 (Accessed on 1 Oct 2020).

[4] Lower Murray Water, "Lower Murray Water Annual Report 2018-2019," https://www.lmw.vic.gov.au/wp-content/uploads/2019/10/Lower-Murray-Water-Annual-Report-2019.pdf, 2019 (Accessed on 1 Oct 2020).

[5] Lower Murray Water, "Lower Murray Water Corporate Plan 2019-2020," https://www.lmw.vic.gov.au/wp-content/uploads/2019/10/2019-032362-2019-20-LMW-Corporate-Plan-FINAL-Version-as-at-31-May-2019.pdf, 2019 (Accessed on 1 Oct 2020).

[6] Y. Wang, K. Too Yok, W. Wu, A. R. Simpson, E. Weyer, and C. Manzie. "Minimizing pumping energy cost in real-time operations of water distribution systems using economic model predictive control." Journal of Water Resources Planning and Management, vol. 147, no. 7, 2021, 04021042.

[7] L. Rossman, "EPANET 2 users manual," National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 2000 (Accessed on 1 Oct 2020).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14104

WATER DISTRIBUTION SYSTEM DESIGN WITH BEHIND-THE-METER SOLAR ENERGY UNDER VARIOUS DISCOUNT RATES

Qi Zhao[1], Angus R. Simpson[2], Wenyan Wu[3] and Ailsa Willis[4]

[1,3]The University of Melbourne, Parkville, Victoria, Australia
[2]The University of Adelaide, Adelaide, South Australia, Australia
[4]Lower Murray Water, Mildura, Victoria, Australia

[1] *qzhao4@student.unimelb.edu.au,* [2] *angus.simpson@adelaide.edu.au,*
[3] *wenyan.wu@unimelb.edu.au,* [4] *Ailsa.Willis@lmw.vic.gov.au*

Abstract

Water distribution systems (WDSs) are essential parts of urban and rural infrastructure systems. The energy consumption including associated costs and GHG emissions for distributing water has increased in recent years. As a result, behind-the-meter (BTM) energy systems such as solar panels and energy storage facilities have been installed by water utilities to reduce energy consumption from the centralised energy grid. There has been a lot of research on the optimisation of the design of WDSs, however, many of these works have not considered BTM energy options or have incorporated BTM energy options in an ad-hoc fashion. Therefore, this study is proposed to optimise the design of WDSs considering BTM solar energy sources. A pressurised irrigation supply system in Victoria, Australia has been used as the case study system. In this paper, the design of a real-world WDS has been formulated as a multi-objective optimisation problem to investigate the trade-offs between the total life cycle cost and total life cycle greenhouse gas (GHG) emissions under various discount rates scenarios used to estimate the operational cost and emissions throughout the service life of the system. It has been found that both the total life cycle cost and GHG emissions have been reduced significantly when BTM solar energy is incorporated. In addition, with the same solar photovoltaic (PV) size, the optimal pipe diameters, as well as the objective function values are sensitive to the discount rate values that have been used.

Keywords
Water distribution systems (WDSs), design, optimisation, pipe sizing, behind-the-meter energy, solar PV.

1 INTRODUCTION

Water distribution systems (WDSs) are essential parts of urban and agricultural infrastructure systems. The energy consumption including associated costs and GHG emissions for distributing water has increased worldwide in recent years due to the rise in water demand resulting from population growth. Therefore, it has been considered by water utilities to optimise the system and find alternative energy solutions, in order to achieve economic and environmental benefits, and promote sustainable development at the same time. In addition, technological development has also contributed to the increased consideration of renewable energy applications, such as on-site renewable energy generation and storage (i.e. "behind-the-meter" (BTM) energy systems). As a result, energy consumption from the centralised energy supply grid and associated GHG emissions can be further reduced.

WDSs are topologically and dimensionally complicated systems, which consist of a range of interconnected components [1]. The optimal design of WDSs commonly involves choosing the best combination of system components (e.g. pipes, pumps and storages) and their sizes and locations to achieve a minimum total life cycle cost [2]. Some other studies have also minimised the GHG emissions simultaneously with multi-objective optimisation approaches [3, 4].

Renewable energy sources in developing BTM energy solutions have been widely considered to either supplement or replace the traditional energy supply [5] in recent years. Among the existing renewable energy sources, solar energy is commonly considered in WDSs [6, 7]. The use of photovoltaic (PV) conversion of solar energy to supply energy to WDSs can significantly reduce the consumption from the electricity grid, and the associated GHG emissions [8], particularly for pumping. Also, in comparison with other forms of renewable energy sources, solar energy supply is more likely to be "behind-the-meter" considering the size and flexibility of the system. Nevertheless, many of the existing studies looking for optimal design and operation solutions for WDSs have not included BTM energy options in the optimisation process.

Considering the complexity, as well as the large costs associated with the life cycle of WDSs, optimisation tools are often used to assist the design of these systems [9, 10]. Deterministic (or classic) and metaheuristic algorithms are two common categories of optimisation techniques used for WDS design optimisation. In comparison with deterministic optimisation methods, metaheuristic algorithms such as Evolutionary Algorithms (EAs), have been found to perform better in solving more complex problems with more decision variables and constraints [11]. Even though they are often associated with a larger computational cost, they have a higher probability of finding optimal or near-optimal solutions due to the exploratory nature of metaheuristics [12].

In this paper, a multi-objective optimisation problem has been formulated to investigate the impact of BTM solar energy on the design of WDSs, considering various discount rates. A fully pipelined irrigation supply system in Victoria, Australia has been chosen as the case study system. Both economic and environmental objectives have been optimised over the system design life.

2 CASE STUDY SYSTEM

The case study system is a pressurised irrigation system located in the Robinvale irrigation district in north-western Victoria, Australia, as shown in Figure 1. A high-pressure pump station on the south bank of the Murray River pumps water from the river throughout the pipeline system to customers for irrigation and domestic water use. Table grapes are mainly planted in this area [13] and they need a large amount of water for irrigation. There are 433 pipes and 244 irrigation outlets in the network. The minimum pressure head required at users' outlets is 35 m. The layout of the network is shown in Figure 2. Pipes diameters considering BTM solar energy are optimised in this study, assuming the locations of pipes are the same as the existing system. Relevant data has been provided by the local water authority Lower Murray Water (LMW). An EPANET model is used as the simulation model for the system [14].

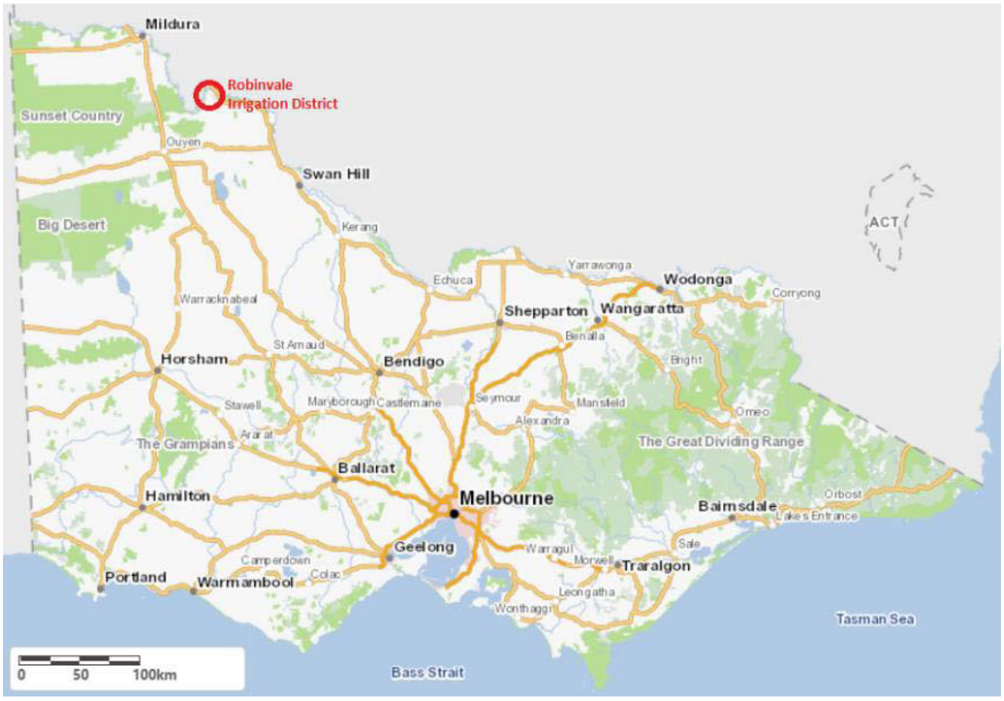

Mildura
Robinvale Irrigation District
Sunset Country
Ouyen
Swan Hill
Big Desert
Kerang
ACT
Yarrawonga
Wodonga
Corryong
Echuca
Shepparton
Wangaratta
Warracknabeal
Benalla
St Arnaud
Horsham
Bendigo
Bright
Stawell
Maryborough
Castlemaine
Seymour
Mansfield
Alexandra
Omeo
Ararat
The Grampians
The Great Dividing Range
Ballarat
Hamilton
Melbourne
Bairnsdale
Orbost
Lakes Entrance
Sale
Geelong
Warragul
Morwell
Traralgon
Camperdown
Colac
Portland
Warrnambool
Leongatha
Wonthaggi
Tasman Sea
Bass Strait
0
50
100km

Figure 1. Location of the study area (in red) within the state of Victoria, Australia

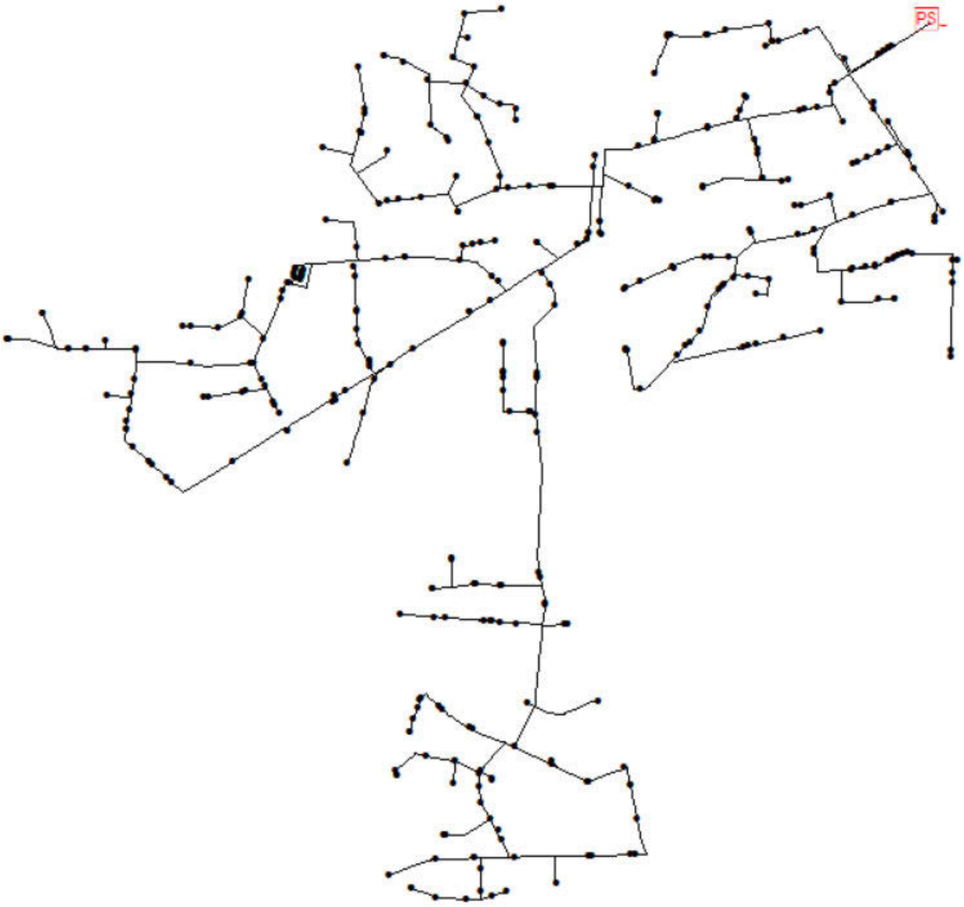


Figure 2. Case study system – a pressurised irrigation system in Robinvale, Victoria, Australia

3 METHODS

3.1 Problem formulation

In this study, a multi-objective optimisation problem has been formulated to minimise total life cycle cost and total life cycle GHG emissions from the case study system. Decision variables are pipe diameters. To improve optimisation efficiency, the 433 pipes in the network have been grouped into 10 decision variables based on energy requirements computed within the network. A minimum pressure head of 50 m (35 m minimum service pressure + head loss of 15 m caused by control valves before the outlet) is used at each node in the EPANET model as the inequality constraint. The algorithm NSGA-II [15] is used in a Python-based optimisation package 'pymoo' [16], with a Python wrapper [17] installed to call functions in EPANET Programmer's Toolkit [18].

Different scenarios have been considered to investigate the impact of solar PV sizes on the optimal design of WDSs. The minimisation of the two objective functions with the incorporation of 0 MW (no solar), 1 MW, 2 MW, 3 MW, 4 MW and 5 MW solar PV has been investigated. A social discount rate of 1.4% [19] is considered initially. The analysis includes the comparison of different Pareto fronts between the six cases of solar PV sizes, trade-off analysis under a specific case (e.g. 5 MW), as well as the comparison of costs and GHG emissions under the 0 MW (no solar) and 5 MW solar case. In addition, other two discount rates (0% and 6%) have been considered to investigate their impact on the optimisation results. The population size and the number of generations for the multi-objective genetic algorithm optimisation runs are both 100 for all optimisation runs.

3.2 Objective Function 1 - Costs

The first objective function (OF1) is to minimise the total life cycle cost, which is given by

$$min\, OF1:\ LCC = CC + OC + PRC + SRC \tag{1}$$

where LCC, CC, OC, PRC and SRC = the total life cycle cost, capital costs, operating costs, pump and solar panel replacement costs respectively. The system is assumed to have the same design life as pipes of 100 years [20]. The capital costs mainly involve purchasing and constructing pipes, pumps and solar panels. The operating costs mainly include spending on electricity purchased from the grid when the energy required is greater than the solar energy production. Pumps and solar panels need to be replaced regularly within 100 years to maintain their performance. Present value analysis (PVA) has been conducted for operating costs, as well as pump and solar PV replacement costs.

Capital cost

The capital cost is given by

$$CC = PiC + PSC + SPC \tag{2}$$

where PiC, PSC and SPC = the capital cost for pipes, the pump station and solar panels. Unit costs of pipes and pumps were sourced from the Department of Primary Industries [21]. A total of 4 different materials and 31 pipe diameters commercially available with corresponding unit costs have been included in the choice table. Pipe lengths were assumed to be the same as the existing system. The capital cost of the pump station has been estimated according to the total power of pumps needed. Average commercial solar panel prices in Victoria as well as solar panel sizes have been taken into account to calculate the solar system capital cost.

Operating cost

The present value (PV) of ongoing electricity operating cost is given below

$$PV(OC) = OC_a\,[\frac{(1+i)^k - 1}{i(1+i)^k}] \tag{3}$$

where i = selected discount rate, OC_a = the annual operating cost (assumed to be unchanged over the design life k = 100 years) which is given as

$$OC_a = \sum_{i=1}^{365} \Big(\sum_{t=1}^{T} c_t \times IE_t\Big)_i \tag{4}$$

where c_t = electricity tariff at time t, IE_t = the energy imported from the electrical grid at time t, which is given as

$$IE_t = (PowerP_t - PowerS_t) \times \Delta t \tag{5}$$

where $PowerP_t$ is the power required for pumping at time t, considering the required flow, pumping head as well as an assumed motor efficiency of pumps; $PowerS_t$ is the power of solar generation at time t. Excess energy will be fed back to the grid when the solar system generates more energy than that needed for pumping. Pump motor efficiency is assumed to be 95%. The solar system is assumed to be able to generate 3.6 kWh of electricity when 1 kW of a solar panel is installed [22].

Pump & solar panel replacement cost

It has been assumed that pump sets and solar panels have a service life of 20 [20] and 25 years [23], respectively. The present value (PV) of pump and solar panel replacement cost is given by

$$PV(PRC) = PuC\,[\frac{1-(1+i)^{(s-k)}}{(1+i)^s - 1}] \tag{6}$$

$$PV(SRC) = SPC\,[\frac{1-(1+i)^{(t-k)}}{(1+i)^t - 1}] \tag{7}$$

where PuC, SPC = capital costs of pumps and solar panels, respectively; k, s and t = 100, 20 and 25 years, namely the service life of the whole system, pumps and solar panels, respectively.

3.3 Objective Function 2 - GHGs

The second objective function (OF2) is to minimise the total life cycle GHG emissions, which is given as

$$min\,OF2\text{: } LCGHG = CGHG + OGHG \tag{8}$$

where $LCGHG$, $CGHG$ and $OGHG$ = the total life cycle, capital and the operating GHG emissions, respectively. Capital GHG emissions are mainly from the process of manufacturing pipes and solar panels. Operating GHG emissions are mainly from consuming electricity imported from the grid (assumed to be drawn from a certain proportion of fossil fuel sources) when solar production is insufficient. The present value analysis is also required to deal with the operating GHG emissions during the entire 100-year life of the system.

Capital GHG emissions

The capital GHG emissions are given as

$$CGHG = (PiEE + SPEE) \times EF \tag{9}$$

where $PiEE$ and $SPEE$ = the embodied energy for manufacturing pipes and solar panels, respectively; EF = the emission factor. Embodied energy coefficients based on pipe materials and the pipe masses based on pipe lengths and unit weights are used to calculate the embodied energy of pipes. Solar panels are assumed to need the embodied energy of approximately 3700 kWh for producing 1 kW capacity of energy generation [24]. An emission factor of 1.09 kg CO_2-e/kWh has been used, as suggested by the Department of Industry Science Energy and Resources [25].

Operating GHG emissions

The present value (PV) of ongoing operating GHG emissions is given by

$$PV(OGHG) = OGHG_a\,[\frac{(1+i)^k - 1}{i(1+i)^k}] \tag{10}$$

where $OGHG_a$ = the annual operating GHG emissions, which are given as

$$OGHG_a = \sum_{i=1}^{365} \left(\sum_{t=1}^{T} IE_t\right)_i \times EF \tag{11}$$

where IE_t = the electricity energy purchased from the grid at time t; EF = the emission factor as above.

4 RESULTS AND DISCUSSION

4.1 Impact of solar panel sizes on optimisation results

As shown in Figure 3, for all six sizes of solar PV considered, trade-offs between the two objective functions have been observed when a 1.4% social discount is used. With an increase of solar PV sizes from top right to bottom left in Figure 3, both objectives have significantly decreased. As indicated in Table 1, the minimum total life cycle cost has decreased by approximately 20% from \$208M to \$167M, with the increase of solar PV size from 0 to 5 MW. Similarly, the minimum total life cycle GHG emissions have been reduced by about 59% from the 0 to 5 MW solar option. Therefore, the consideration of BTM solar PV can bring long-term benefits to the system economically and environmentally.

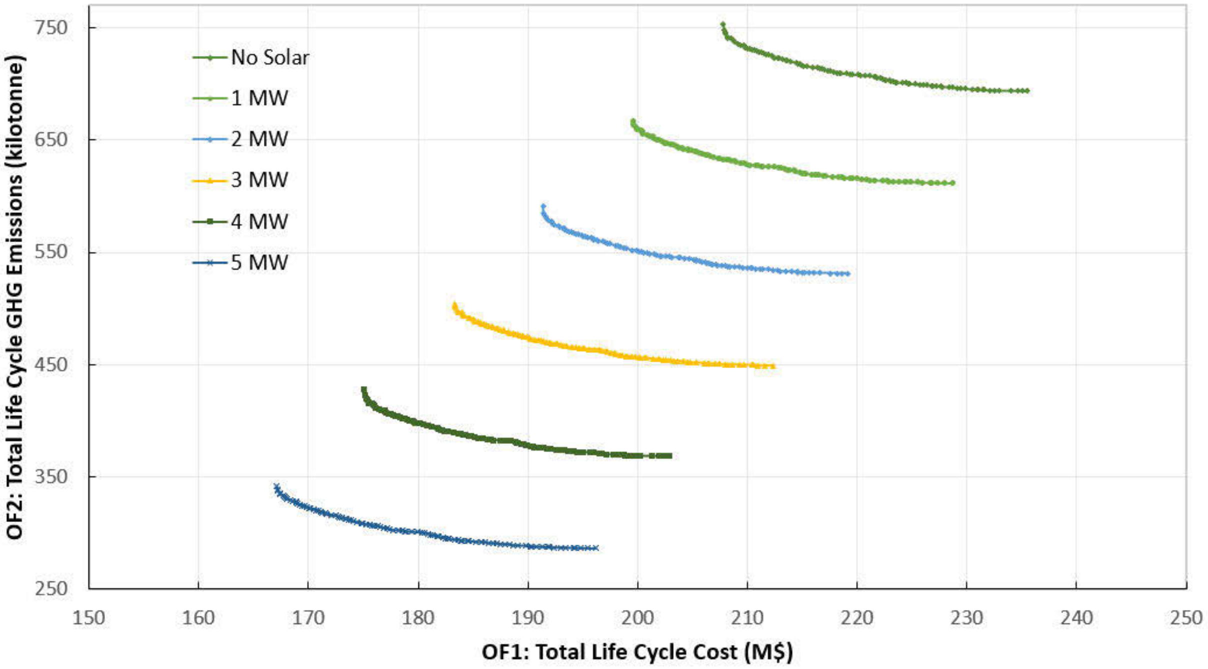


Figure 3. Pareto fronts with 0 MW (no solar), 1 MW, 2 MW, 3 MW, 4 MW and 5 MW solar PV installed (1.4% discount rate)

Table 1. Comparison of minimum objective function values with different sizes of solar PV incorporated (1.4% discount rate)

Solar PV size	0 MW	1 MW	2 MW	3 MW	4 MW	5 MW
Min. LCC (M$)	208	200	191	183	175	167
Min. LCGHG (kt)	694	612	531	449	368	286

A breakdown of costs and GHG emissions for both the least-cost and the least-GHG solutions under the 3 MW solar option as an example is demonstrated in Table 2 and Table 3. In total life cycle costs, the capital cost especially for purchasing and installing pipes represents more than half of the total amount. The operating costs over the 100 years take the second largest percentage. In total life cycle GHG emissions, the operating GHG emissions are much higher than the capital GHG emissions.

Table 2. Breakdown of total life cycle cost for the 3 MW solar option (1.4% discount rate)

Breakdown of LCC (M$)				Total (M$)	% of LCC
Total life cycle cost: 183	Total capital cost	Pipe cost	75	93	51%
		Solar PV cost	3		
		Pump station cost	14		
	Total pump replacement cost over 100 years			24	13%
	Total solar replacement cost over 100 years			5	3%
	Total operating cost over 100 years			62	34%

CC BY NC SA

Table 3. Breakdown of total life cycle GHG emissions for the 3 MW solar option (1.4% discount rate)

Breakdown of LCGHG		Total (kt)	% of LCGHG
Total life cycle GHG emissions (kt): 449	Total capital GHG emissions	91	20%
	Total operating GHG emissions over 100 years	358	80%

Details of the Pareto front obtained from the 0 MW (no solar) option are demonstrated in Figure 4. Clear trade-offs between two objective function values can be observed. An increase in the total cost can reduce GHG emissions and vice versa. A total of 100 optimal solutions have been found along the Pareto front. The least-emission solution has the smallest GHG emissions of 694 kt but the largest cost of $236M. The least-cost solution has the smallest cost of $208M but the largest emissions of 754 kt. As shown in Figure 4, five sample solutions A to E have been chosen for further analysis, including the two points at two ends of the Pareto front. Also, the cost needed to achieve one tonne of GHG reduction between sample points has been calculated. With the decrease of GHG emissions along the Pareto front from top left to bottom right, the cost to reduce 1 tonne of GHG emission increases from $105/tonne to $2,081/tonne of carbon dioxide equivalent (CO_2-e). Further, the detailed Pareto front obtained from the 5 MW solar option is shown in Figure 5. The same calculation steps as above have been applied to this option. It has been observed that the 5 MW solar option needs a slightly higher cost for GHG reduction between every two sample solutions from A through to E, in comparison with the no solar option in Figure 4.

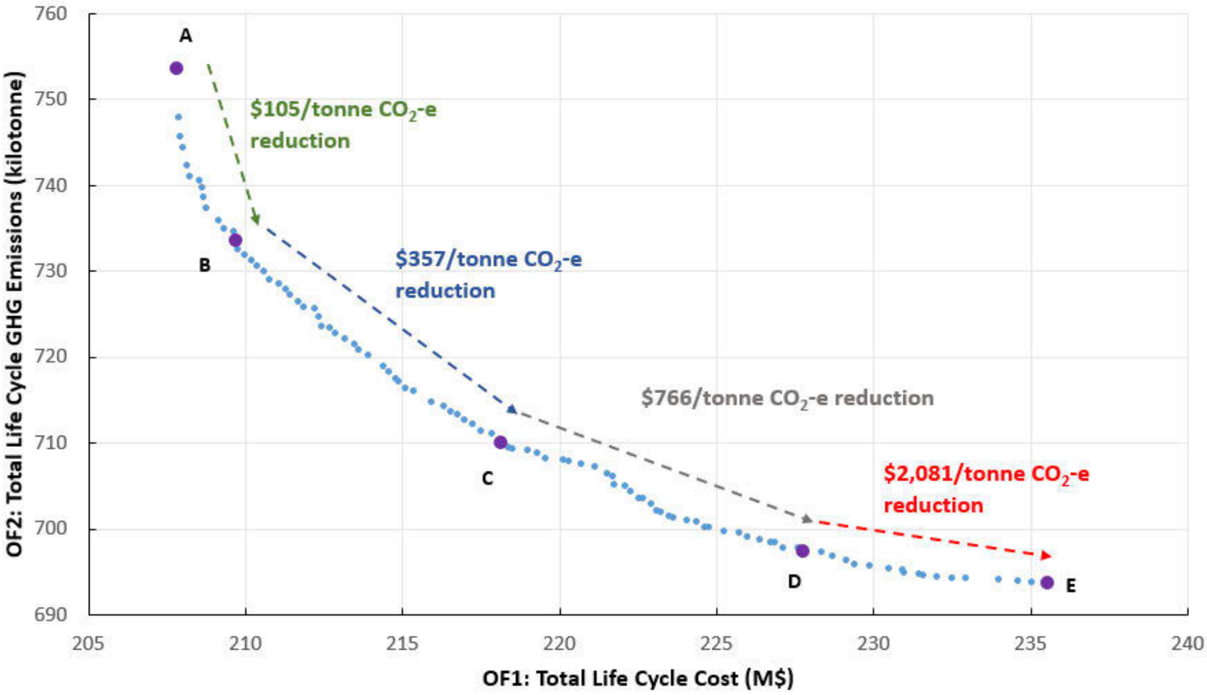


Figure 4. Details of Pareto front under 0 MW (no solar) option (1.4% discount rate)

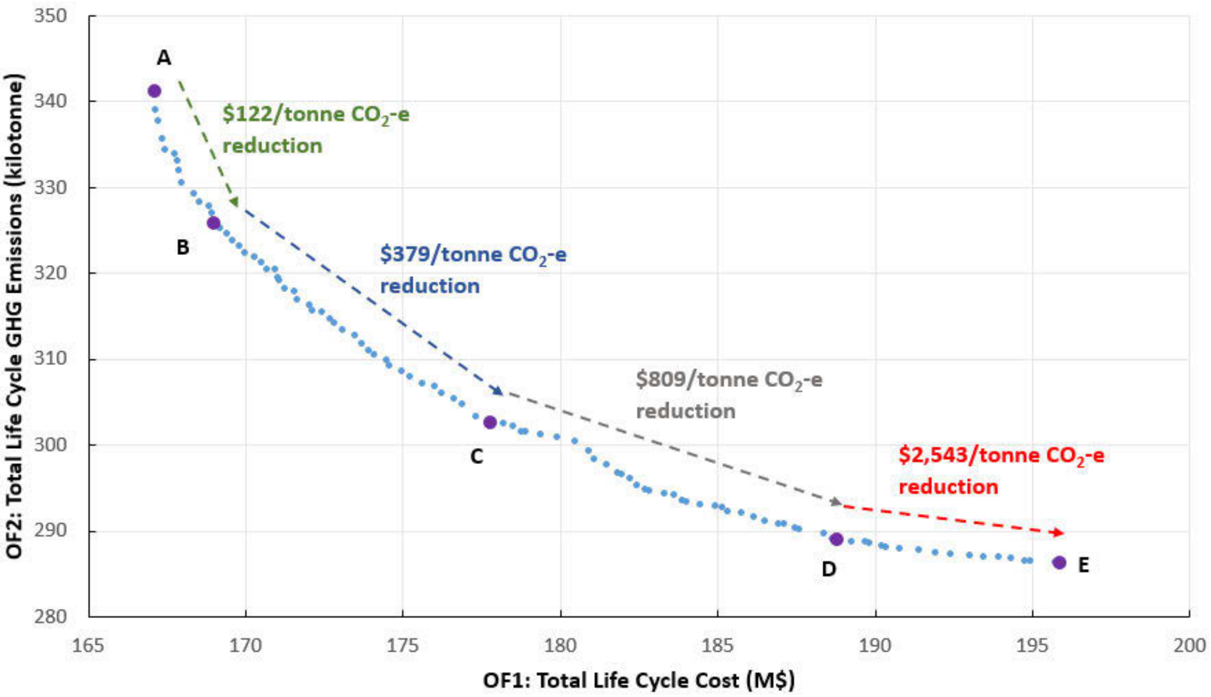


Figure 5. Details of Pareto front under 5 MW solar option (1.4% discount rate)

Further comparison between the 0 MW (no solar) and 5 MW solar options is shown in Table 4. The total life cycle cost has decreased by more than 15% for all five sample solutions after installing 5 MW solar. Also, the total life cycle GHG emissions under the 5 MW solar option have been reduced to less than half of the 0 MW (no solar) option.

Table 4. Comparison of objective function values of the five typical solutions for the 0 MW and 5 MW solar options (1.4% discount rate)

Sample solution	**Total Life Cycle Cost (M$)**			**Total Life Cycle GHG Emissions (kilotonne)**		
	0 MW solar	**5 MW solar**	Change in %	**0 MW solar**	**5 MW solar**	Change in %
A	208	167	20%	754	341	55%
B	210	169	19%	734	326	56%
C	218	178	19%	710	303	57%
D	228	189	17%	697	289	59%
E	236	196	17%	694	286	59%

4.2 Impact of discount rate on optimisation results

Other discount rates have also been considered to investigate the impact of varying the discount rate on the two objective function values. The optimisation process for each scenario has been repeated considering the other two different discount rates: 0% and 6%. The comparison of Pareto fronts among various discount rates of 0%, 1.4% and 6% under the same size of solar PV is shown in Figure 6, Figure 7 and Figure 8. It can be observed from the trend that with the increase

CC BY NC SA

of discount rates from 0% to 6%, both objective function values have decreased for all sizes of solar PV incorporated.

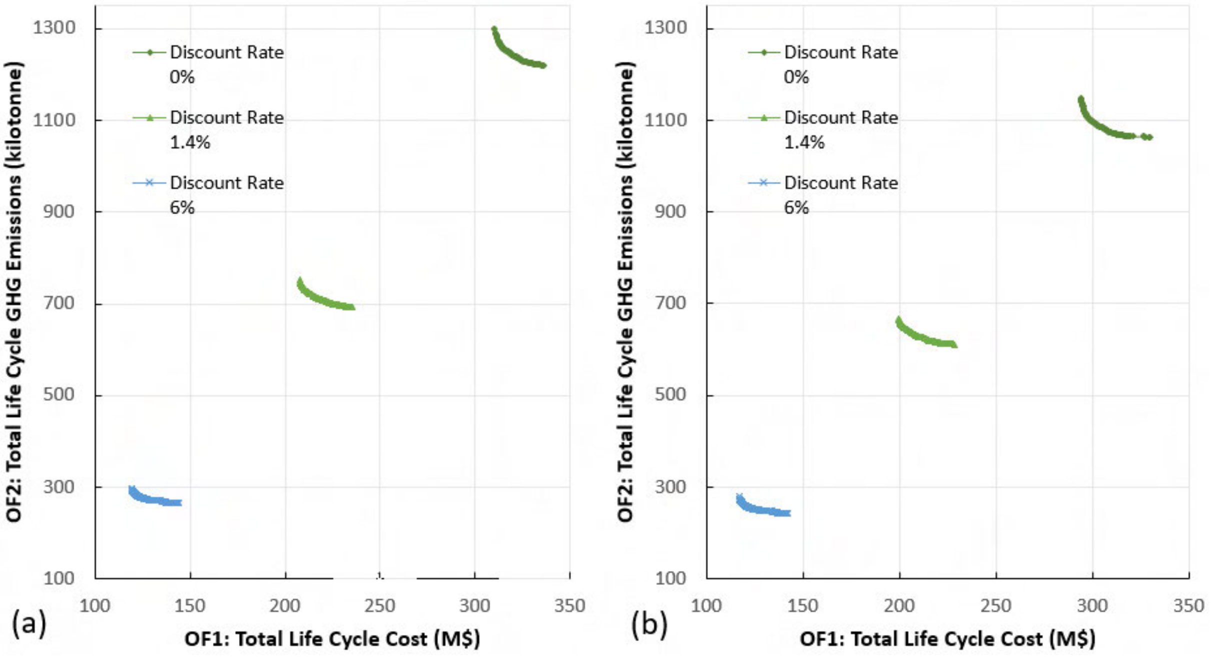


Figure 6. Pareto fronts with variation of discount rates (0%, 1.4% and 6%): (a) no solar; (b) 1 MW solar

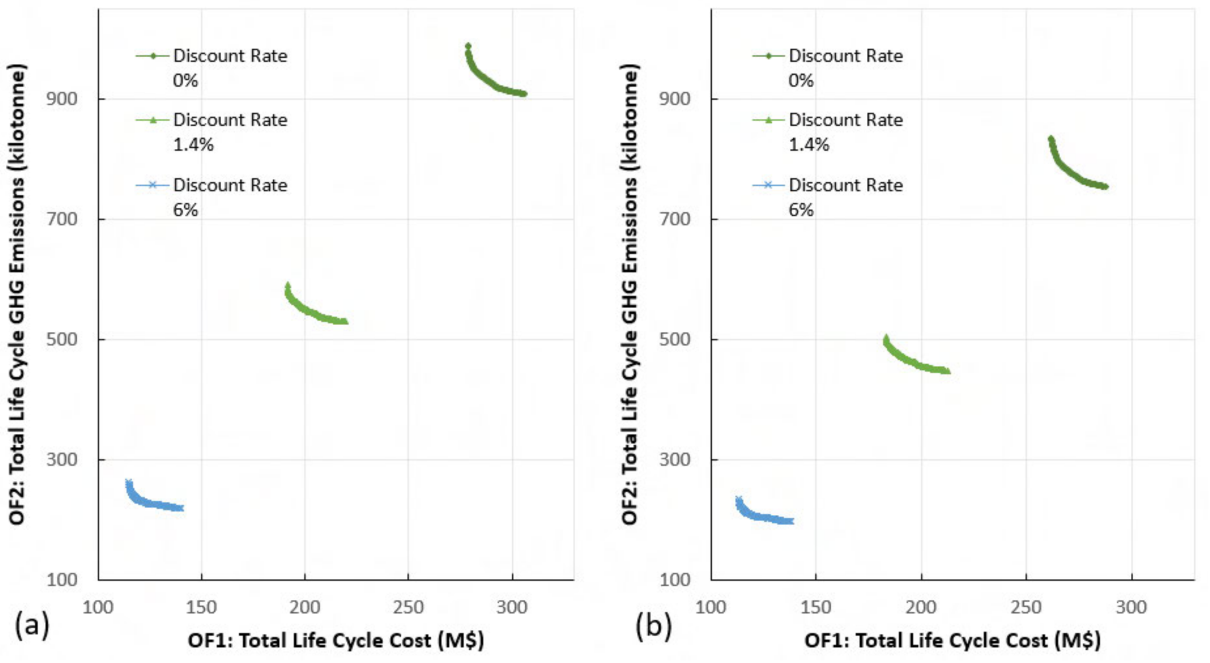


Figure 7. Pareto fronts with variation of discount rates (0%, 1.4% and 6%): (a) 2 MW solar; (b) 3 MW solar

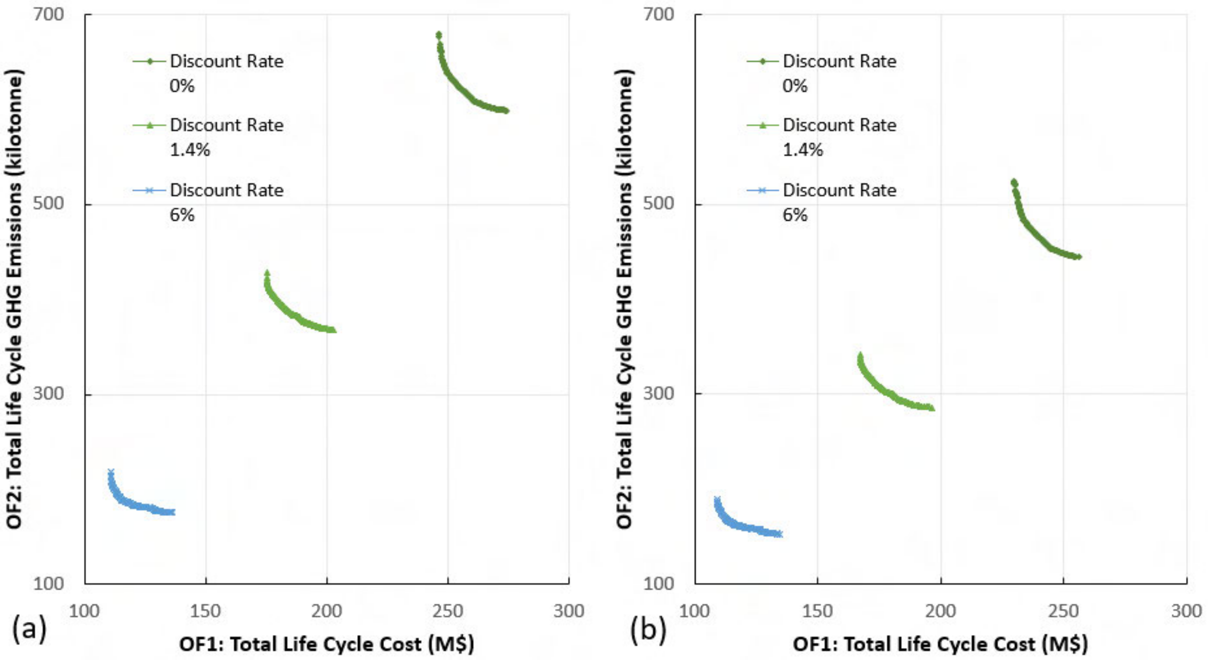


Figure 8. Pareto fronts with variation of discount rates (0%, 1.4% and 6%): (a) 4 MW solar; (b) 5 MW solar

In Table 3 below, d1 to d10 are the optimal pipe diameters obtained for the three discount rates. The total life cycle costs and GHG emissions under different discount rates are shown in Table 4. It can be observed from the results that the optimal pipe diameters (and pipe costs), total life cycle costs and GHG emissions are sensitive to the variation of the discount rate, under the same solar PV size. The increase in the discount rate can lead to smaller pipe diameters (and pipe costs) and less total life cycle costs and GHG emissions, as shown in Table 3 and Table 4.

Table 3. Comparison of optimal pipe diameters and total pipe costs of the least-cost solution for three different discount rates (5 MW solar)

Discount Rate	**d1 (mm)**	**d2 (mm)**	**d3 (mm)**	**d4 (mm)**	**d5 (mm)**	**d6 (mm)**	**d7 (mm)**	**d8 (mm)**	**d9 (mm)**	**d10 (mm)**	**Total Pipe Cost (M$)**
0%	226	412	464	464	520	615	894	1,089	1,388	1,587	**81**
1.4%	226	384	412	412	520	615	894	1,089	1,289	1,488	**75**
6%	202	305	384	412	520	520	778	994	1,089	1,289	**63**

Table 4. Comparison of minimum objective function values for three different discount rates (5 MW solar)

Solar PV size	Min. LCC (M$)			Min. LCGHG (kilotonne)		
	0%	1.4%	6%	0%	1.4%	6%
0 MW	310	208	119	1,220	694	265
1 MW	294	200	117	1,064	612	243
2 MW	278	191	115	909	531	220
3 MW	262	183	113	755	449	198
4 MW	246	175	111	599	368	175
5 MW	230	167	109	445	286	153

5 CONCLUSIONS

In this study, the optimal design of WDSs incorporating behind-the-meter (BTM) solar energy sources has been investigated via a real-world pressurised irrigation system. Both the total life cycle economic cost and total life cycle environmental GHGs throughout the design life of the system have been optimised in the design process. The impact of various discount rates on the optimisation results has also been investigated. Results show that both the total life cycle cost and GHG emissions have been significantly reduced when the BTM solar energy is incorporated. In addition, the optimal pipe diameters, total life cycle costs and GHG emissions are sensitive to the discount rate used while the solar PV size remains the same. The increase in the discount rate can lead to smaller pipe diameters and less total life cycle costs and GHG emissions.

6 ACKNOWLEDGEMENT

We would like to thank Lower Murray Water for supplying data for this paper. Wenyan Wu also acknowledges support from the Australian Research Council via the Discovery Early Career Researcher Award (DE210100117).

7 REFERENCES

[1] T. Abdul Gaffoor, "Real-time Control and Optimization of Water Supply and Distribution Infrastructure," 2017.

[2] S. A. Bagloee, M. Asadi, and M. Patriksson, "Minimization of water pumps' electricity usage: A hybrid approach of regression models with optimization," Expert Systems with Applications, vol. 107, pp. 222-242, Oct 1 2018.

[3] C. S. Stokes, A. R. Simpson, and H. R. Maier, "The cost–greenhouse gas emission nexus for water distribution systems including the consideration of energy generating infrastructure: an integrated conceptual optimization framework and review of literature," Earth Perspectives, vol. 1, no. 1, p. 9, 2014.

[4] W. Wu, A. R. Simpson, and H. R. Maier, "Sensitivity of Optimal Tradeoffs between Cost and Greenhouse Gas Emissions for Water Distribution Systems to Electricity Tariff and Generation," Journal of Water Resources Planning and Management, vol. 138, no. 2, pp. 182-186, Mar-Apr 2012.

[5] W. Wu, H. R. Maier, G. C. Dandy, M. Arora, and A. Castelletti, "The changing nature of the water–energy nexus in urban water supply systems: a critical review of changes and responses," Journal of Water and Climate Change, 2020.

[6] M. N. Sharif, H. Haider, A. Farahat, K. Hewage, and R. Sadiq, "Water–energy nexus for water distribution systems: a literature review," Environmental Reviews, vol. 27, no. 4, pp. 519-544, 2019.

[7] H. Zhang, Z. Lu, W. Hu, Y. Wang, L. Dong, and J. Zhang, "Coordinated optimal operation of hydro–wind–solar integrated systems," Applied Energy, vol. 242, pp. 883-896, 2019.

[8] V. C. Sontake and V. R. Kalamkar, "Solar photovoltaic water pumping system - A comprehensive review," Renewable & Sustainable Energy Reviews, vol. 59, pp. 1038-1067, Jun 2016.

[9] W. Wu, A. R. Simpson, and H. R. Maier, "Multi-objective genetic algorithm optimisation of water distribution systems accounting for sustainability," in Proceedings of Water Down Under 2008, Adelaide, Australia, 2008, p. 1750: Engineers Australia.

[10] W. Wu, A. R. Simpson, and H. R. Maier, "Accounting for Greenhouse Gas Emissions in Multiobjective Genetic Algorithm Optimization of Water Distribution Systems," Journal of Water Resources Planning and Management, vol. 136, no. 2, pp. 146-155, Mar-Apr 2010.

[11] H. R. Maier, S. Razavi, Z. Kapelan, L. S. Matott, J. Kasprzyk, and B. A. Tolson, "Introductory overview: Optimization using evolutionary algorithms and other metaheuristics," Environmental Modelling & Software, vol. 114, pp. 195-213, Apr 2019.

[12] B. Coelho and A. Andrade-Campos, "Using different strategies for improving efficiency in water supply systems," in Proceedings of 1st ECCOMAS Young Investigators Conference, 2012.

[13] Lower Murray Water, "Lower Murray Water Corporate Plan 2019-2020," Lower Murray Water2019, Available: https://www.lmw.vic.gov.au/wp-content/uploads/2019/10/2019-032362-2019-20-LMW-Corporate-Plan-FINAL-Version-as-at-31-May-2019.pdf, Accessed on: August 5, 2021.

[14] L. Rossman, "Epanet 2 users manual," Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 2000.

[15] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182-197, Apr 2002.

[16] J. Blank and K. Deb, "pymoo: Multi-objective Optimization in Python," IEEE Access, vol. 8, pp. 89497-89509, 2020.

[17] Open Water Analytics. owa-epanet 2.2.3 [Online]. Available: https://pypi.org/project/owa-epanet/2.2.3/. Accessed on: October 21, 2021.

[18] L. Rossman, "The EPANET Programmer's Toolkit," Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 1999.

[19] N. H. Stern et al., Stern Review: The Economics of Climate Change. Cambridge University Press, 2006.

[20] Water Supply Code of Australia: WSA 03-2011-3.1, 2011.

[21] Department of Primary Industries, "NSW Reference Rates Manual - Valuation of Water Supply, Sewerage and Stormwater Assets," Department of Primary Industries, Sydney, New South Wales2014.

[22] Clean Energy Council, "Consumer guide to buying household solar panels (photovoltaic panels)," in "Consumer guide to buying household solar panels (photovoltaic panels)," Clean Energy Council, Melbourne19 Dec 2012 2012, vol. 21.

[23] D. Prasad, M. Snow, and M. Watt, "Best Practice Guidelines for Solar Power Building Projects in Australia," The University of New South Wales (UNSW)2005.

[24] A. Reddaway. (2016, October 15, 2021). Energy flows: How green is my solar? Available: https://renew.org.au/renew-magazine/solar-batteries/energy-flows-how-green-is-my-solar/, Accessed on: October 15, 2021.

[25] Department of Industry Science Energy and Resources, "National Greenhouse Accounts Factors," in "Australian National Greenhouse Accounts," Department of Industry Science Energy and Resources,2020.

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14105

GRAPH BASED METHOD FOR CRITICAL PIPE ANALYSIS IN URBAN DRAINAGE NETWORKS AND THE EFFECT OF LOOP DEGREE

Aun Dastgir[1], Martin Oberascher[2], Sina Hesarkazazzi[3] and Robert Sitzenfrei[4]

[1,2,3,4]Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Technikerstrasse 13, Innsbruck, Tirol, Austria

[1] *aun.dastgir@uibk.ac.at,* [2] *martin.oberascher@uibk.ac.at,* [3] *sina.hesarkazazzi@uibk.ac.at,*
[4] *robert.sitzenfrei@uibk.ac.at*

Abstract

Urban drainage networks (UDNs) are important for urban areas, to ensure protection of humans from nature and also to protect nature from anthropogenic impacts. Internal and external pressures on these systems like structural failure (e.g., pipe collapse), and functional failure (e.g., climate change and urbanization, require efficient modelling strategies for its management and maintenance. For proper functioning of the system as a whole, some elements of UDN infrastructure hold more importance than other elements. Identifying these critical elements(pipes) in UDNs is of utmost importance for forming efficient management strategies.

In this study, a graph-based method based on “runoff edge betweenness centrality” is presented for determining the critical elements due to pipe failure of a UDN. In contrast to conventional hydrodynamic modelling method, the proposed graph-based method does not rely on iterative hydraulic simulations. Instead, it takes a mathematical graph representing the structure of the network as starting point and incorporates hydraulic factors to mimic the hydraulic behavior of structural failures in UDNs. Effect of loop degree in the UDNs on this graph-based method is studied by employing the method for a fully branched network and two partially looped networks of different loop degrees. A fully branched network is employed as a case study comprising of 59 circular pipes, draining an area of 175 hectares. Loops are manually created in the same network to study the impact of different number of loops on the accuracy of graph-based method. The results from the graph-based method are compared to the ones from hydrodynamic modelling from SWMM model in terms of accuracy as well as computational time and efforts required.

Results show that the graph-based method is very accurate (98%) in case of a fully branched network but the accuracy decreases with the increase of loop degree in the network. This is due to the complex hydrodynamic effects like back-water phenomenon which cannot yet be modelled using graph-based approach. The method is hence least accurate (92%) in case of fully looped network though still very useful, because the computational time and effort required is much less than hydrodynamic modelling method. The proposed method is also less data intensive and hence can be used by utilities where quick analysis is required or a large number of evaluations have to be made.

Keywords
Complex Network Analysis, Graph Theory, Pipe Criticality, Hydrodynamic Modelling, Edge Betweenness Centrality.

1 INTRODUCTION

Urban Drainage Networks (UDNs) are an important lifeline in urban areas that are built for protection and convenience of humans. These systems are important for public health and for the protection of environment from anthropogenic impacts. Sewer networks have to be maintained and rehabilitated regularly to achieve the serviceability required from these systems [1]. UDNs are large spatial networks consisting of many elements. The overall performance of the network strongly depends on the performance of each individual element. Characteristics and position of a particular element define its importance in the network [2]. Maintaining all the elements to the same high service level is expensive in terms of resources required. Hence, it is important to recognize critical elements of UDNs to prioritize the maintenance and rehabilitation accordingly. The degree of criticality can therefore be used to make such a system more resilient.

There are very few methodologies in literature to identify critical elements in sewer networks. Arthur and Crow [3] described a strategy to find critical elements based on the serviceability loss. The ranking of pipes is done based on the combination of consequence scoring (depending on the consequence of failure) and likelihood scoring (based on likelihood of failure). Another methodology is used for identification of weak points for urban water infrastructure [4, 5]. In that approach systematic Hydrodynamic Modelling Method (HMM) is used, through which the capacity of each conduit is reduced to almost zero to mimic pipe blockage (one at a time) and the hydraulic consequences such as overall system flooding is determined and spatially mapped to the origin of the blockage. For large networks and looped systems this method requires a high computational effort.

Researchers have started working with a new methodology based on graph theory to model UDNs to tackle the problem of high computational efforts in case of large UDNs. Graph theory is a branch of mathematics which employs graphs to model the complex problems of real life, more efficiently [6]. For example, graph theory is used for the layout and design of UDNs [6], functional analysis of UDNs [7], to assess redundancy [8], and assessing the efficiency of different topological positions of combined sewer overflow structures in an UDN [9], in literature. A method has been proposed for finding critical elements using graph theory which focuses on finding the costs of the graph by removing a part of the network [2]. The research uses shortest path length graph measure and head loss as a weight to find the costs associated with removing a particular component (pipe). The cost associated with the graph is calculated after removing an edge (pipe) and the method is repeated for every edge. However, there is a need to extend the research to a graph method that is simpler to implement and is not iterative to save computational time on large networks and which can also be extended to multiple pipe failure or cascading failures.

Recognizing the research gap, this paper proposes a new method based on graph theory to identify the critical individual components (pipes) of a UDN. The method is based on a graph metric called "Runoff Edge Betweenness Centrality" (EBC_e^R). In contrast to conventional hydrodynamic modelling method, the proposed method does not depend on iterative hydrodynamic simulations which is a time-consuming procedure. Instead, it employs the concept of network analysis tailored to UDNs to mimic their hydraulic behaviour. Critical pipes are identified with a graph-based measure (EBC_e^R) which can be interpreted as the volume of flooding when a particular pipe is blocked. The pipes are subsequently ranked according to the caused flood volume when it is blocked. This research focuses on both branched and looped networks and discusses the impact of loop degree on the accuracy of the method. The results from the graph-based method are compared with the results from hydrodynamic modelling method in terms of accuracy as well as computational time. This paper only focuses on single pipe failure/blockage but the methodology can be extended to multiple pipe failure or cascading failure as well due to its high computational efficiency.

2 MATERIALS AND METHODS

First, theory of the graph-based method is explained along with the explanation of existing hydrodynamic method which will be used for comparison. Both methods are then applied for fully branched, semi looped and fully looped network and the results are compared.

2.1 Graph Representation of UDNs

A graph G is a set of vertices (V) and edges (E) and can be represented by G = (V,E) where the set of vertices (nodes) are denoted by V(G) and the set of edges are denoted by E(G). An edge can be represented using its source node i and target node j by an ordered pair (i,j) [6]. A graph is called directed if the edge (i,j) is directed from node i to node j or undirected if the edge has no particular direction in which case edge (i,j) is exactly the same as edge (j,i). In graph analysis, each edge can have a certain weight. A graph can be looped if there exists more than one edge originating from the same vertex or branched if there exists only one path between each pair of vertices (no loops). In UDNs, manholes, storage tanks etc. are represented by vertices and pipes, pumps etc. are represented by edges. A graph can have one or multiple outlets or in case of foul or combined sewer systems, outflows in the receiving water or wastewater treatment plants.

The graph measures that has been used in this study are modified to the functioning and hydraulic properties of UDNs. The first graph measure that has been employed is called shortest path lengths $\sigma_{i,j}$, which can be defined as a path between nodes i and j which has minimum sum of edge weights [10]. Weights can be given to edges according to the requirement of the study and as an attribute of hydraulic behaviour of UDNs. Length is used as an edge weight to find shortest paths in this study. Another graph measure that is employed in this study is, edge betweenness centrality (EBC). Number of shortest paths between every pair of nodes in a graph that pass through a particular edge *e*, represents the EBC of the edge *e* [11]. Considering the hydraulic nature and functioning of UDNs, a modification has been added to EBC graph measure to make it more suitable to the use case. We count the number of shortest paths between every inlet node i to an outlet node j that passes through a particular edge e. Further, instead of counting the number of paths, the contributing runoff area R_i of each inlet node is added to EBC values to compute the cumulative impervious area for every edge of a UDN. This measure is called runoff edge betweenness centrality (EBC_e^R) [12] and is given by:

$$\boldsymbol{EBC_e^R = \sum i, j \ \sigma_{i,j}(e) \ * \ R_i} \tag{1}$$

EBC_e^R values can be multiplied with different rain volumes that gives a surrogate for potential flooding volume.

2.2 Graph Based Method

The graph-based method is used to replicate the flood volume in the system when a pipe blockage is simulated. This is achieved using the concepts of shortest paths and EBC_e^R as explained in the above section. The functionality of our approach is illustrated using a simple branched network consisting of six nodes and five edges as shown in Figure 1.

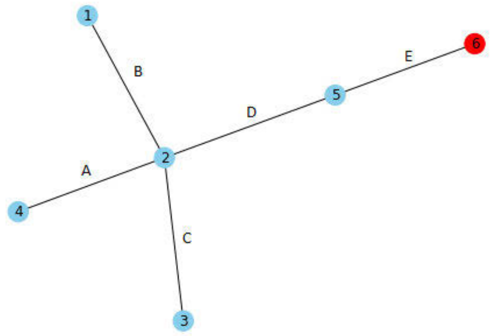


Figure 1: A simple six nodes branched network

In the first step, the shortest paths from all inlet nodes to the outlet node, node 6 shown in red in above network, are determined. Table 1 shows the results from shortest paths calculation for every node. For every node, the pipes that are part of the shortest path get a value of 1 and all

cc BY NC SA

other pipes get a value of 0 forming a vector of zeros and ones. Summing up the rows for each pipe in that matrix, results in a vector which gives us the classical edge betweenness centrality of every pipe shown in Table 1. To account for the functional properties of UDNs, runoff area of each node R_i calculated using the area of the sub-catchments connected to a particular node and their imperviousness, is multiplied with all matrix elements in the corresponding shortest path. When now summing up the rows in that matrix, the EBC_e^R are obtained for each edge.

Additionally, total rainfall volume can also be included in the calculation by multiplying EBC_e^R with the rainfall volume to understand the differences generated by different rainfalls (e.g., different return periods). This can be the indication of maximum possible flooding volume based on EBC_e^R values for every pipe. For a fully branched network, the EBC_e^R value multiplied with the rainfall volume corresponds to the flooding volume when that particular pipe is blocked. This flooding value is then compared with the flood volume obtained from HMM method.

Table 1: Edge betweenness centrality vector for network in Fig.1

Nodes Pipes	**1**	**2**	**3**	**4**	**5**	**EBC**
A	0	0	0	1	0	1
B	1	0	0	0	0	1
C	0	0	1	0	0	1
D	1	1	1	1	0	4
E	1	1	1	1	1	5

In case of a looped network, the aforementioned methodology has to be adapted to include the impact of loops on flooding as well. In that case, a new measure called Capacity Edge Betweenness Centrality (EBC^c) is introduced based on the capacity of pipes.

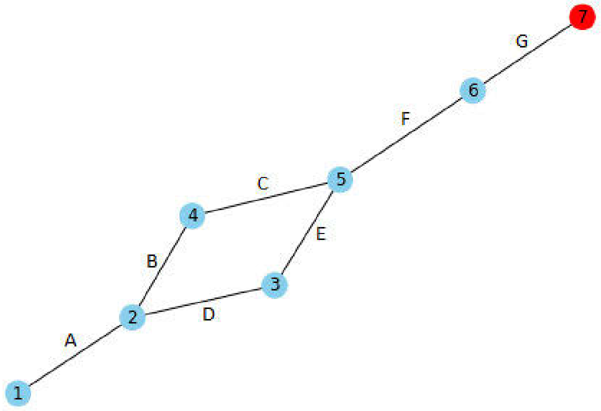


Figure 2: A simple looped network

In the looped network shown in Figure 2, nodes 2,3,4, and 5 are part of loops as they have alternative paths which means even if one of the pipes B, C, D or E are blocked, there is an alternate path for the flow to reach the outlet 7 (shown in red). The amount of flow which can be redirected to the alternate path depends on the capacity and slope of the pipe in the alternate path. For example, when pipe B is blocked, the whole EBC_e^R value of the pipe does not correspond to the flooded volume, rather a portion or all of it can be redirected to the pipe D. To find the amount that can be redirected to D, EBC^c is calculated based on Manning-Strickler formula using diameter and slope of the pipe D. Flooding when the pipe B is blocked can be calculated by subtracting EBC^c

CC BY NC SA

of pipe D from the EBC_e^R of pipe B. Similarly, the procedure can be repeated for all the pipes having alternate paths/loops. The pipes are ranked on the basis of flooding volume in case of blockage or failure. The pipe with the highest flood volume in the system is ranked most critical.

2.3 Hydrodynamic Modelling Method (HMM)

To compare the proposed graph-based method with the state-of-the-art, Achilles approach introduced by [4] is used in this paper to identify critical pipes using hydrodynamic modelling implemented in SWMM. The pipes are ranked on the basis of flooding produced in case the pipe is blocked. The pipe that leads to the most flood volume in the system when it is blocked, is the most critical pipe.

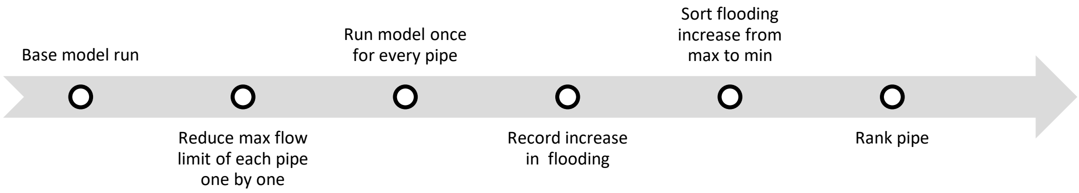

Base model run
Reduce max flow limit of each pipe one by one
Run model once for every pipe
Record increase in flooding
Sort flooding increase from max to min
Rank pipe

Figure 3: Process to determine critical pipes using HMM

Figure 3. describes the process of determining critical pipes using HMM. First, a base run is performed when all pipes function under normal condition, and therefore no flooding occurs in the system. Then, maximum flow limit of pipe is reduced to 0.01 to simulate pipe blockage. A new simulation is carried out for every pipe blockage one by one. The flooding volume is then recorded for every pipe. Pipes are ranked on the basis of flooding produced when it is simulated as blocked.

2.4 Case Studies

A benchmark urban drainage network is selected for this study. The network is designed by [12] for two year return period block rain of 15 minutes duration having total rain volume of 17.1 mm. It is a fully branched network that drains 175 hectares. The network is comprised of 59 circular pipes connected by 60 inlet nodes including one outlet as shown in Figure 4.

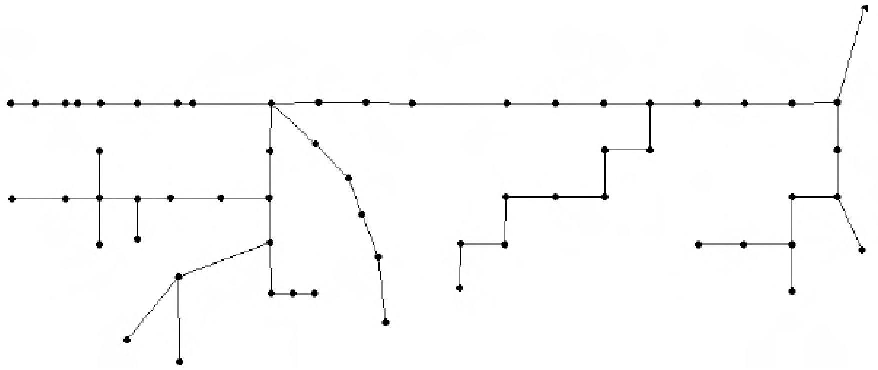


Figure 4: Benchmark Case study network [12]

To extend the study to looped networks, loops are created manually in the benchmark network in Figure 4. Two networks having three loops and six loops respectively are created to study the impact of loop degree on the method. The loops created are not designed.

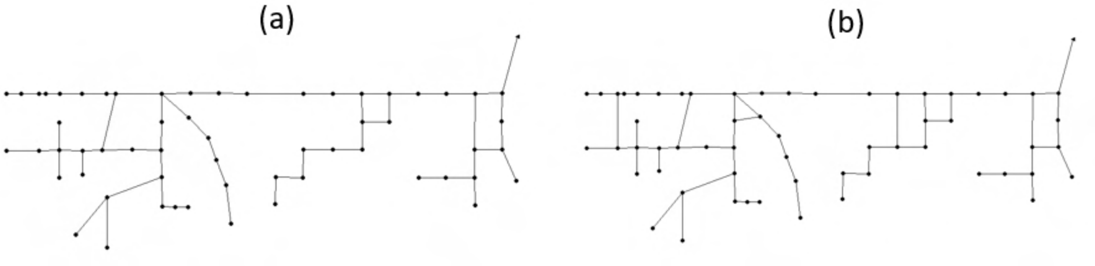


Figure 5: Manually created looped networks (a) Three loops (b) Six loops

2.5 Software and Hardware:

For hydrodynamic modelling, Environmental Protection Agency (EPA's) Stormwater Management Module (SWMM) version 5.1 is used. For graph-based method, Python's module NetworkX version 2.6.2 developed specifically for network analysis is applied. In terms of hardware, a laptop having an Intel® Core™ i7-10610U CPU @ 2.3 GHz processor and 8 GB RAM is used. The system has 64-bit Windows 10 operating system.

3 RESULTS AND DISCUSSION

All three case study networks are analysed using both graph method and HMM and the accuracy of the graph-based method is compared with that of HMM along with the computational time required for both methodologies. 20 most critical pipes given by both methods are compared to give an idea of how well the graph-based method can identify the most critical parts of the sewer networks.

3.1 Fully Branched Network

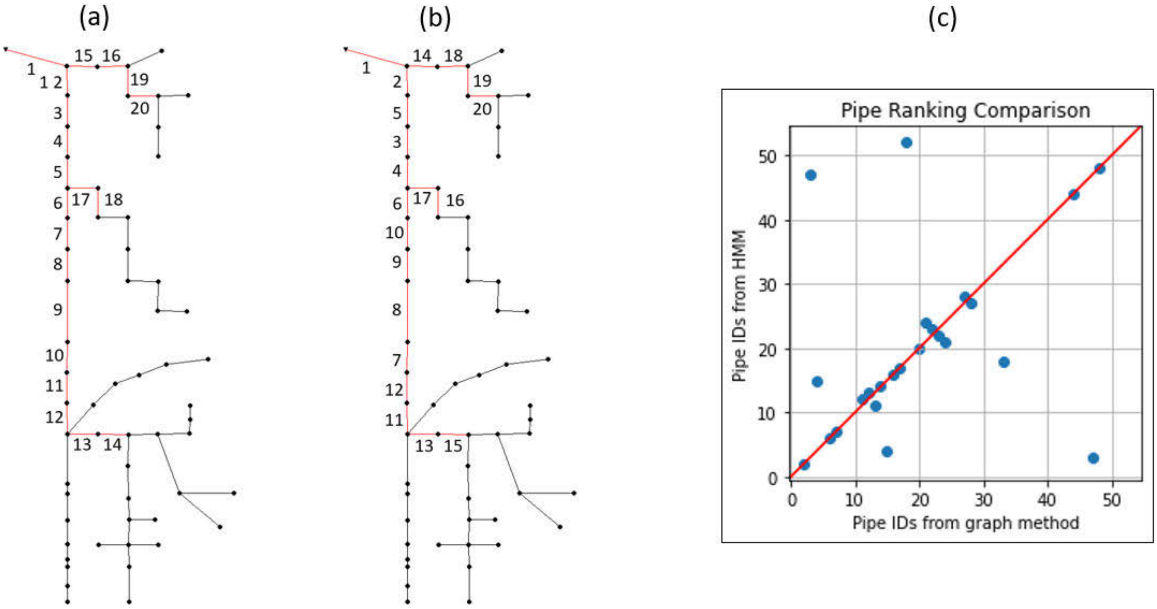


Figure 6: Critical pipes shown in red and in the order of criticality (a) from Graph method, (b) from HMM (c) comparison of pipe rankings

Figure 6 shows the difference between critical pipes identified using graph method and HMM for the branched network. It shows top 20 most critical pipes in red colour and ranked in the order of criticality, (a) obtained from graph method and (b) obtained from HIGA. It can be seen that in

BY NC SA

case of a fully branched network, all 20 pipes identified by both methods are same. The order of criticality is the same for 6 out of these 20 pipes. To better understand the accuracy in terms of order of criticality, a scatter plot comparing the rankings of pipes from these two methods is shown in Figure 6(c). The plot shows the number of pipes having the same rankings out of first 25 pipes. It can be seen that 9 out of 25 most critical pipes are ranked the same by both methods. This shows that the order of criticality of around 35% of the pipes is identified accurately by the graph method.

3.2 Looped Networks:

Critical pipes identified from both methodologies in case of looped networks are shown in figure 7, wherein (a) and (b) of the figure show the results for network with three loops and (c) and (d) show the results for network with six loops. For network with three loops, graph method identifies 18 out of the 20 most critical pipes the same as identified by HMM. The order of criticality is different starting from 3rd most critical pipe. In case of network with six loops, graph method identifies only 15 of the 20 most critical pipes identified by HMM method. This shows that the methodology is not as accurate as in the case of fully branched network. Additionally, the accuracy of the method decreases as the loop degree increases. The reason is that in case of looped networks, graph-based method cannot mimic all the entailed complex hydraulic properties e.g., backwater effect. Also, the graph method considers the impact of loops on flooding for pipes immediately next to the loops but in reality, the impact also propagates further downstream which means that the graph method cannot predict rerouting of water as accurately as HMM method. Increasing the loop degree increases the magnitude of these complexities resulting in less accurate results.

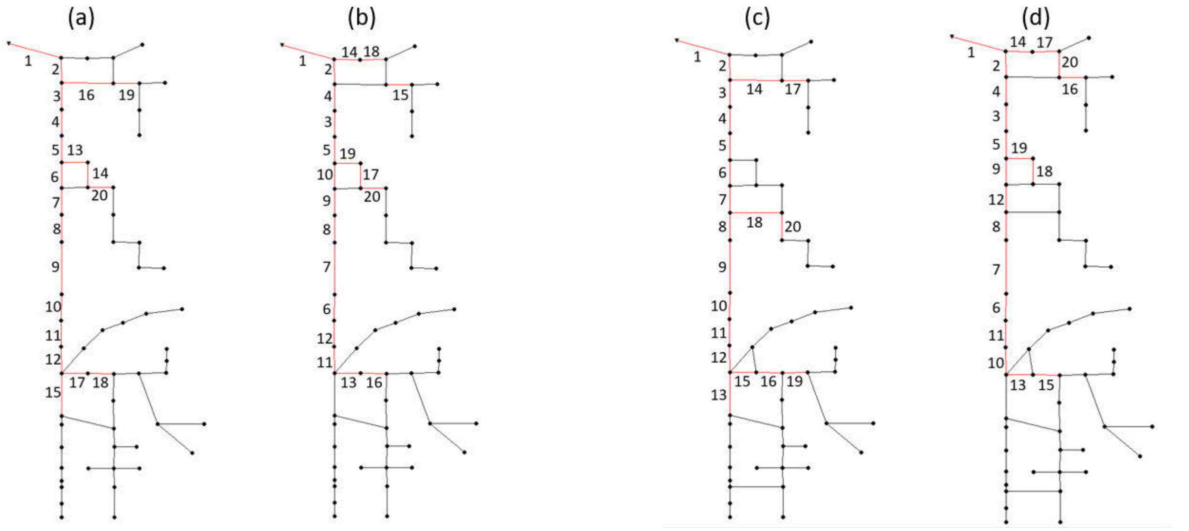


Figure 7: Critical pipes shown in red and in the order of criticality from network with three loops using, (a) graph (b) HMM and from network with six loops using, (c) graph (d) HMM

To get a better idea of the differences in both methods, Figure 8 compares flooding obtained from graph method using EBC_e^R and flooding obtained using HMM, based on a linear regression trendline for every network. In Figure 8(a), the flooding volumes from HMM and based on the graph-based method for the branched networks are shown. It can be seen, that the graph-based method, besides identifying the criticality of the pipes, also quite sufficiently reproduces the flooding volume.

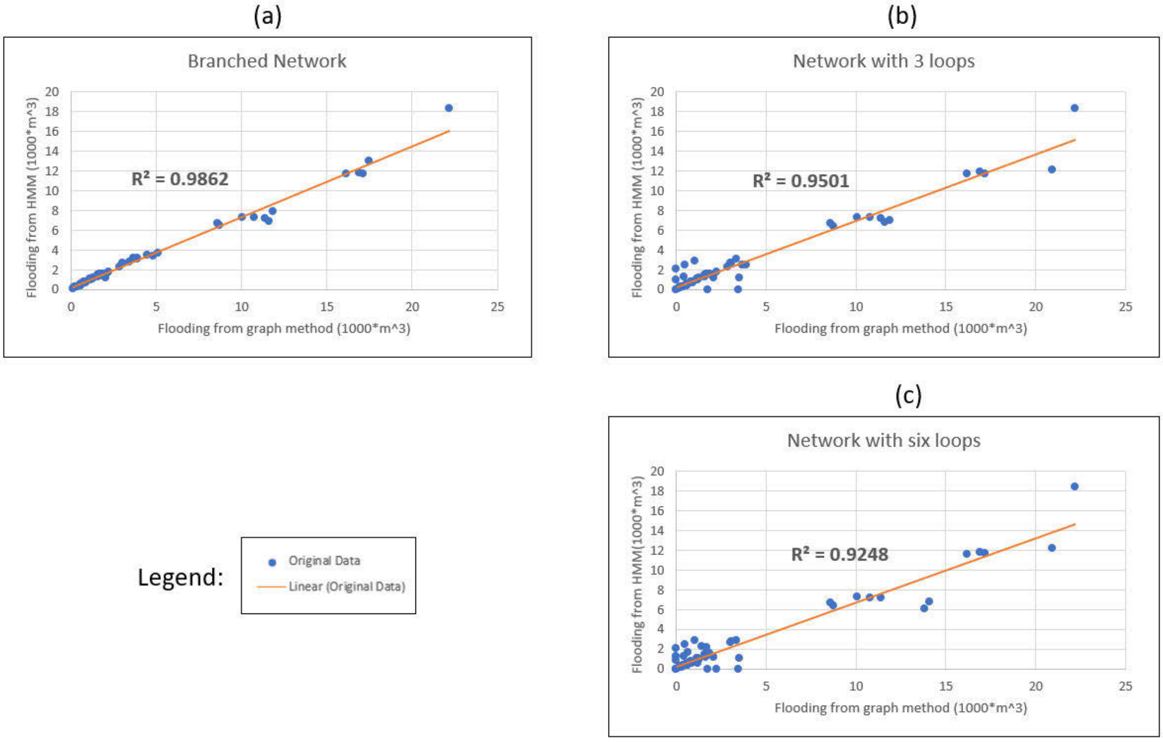


Figure 8: Results of flooding from graph method and HMM for fully branched network (a), network with three loops (b) and six loops (c)

It is evident from this figure that most values are on this line for fully branched network while more values go further from this line as the loop degree increases. Coefficient of determination (R-squared) is calculated for all three networks to check the accuracy of the method. R-squared values clearly describe the impact of loop degree on the accuracy. Value of 0.98 in case of fully branched network shows that the method is very accurate. The values decrease to 0.95 and 0.92 for three loops (Fig. 8(b)) and six loops (8(c)) respectively, showing that the accuracy of the method is decreasing with increasing loop degree.

3.3 Computational time and Data Requirements

The benefit of using the graph method explained in this study, lies in the difference in computational time required to run both methodologies. HMM run time for simulating all pipe blockages is 5 minutes and 20 seconds as the model has to run 59 times, once for each pipe blockage scenario. In comparison, graph method only takes 5 seconds to run thus having a computational gain factor of 64 (320/5) as compared to HMM. This factor will increase with the size of the network. Because for a complex and very large drainage network, HMM method requires a long time to run whereas graph-based method based on time efficient graph techniques does not exhibit the same behavior. To confirm this hypothesis, the method was applied to a real case study having 430 pipes. HMM method in that case took 45 minutes while graph method only took 10 seconds showing graph method having a computational gain factor of 270 (45*60/10). It is thus evident that as the size of UDN increases, the computational gain factor increases non-linearly. Thus, the graph method is much more efficient in case of large UDNs.

Graph based method is also significantly less data intensive as compared to HMM. Methodology presented in this study, although not as accurate as HMM especially for looped networks, can still be used by utilities for a preliminary analysis where all the data is not available. The graph method used in this study has its limitations as well. For example, it cannot model complex hydraulic behaviour (yet), including backwater effect and re-routing of water to parallel pipes as HMM does.

Both these factors contribute to decreasing the accuracy of the method when more loops are introduced in the network, all of which will be tackled in future research.

4 SUMMARY AND CONCLUSIONS

In this study, a new graph-based method was proposed based on network analysis techniques called runoff edge betweenness centrality (EBC_e^R). The method was applied to three case studies with varying loop degree. Results from graph-based method were compared with a hydrodynamic model (HMM) to determine the accuracy of the graph method. Additionally, impact of loop degree on the accuracy of the methodology was investigated. Following main conclusions can be drawn from the study:

- A graph-based method can be used to successfully identify the critical elements of a UDN using network analysis techniques updated to consider the hydraulic behaviour of UDNs.
- The method is applied to three case studies with varying degree of loops. Results indicate that the graph method is almost fully accurate for a branched network. It identifies all 20 critical pipes identified by the HMM method. R-squared value of 0.98 in case of branched network, also shows the high accuracy of the proposed method.
- Accuracy of the graph method decreases with increasing loop degree in the network. The graph method identifies 18 and 15 out of 20 most critical pipes as compared to HMM for three loops and six loops, respectively. R-squares values are also reduced to 0.95 and 0.92 for three loops and six loops.
- The main advantage of using graph method lies in the difference of computational time as compared to HMM. Graph method has a computational gain factor of 64 (5 seconds vs 320 seconds) as compared to HMM. Additionally, graph method is also less data intensive.
- The graph methodology proposed in this study can be used by utilities for a quick analysis where conventional hydrodynamic modelling method becomes too time consuming and where enough data is not available to setup a hydrodynamic model.

5 ACKNOWLEDGEMENTS

This study was funded by the Austrian Science Fund (FWF): P 31104-N29.

6 REFERENCES

[1] Le Gauffre, P., et al., Performance Indicators and Multicriteria Decision Support for Sewer Asset Management. Journal of Infrastructure Systems, 2007. 13(2): p. 105-114.

[2] Meijer, D., et al., Identifying Critical Elements in Sewer Networks Using Graph-Theory. Water, 2018. 10(2).

[3] Arthur, S. and H. Crow, Prioritising sewerage maintenance using serviceability criteria. Proceedings of the Institution of Civil Engineers - Water Management, 2007. 160(3): p. 189-194.

[4] Möderl, M., et al., Identifying weak points of urban drainage systems by means of VulNetUD. Water Science and Technology, 2009. 60(10): p. 2507-2513.

[5] Sitzenfrei, R., et al., Cascade vulnerability for risk analysis of water infrastructure. Water Science and Technology, 2011. 64(9): p. 1885-1891.

[6] Turan, M.E., et al., Feasible Sanitary Sewer Network Generation Using Graph Theory. Advances in Civil Engineering, 2019. 2019: p. 8527180.

[7] Reyes-Silva, J.D., et al., Centrality and shortest path length measures for the functional analysis of urban drainage networks. Applied Network Science, 2020. 5(1): p. 1.

[8] Hesarkazzazi, S., et al., Assessing Redundancy in Stormwater Structures Under Hydraulic Design. Water, 2020. 12(4).

[9] Sitzenfrei, R., et al., Assessing the efficiency of different CSO positions based on network graph characteristics. Water Science and Technology, 2013. 67(7): p. 1574-1580.

[10] Dijkstra, E.W., A note on two problems in connexion with graphs. Numerische Mathematik, 1959. 1(1): p. 269-271.

[11] Girvan, M. and M.E.J. Newman, Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 2002. 99(12): p. 7821.

[12] Hesarkazzazi, S., et al., Implication of Different Pipe-Sizing Strategies for the Resilience of Stormwater Networks, in World Environmental and Water Resources Congress 2021. 2021. p. 244-252.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14107

NRW ESTIMATION AND LOCALIZATION IN WATER DISTRIBUTION NETWORKS VIA HYDRAULIC MODEL CALIBRATION USING 24/7 MONITORING DATA

Alvin Wei Ze Chew[1]; Zheng Yi Wu[2]; Rony Kalfarisi[3]; Meng Xue[4]; Jocelyn Pok[5]; Jianping Cai[6]; Kah Cheong Lai[7], Sock Fang Hew[8], Jia Jie Wong[9]

[1, 3,4,5,6] Bentley Systems Singapore Pte Ltd, 1 Harbourfront Pl, Singapore 098633
[2] Bentley Systems, 76 Watertown Rd, Thomaston, CT 06787, USA
[7,8,9]Water Supply (Network) Department, Public Utilities Board (PUB), Singapore

[1] Alvin.Chew@bentley.com, [2] Zheng.Wu@bentley.com, [3] Rony.Kalfarisi@bentley.com,
[4] Meng.Xue@bentley.com, [5] Jocelyn.Pok@bentley.com, [6] Jianping.Cai@bentley.com,
[7]Lai_Kah_Cheong@pub.gov.sg, [8]Hew_Sock_Fang@pub.gov.sg, [9]Wong_Jia_Jie@pub.gov.sg.

Abstract

Operations of water distribution networks (WDNs) are monitored daily via installed data loggers, where the collated hydraulic data can be leveraged to improve the system's operations over time, and to minimize total economic losses due to non-revenue water (NRW). In collaboration with Public Utility Board (PUB), Singapore's National Water Agency, a practically novel model calibration approach using 24/7 monitoring flow and pressure data has been developed to facilitate PUB's Smart Water Grid (SWG). The approach is developed as a generic integrated solution process to conduct a series of systematic analyses for daily WDN model calibration, namely: (1) estimating the system's daily NRW contributions; (2) performing flow calibration that involves net demand consumption calibration, adjusting pumps operational configurations and localizing NRW sources when the system's daily estimated NRW volume exceeds its assumed background volume; (3) performing energy calibration by rectifying possible drifting in monitored pressure head data and calibrating other physical properties which include, but not limited to, pipe roughness and valve settings, especially during peak-demand hours. The effectiveness of our proposed approach is subsequently tested on three WDN zones in Singapore, having a total pipe length of >100km, that comprises of atypical water usage patterns. The results of model calibration for one of three zones is presented in this paper. The key outcomes derived from the study are: (a) localized a reported leakage event by PUB to less than 100m; (b) calibrated the system's flow balance, to less than 1% average mean absolute percentage error (MAPE), by first identifying and addressing the system's billing data uncertainties, followed by localizing anomaly events that account for the total NRW volume estimated; and (c) calibrated the system's pipe roughness values to improve the total energy balance by achieving an average daily MAPE of 4.0%.

Keywords
water distribution networks; water losses estimation; anomaly localization; demand calibration; hydraulic model calibration; non-revenue water.

1 INTRODUCTION

Potable water is a necessity to sustain the humanity's daily livelihood. With an increasing global population, policy regulations and engineering management are expected to become more stringent to improve and ensure the supply of drinkable water to the public with minimum disruptions which, however, continue to be an engineering challenge to utility companies. For example, in the United States, an estimated volume of 6 billion gallons of treated water is reported to be lost each day [1]. On the other hand, while it may appears that managing underground water distribution networks (WDNs) in smaller countries is less complicated, Singapore, having invested in Smart Water Grid (SWG) [2] management, continues to strive to reduce their yearly non-revenue water (NRW) of around 5% of the total supplied waters. Overall, NRW components can never be fully eradicated during the real-world operations of WDNs due to the system's complexity and the presence of hidden/unknown anomaly events.

Over the years, many engineering approaches have been developed to assist operators to early detect and localize likely anomaly sources during the operations of WDNs, which can be grouped into (1) hydraulic model calibration, and (2) data-driven analytics. The former, that sets the focus of this paper, primarily leverages on physics-based simulations to calibrate the system's hydraulic properties which include flow, pressure, tank levels, pump operations, and demand patterns [3]–[7], where if appropriately calibrated, can represent the baseline operations of the WDN system as part of digital twinning [8]. The latter purely adopts data-driven/statistical methods to train anomaly detection and localization models [9]–[11], where the trained models can be combined with calibrated physics models for digital twin-based decision-supports in near real-time.

To build towards practically effective and useful calibrated hydraulic model(s), this paper identifies and addresses existing shortcomings from published calibration works, namely:

i. A common and inaccurate assumption of no water losses conditions during model calibration. As highlighted above, zero NRW component is never possible for the real-world operations of WDNs, hence the inability to model NRW as part of the calibration step may affect the baseline accuracy in representing the system's actual operations.

ii. Most works are restricted to relatively small networks with high density of sensors per area/pipeline, and leaks are usually simulated under controlled conditions to test the calibrated models for either detection or localization analysis, or both. Quite often, however, the operations of real-world WDNs have limited number of sensors deployed in large supply zones and occurring leaks are usually unknown in their physical characteristics during near real-time.

To address the above-outlined shortcomings, this work, in collaboration with PUB, Singapore, develops a practically novel daily model calibration approach that leverages on continuously monitoring flow and pressure time-series data to estimate daily NRW contributions in large WDNs by emulating near real-time, followed by performing total flow and energy calibration via calibrating the system's daily net demand consumption pattern(s) and importantly pinpointing possible anomaly events in the system which constitute to the everyday NRW volume estimated.

2 METHODOLOGY

2.1 Daily Model Calibration Approach

Figure 1 illustrates the overview of our proposed daily model calibration approach [12] that consists of 3 main systematic analyses for any operational WDN system, namely: (1) estimation of NRW components; (2) flow calibration; and (3) energy calibration. Details of each of the systematic analyses are as follows:

i. **NRW Estimation**: Leveraging on available billing data, collated via traditional metering means or advanced metering infrastructure (AMI), and daily total inflow time-series data to estimate the system's daily NRW volume and its corresponding time-series profile.

ii. **Flow Calibration**: Calibrating the system's net consumption demand pattern due to the real customers which may include accounting for varying water usage patterns due to different types of customers, adjusting available pumps operational configurations, finally performing NRW localization if the total NRW volume exceeds an assumed background NRW volume.

iii. **Energy Calibration**: After flow calibration, any remaining energy discrepancies in the system can be addressed via identifying and rectifying likely sensor drifting over time and adjusting other physical properties such as pipe roughness and valve settings, especially during the peak-demand hours, if justifiable.

2.2 NRW Estimation

For any given day in the operational horizon, estimating its corresponding NRW is performed as follows:

i. Using either historical billing data or metered data derived from AMIs, the average daily consumption rate, termed as $Q_{c,daily}$, is first estimated, followed by approximating the total water consumption ($V_{c,daily}$) volume. For historical billing data, simple averaging techniques can be undertaken to estimate $Q_{c,daily}$ and $V_{c,daily}$ respectively. For example, if the billing data is collated monthly, then $V_{c,daily}$ is derived by averaging the total consumption volume by the total number of days for the specific month. Generally, data collected from AMIs with finer time-resolution are expected to be more accurate to estimate $V_{c,daily}$.

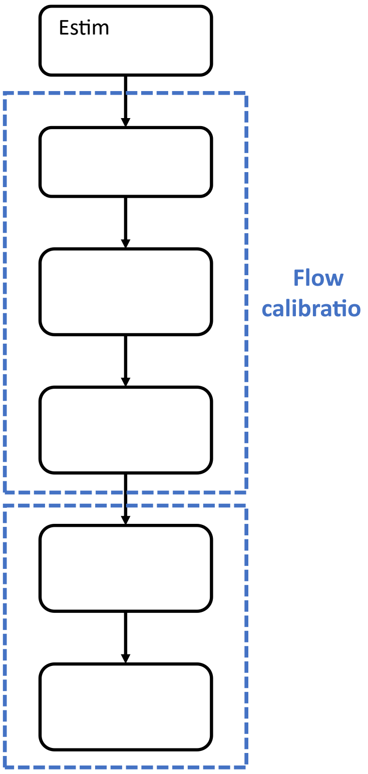


Figure 1. Overview of Daily Model Calibration Approach.

ii. Using daily monitoring data, the corresponding total net inflow time-series profile ($Q_{in,day}(t)$) is derived as follows:

$$Q_{in,day}(t) = Q_{R,day}(t) + Q_{AI,day}(t) - Q_{AO,day}(t) \tag{1}$$

$$V_{in,day} = \int_{t_0}^{t_1} Q_{in,day}(t)\, dt \approx \sum_{i=1}^{M} \frac{Q_{in,day}(t_i) + Q_{in,day}(t_{i+1})}{2} \Delta t_i \tag{2}$$

where $V_{in,day}$ is the total net inflow volume into the system, $Q_{R,day}(t)$ the time-series profile for the total reservoir inflows into the system, $Q_{AI,day}(t)$ the time-series profile for the total additional system inflows from the adjacent zones , and $Q_{AO,day}(t)$ the time-series profile for the total additional system outflows into the adjacent zones, excluding the billed customers in the system, t the time of the day, t_0 the starting time of the day, t_1 the ending time of the day, M total number of intervals along the time axis based on the defined time-step between t_0 and t_1, and i the time index.

iii. The total NRW volume ($V_{nrw,day}$) is then estimated as the difference between the total net inflow and daily water consumption, given as:

$$V_{nrw,day} = V_{in,day} - V_{c,daily} \tag{3}$$

If $V_{nrw,day}$ is estimated to be greater than an assumed background NRW volume, a corresponding NRW time-series profile ($Q'_{nrw}(t)$) is subsequently derived and then further deducted from $Q_{in,day}(t)$ to obtain the net demand consumption profile ($Q'_{in,day}(t)$) via the following solution procedures:

i. For a given number of pressure sensor stations (N) in the WDN system, we first estimate the average time-series pressure head profile ($P_{avg}(t)$) as:

$$P_{avg}(t) = \frac{1}{N}\sum_{i=1}^{N} P_i(t) \tag{4}$$

where $P_i(t)$ is the recorded pressure head of sensor station i at time t.

ii. Estimate an average emitter coefficient ($K_{avg,day}$) value via the well-known pressure-dependent leakage formulation (PDLD) [13] as follows:

$$V_{nrw,day} = K_{avg,day} \int_{t_0}^{t_1} \left(P_{avg}(t)\right)^n dt \tag{5}$$

$$\int_{t_0}^{t_1} \left(P_{avg}(t)\right)^n dt \approx \sum_{i=1}^{M} \frac{\left(P_{avg}(t_i)\right)^n + \left(P_{avg}(t_{i+1})\right)^n}{2} \Delta t_i \tag{6}$$

where n is the exponent, taking the common value of 0.5 for underground waterpipes.

iii. The estimated $K_{avg,day}$ from Eq. (5-6) is then used to construct the $Q'_{nrw}(t)$in Eq. (7), followed by estimating $Q'_{in,day}(t)$ using Eq. (8).

$$Q'_{nrw}(t) = K_{avg,day} \cdot \left(P_{avg}(t)\right)^n \tag{7}$$

$$Q'_{in,day}(t) = Q_{in,day}(t) - Q'_{nrw}(t) \tag{8}$$

2.3 Flow Calibration

Daily flow calibration in the operational WDN system comprising of 3 components, namely: (1) calibrating net demand consumption pattern(s) due to the real customers in the system, (2) calibrating pump operational configurations, and (3) NRW localization to account for the total estimated NRW volume.

2.3.1 Net demand consumption pattern(s) calibration

The system's daily net demand consumption pattern is calibrated via a simple Reference Averaging Approach (RAA) which leverages on the estimated $Q'_{in,day}(t)$ or $Q_{in,day}(t)$ by adhering to the following solution procedures:

i. Compute the average measured inflow value (V_{avg}) from $Q'_{in,day}(t)$ or $Q_{in,day}(t)$ using trapezoidal rule.

ii. Compute the temporal, i.e., at each time-step, ratio values between the respective monitored net inflow (due to real customers) values ($Q_m(t)$) and V_{avg} value as:

$$R(t) = \frac{Q_m(t)}{V_{avg}} \tag{9}$$

iii. The computed $R(t)$ values represent the adjusted demand multiplier values for deriving a new set of model simulated inflow values ($Q_s(t)$).

iv. Compare $Q_s(t)$ with either $Q'_{in,day}(t)$ or $Q_{in,day}(t)$ to compute a new set of temporal ratio values as:

$$R'(t) = \frac{Q_m(t)}{Q_s(t)} \tag{10}$$

v. The adjusted $R'(t)$ values obtained from Eq. (10) are then multiplied with the original $R(t)$ values to derive another set of demand multiplier values, as:

$$R''(t) = R(t) \times R'(t) \tag{11}$$

vi. The computed $R''(t)$ values are again used to derive another set of $Q_s(t)$ profile for comparing with $Q'_{in,day}(t)$ or $Q_{in,day}(t)$. Note that the previously estimated $R''(t)$ values become the newly represented $R(t)$ values.

vii. Repeat steps (iii – vi) till a satisfactory goodness-of-fit is achieved between $Q_m(t)$ and $Q_s(t)$ for the selected day.

Note the above proposed procedures using RAA are used to derive a singular universal demand pattern for all customers in the system. However, there may be cases where there can be multiple water usage patterns in the same system due to different types of customers (e.g., customers who tend to consume more water during the night that differs from the traditional diurnal pattern for water consumptions). To handle such scenarios, there is thus a need to develop daily local demand pattern(s) within a given system via the following solution procedures:

i. For a selected pool of junction nodes which are affiliated to a particular water usage pattern, assign an arbitrary emitter coefficient ($K > 0$) value to them. Repeat this step for N number of possible water usage patterns as determined by the modeller.

ii. Transform the assumed K value(s) into unique demand pattern(s) by following the PDLD formulation from Eq. (6), where the average pressure profile is derived from a singular or selected pool of pressure sensor stations which are situated in the near proximity of the junction nodes affiliated to their corresponding water usage pattern(s).

iii. Leverage on the transformed demand pattern(s) to simulate the system's flow and pressure head profiles, followed by comparing with the monitored individual/average pressure profiles of the selected stations, and $Q'_{in,day}(t)$ or $Q_{in,day}(t)$ for mass balance considerations.

iv. Adjust the multiple demand pattern(s) appropriately by adhering to the simple principle that higher pressures are affiliated to lower water usage pattern, and vice versa.

v. Repeat steps (ii-iv) till good agreement (<< 5% error) is achieved for comparing the simulated and monitored values for $Q'_{in,day}(t)$ or $Q_{in,day}(t)$. At this stage, since energy calibration has not performed, reasonable agreement (~1m) is expected for the individual/average pressure profiles, especially for the high-demand hours.

Practically, it is expected that bulk of the customers in a given supply zone follow a universal calibrated demand pattern, while localized demand pattern(s) are expected to be applied to unique and smaller pool of customers. Hence, it is recommended that the modeller first adopts the proposed RAA method to calibrate a universal demand pattern, before performing the local demand pattern(s) calibration with multiple iterations by ensuring that mass balance for $Q'_{in,day}(t)$ or $Q_{in,day}(t)$ is attained to the highest possible extent.

2.3.2 Pump operational configurations calibration

During the process of calibrating the net demand consumption pattern(s), it is equally important to also check that the internal pump flows within the system are properly calibrated against available monitored pump outflows to ensure the correct distribution of the pump energies to the different junction nodes. To do so, the modeller is required to calibrate the pump operating curves and control statuses. For the latter, it mainly involves adjusting the pumps' operational state of

BY NC SA

either fixed- or variable-speed characteristics. For fixed-speed pumps, their operational states can only be taken as a binary option of either “on” or “off”, while the operational states of variable-speed pumps range between 0.0 (fully switched off) and 1.0 (fully switched on).

2.3.3 NRW localization

If the estimated daily NRW volume is greater the assumed background NRW volume, NRW localization will be performed by using the same PDLD method [13]. The method generally enables the modeler to select and aggregate any combination of junction nodes into a demand group within the network. In each demand group, a given number of the junction nodes will then be identified as potential anomaly hotspots via suitable emitter coefficients which contribute “additional” flow demand to the system, hence emulating the estimated NRW volume for the specific day. Before NRW localization, it is expected that the system’s pipe connectivity, valve settings, pump configurations, if available, and net demand consumption pattern(s) are taken to be calibrated, to the best possible extent.

The PDLD method can be formulated as an implicit non-linear search problem which determines the pool of junction nodes having positive K values to emulate the leakage hotspots in the system. The PDLD method is integrated with the optimization-based model calibration tool [14], [15],which can be executed repetitively for the same anomaly event. The optimization run for the non-linear implicit search problem is then performed with the competent genetic algorithm [16]. To determine the optimal steady-state timings from the minimum night flow (MNF) hours (2am – 4am) for the NRW localization analysis, the computed discrepancies between the model simulated and monitored values for the flow and average pressure head parameters, respectively, are considered to estimate the average hydraulic power ($W_{NRW}(t)$) due to the estimated daily NRW volume in the system, defined as:

$$W_{NRW}(t) = |P_{avg}(t) - P_s(t)| \times (Q_m(t) - Q_s(t)) \tag{12}$$

where $P_{avg}(t)$ and $Q_m(t)$ respectively represent the monitored average pressure head and flow values at a specific steady-state timestamp, while $P_s(t)$ and $Q_s(t)$ respectively represent the simulated average pressure head and flow values at the same timestamp.

Note that $W_{NRW}(t)$ represents the average temporal NRW hydraulic power profile for the selected day. The modeller can then inspect the estimated power values for the same MNF period to identify the top (e.g., top 3) power values and their corresponding timings (e.g., 2.30am) to perform the required NRW localization analysis. The greater the power values, the NRW contributions are expected to be more significant in their hydraulic characteristics.

2.4 Energy Calibration

At this stage, the system’s initial energy discrepancy is addressed, to an extent, by the completed flow calibration. Any remaining energy discrepancy across the multiple stations in the same system are then managed via (1) identifying and rectifying likely drifting(s) in the monitored pressure values of individual sensor stations in the system, and (2) calibrating other physical properties which include, but not limited to, pipe roughness and valve settings within the system.

2.4.1 Rectifying sensor drifting

For any given pressure sensor station, the most obvious indication of sensor drifting is a near-constant deviation value observed between the simulated and monitored pressure values at all timestamps for a specific day. In the practical field context, sensor drifting may be caused by environmental disturbances to the positions of the deployed sensors underground, low operating power of the sensors, and initial calibration of the sensors goes out of range. Rectifying the sensor drifting thus involves adjusting the drifted monitored pressures values, across all hours of the selected day, by an approximated offset value.

2.4.2 Calibrating physical properties

After completing the NRW localization analysis, the WDN system's average pressure head discrepancy is expected to be most minimum for the MNF hours as compared to the peak-demand hours (e.g., 9am-11am) as the observed velocities in the underground pipes during the former hours are expected to be much less than 1.0 m/s, hence friction losses in the pipes are expected to be relatively insignificant to that of the latter hours.

To appropriately calibrate physical properties which account for the remaining energy discrepancies in the system, especially during the peak-demand hours, we progressively calibrate the pressure profiles for pressure sensor stations located nearest to the upstream reservoirs/tanks and gradually moving towards the furthest stations. Typically, for real-world WDN systems, we would expect that the valve settings to be well-calibrated with respect to the available Geographic Information System (GIS) information as provided by the utility company. Hence, the most typical physical property to be calibrated is the system's pipe roughness by systematically adjusting the initial C-factor values for the different pool of connected pipes between the reservoirs/tanks and the respective pressure sensor stations. The final adjustments to the different segments of connected pipes in the system are then evaluated via the level of goodness-of-fit between the model simulated and monitored pressure head values for all stations. Finally, we note that calibration of the relevant physical properties in a given system is usually conducted once for the initial hydraulic model having not undergone any prior calibration.

3 CASE STUDY

3.1 Description of WDN system

In collaboration with PUB, Singapore's National Water Agency, a real-world WDN system that serves industrial consumers is undertaken to verify our proposed calibration approach. The selected system consists of underground water pipes having a total length of 331.3km, 1 service reservoir, 11796 junction nodes, 35 pressure sensor stations, as illustrated in Figure 2. No operating pumps are installed in this network. The pipe diameters in the system range between 15.0mm and 1400.0mm, with an average value of 204.0mm. Due to data confidentiality information, the exact naming, and locations of the different hydraulic properties in the system cannot be revealed.

Historical flow and pressure SCADA data for the operational week of 19-25 Apr 2021 is thus leveraged to verify our proposed approach by emulating the near real-time context. For the selected week, a single leak event is reported on 21 Apr 2021. Again, we underline that the leak sizes of are often unknown in the near real-time context. Also, the initial hydraulic model configuration as provided by PUB has already been pre-constructed, to an extent, to represent the actual pipe connectivity and valve settings in the system.

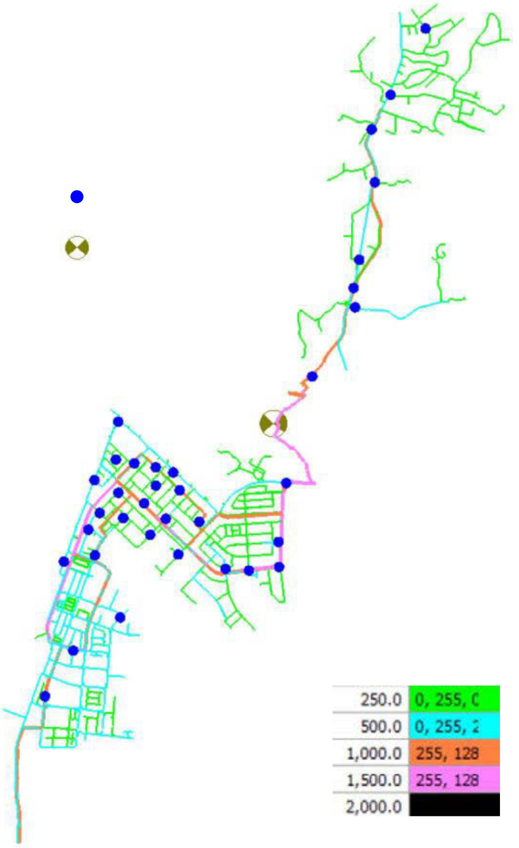
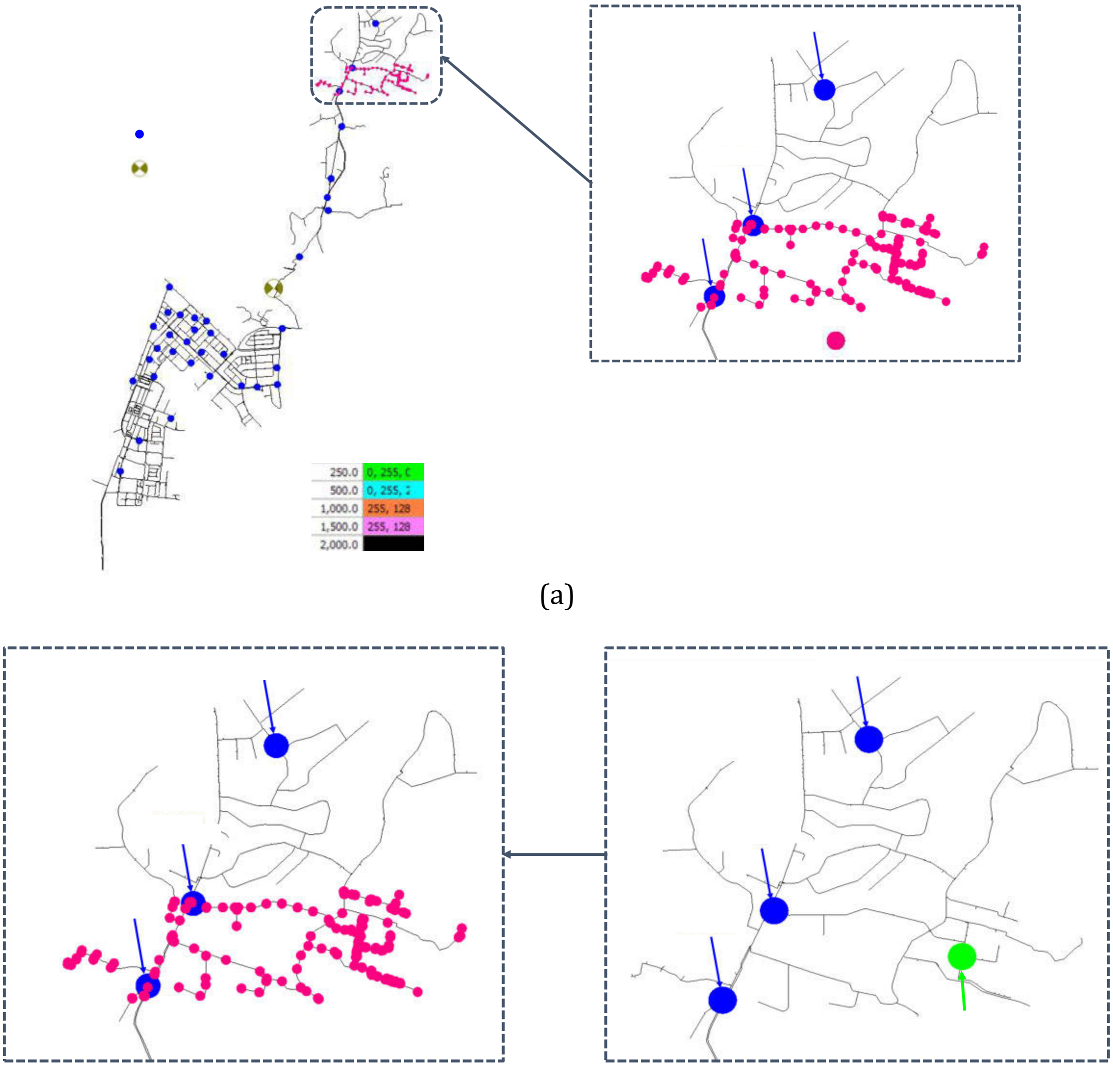


Figure 2. Case study WDN system in Singapore

CC BY NC SA

For the selected system, it has been found that there are 2 sets of unique demand consumption patterns where one follows the traditional diurnal (TD) usage pattern, i.e., low usage during the night and high usage during the day, and the other follows the exact opposite trend (NTD). The latter appears to correspond to a segment of farms located towards the north of the network itself, as shown in Figure 3a, where STN_J pressure readings has been found to best represent the observed water usage pattern for the identified farms in that area. For engineering simplicity, a singular junction node, as indicated in Figure 3b, is thus assumed to aggregate the total demand consumption by the known farms in the same area, where the assigned node is located in the near center of the available farms. The junction nodes (majority) in the other segments of the system, which are not situated near to the farms, are regarded to follow the TD pattern.

For illustrations, Figure 4a and 4b represent the typical normalized pressure profiles under the TD and NTD water usage patterns, respectively, for the system's Monday operations. As discussed, the lower normalized pressures for NTD's pattern during the MNF hours are due to the higher water usage by the available farms in the identified area (Figure 3a-3b), and vice versa during the typical peak-demand hours when compared to that of the TD's pattern.



(a)

(b)

Figure 3. Representation of junction nodes affiliated to farms' water usage pattern: (a) identified pool of nodes; and (b) aggregated junction node for engineering simplicity.

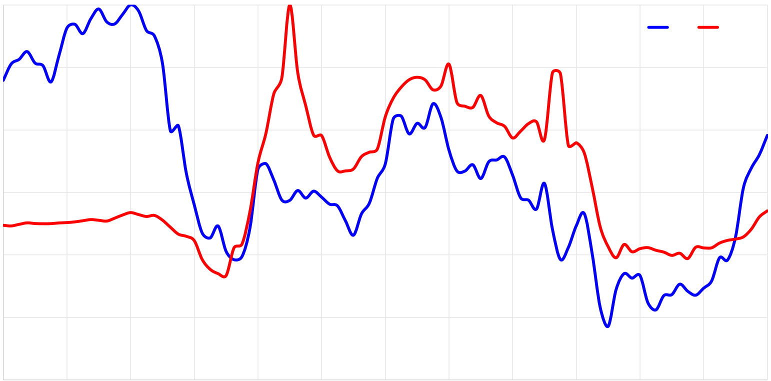


Figure 4. Normalized pressure profiles for TD vs NTD water usage patterns (Monday's example).

3.2 Flow Calibration

By following our proposed solution procedures to calibrate TD (via RAA) and NTD (localized) demand consumption patterns, Figure 5 represents their respective final calibrated patterns for Monday-Sunday in the selected week of 19 Apr 2021. These calibrated patterns are associated with the real customers in the system, by removing the NRW contributions at this stage. Following on, by developing the unique NTD pattern, we can better estimate the actual NRW volume, in percentage values, for the respective days as shown in Figure 6, by leveraging on the total net inflow and historical billing data records for the selected month.

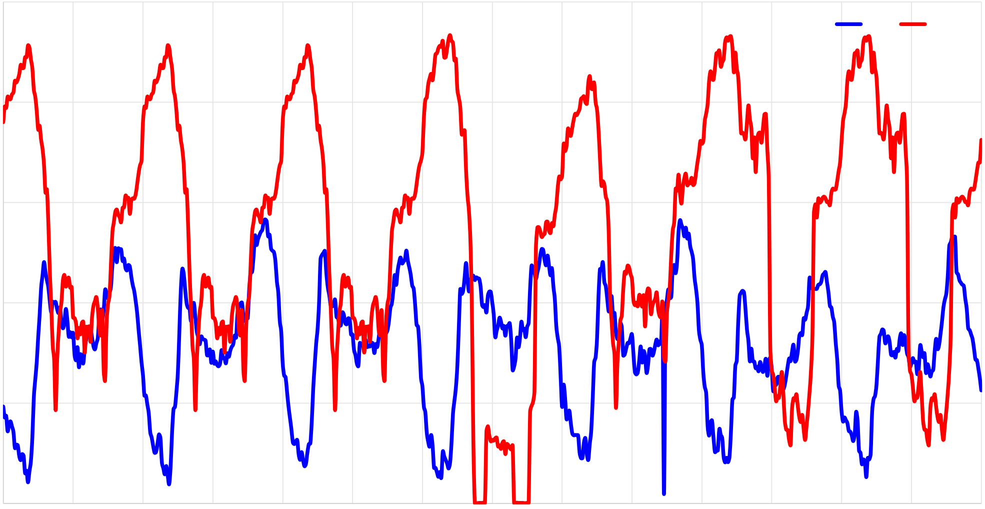


Figure 5. Calibrated demand consumption patterns for TD – bulk of customers in system, and NTD – local farms.

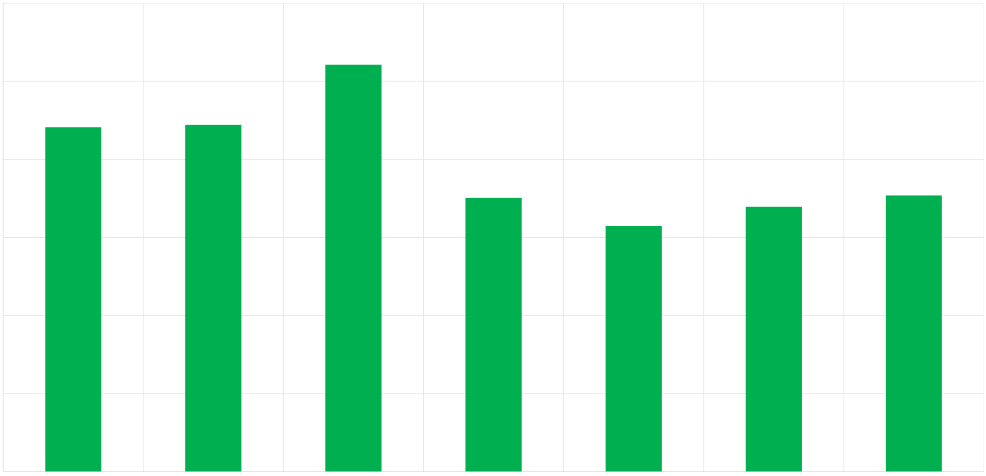
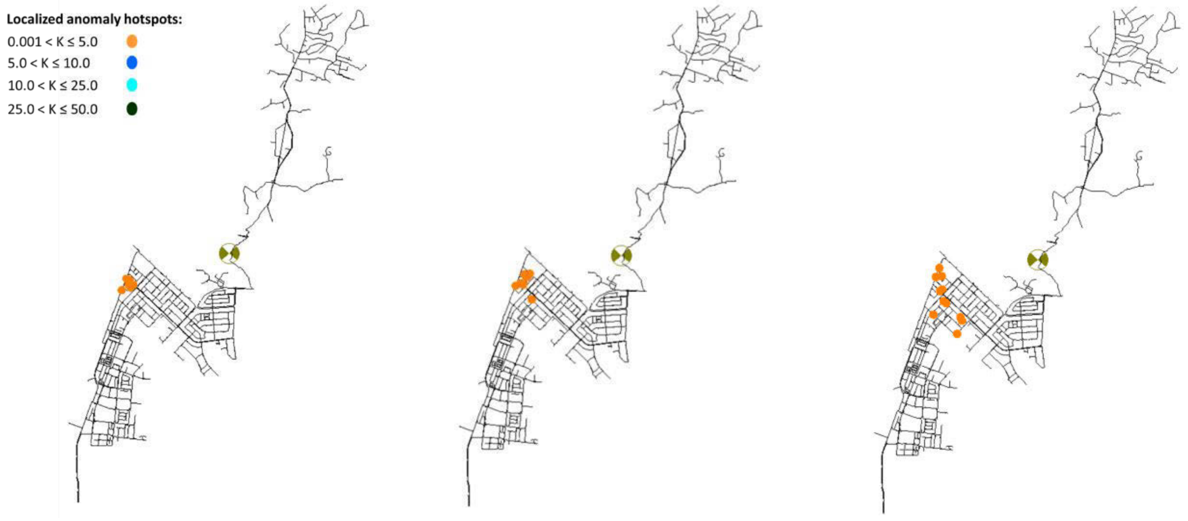


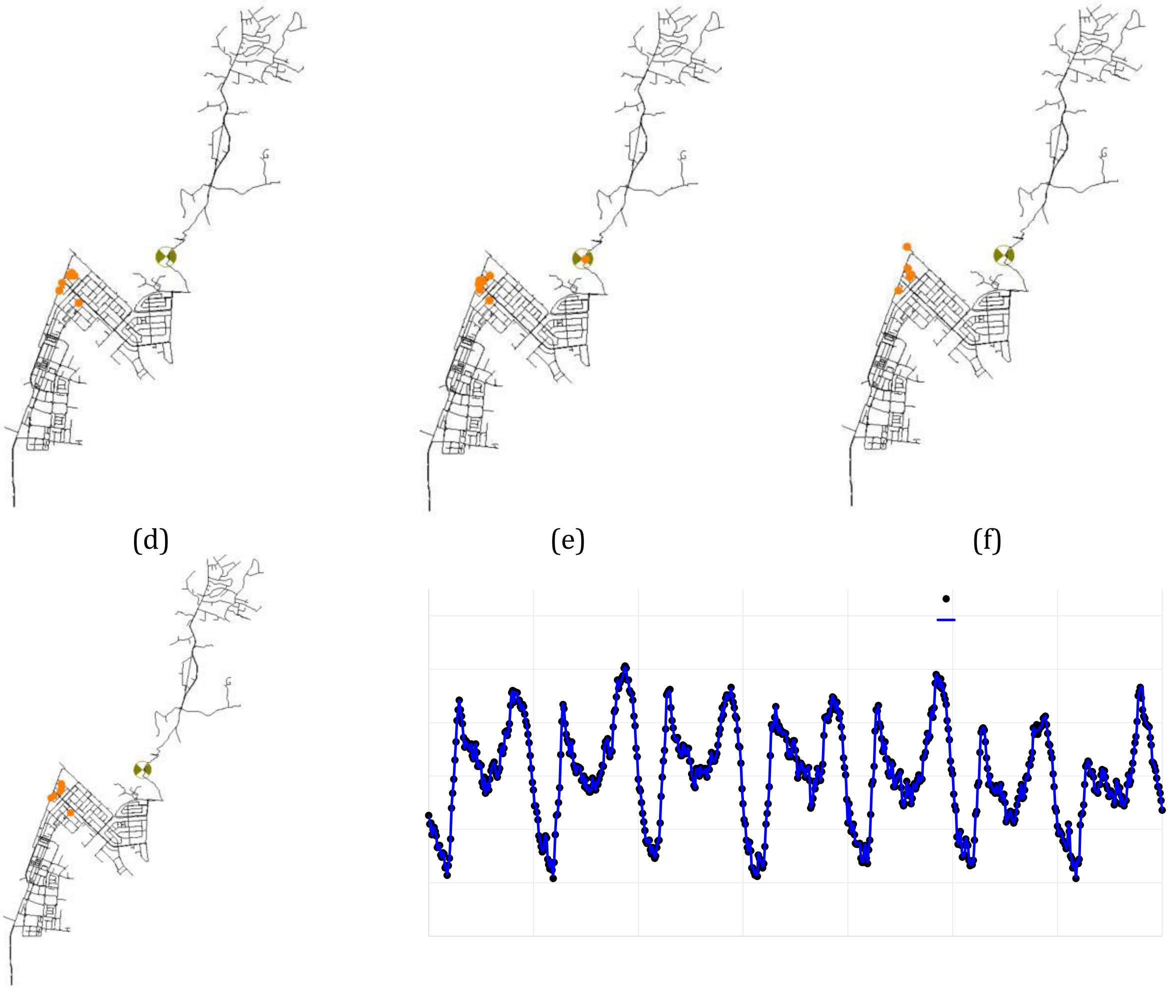
Figure 6. Estimated NRW volume in system for selected week of 19 Apr 2021.

The estimated NRW volumes are then localized as possible anomaly events which may include, but not limited to, hidden/unreported and background leak events and billing data uncertainties. Figures 7a-7g summarize the localized nodes, with a range of estimated emitter coefficient (K) values, within the network for the respective days (19-25 Apr) by using their corresponding MNF timestamp having the highest anomaly hydraulic power, with respect to Eq. (12). At this stage, the following key observations can be derived, namely:

- Figure 7h compares the model simulated and monitored total net inflow values for the selected system, where an average of 0.5% MAPE can be derived across the selected week.
- Across all 7 days, the NRW localization analysis constantly localize several anomaly nodes in around the same area of the network, as shown in Figures 7a-7g, hence indicating a strong likelihood of anomaly events (hidden leaks, demand uncertainties, etc.) taking place in that localized area.
- The approach could localize an actual reported event by PUB to less than 100m on 23 Apr 2021 with a maximum delayed time of 1 day, while also localizing the other possible anomalies which constitute to the respective NRW (%) volume estimated daily.



(a) (b) (c)



(d) (e) (f)

(g) (h)

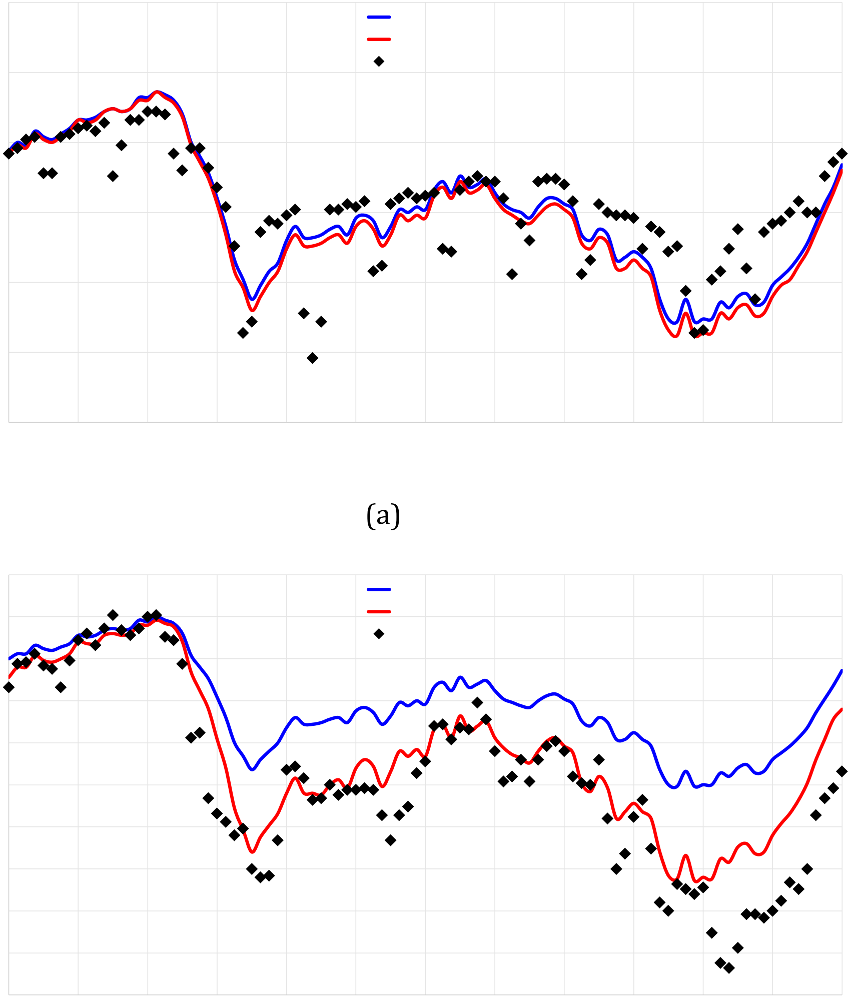
Figure 7. Localized anomaly nodes in network across different days of selected 19 Apr 2021 week: (a) 19 Apr – 3.30am; (b) 20 Apr – 2.45am; (c) 21 Apr – 2.30am; (d) 22 Apr – 2.15am; (e) 23 Apr – 2.15am; (f) 24 Apr – 3.15am; (g) 25 Apr – 2.15am; (h) total net inflow comparison after demand calibration and NRW localization.

3.3 Energy Calibration

Upon completing the NRW localization analysis for each day in the selected week, we proceed to first identify and rectify any likely drifting(s) in the recorded pressure data for the available sensor stations in the network. For the present analysis, it has been found that rectifying any sensor drifting(s) is done once on Monday (19 Apr 2021), and the same adjusted pressure head values can subsequently be maintained for the remaining of the same week. As discussed, since the valve settings are expected to well-calibrated by PUB beforehand, emphasis is thus placed on calibrating the system's pipe roughness values on the same Monday while maintaining the calibrated roughness values for the remaining days.

We thus progressively calibrate the system's pipe roughness from the nearest to the furthest pressure sensor stations from the upstream reservoir. Building upon the completed NRW localization scenario from Monday (19 Apr 2021), Figures 8a and 8b illustrate the before- and aftereffects of the pipe roughness calibration for the nearest and furthest stations respectively. Performing the systematic pipe roughness calibration for all available stations, excluding stations having "bad" monitored data such as missing data, negative/zero pressure values for the selected week, Figure 9 summarizes the MAPE (%) scores for the respective stations across the entire week where the average daily MAPE score is approximately 4%. Likewise, several key observations can be derived, namely:

- Stations, such as STN_D, STN_E and STN_Y, located furthest away from the reservoirs, tend to have higher daily average MAPE scores of between 4-6%. This could be caused by the accumulated calibration errors as stations move away from the reservoirs.
- The reasonably good agreement (~2% MAPE on average) obtained for the energy comparison for STN_I, STN_J especially, and STN_K, together with the prior good agreement achieved for the flow calibration, justifies the proposed local demand calibration by iterating against the daily monitored pressure values from STN_J, as the reference station, and a singular junction node to aggregate the demands for the local farms in the identified area.



(a)

(b)

Figure 8. Before- and aftereffects of pipe roughness calibration on pressure head values for Monday's (19 Apr 2021) analysis: (a) STN_M-nearest station; and (b) STN_E-furthest station.

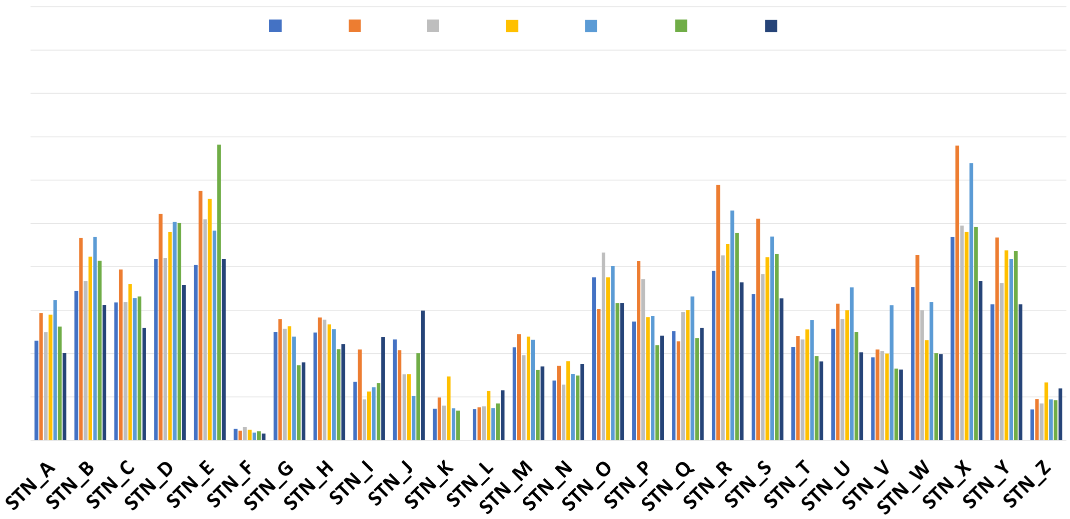


Figure 9. MAPE (%) scores for pressure head comparison across stations (excluding stations having "bad" data such as missing data, negative pressure, etc.) in WDN system for 19 Apr 2021 selected week.

4 CONCLUSIONS

This paper develops a practically novel daily model calibration approach for the real-world operations of water distribution networks (WDNs) by encompassing 3 main systematic components of (1) estimation of NRW contributions, (2) flow calibration, and (3) energy calibration. It is believed that the proposed approach is capable to calibrate real-world large WDN systems by leveraging on monitoring flow and pressure data which are collected 24/7 in the underground water pipelines. In collaboration with PUB, Singapore's National Water Agency, the hypothesis has since been verified by testing the approach on three WDN zones in Singapore having more than 1000km of underground pipes with varying demand consumption pattern(s), pertaining to different groups of customers, where the calibrated hydraulic model achieves daily average mean absolute percentage error (MAPE) scores of <1.0% and 4.0% approximately for the total flow and energy calibrations, respectively. Overall, the calibration approach serves as integral component to PUB's Smart Water Grid management, where a resulting well-calibrated model provides the baseline physics-based environment to facilitate a two-way data/information communication between the physical and digital working environments which can enhance the daily operations of WDNs. Acknowledgements

5 ACKNOWLEDGEMENTS

This research is supported by the Singapore National Research Foundation under its Competitive Research Program (CRP) (Water) and administered by PUB (PUB-1804-0087), Singapore's National Water Agency.

6 REFERENCES

[1] ASCE, "2021 report card for America's infrastructure," https//infrastructurereportcard.org/wp-content/uploads/2020/12/Drinking-Water-2021.PDF, accessed July 2021, pp. 34–43, 2021, doi: 10.1061/9780784478851.

[2] P. U. B. Singapore, "Managing the water distribution network with a Smart Water Grid," Smart Water, vol. 1, no. 1, p. 4, 2016, doi: 10.1186/s40713-016-0004-4.

[3] Z. Y. Wu and T. Walski, “Progressive Optimization Approach for Calibrating EPS Hydraulic Models,” Water Distribution Systems Analysis 2010. pp. 1558–1572, Jun. 25, 2010, doi: doi:10.1061/41203(425)139.

[4] A. Ostfeld et al., “Battle of the Water Calibration Networks,” J. Water Resour. Plan. Manag., vol. 138, no. 5, pp. 523–532, 2012, doi: 10.1061/(asce)wr.1943-5452.0000191.

[5] S. Alvisi and M. Franchini, “Calibration and Sensitivity Analysis of the C-Town Pipe Network Model,” Water Distribution Systems Analysis 2010. pp. 1573–1584, Jun. 25, 2010, doi: doi:10.1061/41203(425)140.

[6] T. Koppel and A. Vassiljev, “Calibration of Water Distribution Network for BWCN,” Water Distribution Systems Analysis 2010. pp. 1599–1609, Jun. 25, 2010, doi: doi:10.1061/41203(425)142.

[7] T. Koppel and A. Vassiljev, “Calibration of a model of an operational water distribution system containing pipes of different age,” Adv. Eng. Softw., vol. 40, no. 8, pp. 659–664, 2009, doi: 10.1016/j.advengsoft.2008.11.015.

[8] M. Grieves, Product lifecycle management : driving the next generation of lean thinking. New York: McGraw-Hill, 2006.

[9] M. Xue, A. W. Z. Chew, J. Cai, J. Pok, R. Kalfarisi, and Z. Y. Wu, “Improving near real-time anomaly event detection and classification with trend change detection for smart water grid operation management,” Urban Water J., pp. 1–11, Apr. 2022, doi: 10.1080/1573062X.2022.2058565.

[10] W. Jiajia, M. Donghui, and W. Wei, “Leakage Identification in Water Distribution Networks Based on XGBoost Algorithm,” J. Water Resour. Plan. Manag., vol. 148, no. 3, p. 4021107, Mar. 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001523.

[11] R. Luis, B. Joaquim, P. Vicenç, and C. Gabriela, “Clustering-Learning Approach to the Localization of Leaks in Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 148, no. 4, p. 4022003, Apr. 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001527.

[12] A. W. Z. Chew et al., “Daily Model Calibration with Water Loss Estimation and Localization Using Continuous Monitoring Data in Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 148, no. 5, pp. 1–14, 2022, doi: 10.1061/(asce)wr.1943-5452.0001546.

[13] Z. Y. Wu, S. Paul, and T. David, “Pressure-Dependent Leak Detection Model and Its Application to a District Water System,” J. Water Resour. Plan. Manag., vol. 136, no. 1, pp. 116–128, Jan. 2010, doi: 10.1061/(ASCE)0733-9496(2010)136:1(116).

[14] Bentley Systems Incorporated, “WaterGEMS user manual,” 2002.

[15] Bentley Systems Incorporated, “Darwin Calibrator Technique Reference,” 2002.

[16] W. Z. Y. and S. A. R., “Competent Genetic-Evolutionary Optimization of Water Distribution Systems,” J. Comput. Civ. Eng., vol. 15, no. 2, pp. 89–101, Apr. 2001, doi: 10.1061/(ASCE)0887-3801(2001)15:2(89).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14109

ACOUSTIC DATA ANALYSIS FRAMEWORK FOR NEAR REAL-TIME LEAKAGE DETECTION AND LOCALIZATION FOR SMART WATER GRID

Alvin Wei Ze Chew[1]; Zheng Yi Wu[2]; Rony Kalfarisi[3]; Meng Xue[4]; Jocelyn Pok[5]; Jianping Cai[6]; Kah Cheong Lai[7], Sock Fang Hew[8], Jia Jie Wong[9]

[1, 3,4,5,6] Bentley Systems Singapore Pte Ltd, 1 Harbourfront Pl, Singapore 098633
[2] Bentley Systems, 76 Watertown Rd, Thomaston, CT 06787, USA
[7,8,9]Water Supply (Network) Department, Public Utilities Board (PUB), Singapore

[1] *Alvin.Chew@bentley.com,* [2] *Zheng.Wu@bentley.com,* [3] *Rony.Kalfarisi@bentley.com,*
[4] *Meng.Xue@bentley.com,* [5] *Jocelyn.Pok@bentley.com,* [6] *Jianping.Cai@bentley.com,*
[7]*Lai_Kah_Cheong@pub.gov.sg,* [8]*Hew_Sock_Fang@pub.gov.sg,* [9]*Wong_Jia_Jie@pub.gov.sg.*

Abstract

Acoustic sensors are widely used for monitoring urbanized water distribution networks (WDNs) to detect and localize pipe leakages. Since their inception, few research studies have focused on developing a generic, effective, and practical methodology to analyse complex acoustics signals for leakage detection and localization in large-scale WDNs. In collaboration with PUB, Singapore's National Water Agency, a generic acoustic data analysis approach has been developed to facilitate PUB's present Smart Water Grid (SWG) management. The proposed approach encompasses multi-stage systematic analyses, namely: (1) data quality assessment; (2) data pre-processing; (3) near real-time leakage event detection and classification; and finally (4) near real-time leakage localization. Our proposed approach is then tested in major WDNs in Singapore having more than 1100km of underground water pipelines and 82 permanently installed hydrophone acoustic sensors between 1 Aug 2019 and 31 Aug 2020, where multiple historical leakage events were reported to within 600m, or less, from neighbouring hydrophones across the large complex networks. By emulating the near real-time detection and localization analyses daily, our proposed methodology could localize reported leakage events to an error range of around 150m on average, while demonstrating significant and stable acoustic leakage power rate over the temporal size of the leakage event cluster(s).

Keywords
Water distribution networks; acoustic signals; leakage detection and localization; acoustic energy analysis; autocorrelation analysis; peaks finding and pairing.

1 INTRODUCTION

Drinkable water is an important resource for humanity's livelihood. With rising uncertainty due to climate change and a growing global population, utility companies are facing increasing challenges to protect and ensure the supply of potable water to the public with minimum disruptions. For example, in the United States, an estimated volume of 6 billion gallons of treated water is reported to be lost each day where approximately 240,000 water mains breaks occur yearly [1]. In the context of Singapore, despite the continual investment in Smart Water Grid (SWG) management [2] by the state government, further reducing non-revenue water (NRW) losses continue to be major challenge due to the complexity of the real-world operations and hidden leakage events which can occur unexpectedly in the underground water distribution networks (WDNs) over time.

Over the last decade, acoustics sensors have been increasingly deployed by utility companies as part of their 24/7 permanently monitoring or temporary leakage program(s) due to low capital cost involved and ease of use. It is believed that acoustic sensors can complement with traditional

BY NC SA

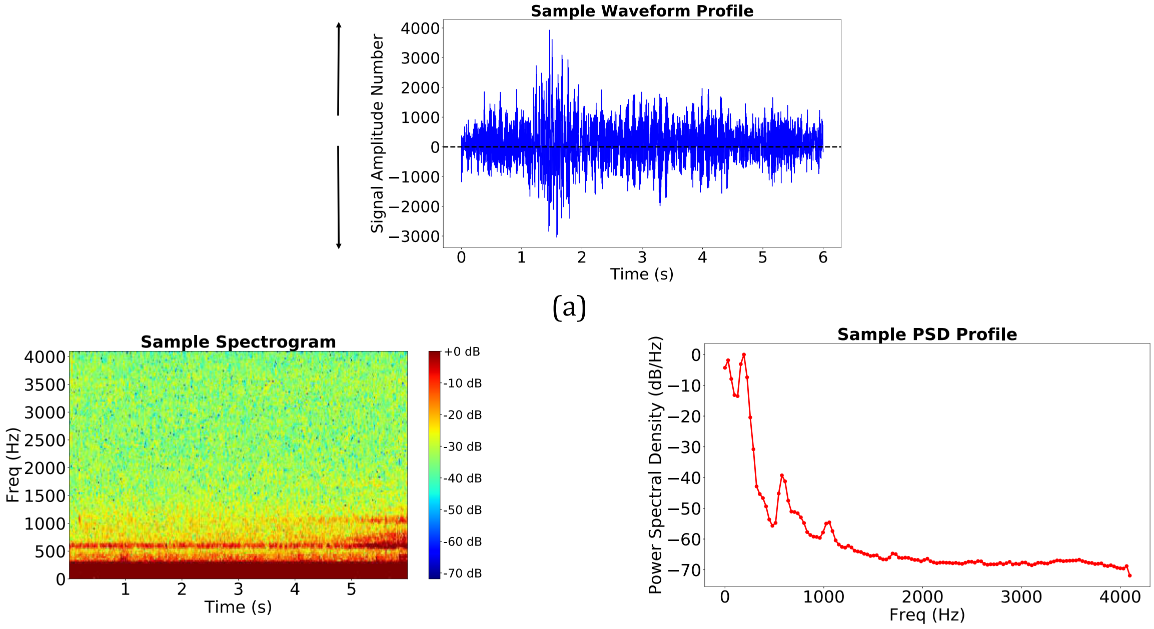
flow and pressure sensors to readily detect and localize hidden and insidious leakage events, before becoming disruptive events to the local public. Typically, an acoustic signal, as caused by pipe leakages, is due to the complex interaction between the flowing water and the interiors of the underground pipe wall(s), hence generating random wave signals with both short-term nonstationary and long-term stationary components [3]. Since their inception, many research works have been done to develop different engineering approaches for leveraging on acoustic data signals to perform leakage detection and localization in WDNs which include, but not limited to, traditional experimental analysis to perform signal-based processing [4]–[9], and advanced data-driven and deep learning methods [10]–[12]. Multiple notable works have also been performed using controlled field tests in reasonably large networks (≥120km of pipelines) with high density of acoustic sensors (≥300 sensors) per area [13], [14]. To the very best of our knowledge, while significant research has been done over years, we note that most of the proposed approaches are unlikely to be applicable for the real-world practical context due to the following reasons:

- The conducted works which have achieved high detection and localization accuracies are confined to networks having very high density of sensors per area or pipeline, whether under controlled experimental or field tests. For example, the case studies performed by [13], [14] in the context of Adelaide involved the deployment of high density of acoustic sensors per unit area, in order to detect and localize leakage events to within a limited spatial distance range. However, this requirement may not be generalized to all real-world WDNs as there can be cases having sparse number of acoustic sensors installed permanently, hence there is a strong likelihood that leakage events may occur at reasonably far locations from the nearest acoustic sensor(s), unlike from traditional experiments where sensors are often deployed less than 5-10m away from the simulated leaks in the networks.
- Limited number of acoustic datasets collected in real large-scale systems, pertaining to historical leakage events, available for training detection and localization models via deep learning. In the practical field context, it is not possible to collate large quantity of reported leakage events in well-managed WDNs, thus leading to imbalanced datasets in terms of the total number of leakages to non-leakage acoustic data records for training leakage detection and localization models. This limitation thus challenges the necessity for deploying advanced data-driven and deep learning methods for any engineering modelling objectives, especially in cases having sparse datasets.

To address the above-outlined shortcomings, this work, in collaboration with PUB, Singapore, develops a practically novel and generic acoustic data analysis methodology for analysing complex acoustic data signals collected under uncontrolled field conditions for detecting and localizing historical leakage events in more than 1100km of underground water pipelines with 82 permanently installed hydrophone sensors. This practical setting results in around 1 hydrophone availability for every 15km of pipeline in the combined network that is in stark contrast from the other notable reported field tests [13], [14]. By emulating the near real-time context to analyze acoustic data signals collected from the deployed hydrophones, the proposed methodology comprises of a series of systematic analyses which include: (1) data quality assessment, (2) features generation, (3) data pre-processing, (4) near real-time leakage detection, followed by (5) near real-time leakage localization.

2 DATA DESCRIPTION

Acoustic signal is usually stored as raw audio file which represents the signal's waveform amplitude profile. A typical waveform amplitude profile is illustrated in Figure 1a, which can be converted into its corresponding spectrogram and power spectral density (PSD) profiles as shown in Figure 1b and 1c respectively. Spectrogram (Figure 1b) comprises of two dimensions where the x-axis represents time (seconds), while the y-axis represents the frequency (Hz) of the acoustic signal. An additional 3rd dimension, as represented by normalized color intensity values in decibels (Db), quantifies the signal strength at a specific frequency value. PSD (Figure 1c) analyses the power density distribution of the same signal over its frequency range, where its x-axis represents the signal's frequency (Hz), and the y-axis represents the corresponding power density (db/Hz) at a specific frequency value.



(a)

(b)

(c)

Figure 1. Basics of acoustic signals data: (a) time-series waveform (amplitude) profile; (b) spectrogram; and (c) power spectral density (PSD).

3 METHODOLOGY

Figure 2 summarizes the key systematic procedures involved in our proposed generic leakage detection and localization approach, for every available acoustic sensor station(s) in the WDN system, that comprises of the following components:

i. **Data Quality Assessment:** Remove acoustic audio data files of "bad" quality characteristics.

ii. **Features Generations:** Generate acoustic power features for leakage detection and localization analysis.

iii. **Data Pre-processing:** Remove large transient power values for each acoustic sensor station.

iv. **Near real-time leakage detection and clustering:** Perform outlier detection using pre-processed acoustic power data, followed by clustering the detected outliers into leakage event clusters.

v. **Near real-time leakage localization:** By linking to the detected event clusters, perform leakage localization using autocorrelation analysis for power-peaks finding and pairing.

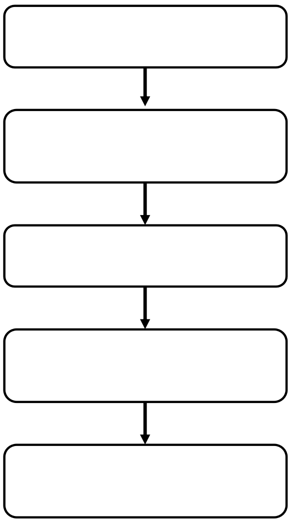


Figure 2. Overview of proposed leakage detection and localization using acoustic signals

3.1 Data Quality Assessment

During the operations of WDNs in the practical field context, a mixture of unknown environmental noises is expected to be embedded in the acoustic signals. It is also common for long-term and permanently installed sensors to not function correctly in the field at all times, hence resulting in numerical errors to be introduced into the recorded acoustic readings over time. Therefore, it is necessary and imperative to assess the initial data quality of each acoustic audio file before further signal analysis. The overall acoustic data quality assessment is performed using several metrics, as summarized in Table 1, while following a proposed screening protocol in Figure 3. An acoustic data file is only classified as "good" quality if it passes all 3 criteria.

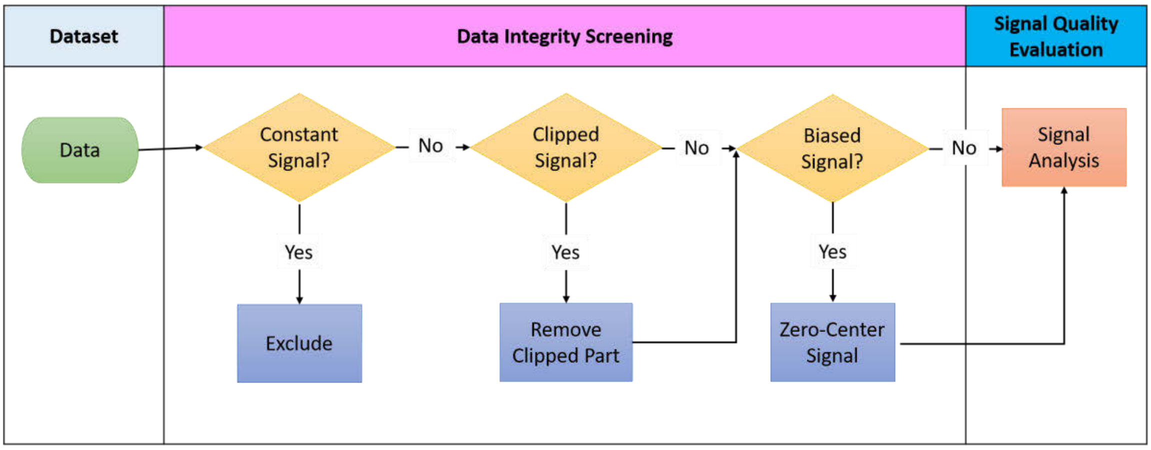


Figure 3. Protocol to systematically preprocess each acoustic data file for constant signals, clipped signals, and offset/biased signal.

Table 1. Descriptions of metrics adopted for performing acoustics data quality assessment.

Data Quality Issue	Problem Descriptions	Rectification measure
Constant Signal	Zero or constant amplitude values for a given datetime	Exclude from analysis
Clipped Signal	Waveform profile is being clipped as only amplitude values within the known upper- and lower bounds for a given bit depth can be recorded	Removal of clipped component
Drifted Signal	Signal amplitude values are not centered along the zero-axis.	Zero-centering of signal values

3.2 Feature Generation

After performing data quality assessment for each acoustic sensor station in the WDN system, we proceed to leverage on the "good" quality audio files to generate their corresponding total power (P_{all}) data representations. Note that P_{all} represents the basic case by using the original waveform amplitude profiles, as previously illustrated in Figure 1a. In general, it is computed by first converting the waveform into its PSD profile (Figure 1c), followed by summing the power density values (normalized or un-normalized form) across their entire or selected frequency range (e.g., 100-750Hz). Note this summation is similar to the traditional power root-mean-square (RMS) computations.

3.3 Data Pre-processing

For each acoustic sensor station, the generated P_{all} values then undergo a series of data pre-processing procedures which encompass the following:

CC BY NC SA

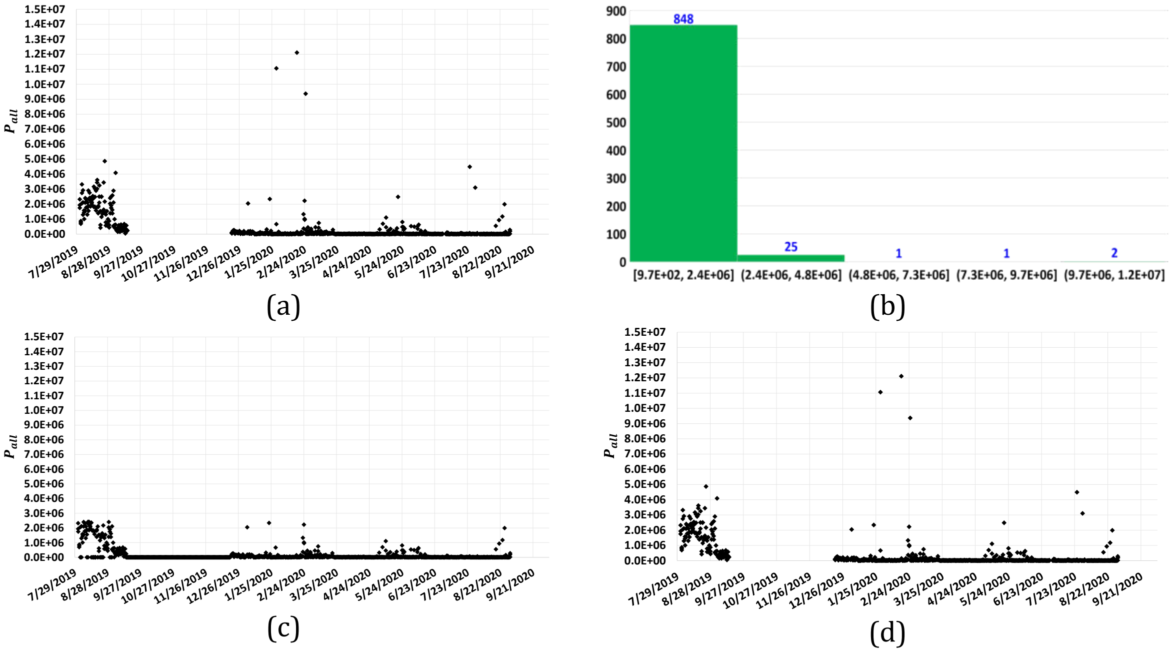
i. **Histogram analysis (1^{st} level of filtering)**: Distribute the data instances with a defined number of bins (N_B) where the data instances from the 1^{st} bin are retained and the remaining data instances from the other bins, i.e., 2^{nd} bin and beyond, are collated together.

ii. **Data restoration**: For each of the data instances in the combined 2^{nd} bin and beyond, compute the ratio between available adjacent data instances, namely: (1) between $P(t)$ and $P(t-1)$; and (2) between $P(t)$ and $P(t+1)$, where $P(t-1)$ and $P(t+1)$ are the original data instance values. If either of the computed ratio values from the above (1) and (2) computations are within a defined threshold scaling value (S_{thres}), $P(t)$ data instance will be formally restored back into the original 1^{st} bin of data instances, else $P(t)$ will formally remain in the 2^{nd} bin and beyond. At this stage, the original 1^{st} bin of data instances has been updated with any restored data instances. Finally, for the data instances, in the 2^{nd} bin and beyond, without any adjacent neighboring values, they will formally remain in the 2^{nd} bin and beyond.

iii. **Histogram re-analysis (2^{nd} level of filtering)**: Re-distribute the data instances in the final 1^{st} bin of data instances from the preceding step (iii) with the same number of bins (N_B), as previously defined in step (i), to obtain a new 1^{st} bin of data instances.

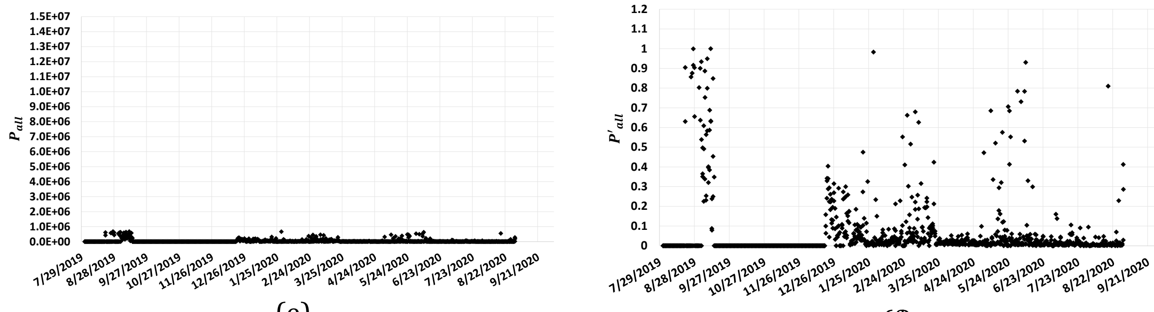
iv. **Data normalization**: For the new 1^{st} bin of data instances after the data re-distribution from preceding step (iii), normalize each data values accordingly (e.g., max-normalization or min-max-normalization).

To illustrate the above-proposed data pre-processing procedures, Fig. 4a shows a typical time-series profile for P_{all}. Fig. 4b plots the distribution of the power values across 5 bins (N_B). Fig. 4c shows the collated data instances (total of 848) from the 1^{st} bin as derived from Fig. 4b, while Fig. 4d illustrates the updated time-series profile after data restoration. By performing another round of histogram re-analysis using the collated data values from Fig. 4d, Fig. 4e plots the time-series profile for the final 1^{st} bin of data values. Finally, Fig. 4f illustrates the normalized power values from Fig. 4e by using max-normalization method.



(a) (b)

(c) (d)



(e) (f)

Figure 4. Example for pre-processing procedures for P_{all} values: (a) original time-series profile; (b) histogram distribution for original profile; (c) time-series profile after 1st filtering; (d) time-series profile after data restoration; (e) final time-series profile after 2nd filtering; (f) normalized final time-series profile

3.4 Near real-time leakage detection & clustering

Using their corresponding normalized time-series profiles for P_{all}, leakage detection and classification is then performed for each available acoustic sensor station in the system by adopting temporal-based clustering which depends on several key model parameters, namely: (a) universal reference value (P_{ref}) to perform power-based outlier detection; $0 \leq P_{ref} \leq 1.0$, (b) minimum number of detected outliers ($N_{outlier}$) between 2am-4am required for each station daily; $1 \leq N_{outlier} \leq 3$; (c) minimum number of consecutive days (N_{cons}) for each station to form an anomaly cluster where each of the days fulfilled the prior $N_{outlier}$ value defined; $1 \leq N_{cons}$.

For each selected combination of P_{ref}, $N_{outlier}$, and N_{cons} (from model training) in the near real-time context, the following set of systematic analyses is performed for each acoustic sensor station.

i. **Power-based outlier detection:** Compare each of the normalized data values with the universally defined P_{ref} value. If the normalized data value is greater than P_{ref}, then the corresponding timestamp is marked as a detected outlier.

ii. **Leakage Event Clustering:** The detected outliers are then aggregated along the time horizon and subsequently clustered or classified into leak events, where each of the detected events is required to fulfill the following criteria:

    a. **Basic Criterion 1 (BC-1)**: The total number of outliers detected during the MNF hours (e.g., 2am-4am) daily ≥ $N_{outlier}$.

    b. **Advanced Criterion 1 (AC-1):** The detected outliers are then aggregated over consecutive number of days into a single anomaly cluster event, where its corresponding size ($S_{current}$) ≥ N_{cons}.

iii. **Near real-time analysis:** Referring to Figure 5, on a daily basis, **BC-1** must be first fulfilled, followed by adding the identified number of outliers into $S_{current}$. If **BC-1** is not fulfilled on any given day, **AC-1** is then triggered to check if the respective criterion is fulfilled for $S_{current}$. $S_{current}$ is then reset to 0 on the following day for continuing the near real-time analysis.

Note that the selected P_{ref}, $N_{outlier}$, and N_{cons} parameters for the near real-time analysis are usually determined/optimized from the model training phase. Figure 6 exemplifies our proposed leakage event detection and clustering by using $P_{ref} = 0.4$, $N_{outlier} = 1$, and $N_{cons} = 3$, which results in $S_{current}$ to be 11 days where no outliers are detected after Day 11th as shown.

BY NC SA

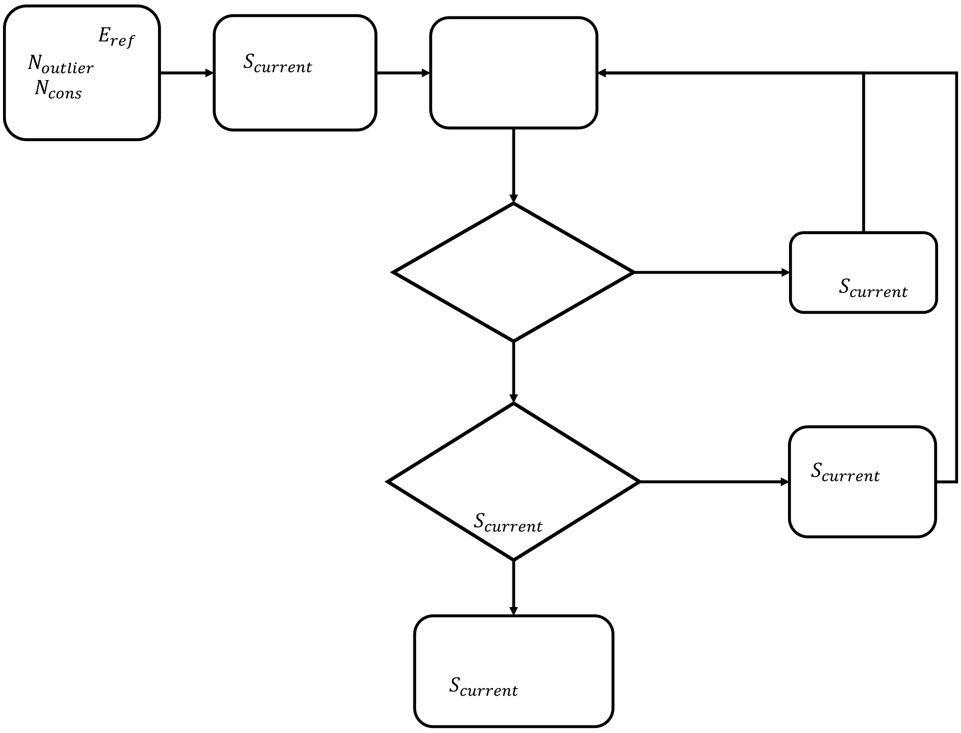


Figure 5. Proposed procedures for performing near real-time leakage detection and clustering.

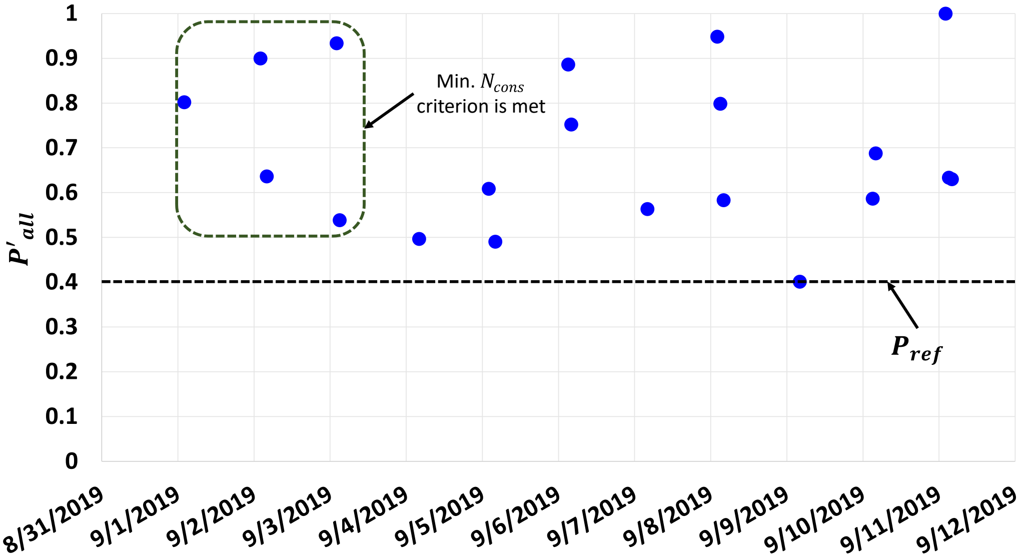


Figure 6. Example for anomaly event detection and clustering using $P_{ref} = 0.4$, $N_{outlier} = 1$, and $N_{cons} = 3$.

3.5 Near real-time leakage localization

Upon detection of leakage event clusters based upon any selected combination of P_{ref}, $N_{outlier}$, and N_{cons}, leakage localization is performed with the corresponding audio files for the associated detected outliers within the event cluster in the same near real-time context by adopting a set of mathematical procedures as follows:

i. **Bandpass filtering:** For each detected outlier that corresponds to a specific audio file, apply bandpass filter of a defined frequency range (e.g., 100-750Hz) to its original waveform profile to extract a filtered waveform profile, followed by generating its corresponding spectrogram.

ii. **Averaged spectral amplitude:** For the same selected frequency range from (i), derive the averaged spectral amplitude profile from the generated spectrogram using a defined number of moving average points (N_{avg}).

iii. **Autocorrelation analysis:** Apply autocorrelation function to averaged spectral amplitude profile that analyses all possible time-lags for the total time-length of the audio file.

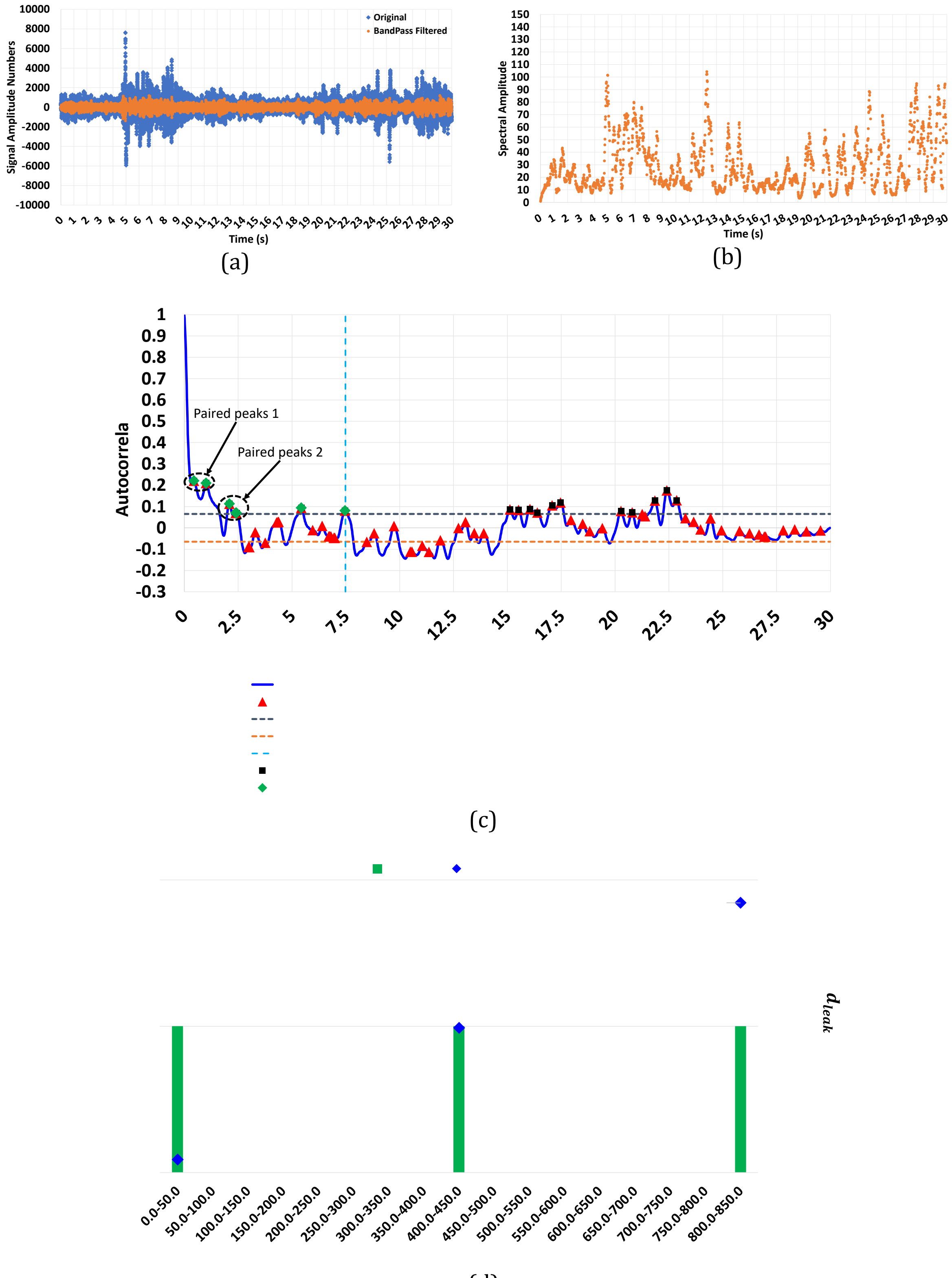
iv. **Power-peaks finding & pairing:** Perform peak finding and pairing on the derived autocorrelation plot from (iv) using suitable confidence intervals (e.g., 95%) for multiple tolerance bounds as follows:

   a. **Horizontal upper- and lower-bounds for autocorrelation values:** $\pm\frac{Z}{\sqrt{N}}$, where N represents the total number of data points in the averaged spectral amplitude time-series profile, and Z represents the t-statistic score for a defined confidence interval. In principle, the autocorrelation values which are outside of the upper- and lower-bounds are retained for the subsequent analysis.

   b. **Vertical bound for time-lag values:** For the same defined confidence interval, estimate the vertical time-lag bound using $\mu + n\frac{\sigma}{\sqrt{M}}$, where μ and σ represent the harmonic mean and standard deviation of the time-lag values corresponding to the M number of retained autocorrelation values from (a), and n refers to the total number of sigmas to be considered. For example, for 95% confidence interval, n equates to 3.

v. **Localization distances estimation:** For each pair of the power peaks identified from (iv), check if (1) their corresponding time-offset (Δt) is within a defined threshold time-offset (T_{thres}), and (2) the corresponding absolute difference between the pair of autocorrelation values is within a defined threshold autocorrelation tolerance (A_{thres}). If both criteria (1) and (2) are fulfilled, proceed to estimate the localization distance (d_{leak}) for each pair by multiplying Δt with the speed of sound in water (v_{sound}). Collate and distribute all estimated d_{leak} values via histogram analysis with a defined distance width (d_{width}), followed by determining the average localization distance (D_{leak}) from the specific bin having the highest frequency count.

Figure 7 exemplifies the key procedures involved to estimate the localization distance(s) for a singular detected outlier and its associated audio file, where Figure 7a represents the original and bandpass filtered waveform profile for the audio file. Figure 7b then illustrates the averaged spectral amplitude profile derived from the bandpass filtered waveform, followed by using the filtered waveform to generate its autocorrelation plot and the associated peaks as shown in Figure 7c. The same figure illustrates the resulting upper- and lower-horizontal bounds using 95% confidence interval to first identify the statistically significant peak autocorrelation values, i.e., those outside of the two bounds, followed by plotting the vertical bound, as shown, to isolate the key peak values to within a certain time-lag (s). Finally, Figure 7d shows the histogram plot for 3 estimated d_{leak} of 40.5m, 445.2m, and 829.6m for 3 unique pairs of peaks identified from Figure 7c, where each distance has a count of 1 when d_{width} is fixed at 50.0m and v_{sound} = 1480m/s. In cases where there are multiple dominant localization distance ranges having the same counts, then we estimate their harmonic mean value among them, which will result in a D_{leak} value of 106.6m for the current selected example.



(a)

(b)

(c)

(d)

Figure 7. Example for estimating localization distance for each detected outlier: (a) applying bandpass filter to original waveform profile; (b) derivation of averaged spectral amplitude profile from filtered profile; (c) derivation of autocorrelation plot with paired peaks; (d) localization distances estimation and histogram analysis

cc BY NC SA

4 CASE STUDY

4.1 Description of WDN systems

In collaboration with PUB, Singapore's National Water Agency, our proposed leakage detection, and localization methodology is verified with large-scale WDNs, which encompass three water supply zones (Zone-1, Zone-2, Zone-3) in Singapore, as shown in Figure 13. All three zones consist of underground water pipes having a total length of 1100km and 82 permanently installed hydrophone sensors. Table 2 summarizes the details of the audio files collected across all hydrophones in all three zones for the period between 1 Aug 2019 and 31 Aug 2020, where 18 historical leakage events were reported to within 600m, or less approximately, from neighboring hydrophone acoustic sensors, as summarized in Table 3 for the respective zones.

Table 2. Details pertaining to acoustic data files collected in Zone-1, Zone-2, and Zone-3.

Detail	Zone-1	Zone-2	Zone-3
Date range	1 Aug 2019 – 31 Aug 2020		
Total quantity of hydrophones	27	47	8
Total quantity of .WAV files	156734	232026	44374
Total quantity* of .WAV files from MNF hours	21734	25518	6109
Bit depth of .WAV files	16		
Sampling rates (Hz)	2048-8192		
Time length of .WAV files (s)	6.0-30.0		
No. of channels	1 (mono)		

**** After undergoing data quality assessment***

Table 3. Summary of reported leakage events in Zone-1 to Zone-3, and their associated nearest hydrophone stations.

Zone	Reported Leak Dates	1st detected date (nearest leakage cluster to reported leak)	Nearest station	Pipeline distance between leak & nearest station (m)	No. of detected leakage clusters	Avg. Predicted localization dist. (m)
1	9/11/2019	9/1/2019	STN_A	72	2	174
	9/9/2019	9/5/2019	STN_B	419	12	263
	8/16/2019	8/12/2019	STN_C	587	12	517
2	1/9/2020	1/5/2020	STN_A	36	3	157
	5/22/2020	5/20/2020	STN_B	111	7	363
	8/1/2019	8/2/2019	STN_C	202	3	234
	8/5/2020	8/3/2020	STN_D	324	7	146
	5/6/2020	5/7/2020	STN_E	530	7	473
	3/13/2020	3/7/2020	STN_F	600	4	280
	3/12/2020	3/7/2020	STN_F	600		280
3	3/6/2020	3/2/2020	STN_A	353	9	654
	1/23/2020	1/24/2020	STN_B	453	19	415
	3/28/2020	3/24/2020	STN_C	518	13	450
	6/28/2020	6/24/2020	STN_C	627		514
	1/21/2020	1/17/2020	STN_D	613	21	625
	1/20/2020	1/17/2020	STN_D	736		625

4.2 Near real-time leakage detection and localization

4.2.1 Leakage event detection

By emulating the near real-time context using the historical reported leakage events, we first perform the leakage detection and clustering analysis by adopting the model parameters of $P_{ref} = 0.4$, $N_{outlier} = 1$, and $N_{cons} = 3$. Table 4 summarizes the total number of detected event clusters for the respective station in each zone, and the 1st detected date of the event cluster located closest to the reported event temporally. For example, for STN_A in Zone-1, there are 2 detected event clusters for that station during 1 Aug 2019 and 31 Aug 2020, where the detected cluster nearest to the reported event on 9/11/2019 is first formed on 9/1/2019 and lasted for a total 11 days till 9/11/2019. We do, however, note that some cases may have the nearest detected event cluster(s) to be formed after the leakage event is reported with a maximum delayed time of 1 day, as demonstrated in the examples (see Table 3) for STN_B and STN_C in Zone-2, and STN_B in Zone-3. For all stations in Table 3, it is worth highlighting that the other leakage event clusters, which are not located temporally close to their respective reported events, may or may not represent hidden and unreported leakage events that require further field investigations.

4.2.2 Leakage localization

For each of the nearest event cluster detected to the reported events in Table 3, we proceed to estimate their dominant localization distances (D_{leak}) for every detected outlier within the event cluster by following our proposed mathematical procedures as summarized previously. Figures 8a-8c illustrates the resulting localization distances computed over the temporal size of the event cluster for a single leakage scenario from each of the zones, respectively:

- STN_A from Zone-1 for leakage event reported on 9/11/2019 (see Figure 8a)
- STN_A from Zone-2 for leakage event reported on 1/9/2020 (see Figure 8b)
- STN_C from Zone-3 for leakage event reported on 3/28/2020 (see Figure 8c)

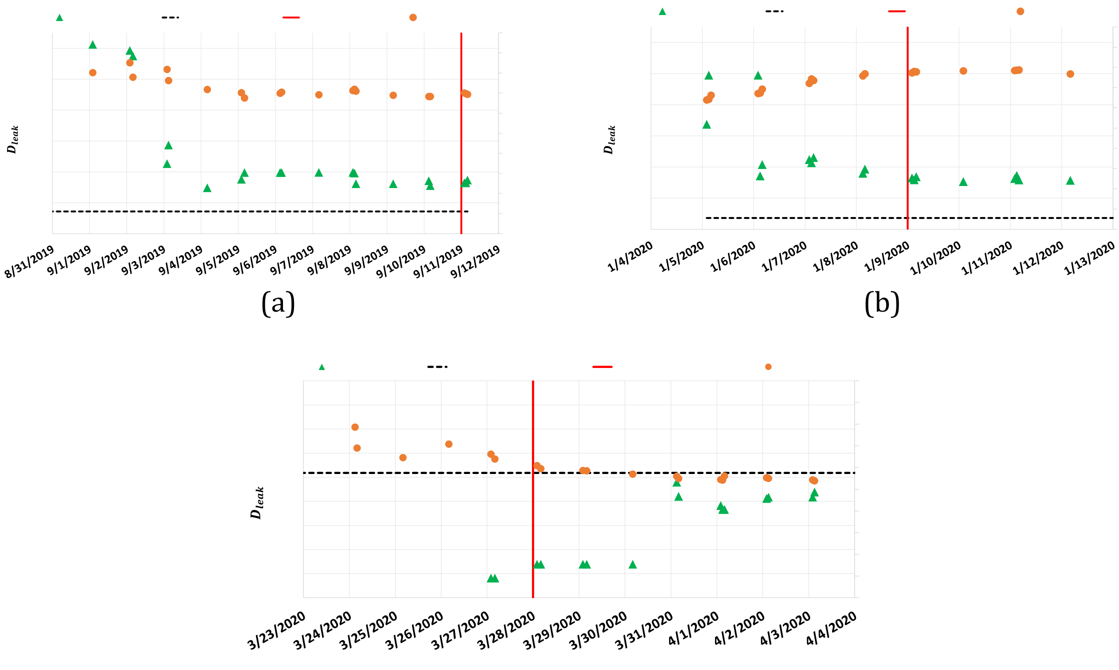
In each of the Figures 8a-8c, several important pointers must be noted, namely: (1) the estimated D_{leak} values, as represented in the respective primary axis, over the temporal size of the event cluster are based upon harmonic mean computations in a rolling-forward temporal basis to emulate the near real-time context, (2) same harmonic mean computation principle is applied to compute the acoustic power over time in near real-time as represented in the corresponding secondary axis, and (3) the localization distance estimated for the final outlier of the event cluster, before the cluster breaks off, is then taken as the final average D_{leak} for the analysis.

By adopting the universal model configurations of 100-450Hz frequency range, N_{avg} = 12, 95% confidence interval for the upper-, lower- and vertical-bounds, $T_{thres} \approx 0.676$s, A_{thres} = 0.25, v_{air} = 1480m/s, and d_{width} = 50.0m, the final column of Table 3 summarizes the final average predicted D_{leak} value for each of the reported events. In summary, the following key observations can be made at this stage, namely:

- For most cases, the average error discrepancy between the reported and predicted pipeline distances is approximately 150m in absolute value, except for STN_B (Zone-1), STN_E and STN_F (Zone-2), and STN_A (Zone-3) where the bulk of their predicted distances are underestimated by more than 150m from the actual reported distances.

- Conservatively, the minimum detection criteria of $P_{ref} = 0.45$, $N_{outlier} = 1$, and $N_{cons} = 3$ provide sufficiently high confidence level that a pipeline leakage event is most likely taking place in the near proximity (≤ 600m) of the respective hydrophone station that is detecting and reporting the event cluster to the operator.

- As demonstrated in Figures 8a-8c, the normalized acoustic power values gradually increase over time, or at the very least, maintain a near-constant value above the minimum required power of 45% and above. This temporal observation can serve as an additional indication that a pipeline leakage event is taking place in the near proximity of the respective hydrophone station. Another common observation is that the power values may first rise to high value (e.g., to around 70-80%), followed by approaching a near-plateau power percentage value (> 45%) over the temporal size of detected event cluster.



(a) (b)

(c)

Figure 8. Estimated localization distances over temporal sizes of detected cluster located closest to respective reported leakage events: (a) STN_A from Zone-1 for leakage event reported on 9/11/2019; (b) STN_A from Zone-2 for leakage event reported on 1/9/2020; and (c) STN_C from Zone-3 for leakage event reported on 3/28/2020.

5 CONCLUSIONS

This paper develops a generalized acoustic data analysis methodology to perform near real-time leakage detection and localization in underground water distribution networks (WDNs). In collaboration with PUB, Singapore's National Water Agency, our proposed methodology comprises of systematic procedures for analysing acoustic data signals, namely: (1) data quality assessment; (2) features generations; (3) data pre-processing; (4) near real-time leakage detection and clustering; and (5) near real-time leakage localization. It is believed that the methodology can detect and localize leakage events in large WDNs having permanently installed acoustic sensors and has since been verified with 3 WDN zones in Singapore having 82 permanently installed hydrophone sensors across the networks. By emulating the near real-time context using historically available reported leakage events, our approach could successfully detect leakage events, as reported to within 600m or less from a neighbouring hydrophone station, with a maximum delayed time of 1 day in all 3 zones. The detected leakage event clusters are then further analysed to predict the likely locations of the leakage events from the nearest hydrophone stations, where the bulk of the events can be localized to within 150m error discrepancy, on average, with significant detected acoustic power of more than 45% over the temporal size of the event clusters.

6 ACKNOWLEDGEMENTS

This research is supported by the Singapore National Research Foundation under its Competitive Research Program (CRP) (Water) and administered by PUB (PUB-1804-0087), Singapore’s National Water Agency.

7 REFERENCES

[1] E “2 2 p ’ ” p // p . / p -content/uploads/2020/12/Drinking-Water-2021.PDF, accessed July 2021, pp. 34–43, 2021, doi: 10.1061/9780784478851.

[2] . U. . p “M b W ” W . 1, no. 1, p. 4, 2016, doi: 10.1186/s40713-016-0004-4.

[3] . M. . J p . W . . “ oustic Filtering of the Pipe and Sensor in a Buried Plastic Water Pipe and its Effect on Leak Detection: An Experimental I ” . 4 . 3. pp. 5595–5610, 2014, doi: 10.3390/s140305595.

[4] . M. “ f an In- p W L ” M Institute of Technology, 2010.

[5] . . K . K R. M M. . H b b “ L W p U M p ” J. p . Eng. Pract., vol. 3, no. 2, pp. 47–54, 2012, doi: 10.1061/(asce)ps.1949-1204.0000089.

[6] M. M. . “L p p b p p ” I . . . 66 2 5 : 10.1063/1.4915721.

[7] . Z. . M J. W “ b x p p ” p p . . . 82 p. 108255, 2021, doi: 10.1016/j.apacoust.2021.108255.

[8] M. K Z. L M. . E . H. “ p I p p –Based Cross-Correlation Techniques along with Empirical Mode Decomposition for Water Pipeline Leakage Localization Utilizing Acousto- p ” J. p . E . . . . 3 p. 4 2 27 2020, doi: 10.1061/(asce)ps.1949-1204.0000471.

[9] . H. M . . R. “I p L Iron and Copper Water- b p ” J. p . E . . . 8 . 3 p. 5 7 2 7 doi: 10.1061/(asce)ps.1949-1204.0000257.

[10] R. . . . T J. “ L W b p U p Autoencoder and H p ” J. p . . E . . 34 . 2 p. 4 2 2 2 doi: 10.1061/(asce)cp.1943-5487.0000881.

[11] J. K . . M b . J. L . W . E “ L b E b ” IEEE Trans. Ind. Electron., vol. 65, no. 5, pp. 4279–4289, 2018.

[12] . . “L W b T –Frequency Convolutional ” J. W R . . M . . 47 . 2 p. 4 2 2 2 : 10.1061/(asce)wr.1943-5452.0001317.

[13] . M J. . Z . M L. x M. . L b “L -Before-Break Main Failure W b p U W T : ” J. Water Resour. Plan. Manag., vol. 146, no. 10, p. 05020020, 2020, doi: 10.1061/(asce)wr.1943-5452.0001266.

[14] . Z M. . L b M. L. p J. . . “ p b prevention by permanent acoustic noise level monitoring in smart ” U b W J. . 7 no. 9, pp. 827–837, 2020, doi: 10.1080/1573062X.2020.1828501.

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14111

MODELLING OF AIR POCKET ENTRAPMENT DURING PIPE FILLING IN INTERMITTENT WATER SUPPLY SYSTEMS

João P. Ferreira[1], David Ferras[2], Dídia I. C. Covas[3], Zoran Kapelan[4]

[1]Dept. of Water Management, Faculty of Civil Engineering and Geosciences, Delft, Netherlands,
[2]Department of Water Supply, Sanitation and Environmental Engineering, IHE Delft Institute for Water Education, Delft, Netherlands.
[3]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.
[4]Dept. of Water Management, Faculty of Civil Engineering and Geosciences, Delft, Netherlands,
[1] j.p.ferreia@tudelft.nl, [2] d.ferras@un-ihe.org, [3] didia.covas@tecnico.ulisboa.pt,
[4] z.kapelan@tudelft.nl

Abstract

Intermittent water supply (IWS) operations frequently involve water filling and emptying cycles that are strongly influenced by air-water interaction. Storm Water Management Model (SWMM) software has been recently proposed to simulate pipe filling events in IWS systems. As the tool is conceived to simulate both free surface and pressurized flows, it also has the potential to analyse the air entrapment. However, there is no numerical model capable of accurately and efficiently simulating the air behaviour during these events, nor of predicting the locations of air pockets created during the pipe filling process.

An experimental pipe rig was assembled to better understand the pipe filling process in terms of pressure variation, the propagation of the filling wave and the air entrapment locations under different initial conditions. The pipe rig has a classic reservoir-pipe-valve configuration. Different behaviours are observed in this experimental setup during the pipe filling tests. An entrapped air pocket is created at the high point for lower flow rates, which is not dragged when the pipe is full. This air pocket can go from a similarly free surface flow inside the pipe to a complete water filled flow, depending on the flow rate. For low flow rates, a high head loss is introduced due to a hydraulic jump inside the pipe. For higher flow rates, the air is dragged, no air is entrapped and only the local head losses from the change of direction at the high point are observable.

Following the collected experimental data, SWMM is used to assess to which extent it can predict entrapped air pockets location and their volume. Different filling processes can occur and an air model should be included to simulate the tests carried out in the pipe rig. The results obtained show that SWMM is capable of predicting air pocket locations but not the air pocket volumes. Further research is necessary to improve SWMM in this context.

Keywords
Intermittent water supply, pipe filling, air pockets, experimental test, numerical modelling, SWMM.

1 INTRODUCTION

Water supply is of major importance for our daily routines and activities. Water is expected to immediately come out of the taps whenever these are opened by the consumers, giving it the name of continuous water supply (CWS). This assumption is not always true outside of developed countries. Around 40% of the world population is supplied by a piped network that suffers from what is called intermittent water supply (IWS) [1]. Such supply is characterized by a distribution system not being pressurized 24/7, like CWS, with different degrees of reliability. Three different stages can be identified in IWS systems operation: pipe filling, system supply and pipe emptying.

The system starts empty until the utility decides to open valves to start the pipe filling stage. Water goes into the pipe and air needs to be release through taps, tank valves and air valves. However, air can cause severe damage in distribution systems due to unexpected pressure variations in the systems resulting in supply disruptions [2]. Air can compress and expand to the point that it generates pressures that were not accounted for in the design stage, leading to increased pipe failures and higher leakage levels as observed by Christodoulou and Agathokleous [3]. During the filling stage, the system is prone to accumulate entrapped air pockets due to the dynamic behaviour of the flow, but current engineering practice doesn't enable the description of where and under which circumstances air pockets get entrapped other than a rule-of-thumb based on the "high points".

When the pressure is sufficient to deliver water to the service connections, the supply stage starts at that location. The air that was initially in the pipe may or not become entrapped in the system at high points and sudden changes in slope and direction depending on the flow velocity and the pipe slope [4]. Another consequence of having air inside the pipes is the inherent local head losses, due to a reduction of the effective pipe cross section area, which are not negligible in low pressure system due to the high flow velocities as in IWS systems [5, 6]. Thus, air should be released in a controlled way so that lower head losses and pressure variations would occur. However, there is no model to determine air pocket location and volume during/after the pipe filling stage.

The pipe emptying stage begins when the utility closes the connection between the supplying storage tank and the pipe system. Water is still delivered for a short period of time thanks to the reminiscences of pressure in the system and to elevation differences. During this period, air needs to go into the pipes to allow the water flowing out. Either air goes into the pipes by leaks and air valves or sub atmospheric pressures occur in the pipe (in extreme cases leading to pipe buckling). Sub atmospheric pressures are bad not only from the pipe perspective, since they can lead to pipe buckling [7], but also from the water quality perspective, since these allow for losses of untreated water and for contaminants intrusion into the pipe system, putting at risk the supply of safe water [8].

Numerical modelling of water systems can be very useful for predicting air water behaviour during pipe filling and emptying processes. These can be categorized in four types of models: extended period simulation models in pressurized pipes (e.g. EPANET model), free surface flow models in open-pipes (e.g. Stormwater Management Model, SWMM), rigid water column (RWC) model and elastic water column model (EWC).

Extended period simulation models are amply used for pressurized pipes simulation, providing accurate results as long as demands, valves and pumps settings and other boundary conditions are well defined. One of the main assumptions is that the pipes are pressurized 24/7 which is not verified in IWS systems. This assumption also does not make possible to track and detect locations prone to air pocket locations.

Free surface flow models in open-pipes, like SWMM model, has also been used to analyse water systems but in a water drainage perspective. Nevertheless, free surface flows can be accurately simulated by this model, since the model numerically solves Saint-Venant equations (Eqs. 1 and 2) to calculate water depths:

$$\frac{\delta A}{\delta t}+\frac{\delta Q_w}{\delta x}=0 \tag{1}$$

$$\frac{\delta Q_w}{\delta t}+\frac{\delta (Q_w^2/A)}{\delta x}+gA\frac{\delta H}{\delta x}+gAS_f=0 \tag{2}$$

where A is the flow cross section, Q is the water flow rate, g is the gravitational acceleration, S_f is the friction factor, t is time and x is the length. Specifically using SWMM, the solver features two surcharge methods to when pipes become pressurized: EXTRAN and SLOT surcharge methods. The former relies on mass and momentum equations for presurized pipes and the second creates an artificial slot on top of the pipe for the assumptions of Saint-Venant equations to remain valid.

Rigid water column models are easy to use and to implement for single pipe systems [9]. Mass continuity and head loss equations are used to estimate flow rates and pressures. Though, the assumptions are incompatible with the existence of entrapped air pockets in the pipes and that of free surface flow existing during the filling process as observed by Guizani *et al.* [10].

Elastic water column models can be applied to pressurized and free surface flows [12, 13] and have been proposed to simulate pipe filling events with water column separation [14]. Accurate results are obtained for both situations. However, lower time steps and higher spatial discretization are required making these models computationally demanding and inefficient for the modelling of pipe networks.

The study of air in pipes has been of interest withing the scientific community but with little developments in the IWS domain. Some work has been carried out to determine entrapped air pockets head losses [4] and breakout rates [15] in steady state flows, but such research was never carried for pipe filling events when the dynamic effects are more severe. Also, no hydraulic solver is known to accurately predict air pocket volumes and related pressure variations due to their existence in pipe systems, being such limitation pointed out as one of the main disadvantages of SWMM [16]. SWMM might be used to check the possibility of existence of entrapped air pockets since it simulates open channel flows. Though, since the software was originally developed for drainage purposes, the air pockets are not considered as soon as the cross section prone to air entrapment is pressurized. SWMM has already been suggested as a possible alternative to model IWS systems [11]. Yet, no reference was made to the air behaviour, which is of utmost importance in IWS.

This paper aims at assessing to which extent SWMM can accurately predict air pockets' location and volume in pipe filling events. Experimental results are collected in a laboratory setup to obtain pressure signals and video recordings to better understand the pipe filling process and to determine the air pocket volume. Collected data are compared with numerical results from SWMM to assess if the model can predict the air pocket location and volume.

This paper is organised as follows: section 2 presents the pipe-rig set up and describes the experimental tests that were carried out as well as the conducted test procedure; section 3 summarises the main results from the experimental tests; section 4 briefly explains the numerical model and presents the comparison between numerical and experimental data; and section 5 summarizes the results, presents the main conclusions and further research topics.

2 EXPERIMENTAL DATA COLLECTION

2.1 Experimental rig

The experimental rig, depicted in Figure 1, is composed of a horizontal acrylic pipe with a horizontal length of 12.4 m and a 21 mm inner diameter. An elevated tank supplies the system and a pneumatically actuated ball valve with an internal diameter of 20 mm controls the pipe filling start at the upstream end of the pipe. A high point is installed at 5.30 m from the upstream valve, having a triangular shaped with a 45° slope with a rising pipe and a downwards pipe in the direction to the downstream end. The height of the high point is 0.30 m. An acrylic plate with a drilled orifice with a diameter d = 4.5 mm was installed to replicate a contraction in the flow cross section at the downstream end of the pipe, just before its discharge into the atmosphere.

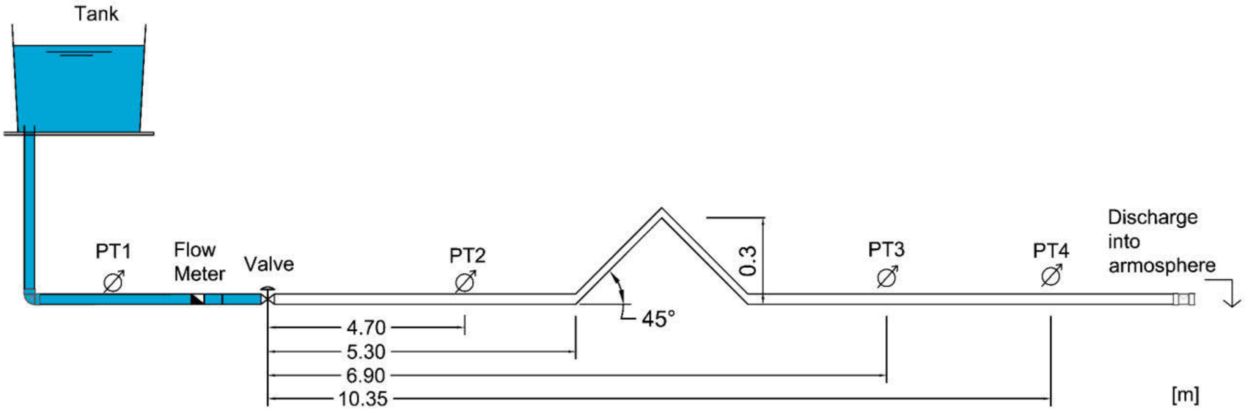


Figure 1. Experimental rig layout

Pressure measurements were carried out using Siemens SITRANS P pressure transducers series Z with a maximum measuring range of 0-2.5 m, a full scale accuracy of 0.5% and a time response lower than 0.1 s. Four pressure transducers were installed in the pipe: (i) the first at the upstream end of the valve to control the tank head and used to determine the water tank level as a boundary condition (PT1); (ii) the second (PT2) and the third (PT3) located before and after the system high point at a distance from the upstream valve of 4.7 and 6.9 m, respectively, and (iii) the fourth (PT4) at 10.35 m from the upstream valve. Pressure measurements were acquired at a 1 kHz frequency. Final steady state flow rate measurements were carried out by an Dynasonic ultrasonic flow meter with a full range accuracy of 1%. Video recording were carried out using a GoPro 7 with a resolution of 2074 x 1520 pxs and a frame rate of 24 fps.

2.2 Data collection

A number of tests was carried out to estimate the entrapped air pockets volume in the high point of the pipe system. Changing the water tank level and installing an orifice at the downstream end allowed obtaining different filling and final steady state flow rates which resulted in different air pocket sizes. The initial water tank levels, H_{ini}, were 0.5 and 1.5 m and the two different orifice sizes were tested: one without any orifice and one with the 4.5 mm orifice, corresponding to 10% of the pipe cross sectional area.

Tests are carried out by starting with the pipe completely empty and the orifice to be tested at the downstream end of the pipe. The upstream ball valve is opened and water coming from the tank starts filling the pipe. Air is released from the pipe as the water front wave advances. Once the water front wave reaches the downstream end, a transient is generated since the orifice acts a blockage in the pipe. Once the final steady state is reached (after the transient dissipates), the flow rate is measured. When the test is completed, the pipe is flushed with compressed air to ensure the pipe is emptied for the next test.

3 EXPERIMENTAL RESULTS

Pressure head, final steady state flow rates and measured air pocket volumes are used to analyse the influence of boundary conditions in air pocket formation. Piezometric head, H, and resulting steady state air pocket volume are shown in Figure 2 for the H_{ini} = 1.5 m and for the two orifice configurations. Video recordings are analysed to obtain the air pocket volumes, V_{air} at the final steady state flow. Steady state flow rates and air pocket volumes are presented in Table 1. The higher the flow rate is, the lower the air pocket volume also is, since high flows can drag the air pockets to downstream, as observed by Pothof and Clemens [4].

The measured pressure signal for the downstream configuration with d = 4.5 mm is presented in Figure 2a. A small air pressurization is observed, once the valve opens at time t = 0 s. The air pressurization is equal in all transducers, since the water that enters the system is small comparatively to the pipe volume. The water front wave progresses and reaches each transducer when the pressure signal detaches from the air pressure front the start at t = 3.8, 5.4 and 7.0 s for PT2, PT3 and PT4, respectively. PT2 shows a much higher increase in piezometric head when compared to the other transducers, due to the elevation of the high point. When the wave arrives at the downstream end, a pressure variation is observed since the orifice behaves like a blockage and has an inherent local head loss. The steady state piezometric head is progressively lower in the transducers due to head losses. Figure 2b shows the resulting air pocket and there is a clear air pocket formed at the high point.

The pressure signal for the downstream configuration with d = 21 mm in Fig 2c does not show any air pressurization since air can be freely released at the downstream end. The water front wave arrives at PT2 at t = 3.7s and the pressure signal shows the same behaviour as in d = 4.5 mm due to the high point. The wave arrives at PT3 at t = 5.2 s, sooner than in the previous configuration since the air pressurization slightly decreases the filling flow rate, and at PT4 at t = 6.4 s. No pressure variation is observed for this configuration and the piezometric head differences between transducers is much higher due to the higher flow rate presented in Table 1. Figure 2d shows the resulting air pocket for this orifice configuration. The air pocket is considerably smaller and does not create a flow similar to free surface flow on the tipping point of the high point. Even though this air pocket is smaller than the first one, higher head losses are observed due to the higher flow rate as well. Some variability is observed in the air pocket volumes from Table 1 for first and second test tests for the same initial conditions and for d = 4.5 mm.

Table 1. Steady state flow rate and air pocket volume for each tank water level and downstream orifice combination

d (mm)		H_{ini} (m)			
		0.5		1.5	
		First test	Second test	First test	Second test
4.5	Q (l/h)	118	90	204	216
	V_{air} (mm3)	2246	18181	472	326
21	Q (l/h)	672	666	1263	1232
	V_{air} (mm3)	0	0	0	0

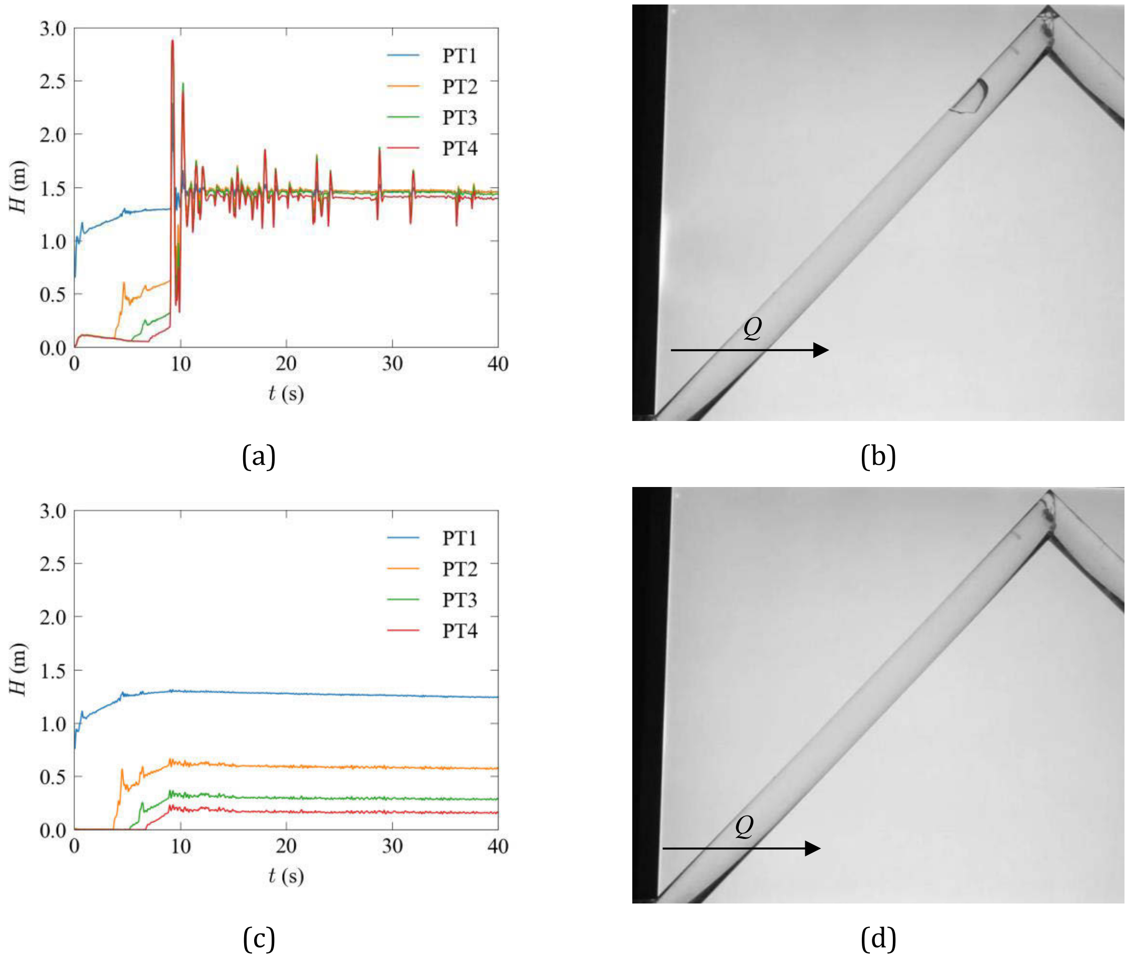


Figure 2. Pressure signal for each orifice diameter for H_{ini} *= 1.5 m and respective air pocket for the orifice configuration of: a, b) d = 4.5 mm and c, d) d = 21 mm*

To assess the air pocket volume variability for repeatability purposes, a set of 10 tests was carried out for the configuration H_{ini} = 1.5 m and d = 4.5 mm under the same initial conditions. Figure 3 shows the air pocket volume as a function of the final steady state flow rate. A large variability of air pocket volumes is observed with an average value of 946 mm^3 and a sampling standard deviation of 460 mm^3. However, despite the air pocket volume variability, a consistent flow behaviour is observed in all tests in video recordings: air pockets are always formed at the downwards slopping pipe, slightly after the high point and the flow rate is no capable of clearing the air pocket for this pipe slope. Having this air pocket volume variability, this volume cannot be precisely predicted for a given boundary conditions, but the air pocket location and flow behaviour is expected be similar. Also, air volumes are estimated based on video records of the tests, thus, there is always some uncertainty in their estimation, though not of the order of magnitude of the observed volume variation.

BY NC SA

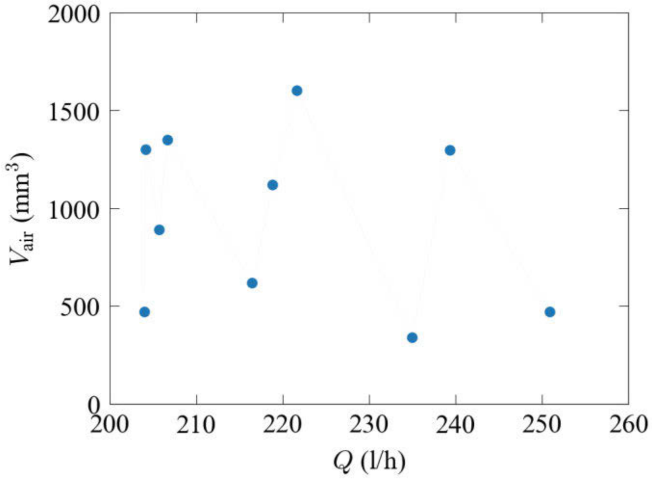


Figure 3. Repeatability analysis of entrapped air pocket volume for H_{ini}=1.5 m and d=4.5 mm

4 SWMM MODEL

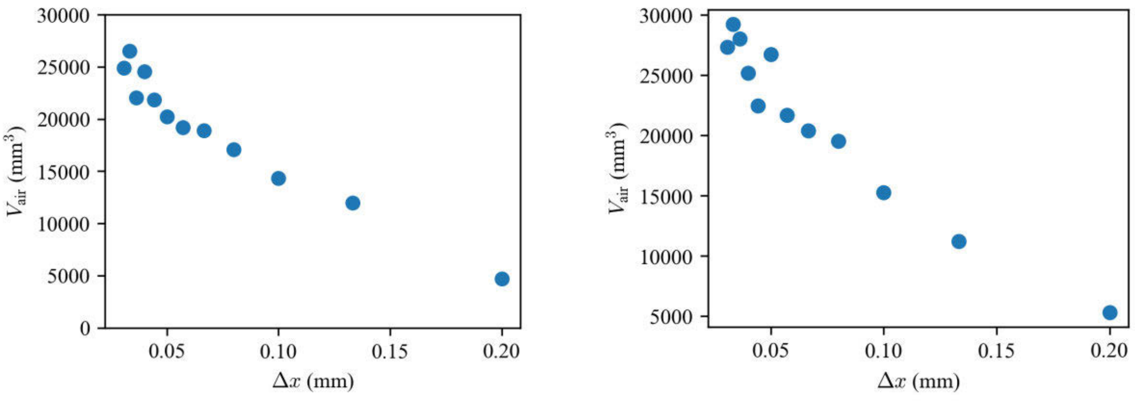
The existing SWMM software (v5.1.015) is used to estimate the entrapped air pocket volumes. The model consists of a storage tank at the upstream end, a single pipe simulated by means of pipe sections with a spatial discretization of Δx = 0.2 m (to improve dynamic event's results [17]), an orifice at the downstream end of the pipe and a free outfall to simulate the discharge into the atmosphere. Pipe sections from the spatial discretization have a diameter of 0.021 m, are connected to manholes with surcharge depths equal to 100 m so that no node gets ponded [11]. Minimum surface area of the manholes is set to 1x10^{-5} mm^2 to ensure no storage occurs in the manholes. EXTRAN and SLOT surcharge methods included in SWMM are tested. An additional function was added to SWMM source code to track and quantify air pockets' location and volume.

A sensitivity analysis of the spatial discretization was carried out to assess its influence on the simulated air pocket volume in SWMM. For that, the configuration of H_{ini} = 0.5 m and d = 4.5 mm was simulated in SWMM. Simulations for progressively smaller space steps, Δx, in the high point region (0.3 m upwards and downwards) were carried out to determine the value that air pocket volume would converge to, being the remaining pipe with Δx =0.3 m. The spatial steps and the corresponding time steps to ensure stable numerical results are presented in Table 2.

Table 2. Tested spatial discretization and routing time step to assess air pocket volumes

Δx (mm)	200	130	100	80	66	57	50	44	40	36	33	31	29	26
Δt (ms)	44	29	22	17	14	12	11	9	8	8	7	6	6	5

Numerical results obtained for the air pocket volume for these space steps are presented in Figure 4 for the two surcharge methods. EXTRAN surcharge method results show an increasing air pocket volume until Δx = 0.05 m, reaching a maximum value of air volume around V_{air} = 25,000 mm^3 (Figure 4a). SLOT surcharge method results are slightly different, varying from V_{air} = 20,000 and 30,000 mm^3 for space step lower than 0.05 m, but still around the results range for EXTRAN method (Figure 4b). It should be highlighted that the space step that started providing convergent air pocket volume values is four times lower than the recommended by Pachaly *et al.* [17] (Δx = 0.2 m) to better describe pipe filling events.



(a) (b)

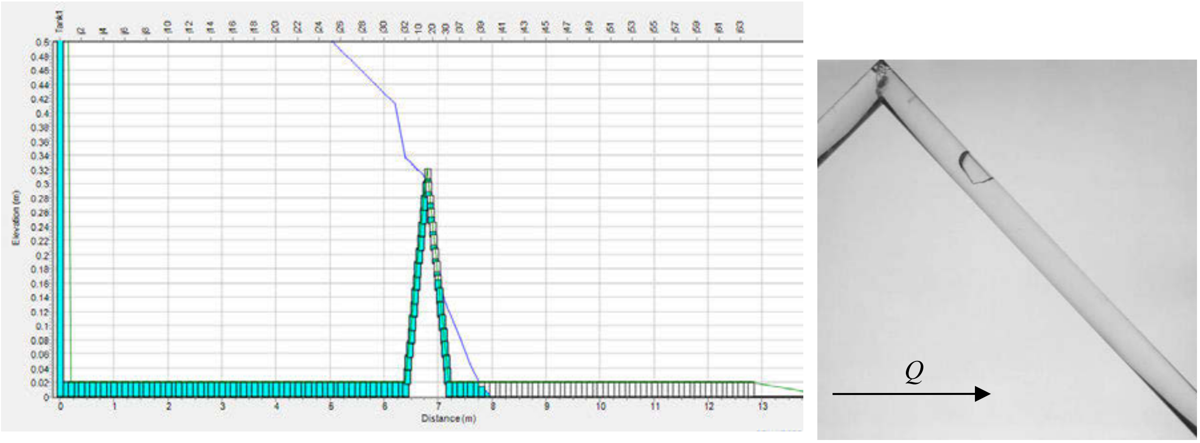
Figure 4. Entrapped air pocket volume for H_{ini}=0.5 m and d = 4.5 mm using: a) EXTRAN and b) SLOT surcharge methods

Tested configurations were simulated in SWMM using a spatial discretization of 0.026 mm, almost ten times lower than the recommended. Table 3 shows the comparison between the SWMM results using EXTRAN surcharge method and the experimental average air pocket volume. Numerical results for d = 4.5 mm do not accurately reproduce experimental data.

Table 3. Experimental vs numerical air pocket volumes for the tested conditions

d (mm)		H_{ini} (m)	
		0.5	1.5
4.5	Experimental V_{air} (mm^3)	10213	399
	Numerical V_{air} (mm^3)	25000	19400
21	Experimental V_{air} (mm^3)	0	0
	Numerical V_{air} (mm^3)	0	0

However, differences between video recordings and SWMM filling process might explain the difference between air pocket volumes. Figure 5 shows the numerical results from SWMM and an image from the video recording for H_{ini} = 1.5 m and d = 4.5 mm. Figure 5a) presents a snapshot of the pipe filling in SWMM graphic user interface and Figure 5b) shows the air pocket in steady state flow for the aforementioned tested conditions. As observed in Figure 5a), the entrapped air pocket in SWMM is created because the pipe at immediately downstream the high point becomes pressurized. A clear water flow separation occurs between the initial air volume and dowsnteram air volume, leading to the creation of an entrapped air pocket. Video recordings (snapshot in Figure 5b) show a different behaviour: the water filling wave is relatively perpendicular to the pipe cross section and the air pocket only reaches the high point due to air movement when a steady state is reached due to the pressure variations. Thus, the air pocket location is accurately represented in SWMM, but the calculated air volume is much higher than the observed in the experimental tests. Lower flow rates should be tested and an air model should be incorporated in SWMM in order to better describe this filling process and to better quantify the air volume entrapped in the pipe.



(a) (b)

Figure 5. Comparison of air pocket numerically calculated and observed for the test H_{ini}=1.5 m and d=4.5 mm from: a) SWMM pipe filling graphic user interface on the pipe filling process, where blue line is the hydraulic grade line and high point can be observed and b) video recordings

5 CONCLUSIONS

Air pockets can compromise the safety of pipe systems upon the occurrence of a transient event. Intermittent water supply systems have a strong air dynamics due to pipe filling and emptying processes. No water supply hydraulic solver is able to accurately describe the phenomena in IWS systems, since these models assume the pipes are already pressurized. SWMM has been proposed as an alternative, since it solves the open-channels equations allowing to simulate pipe filling and emptying stages. However, one of SWMM's main disadvantages is that air pocket dynamic model is not a feature incorporated in the SWMM solver.

This study concludes that, whilst SWMM seems to be able to accurately predict the air pocket location, it is not able to reproduce the air pocket volumes in high sloped pipes. An air model should be incorporated in the SWMM to better estimate the air pocket pressures and volumes. A spatial discretization four times smaller than the recommended in literature was shown to be required to obtain convergent air pocket volumes, but this still requires further research.

REFERENCES

[1] B. Charalambous and C. Laspidou Dealing with the Complex Interrelation of Intermittent Supply and Water Losses, 1st Edition, IWA Publishing, 2017.

[2] L. Ramezani, B. Karney and A. Malekpour, "Encouraging Effective Air Management in Water Pipelines: A Critical Review", Journal of Water Resources Planning and Management, vol. 142, no. 12. 2016, pp. 04016055.

[3] S. Christodoulou and A. Agathokleous, "A study on the effects of intermittent water supply on the vulnerability of urban water distribution networks", Water Science and Technology: Water Supply, vol. 12, no. 4. 2012, pp. 523-530.

[4] I. W. M. Pothof and F. H. L. R. Clemens, "Experimental study of air–water flow in downward sloping pipes", International Journal of Multiphase Flow, vol. 37, no. 3. 2011, pp. 278-292.

[5] I. Pothof and F. Clemens, "On elongated air pockets in downward sloping pipes", Journal of Hydraulic Research, vol. 48, no. 4. 2010, pp. 499-503.

[6] S. Mohapatra, A. Sargaonkar and P. K. Labhasetwar, "Distribution Network Assessment using EPANET for Intermittent and Continuous Water Supply", Water Resources Management, vol. 28, no. 11. 2014, pp. 3745-3759.

[7] Ó. E. Coronado-Hernández, M. Besharat, V. S. Fuertes-Miquel and H. M. Ramos, "Effect of a Commercial Air Valve on the Rapid Filling of a Single Pipeline: a Numerical and Experimental Analysis", Water, vol. 11, no. 9. 2019, pp.

[8] C. M. Fontanazza, V. Notaro, V. Puleo, P. Nicolosi and G. Freni, "Contaminant Intrusion through Leaks in Water Distribution System: Experimental Analysis", Procedia Engineering, vol. 119, no. 2015, pp. 426-433.

[9] J. Parmakian Water Hammer Analysis, 1963.

[10] M. Guizani, J. Vasconcelos, S. J. Wright and K. Maalel, "Investigation of Rapid Filling of Empty Pipes", Journal of Water Management Modeling, vol. R225-20,, no. 2006, pp. 463-482.

[11] A. Campisano, A. Gullotta and C. Modica, "Using EPA-SWMM to simulate intermittent water distribution systems", Urban Water Journal, vol. 15, no. 10. 2019, pp. 925-933.

[12] M. H. Chaudhry Open-Channel Flow, 2n Edition, Springer, 2008.

[13] M. H. Chaudhry Applied Hydraulic Transients, 3rd Ed., Springer-Verlag New York, 2014.

[14] A. Malekpour and B. W. Karney, "Profile-Induced Column Separation and Rejoining during Rapid Pipeline Filling", Journal of Hydraulic Engineering, vol. 140, no. 11. 2014, pp. 1-12.

[15] C. L. Lubbers and F. Clemens, "Breakdown of air pockets in downwardly inclined sewerage pressure mains", Water Science Technology, vol. 54, no. 11-12. 2006, pp. 233-40.

[16] R. L. Pachaly, J. G. Vasconcelos and D. G. Allasia, "Surge predictions in a large stormwater tunnel system using SWMM", Urban Water Journal, vol. 18, no. 8. 2021, pp. 577-584.

[17] R. L. Pachaly, J. G. Vasconcelos, D. G. Allasia, R. Tassi and J. P. P. Bocchi, "Comparing SWMM 5.1 Calculation Alternatives to Represent Unsteady Stormwater Sewer Flows", Journal of Hydraulic Engineering, vol. 146, no. 7. 2020, pp. 1-16.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14112

A DECISION-MAKING APPROACH TO ASSESS AND PRIORITISE INTERVENTION SOLUTIONS IN WATER DISTRIBUTION SYSTEMS

Marta Cabral[1], Dália Loureiro[2] and Dídia Covas[3]

[1]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon (Portugal)
[2]Urban Water Unit, National Civil Engineering Laboratory, Lisbon (Portugal)
[3]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon (Portugal)

[1] *marta.f.cabral@tecnico.ulisboa.pt,* [2] *dloureiro@lnec.pt,* [3] *didia.covas@tecnico.ulisboa.pt*

Abstract

The present paper aims at proposing and demonstrating the application of a decision-making approach for the identification, assessment, comparison and prioritisation of different intervention solutions. Infrastructural, financial, performance and economic perspectives are considered in the study. The proposed approach is composed of three main modules: 1) Analysis planning and database construction; 2) Infrastructure, asset or component diagnosis and prioritisation; and 3) Study of intervention solutions for priority assets or components. Three assessment levels – macro, meso and micro – are proposed, and the decision-making approach is adapted to each level. A water distribution system located in Portugal is used to demonstrate the proposed approach. This case study comprises five water subsystems, including different assets, such as water storage tanks, pumping and booster stations and water distribution pipes. Five intervention solutions are defined by identifying the main problems associated with the priority subsystem and the respective causes. The intervention solutions are compared considering the financial metrics and performance indicators, such as standardised energy consumption, energy in excess per unit of the authorised consumption, infrastructure value index and non-revenue water. New metrics regarding the assets' physical condition are also incorporated in the assessment system. Results have shown the influence of considering different assessment criteria and performance indicators in the solutions' prioritisation, highlighting that the solution with the lowest capital cost does not always correspond to the solution with the highest overall performance.

Keywords
Decision-making approach, Intervention solutions, Prioritisation process, Water distribution systems.

1 INTRODUCTION

Decisions on the rehabilitation of urban water systems are traditionally made considering only economic indicators or subjective analysis perspectives. Nowadays, it is well recognised that the decision-making process based on a single perspective is limited and inappropriate, as other essential aspects, such as the technical performance, the physical condition, water quality issues and the risk of failure, are not taken into account [1].

Decision-making approaches aim to identify and prioritise all potential solutions with consideration of different perspectives (e.g., performance, condition, risk assessment, financial analysis) and to establish trade-offs [2]. In these approaches, intervention solutions are identified and compared with the *status quo* situation, corresponding to maintaining the current O&M practices and not making any investment.

Solutions can be classified as infrastructural, O&M or non-infrastructural [3]. Infrastructural solutions include investment interventions, such as rehabilitation works, as well as any expansion interventions. O&M solutions are considered due to deficiencies or potential improvements of

O&M or new operating and maintenance needs associated with the implementation of infrastructural solutions. O&M interventions can be divided into localised (e.g., flow-meters installation), permanent (e.g., changes in the operating mode of pump group) or systemic (e.g., inspection and repair of storage tanks). Examples of non-infrastructural solutions are pressure management or the implementation of efficient water-use measures.

Different techniques can be used to improve the decision-making process and the respective outcomes, such as multicriteria analysis (e.g., [1], [4], [5] and [6]), cost-benefit analysis (e.g., [7], [8] and [9]) and life-cycle cost analysis (e.g., [10], [11], [12] and [13]). Although there are several decision-making approaches, most are too complex to be used by water utilities hindering their use and the interpretation of results. Therefore, the use of simplified approaches, considering only one point of view, or of too many unjustified assumptions can make the purpose of implementing these approaches unfeasible. Many decisions are not sufficiently explored and can cause a premature end of asset life concerning their physical condition [14]. Besides, the selection of solutions is inherent to the rehabilitation or maintenance strategy of the water utility. Attention is now moving away from reactive strategies, which involve none or minor long-term planning, towards pro-active approaches based on predictive analyses to provide a sustainable service in the long-term [15].

The present paper aims at proposing and demonstrating the application of a decision-making methodology that allows the identification, assessment, comparison and prioritisation of different intervention solutions considering infrastructural, financial, performance and economic perspectives. Firstly, a description of the proposed decision-making approach is presented in Section 2. The case study is presented in Section 3 and the application of the proposed approach to this case study is presented in Section 4. Finally, conclusions are drawn and further research is presented in section 6.

2 DECISION-MAKING APPROACH

2.1 General approach

The proposed decision-making approach is composed of three main modules (Figure 1): 1) Analysis planning and database construction; 2) Infrastructure, asset or component diagnosis and prioritisation; and 3) Study of intervention solutions for priority assets or components. The aim of this approach is to identify, assess, compare and prioritise intervention solutions in urban water infrastructures considering infrastructural, financial, performance and economic perspectives. This approach is aligned with the infrastructure asset management approaches proposed by the ISO 5000x standards and the ISO 24512 [16].

The main innovative contributions from this approach compared to existing approaches are the proposal of three assessment levels – macro, meso and micro – and the adaptation of the decision-making approach to each level. Depending on the chosen assessment level, necessary data may vary in terms of detail, models in terms of complexity and results in terms of applicability. This approach can be applied by utilities with different levels of maturity and of infrastructural and operational knowledge and for different scopes and purposes of the analysis. A detailed explanation of each module will be presented, highlighting the differences between the three assessment levels.

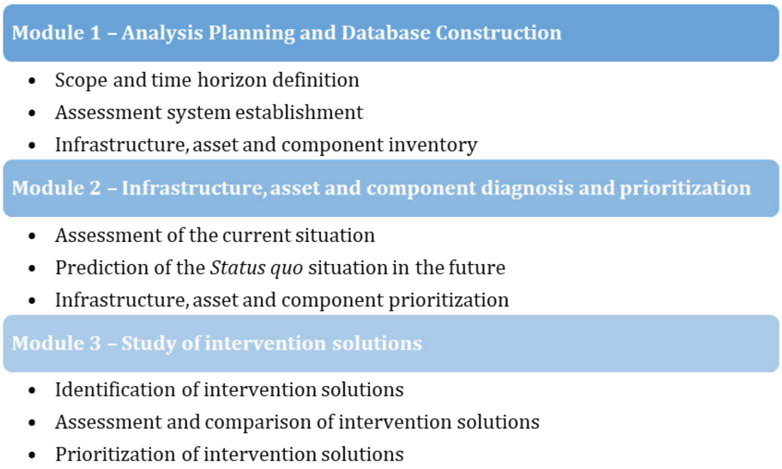

Module 1 – Analysis Planning and Database Construction
Scope and time horizon definition
Assessment system establishment
Infrastructure, asset and component inventory
Module 2 – Infrastructure, asset and component diagnosis and prioritization
Assessment of the current situation
Prediction of the *Status quo* situation in the future
Infrastructure, asset and component prioritization
Module 3 – Study of intervention solutions
Identification of intervention solutions
Assessment and comparison of intervention solutions
Prioritization of intervention solutions

Figure 1. Decision-making approach to study intervention solutions in urban water infrastructures.

2.2 Analysis planning and database construction

Module 1 aims at defining the scope and time horizon of the analysis, to establish the assessment system and to construct a database per infrastructure, asset or component. The scope of the analysis is related to the assessment level. The macro assessment level aims at planning investment and maintenance tactics through the analysis of different scenarios and establishing policies and future directions at the infrastructure level. The meso assessment level corresponds to an intermediate analysis at the subsystem or asset level. The micro assessment level aims at planning and comparing intervention solutions at the asset and component level. These three assessment levels are carried out at the tactical planning level for a time horizon of 3-5 years; however, the period of analysis, in which the impacts of each solution are evaluated, should be larger than time horizon (e.g., 20-30 years). The period of analysis can significantly influence the results, thus, it should be defined taking into consideration different factors, such as the assets' service lives, the utility concession period and the assessment level.

The definition of the assessment system includes the establishment of objectives, criteria, performance indicators and reference values. The ISO 24512 [16] establishes a set of main objectives for drinking water utilities and all objectives must be measurable to ensure their monitoring over time by defining assessment criteria, performance indicators and reference values. For each assessment criteria, different performance indicators (PI) are defined according to the utility objectives and viewpoints. Some PI can be more disaggregated and applied to the entire system and subsystems to prioritise areas of analysis. Reference values allow to classify the obtained PI as good (represented by the green colour), average (represented by the yellow colour) or unsatisfactory (represented by the red colour).

The database construction includes the collection and processing of asset data for each urban water infrastructure. Three data categories are defined: technical, operational and economic. Technical data are associated with the physical characteristics of each asset, operational data with the asset operating mode and condition and economic data with CAPEX (capital expenditures) and OPEX (operating expenditures). The information detail may vary depending on the chosen assessment level. The macro assessment level requires basic information to study different investment and maintenance solutions, while the micro assessment level needs basic and complementary information to assess and select intervention solutions. Meso assessment level corresponds to an intermediate level of data needed. At the macro and meso levels, some basic information can be qualitative, such as the asset condition (i.e., the condition can be defined using a qualitative scale: excellent, good, reasonable, poor and unsatisfactory condition). However, at the micro assessment level, the information should be more robust and accurate and, therefore, it is preferable to be quantitative.

2.3 Infrastructure, asset or component diagnosis and prioritisation

Module 2 aims at assessing the current situation and predicting the time evolution of the *status quo* situation (i.e., not considering rehabilitation interventions and maintaining the current O&M practices) and prioritising infrastructures, assets or components in terms of intervention needs. The assessment of the current situation is carried out by calculating the defined PI in the previous module and is used to identify the main operational and physical problems. The diagnosis and the prioritisation processes can be carried out in the infrastructure, asset or component depending on the chosen assessment level, respectively for macro, meso or micro.

Condition assessment is one of the most important steps in the assessment of the current situation however, the complexity of the process and the robustness of the results can vary with the assessment level. At the macro assessment level, the condition assessment is carried out using a value-based approach, corresponding to the calculation of the Infrastructure Value Index (IVI) not requiring inspection. This is one of the most appropriate approaches for the macro assessment level since it is easy and straightforward to calculate and less resource-consuming. At the meso assessment level, the condition assessment is carried out by using a direct rating-based approach, consisting of visually inspecting each component, evaluating it against the rating criteria and selecting the appropriate rating as: excellent, good, reasonable, poor and unsatisfactory condition. At the micro assessment level, the condition assessment is carried out using a distress-based approach to assess the physical condition of urban water assets through visual inspection for the identification and classification of anomalies. The direct rating-based approach and the distress-based approach were proposed by Cabral et al. [17]. A prediction of the *status quo* situation in the future is also developed to verify the evolution of the current situation and to predict future problems not yet identified considering future changes in terms of consumption demand, new regulatory requirements and asset deterioration.

After the assessment of the current and future situations, infrastructures, assets or components are prioritised for intervention, considering the assessment results of PI and their criticality. The asset criticality is associated with the failure consequences, that is assets with high failure consequences are considered critical. Five prioritisation levels can be defined: extreme low, low, moderate, high and critical. If more than one PI is used in the prioritisation process (for example, condition assessment and supplied energy index), weighting factors for each PI may be used to obtain an overall assessment.

2.4 Study of intervention solutions

Module 3 aims at identifying, assessing, comparing and prioritising intervention solutions considering infrastructural, financial, performance and economic perspectives to solve the problems identified in the previous module.

At the macro assessment level, a set of long-term rehabilitation solutions at the system or subsystem level is identified considering the future uncertainties. Examples of these solutions are: the asset replacement at the end of service life; the assets' replacement for maintaining a constant IVI at 0.50; and the assets' rehabilitation at a 5% rate (in terms of costs). At the micro assessment level, solutions can be divided into infrastructural, O&M or non-infrastructural. Infrastructural solutions can be replacing equipment or replacing/renewing civil work components. The modern engineering equivalent replacement asset (MEERA) approach should be used to assess the intervention solutions, in which technologically similar assets are selected. This approach should be applied whenever the assets are no longer available in the market (e.g., asbestos cement pipes have fallen into disuse) or the assets have significant changes in technology (e.g., pump groups with higher efficiency). Besides, different technologies of replacement and renewal should be considered since it may have different costs and benefits. At the meso assessment level, it is possible to develop a detailed assessment of long-term rehabilitation solutions or a simplified assessment of intervention solutions.

The assessment of each solution presupposes an analysis of costs and benefits during a time horizon. Infrastructural, economic, performance and financial perspectives are considered to assess the benefits of each solution using different PI and metrics. Regarding financial metrics, the following metrics are calculated for each analysed solution: net present value (NPV), internal rate of return (IRR), payback period (PBP) and the profitability index.

The comparison of solutions is carried out by comparing each one with the *status quo* situation in terms of the different PI. The comparison of solutions allows their prioritisation for each PI or studied metric. If more than one PI is used, an overall prioritisation process is carried out to select the solution that maximises the investment recovery. This prioritisation process is based on the methodology proposed by Alegre and Coelho [18], in which a weighting of each PI to express its relative importance in the assessment system is assigned. Besides, a normalisation process is necessary since each PI is expressed in different units. A continuous scale varying from 0 to 3 is used and divided into three levels:

- [2; 3]: good assessment, represented by the green colour;
- [1; 2[: average assessment, represented by the yellow colour;
- [0; 1[: unsatisfactory assessment, represented by the red colour.

A ranking of solutions based on the overall assessment is carried out allowing to select the solution with the best performance assessment considering the utility budget.

3 CASE-STUDY

The decision-making approach is applied to a water distribution system located in a touristic area in the Algarve region of Portugal at the micro assessment level, corresponding to the most detailed and complex assessment level. This water distribution system is characterised by a high seasonal water consumption variation, with significantly higher consumption during the summer period than in the winter period (i.e., four times higher). The water distribution system is divided into five water subsystems (1 – 5), including different assets, such as water storage tanks, pumping and booster stations and water distribution pipes. The general characteristics of the subsystems are presented in Table 1. The characteristics of water distribution pipes of Subsystem 1 were not provided by the utility, thus, these will not be considered in the subsequent analysis.

Table 1. General characteristics of the five subsystems of water distribution case study.

Subsystem	Asset inventory
1	1 water storage tank: Capacity = 2 250 m^3 1 booster station: Total hydraulic power = 4.4 kW Characteristics of the network pipes: unknown
2	2 water storage tanks: Capacity = 800 m^3 and 125 m^3 3 booster station: Total hydraulic power = 0.3 kW, 1.8 kW and 0.3 kW 49 203 km of pipes in AC, PVC, DI; HDPE: DN = [60; 350] mm
3	1 booster station: Total hydraulic power = 1.3 kW 2 476 km of pipes in AC and PVC: DN = [80; 110] mm
4	1 booster station: Total hydraulic power = 1.6 kW 3 891 km of pipes in AC and PVC: DN = [60; 110] mm
5	2 water storage tanks: Capacity = 10 200 m^3 and 500 m^3 1 pumping station: Total hydraulic power = 11.6 kW 1 booster station: Total hydraulic power = 1.2 kW 38 169 km of pipes in AC, PVC, DI; HDPE: DN = [60; 500] mm

Notes: AC – Asbestos Cement; PVC – Polyvinyl Chloride; DI – Ductile Iron.

BY NC SA

4 RESULTS

4.1 Analysis planning and database construction

The application of the decision-making approach aims at analysing and comparing different intervention solutions at the asset and component level considering a time horizon of 20 years. The asset diagnosis and prioritisation process and the study of solutions are developed by calculating different PI associated with the occurrence of failures, the economic and infrastructure sustainability and integrity and the energy use efficiency (Table 2).

Table 2. Assessment system, including criteria, performance metrics and respective reference values.

Criteria	Performance indicators (units)	Reference values
Occurrence of supply failures	Failures of service connections (no./(1000 service connections.year)) [19]	🟢 [0.0; 1.0] 🟡]1.0; 2.5] 🔴]2.5; +∞[
Economic and financial sustainability of the utility	Non-revenue water (%) [19]	🟢 [0; 20] 🟡]20; 30] 🔴]30; 100]
Infrastructure sustainability and integrity	Infrastructure value index* (IVI) (-) [20]	Long-term planning: 🟡]0.60; 1.0] 🟢 [0.40; 0.6] 🔴 [0.0; 0.40[Condition assessment: 🟢]0.60; 1.0] 🟡 [0.40; 0.6] 🔴 [0.0; 0.40[
	Residual life ratio* (RLR) (-) [20 [21]]	🟢]0.60; 1.0] 🟡 [0.40; 0.6] 🔴 [0.0; 0.40[
	Asset condition rating** (-) [17]	🟢 {5}, {4} 🟡 {3} 🔴 {2}, {1}
	Infrastructure average and maximum deterioration index*** (IDI) (-) [17]	🟢 [0; 40] 🟡]40; 60] 🔴]60; 100]
	Asset average and maximum deterioration index*** (ADI) (-) [17]	🟢 [0; 40] 🟡]40; 60] 🔴]60; 100]
	Component average and maximum deterioration index*** (CDI) (-) [17]	🟢 [0; 40] 🟡]40; 60] 🔴]60; 100]
Energy use efficiency	Standardised energy consumption (kWh/(m^3.100m)) [19]	🟢 [0.27; 0.40] (average efficient between 68 and 100%) 🟡]0.40; 0.54] (average efficient between 50 and 68%) 🔴]0.54; 5] (average efficient lower than 50%)
	Energy in excess per unit of the authorised consumption (kWh/m^3) [21]	🟢] 0; 0.15] 🟡]0.15; 0.30] 🔴]0.30; +∞[

Notes: *Calculated at the macro assessment level; **Calculated at the meso assessment level; ***Calculated at the micro assessment level.

The Component Deterioration Index (CDI), Asset Deterioration Index (ADI) and Infrastructure Deterioration Index (IDI) are new metrics of the assessment system that were proposed by Cabral et al. [17]. These metrics allow to obtain a more robust condition value of components, assets and infrastructures than the existing metrics that presuppose the use of reference service lives (e.g., IVI and RLR).

4.2 Infrastrcuture, asset and component diagnosis and prioritisation

The assessment system results for the current situation (reference year of 2018) are presented in Table 3. All subsystems have an opportunity to improve pump groups in terms of energy efficiency, since the standardised energy consumption presents an average (Subsystems 1, 4 and 5) or unsatisfactory performance (Subsystems 2 and 3). The PI of energy in excess per unit of the authorised consumption corroborates a potential improvement, especially in Subsystem 3, with unsatisfactory performance, and in Subsystems 4 and 5, where the indicator shows an average performance.

Table 3. Subsystem diagnosis using performance indicators defined in Table 2 (reference year of 2018).

Criteria	**Performance indicator**	**Subsystem**				
		1	**2**	**3**	**4**	**5**
Occurrence of supply failures	Failures [no./(1000 service connections.year)]	🔴 5.35	🟢 0.00	🟢 0.00	🟢 0.00	🟢 0.00
Economic and financial sustainability of the utility	Non-revenue water (%)	🟢 2.4	🟢 15.0	🟢 15.0	🟢 15.0	🟢 15.8
Infrastructure sustainability and integrity	IVI (-)	🟢 0.65	🔴 0.39	🔴 0.28	🔴 0.14	🔴 0.26
	Average IDI (-)	🟢 13	🟢 15	🟢 21	🟢 28	🟢 27
	Maximum IDI (-)	🟢 21	🟡 49	🟢 33	🟢 32	🟡 49
Energy use efficiency	Standardised energy consumption [kWh/(m^3.100m)]	🟡 0.49	🔴 0.85	🔴 1.35	🟡 0.53	🟡 0.42
	Energy in excess per unit of the authorised consumption (kWh/m^3)	🟢 0.15	🟢 0.11	🔴 0.41	🟡 0.21	🟡 0.16
Normalised global assessment (0-3; 0 corresponds to the lowest performance)		🟡 **1.93**	🟡 **1.84**	🟡 **1.67**	🟡 **1.70**	🟡 **1.83**
Ranking (1-5; 1 corresponds to the highest priority)		**5**	**4**	**1**	**2**	**3**

Subsystem prioritisation is carried out considering an equal weight (default value of 1) for each indicator, representing the relative importance of the assessment system's metric. Subsystem 3 is considered the highest priority with an overall evaluation of 1.67, representing the lowest value of the five studied subsystems. This subsystem is composed of one booster station and 2 476 km of pipes in Asbestos Cement (AC) and Polyvinyl Chloride (PVC) with nominal diameters varying from 80 and 110 mm. The booster station includes four pump groups installed in parallel and a variable speed driver with a rated flow rate of 16 m^3/h, rated head of 58.4 m and a total hydraulic power of 7.63 kW.

CC BY NC SA

The diagnosis for each pump group of the booster station is presented in Table 4. The pump groups present an unsatisfactory efficiency (24%) and lower than expected (provided by the manufacturer), which indicates a performance degradation. This efficiency was estimated by the ratio between the energy supplied and the billed electric energy. Thus, it represents the global efficiency of the pumping station, not being possible to distinguish the efficiency of each group. Moreover, according to the average pumped flow rate and rated flow rate ratio (Q/Q_R), pump groups are also operating, on average, away from their rated conditions (i.e., point of maximum efficiency). Values of RLR (i.e., ratio between the residual life and the service life) for each pump group show that groups have reached the end of their expected service lives, considering a service life of 20 years. However, the calculated average CDI (component deterioration index obtained through the identification and classification of anomalies during the assets' visual inspection) presents the same value of 21 for all pump groups. This means that the four pump groups are in good condition, according to the visual inspections, despite having a low power efficiency and also reached the end of their service life.

Table 4. Pump groups diagnosis (reference year of 2018).

Pump ID	Flow rate (l/s)	Head (m)	RLR (-)	CDI (-)	Pump efficiency* (%)	Q/Q_R** (%)
1	2.67	27	● 0	● 21	● 24	● 118
2	1.64	33	● 0	● 21		
3	1.25	33.4	● 0	● 21		
4	1.28	33.3	● 0	● 21		

Notes: *Pump efficiency: Good assessment [68%, 100%], Average assessment [50%, 68%], Unsatisfactory assessment [0%, 50%]. **Q/Q_R: Good assessment [90%, 105%], Average assessment [70%, 90%] and [105%, 120%], Unsatisfactory assessment [0%, 70%] and [120%, 150%].

The diagnosis for each distribution pipe of Subsystem 3 is presented in Table 5. Physical characteristics, including the material, nominal diameter and length of each pipe are presented, as well as the RLR, considering a service life of 40 years for the two pipe materials. The studied distribution pipes are reaching the end of their service life, representing a high investment in the short-term. The unit head losses for the winter and summer periods were obtained by simulation of the hydraulic models using EPANET. The first two water distribution pipes (with ID 1 and 2) present high unit head losses in the summer period due to their poor design (i.e., small diameter for summer operation). Note that only some distribution pipes are presented in the table to illustrate the diagnostic process

Table 5. Water distribution pipes diagnosis (reference year of 2018).

Pipe ID	Material	Nominal diameter (mm)	Length (m)	RLR (-)	Unit head losses (-)	
					Summer period	Winter period
1	PVC	110	107.40	● 0.13	6.44	0.24
2	AC	100	787.93	● 0.13	6.15	0.23
3	AC	80	116.40	● 0.13	1.76	0.14
4	PVC	80	168.77	● 0.13	0.09	0.00
…	…	…	…	…	…	…
19	AC	80	47.33	● 0.05	0.04	0.00
20	AC	80	33.51	● 0.05	0.00	0.00

Notes: AC – Asbestos Cement; PVC – Polyvinyl Chloride. RLR – Residual Life Ratio.

The prioritisation process aims at identifying the assets (i.e., water pipes) and components (i.e., pump groups) for intervention. Figure 2 depicts the prioritisation matrix, according to the assets and components criticality and the RLR. The component critically was defined considering the importance of each asset and component in the studied subsystem. A scale that varies between 1 (less important functional components) and 6 (very important functional components) was considered to assign the criticality of each asset and component. The four pump groups and the water pipes with the ID 1 and 2 were classified with a criticality 6, being essential components to the functioning of the subsystem. The remaining water pipes were classified with a criticality of 5.

The five points represent the set of assets/components with the same criticality and RLR. The four pump groups are represented by one point with a RLR of zero and a criticality of 6. The two water pipes with ID 1 and 2 are represented by one point with a RLR of 0.13 and a criticality of 6. Thus, these assets/components are considered critical priorities for intervention. The remaining water pipes are considered high priorities for intervention, due to the low values of residual life. Most pipes have their assessment overlapped in the figure.

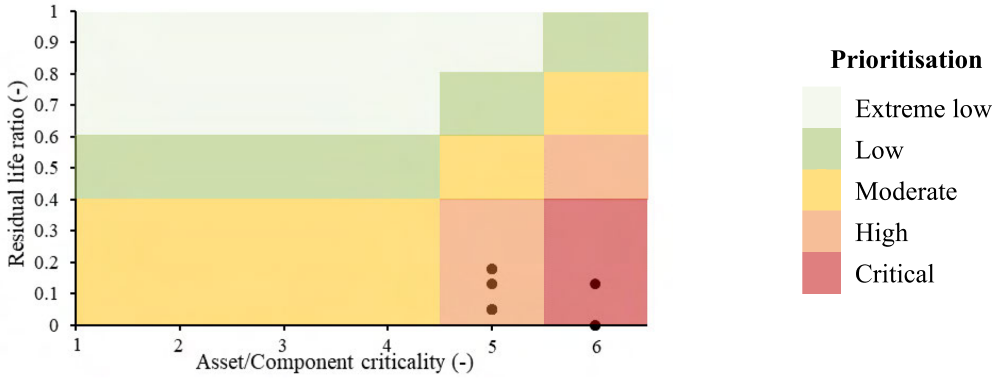


Figure 2 – Asset (water pipes) and component (pump groups) prioritisation for intervention (reference year of 2018).

4.3 Study of intervention solutions

The identification of intervention solutions is carried out through the identification of the main problems associated with the subsystem and the respective causes. The main problems were identified through the subsystem and component diagnosis and can be divided into three categories: pump group inefficiencies and water and head losses. Five intervention solutions are defined to attain the main problems identified:

- Solution 0 (S0): *Status quo*, not considering interventions and maintaining the current O&M practices. This solution is used only to compare with the other options to exemplify the benefit of each solution with the situation of non-intervention.
- Solution 1 (S1): Replacement of the four pump groups with the same rated characteristics (like-for-like replacement). An efficiency of 50% for the new pump groups is considered including the pump, the motor and the variable speed driver.
- Solution 2 (S2): Replacement of the four pump groups properly designed, considering the opportunity to improve the operating points of the pump groups. Hydraulic models provided by the water utility with consumption patterns are used to compare the average pressure in winter and summer periods with the minimum required pressure. A reduction of the operating head of the pump groups is considered, representing an average reduction of 13 m.

BY NC SA

- Solution 3 (S3): Replacement of 895.33 m of water pipes (corresponding to water pipes ID 1 and 2) properly designed. Both water pipes are replaced by PVC pipes since it corresponds to the most predominant material in this subsystem and the nominal diameter considered for both replaced pipes is 125 mm.
- Solution 4 (S4): a combination of Solutions 1 and 3, including the replacement of pump groups with the same rated characteristics and the replacement of water pipes with high head losses.
- Solution 5 (S5): a combination of Solutions 2 and 3, including the replacement of the four pump groups and the replacement of water pipes with high head losses with an adequate design. This solution does not represent a complete combination of Solutions 2 and 3, since the reduction of the head losses in the water pipes allows to reduce even more the operating head of the pump groups.

A set of assumptions were established to assess the defined intervention solutions related to water and energy price, service lives, maintenance costs, water losses and efficiency degradation. The intervention solutions are compared considering the following PI: standardised energy consumption, energy in excess per unit of the authorised consumption, infrastructure value index and non-revenue water.

The standardised energy consumption allows to assess and to compare the pumping energy efficiency for a single pump group or the whole pumping or booster station (Figure 3a). Solutions 0 and 3 present the same evolution during the time horizon of the analysis since in these solutions no pump groups are replaced and the pump groups' efficiency continues to degrade until reach the minimum limit of 10 % (in the year 2032). After that year, the efficiency of the pump groups remains constant, causing the same behaviour in the shaft input energy. Solutions 1 and 4 also present the same results for this PI, since the rated conditions of the replaced pump groups are the same in the two solutions. An outstanding improvement of the standardised energy consumption in these two solutions is caused by the new pump groups' efficiency contributing to the decreases of the shaft input energy. However, results still indicate an unsatisfactory assessment. Finally, Solutions 2 and 5 present the best results for this PI (representing an average assessment), although their operating points are slightly different.

The comparison of intervention solutions by the energy in excess per unit of the authorised consumption is presented in Figure3(b). The increase of the pump groups' efficiency from Solution 0 to 1 and to 2 contributes to the improvement of this PI in these solutions, changing from an unsatisfactory assessment (S0) to an average assessment (S1) and, finally, to a good assessment (S2). A slight difference between Solutions 0 and 3, 1 and 4 and 2 and 5 is verified, mainly, due to the replacement of the two pipes (in Solutions 3, 4 and 5), which allows to reduce the water losses and, consequently, the pumped volume. Thus, the shaft input energy is lower in these solutions. Furthermore, the solutions that include the replacement of the water pipes consider lower total annual water losses. Therefore, the energy in excess in these solutions does not show such a pronounced increase.

The comparison of intervention solutions by the infrastructure value index (IVI) is presented in Figure3(c). Solution 0 presents the lowest IVI since no assets are replaced in this solution, achieving a null IVI in 2026 and compromising the correct functioning of the subsystem. This solution represents the IVI of the water pipes since the pump groups have already reached the end of service life in 2018. Solutions 1 and 2 present similar IVI since both solutions include the replacement of the four pump groups with the same service life and the replacement costs are very similar (being slightly lower for Solution 2). However, these two solutions still present an unsatisfactory IVI, due to the ageing water pipes that reach the end of service in 2026. Better

CC BY NC SA

results are achieved for the remaining solutions, with the replacement of two water pipes (S3) and the replacement of two water pipes and pump groups (S4 and S5). The greater the number of assets replaced is, the higher the obtained IVI becomes. In the three solutions, a null IVI is not reached during the time horizon of the analysis. Solutions 4 and 5 achieve the same IVI as well as Solution 3 in the last year of the analysis, due to the end of service life of replaced pump groups in the former solutions.

In the case of non-revenue water, the solutions are divided into two different results, depending on whether the solutions include the replacement of the two water pipes (Figure 3d). The non-revenue water is equivalent to the water losses, assuming that all consumption is billed. Results obtained for this PI are not as significant as in the previous ones, since water losses are not a problem in this subsystem.

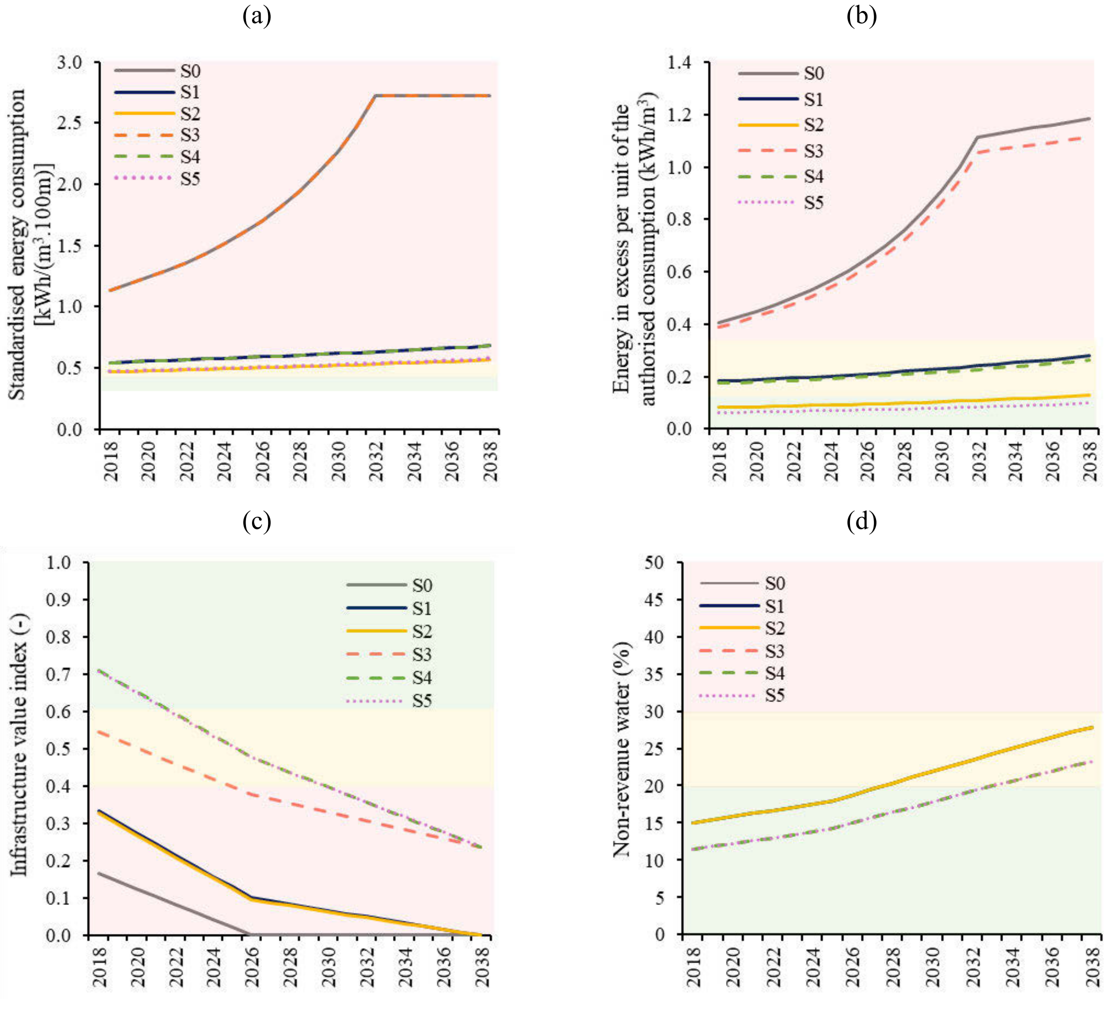


Figure 3 – Impact of intervention solutions on performance indicators between 2018 and 2038: (a) standardised energy consumption; (b) energy in excess per unit of the authorised consumption; (c) infrastructure value index; and (d) non-revenue water.

The comparison of intervention solutions by the infrastructure value index (IVI) is presented in Figure3(c). Solution 0 presents the lowest IVI since no assets are replaced in this solution, achieving a null IVI in 2026 and compromising the correct functioning of the subsystem. This solution represents the IVI of the water pipes since the pump groups have already reached the end of service life in 2018. Solutions 1 and 2 present similar IVI since both solutions include the replacement of the four pump groups with the same service life and the replacement costs are very similar (being slightly lower for Solution 2). However, these two solutions still present an unsatisfactory IVI, due to the ageing water pipes that reach the end of service in 2026. Better results are achieved for the remaining solutions, with the replacement of two water pipes (S3) and the replacement of two water pipes and pump groups (S4 and S5). The greater the number of assets replaced is, the higher the obtained IVI becomes. In the three solutions, a null IVI is not reached during the time horizon of the analysis. Solutions 4 and 5 achieve the same IVI as well as Solution 3 in the last year of the analysis, due to the end of service life of replaced pump groups in the former solutions.

In the case of non-revenue water, the solutions are divided into two different results, depending on whether the solutions include the replacement of the two water pipes (Figure 3d). The non-revenue water is equivalent to the water losses, assuming that all consumption is billed. Results obtained for this PI are not as significant as in the previous ones, since water losses are not a problem in this subsystem.

A comparison of the intervention solutions considering the cumulative cash flows is carried out and presented in Figure 4. All the studied solutions present significant cumulative cash flows, representing good potential solutions to implement. Solution 3 presents the highest payback period and the lowest cumulative cash flow, since the replacement of water pipes reduces water losses (and non-revenue water). However, this subsystem already has a good assessment for this PI.

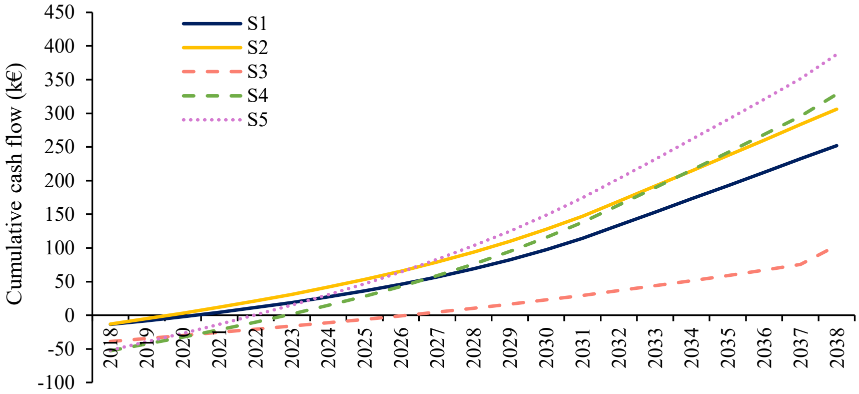


Figure 4 – Cumulative annual cash flow: comparison of intervention solutions between 2018 and 2038

The prioritisation of the intervention solutions is carried out, including different metrics. Regarding financial metrics, Table 6 presents the results obtained for the net present value (NPV), the payback period (PBP), the internal rate of return (IRR) and the profitability index for each analysed solution. Three discount rates are used (i.e., 5%, 7% and 10 %) since this rate represents one of the most uncertain variables of the analysis. All the studied solutions show positive NPV for the three discount rates, representing potential solutions to implement. Solution S5 presents the highest NPV, although it also corresponds to one of the highest investments. All solutions present extremely low values regarding the payback period, varying between 2 (S2) and 9 years (S3).

The IRR corresponds to the discount rate that gives a zero NPV. Solution 3 presents the lowest IRR (13%). If a discount rate of 13% or higher is considered, Solution 3 will present a null or negative NPV, respectively, and is no longer viable. The profitability index is the ratio between the NPV and the initial investment, allowing to quantify the amount of value created per unit of investment and, therefore, rank intervention solutions. Solution 2 presents the highest profitability index for the three discount rates, the lowest PBP and the highest IRR, being considered the best solution using these metrics.

CC BY NC SA

Table 6 – Prioritisation of intervention solution using financial metrics.

Solution	**Investment (€)**	**NPV (€)**			**PBP (years)**	**IRR (%)**	**Profitability index (-)**		
		5%	**7%**	**10%**			**5%**	**7%**	**10%**
S1	13 700	133 437	105 659	75 911	3	49	10	8	6
S2	12 896	166 443	133 487	98 050	2	67	13	10	8
S3	38 835	40 732	25 918	10 562	9	13	1	0.7	0.3
S4	52 535	161 580	122 386	80 282	5	24	3	2	2
S5	51 731	197 715	152 869	104 548	4	29	4	3	2

The prioritisation of intervention solutions is also carried out using the PI used to compare the solutions (Table 7). Thus, the same normalisation process applied in the subsystem diagnosis is considered, using a continuous scale varying from 0 to 3 and divided into three levels: [2; 3] – good performance (🟢), [1; 2[– average performance (🟡); and [0; 1[– unsatisfactory performance (🔴). Moreover, a weight to each PI is necessary to assign its importance in the assessment system to calculate the overall assessment of each intervention solution, in this study equal weights for the PI are considered (default value of 1).

Regarding the PI associated with the energy use efficiency (standardised energy consumption and energy in excess), an improvement in the solutions with the replacement of the pump groups is verified, especially if properly designed (Solutions 2 and 5). The majority of solutions still present an unsatisfactory performance for IVI, since the assets of Subsystem 3 are reaching the end of their service lives, corresponding to an overall IVI of Subsystem 3 during the time horizon of the analysis of 0.12 (Solution 0). Solutions 4 and 5 include the replacement of more assets (pump groups and water pipes), allowing to obtain an average performance. The non-revenue water has not been identified as a problem in this subsystem and all solutions present good results, although Solutions 0, 1 and 2 show an average performance (even though very close to good performance). All studied solutions present an average performance for the overall assessment; however, Solution 5 has the highest overall assessment and a value close to 2, representing good performance.

Table 7– Prioritisation of intervention solutions using normalised performance indicators.

Normalised performance indicators	**Solutions**					
	S0	**S1**	**S2**	**S3**	**S4**	**S5**
Standardised energy (-)	🔴 0.71	🟡 1.03	🟡 1.28	🔴 0.71	🟡 1.03	🟡 1.25
Energy in excess (-)	🟡 1.48	🟢 2.56	🟢 2.88	🟡 1.52	🟢 2.60	🟢 2.95
IVI (-)	🔴 0.12	🔴 0.42	🔴 0.40	🔴 0.99	🟡 1.28	🟡 1.28
Non-revenue water (-)	🟡 1.95	🟡 1.95	🟡 1.95	🟢 2.25	🟢 2.25	🟢 2.25
Overall assessment (-)	🟡 **1.06**	🟡 **1.49**	🟡 **1.63**	🟡 **1.36**	🟡 **1.79**	🟡 **1.93**
Ranking (1-5; 1 corresponds to the best solution)	**-**	**4**	**3**	**5**	**2**	**1**

Notes: [2; 3] – Good performance (🟢); [1; 2[– Average performance (🟡); [0; 1[– Unsatisfactory performance (🔴).

The comprehensive evaluation of the intervention solutions and the corresponding investment is presented in Figure 5. Properly designed asset replacement solutions (S5 and S2) are preferable to asset replacement like-for-like solutions (S4 and S1). In addition, Solutions 5 and 2 show a higher overall performance assessment and slightly lower investment costs than Solutions 4 and 1, respectively. Furthermore, Solution 3 presents the lowest overall assessment; however, the investment cost is higher than for Solutions 1 and 2. For that reason, this solution should not be considered for implementation. The choice of the solution to implement will always depend on the available water utility budget, in which the chosen solution presents the best overall performance assessment and investment values lower or equal to the defined budget.

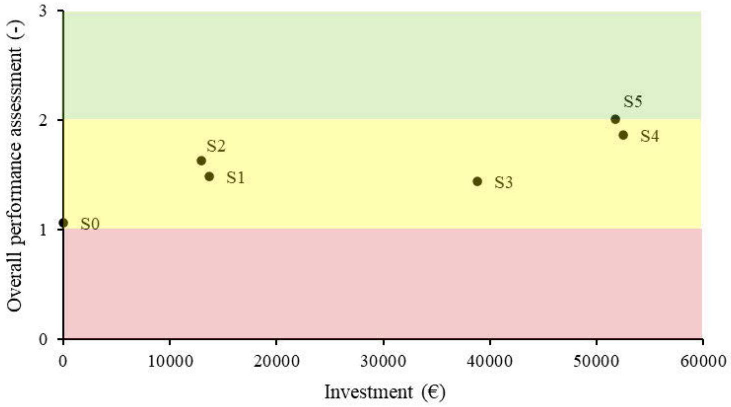


Figure 5 – Overall performance assessment of the intervention solutions in Subsystem 3 considering equal weights for the performance indicators.

5 CONCLUSIONS

The present paper aimed at proposing and demonstrating a decision-making approach to identify, assess, compare, and prioritise different intervention solutions. A three-module decision-making approach was applied to a water distribution system in Portugal. Five intervention solutions were defined and compared considering the financial metrics and performance indicators, such as standardised energy consumption, energy in excess per unit of the authorised consumption, infrastructure value index and non-revenue water.

The application of the proposed approach has demonstrated the influence of considering different assessment criteria and PIs in the solutions' prioritisation, highlighting that not always the solution with the lowest associated capital cost corresponds to the solution with the highest overall performance (i.e., the most technically and operationally recommendable). The study of less-conventional interventions (e.g., the replacement of pump groups by other adequately designed and operated) was carried out. These solutions were compared with conventional interventions (e.g., the like-for-like pump replacement, replacing pumps by others technologically equivalent and with the same operation conditions). Results have shown that less-conventional interventions may result in a higher overall performance, especially in the long-term, and may even lead to a lower capital cost.

This proposed approach was a step forward in the implementation of IAM approaches in urban water infrastructures since it allows to integrate different perspectives, which is important for the establishment of policies and future rehabilitation solutions at the infrastructure or asset level and for the improvement of intervention solutions prioritisation necessary for the development of sounder tactical level plans. The definition of three assessment levels – macro, meso and micro – and the incorporation of new metrics associated with the assets' physical condition were the main novel contributions of this research.

Further research should include the study of different intervention solutions, such as system zoning, leak detection and changes in the system layout. Addicionally, a sensitivity analysis to help in the prioritisation process and to study the variables of analysis with the highest uncertainty should be carried out, such as the discount rate, the O&M costs and the time horizon.

6 ACKNOWLEDGEMENTS

The authors would like to thank the Foundation for Science and Technology and to Fundo de Apoio à Inovação for funding this research as well as to the Portuguese water utilities for providing the case studies and the required data. The author is grateful for the Foundation for Science and Technology's support through funding UIDB/04625/2020 from the research unit CERIS.

7 REFERENCES

[1] N. Carriço, D. Covas, M. C. Almeida, J. P. Leitão, H. Alegre, "Prioritization of rehabilitation interventions for urban water assets using multiple criteria decision-aid methods," Water Science and Technology, vol. 66. 2012, pp. 1007-1014. Doi:10.2166/wst.2012.274.

[2] INGENIUM, IPWEA, International Infrastructure Management Manual, 2006 Edition, Association of Local Government Engineering NZ Inc and Institute of Public Works Engineering of Australia, 2006.

[3] H. Alegre, D. Covas, Infrastructure asset management in urban water systems: an approach focuses on rehabilitation (in Portuguese), 1st ed., ERSAR, LNEC, 2010.

[4] M. S. Morley, D. Vitorino, K. Behzadian, R. Ugarelli, Z. Kapelan, S. T. Coelho, M. C. Almeida, "Decision support system for the long-term city metabolism planning problem," Water Science and Technology: Water Supply, vol. 16, no. 2, 2016, pp. 542-550. Doi: 10.2166/ws.2015.167.

[5] C. Royce, B. Neijens, "Value-based decision making," in Leading Edge Conference on Strategic Asset Management, Trondheim, Norway, 2017.

[6] M. Grimaldi, M. Sebillo, G. Vitiello, V. Pellecchia, "Planning and managing the integrated water system: a spatial decision support system to analyze the infrastructure performances," Sustainability, vol. 12, no. 16, 2020. Doi: 10.3390/su12166432.

[7] G. Hutton, L. Haller, J. Bartram, "Global cost-benefit analysis of water supply and sanitation interventions," Journal of Water and Health, vol. 5, no. 4, 2007, pp. 481-502. Doi: 10.2166/wh.2007.009.

[8] D. Sartori, G. Catalano, M. Genco, C. Pancotti, E. Sirtori, S. Vignetti, C. Del Bo, Guide to cost-benefit analysis of investment projects, Economic appraisal tool for Cohesion Policy, 2014.

[9] K. Sjöstrand, A. Lindhe, T. Söderqvist, L. Rosén, "Cost-benefit analysis for supporting intermunicipal decisions on drinking water supply," Journal of Water Resources Planning and Management, vol. 145, no. 12, 2019. Doi: 10.1061/(ASCE)WR.1943-5452.0001121.

[10] N. Jayaram, K. Srinivasan, "Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing," Water Resources Research, vol. 44, no. 1, 2008. Doi: 10.1029/2006WR005316.

[11] R. Ugarelli, Asset management of wastewater networks, PhD Thesis, Alma Mater Stusiorum, Universita' di Bologna, 2008.

[12] L. Waghmode, A. Sahasrabudhe, P. Kulkarni, "Life cycle cost modelling of pumps using an activity based costing methodology," Journal of Mechanical Design, vol. 132, no. 12, 2010. Doi: 10.1115/1.4002970.

[13] T. S. Bixler, J. Houle, T. P. Ballestero, W. Mo, "A spatial life cycle cost assessment of stormwater management systems," Science of The Total Environment, vol. 728, 138787, 2020. Doi: 10.1016/j.scitotenv.2020.138787.

[14] S. Bruaset, S. Sægrov, R. Ugarelli, (2018). "Performance-based modelling of long-term deterioration to support rehabilitation and investment decisions in drinking water distribution systems," Urban Water Journal, vol. 15, no. 1, 2018, pp. 46-52. Doi: 10.1080/1573062X.2017.1395894.

[15] M. Engelhardt, P. J. Skipworth, D. A. Savic, A. J. Saul, G. A. Walters, G. A. (2000). "Rehabilitation strategies for water distribution networks: a literature review with a UK perspective," Urban Water, vol. 2, no. 2, 2000, pp. 153-170. Doi: 10.1016/S1462-0758(00)00053-4.

[16] ISO 24512:2007. Activities relating to drinking water and wastewater services – Guidelines for the management of drinking water utilities and for the assessment of drinking water services. International Organization for Standardization, Genebra.

[17] Cabral et al., 2022 distress-based WRM

[18] H. Alegre, S. T. Coelho, Infrastructure asset management of urban water systems, Water supply system analysis-selected topics, 2012. Doi: 10.5772/52377.

[19] H. Alegre, J. M. Baptista, E. Cabrera, F. Cubillo, P. Duarte, W. Hirner, R. Parena, Performance Indicators for Water Supply Services, 3rd ed., London, UK: IWA Publishing, 2016, ISBN 9781780406329.

[20] H. Alegre, Infrastructure asset management of water supply and drainage and treatment of wastewater (in Portuguese), Research program and graduate program presented to obtain the title of “Qualified for the Exercise of Scientific Research Coordination Functions”, LNEC, Lisbon, Portugal, 2008.

[21] P. Duarte, D. Covas, H. Alegre, “PI for assessing effectiveness of energy management processes in water supply systems,” in IWA International Conference PI09: Benchmarking water services–the way forward, Amsterdam, Netherlands, 2009, pp. 11-13

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14116

MEASURING DRINKING WATER TEMPERATURE CHANGES IN A DISTRIBUTION NETWORK

Mirjam Blokker[1], Joost van Summeren[2], and Karel van Laarhoven[3]

[1,2,3]PhD, KWR Water Research Institute, Groningenhaven 7, 3433 PE, Nieuwegein (the Netherlands)

[1] *mirjam.blokker@kwrwater.nl,* [2] *joost.van.summeren@kwrwater.nl,* [3] *karel.van.laarhoven@kwrwater.nl*

Abstract

During distribution the drinking water temperature changes due to the temperature of the soil surrounding the drinking water distribution network (DWDN). A drinking water temperature below 25 °C at the tap is required to meet Legionella prevention standards and/or drinking water standards. With climate change, urbanisation and the energy transition towards more district heating networks, the urban subsurface will heat up further, and more exceedances of the 25 °C threshold are expected.

To understand the effectiveness of various measures to keep drinking water temperature below the threshold a modelling approach was followed. The drinking water temperature model (WTM) calculates drinking water temperatures at each customer from a hydraulic network model and heat conduction from the outside of the pipe wall, where the soil temperature is kept constant and the soil is thus assumed to be an infinite heat source. The WTM was validated with measurements in a DWDN on relatively small diameters (150 mm or smaller) and at locations far enough from the source so that equilibrium with the soil temperature was already reached (i.e. the influence of residence time was not validated). In reality, the soil is not an infinite heat source, but is affected by the drinking water temperature. We therefor developed an enhanced WTM (called WTM+) which uses an extra insulation layer (of soil material) around the drinking water pipe to account for the soil which is affected by the drinking water temperature. In order to determine a suitable length scale for this extra insulation layer, and to validate the WTM+, we measured drinking water temperatures in two real case studies.

Case study 1 is a single 1 km pipe where we manipulated flows, and thus residence time. Case study 2 is a DWDN with a variation in diameters and residence times. Designing and executing these measurements proved to be quite a challenge. In case study 1 the incoming drinking water temperatures and soil temperatures were not stable during the three weeks of measurements. The temperature changed typically less than 1.5 °C over 1 km, and the sensors had a resolution of only 0.1 °C. In case study 2 the measurements were done on a single day (morning, and repeated in the afternoon), and were quite stable. The drinking water temperature changed up to 8 °C over the course of the residence time. However, as the DWDN hydraulic network model is never perfect, the residence times are not all known accurately, and the surrounding soil temperatures, that may have varied quite a lot over the DWDN, were not measured. Nevertheless, the case studies did prove to be suitable for validating the WTM+, including the effect of residence time.

Keywords

Drinking water temperature, measurements, real DWDN.

1. INTRODUCTION

During distribution the drinking water temperature changes due to the temperature of the soil surrounding the drinking water distribution network (DWDN). In the Netherlands and many other countries, a drinking water temperature below 25 °C at the tap is required to meet Legionella prevention standards and/or drinking water standards [1]. With climate change, urbanisation and the energy transition towards more district hearing networks, the urban subsurface will heat up further, and more exceedances of the 25 °C threshold are expected [1].

To understand the effectiveness of various measures to keep drinking water temperature below the threshold a modelling approach was followed. The drinking water temperature model (WTM, [2]) calculates drinking water temperatures at each customer from a hydraulic network model and heat conduction from the outside of the pipe wall, where the soil temperature is kept constant and the soil is thus assumed to be an infinite heat source. The WTM was validated with measurements in a DWDN on relatively small diameters (150 mm or smaller) and at locations most likely beyond the maximum heating time (i.e. the influence of residence time was not validated).

The WTM equations [2] show that in a Ø100 mm pipe the drinking water temperature increases with 90% of the initial temperature difference between drinking water and the pipe wall within less than 2 hours. However, this approach assumes infinite heat capacity of the soil and does not take into account heat exchange from drinking water to soil. In reality the soil is not an infinite heat source, but is affected by the drinking water temperature. As drinking water flows through these pipes year after year, the influence on the surrounding soil may not be neglected. This would mean that in the example above, the time for the drinking water temperature to increase may be (much) longer than 2 hours. We therefor introduced an enhanced WTM (called WTM+) which uses an extra insulation layer (of soil material) around the drinking water pipe to account for the soil which is affected by the drinking water temperature. In order to determine the size of this extra insulation layer, and to validate the WTM+ we measured drinking water temperatures in two real case studies. Case study 1 is a single pipe stretch of ca. 1 km where we manipulated flows, and thus residence time. Case study 2 is a DWDN with a variation in diameters and residence times. Designing and performing these measurements proved to be quite a challenge. This paper shows the results of the validation measurements.

2. BACKGROUND ON MODELLING AND VALIDATION APPROACH

The goal of the measurements was the validation of the WTM as a function of residence time [2, 3]:

$$T_{water}(t,\tau) = T_{boundary}(t) + (T_{water,0}(t) - T_{boundary}(t))exp(-k\tau) \tag{1}$$

$$k = \frac{4 \cdot \alpha_{water}}{D_1{}^2 \left(\frac{1}{Nu} + \frac{\lambda_{water} \cdot ln\left(\frac{D_2}{D_1}\right)}{2\lambda_{pipe}} + \frac{\lambda_{water} \cdot ln\left(\frac{D_3}{D_2}\right)}{2\lambda_{soil}}\right)} \tag{2}$$

Where k is the overall heat transfer coefficient [2], which depends on characteristics of the water, and the insulating material of the pipe and pipe surroundings, and in the case of a flowing medium on the Nusselt number (Nu). For the boundary conditions it is assumed that $T_{water}(\tau=0) = T_{water,0}$ and $T(\tau=\infty) = T_{boundary}$,. Conditions are time variable, where t is time and τ is the travel time. Of course, k can also be time dependent when Nu changes over time. We will assume that the equation is still valid when $T_{boundary}$ is not uniform over the pipe circumference.

With D_1 the inside pipe diameter, D_2 outside pipe diameter ($D_2 = D_1 + 2 \times d_{\text{pipe}}$, with d_{pipe} the pipe wall thickness), e.g. D_1 = 152.0 mm, D_2 = 160 mm. For pipes, Nu can be described as a function of the dimensionless Reynolds number (Re) and Prandtl number (Pr): $Nu = 0.027\ Re^{0.8}\ Pr^{0.33}$ for turbulent flows, and Nu = 3.66 for laminar flows. Furthermore, α_{water} is the thermal diffusion coefficient [0.14 m^2/s]; $\alpha_{water} = \lambda_{water}/\rho_{water}\ C_{p,\ water}$; λ_{water} is the thermal conductivity of water [0.57 $W.m^{-1}.K^{-1}$]; ρ_{water} is the density of water [1000 $kg.m^{-3}$]; $C_{p,water}$ is the heat capacity of water [4.19 $J.kg^{-1}.K^{-1}$] – with the parameter values given at 20 °C. In this paper we only consider plastic pipes, for which λ_{pipe} is the thermal conductivity of PVC [= 0.16 $W.m^{-1}.K^{-1}$]. D_3 the diameter where the boundary condition is valid, equal to the outside pipe diameter plus the surrounding soil ($D_3 = D_2$ + GpD). What the best value for D_3 is, is still to be determined. In this paper we only consider pipes that are installed in sand, for which λ_{soil} is the thermal conductivity of dry sandy soil [= 1.6 $W.m^{-1}.K^{-1}$]. Figure 1 shows the solution for Eq. (1) and (2) for a range of values for D_3.

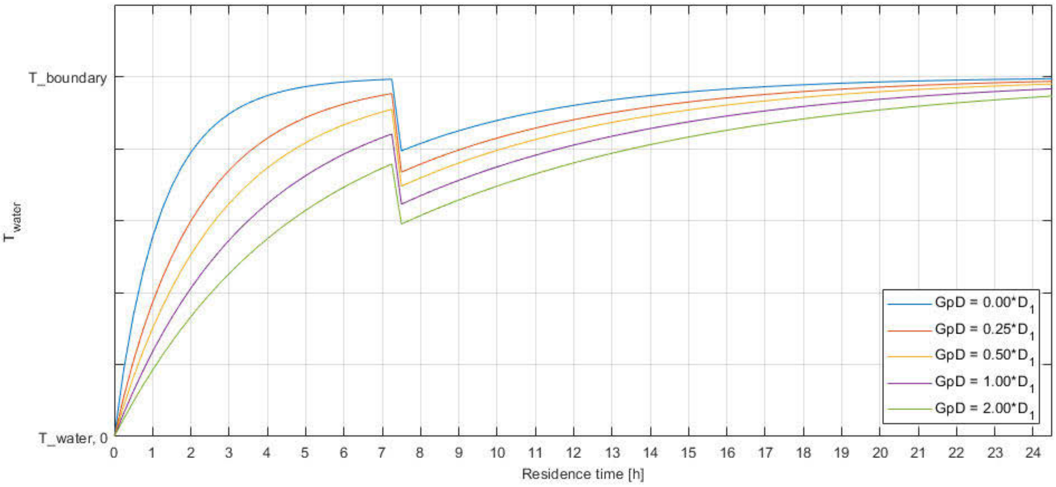


Figure 1. T_{water} against travel time (τ) for a Ø160 mm PVC pipe with various potential length scales for the extra insulation layer of soil.

In practice, time t implies a distance x ($=v/t$). The challenge in validating the WTM+ in practice is to find a DWDN where $T_{\text{boundary}}\ (t)$ is known over length x and time t. This is a potential problem because a) over the pipe length circumstances above and below ground are not constant over both time (under the influence of changing weather) and space (e.g. district heating pipes or electricity cables are installed alongside only part of the drinking water pipe, the drinking water pipe is installed under an incline and thus surrounding temperature changes over the pipe length), and b) it is not clear how to exactly determine the undisturbed soil temperature that is $T_{\text{boundary}}\ (t)$. The undisturbed soil temperature (i.e. not influenced by temperature of the drinking water pipe itself) can only easily be measured when there is no drinking water pipe, and by definition this can therefor not be measured at the location of the pipe under investigation. Another problem is that the temperatures can not be controlled, they can only be measured.

We tried to overcome these issues by following two paths on two case studies, which are described further in Section 3. The first case study entailed measurements of drinking water temperature in one single stretch of a drinking water pipe, without demand along the length. At this location we also measured soil temperatures at three locations along the pipe length, measured soil properties, mapped electricity cables and district heating pipes and a 2D soil temperature model [STM+, 4] was built to validate the STM+. After the validation the STM+ was used to describe $T_{\text{boundary}}\ (t)$. The length x was a constant, but with a variation in flow, τ was changed. Measurements took several weeks, and none of the parameters in Eq. (1) were kept constant. The second case study entailed measurements of drinking water temperature in a DWDN with various pipe diameters. At this location we did not measure soil temperatures, but used the STM [2, 3] to estimate T_{boundary}. Measurements were done at various distances (x) and thus times (τ), but τ was not controlled, but instead was estimated based on the hydraulic network model. Measurements took only one day, and we assumed that the parameters in Eq. (1) were constant over time, but not necessarily over space.

3. CASE STUDY DESCRIPTIONS

3.1 Case study 1: Single pipe system

The first case study is a single pipe, a Ø160 mm PVC pipe (D_1 = 152 mm, D_2 = 160 mm) in Rotterdam with a length of 925 m. This pipe is fed from a surface water PS (pumping station) through a stretch “S1” of 1650 m (D_1 = 1569 mm), “S2” of 1350 m (D_1 = 1369 mm), and”S3”of 570 m (D_1 = 150,6 mm), see Figure 2. The water in the measured pipe flows from location L1 to location L2 (stretch “S4” - 415 m) to location L3 (stretch “S5” 510 m). The flow at L3 was controlled with a hydrant. During a two weeks measurement period (19 May 2020 – 03 June 2020) the flow rate was regulated in order to get measurements for residence times of the water in the pipe from 1 to 24 hours.

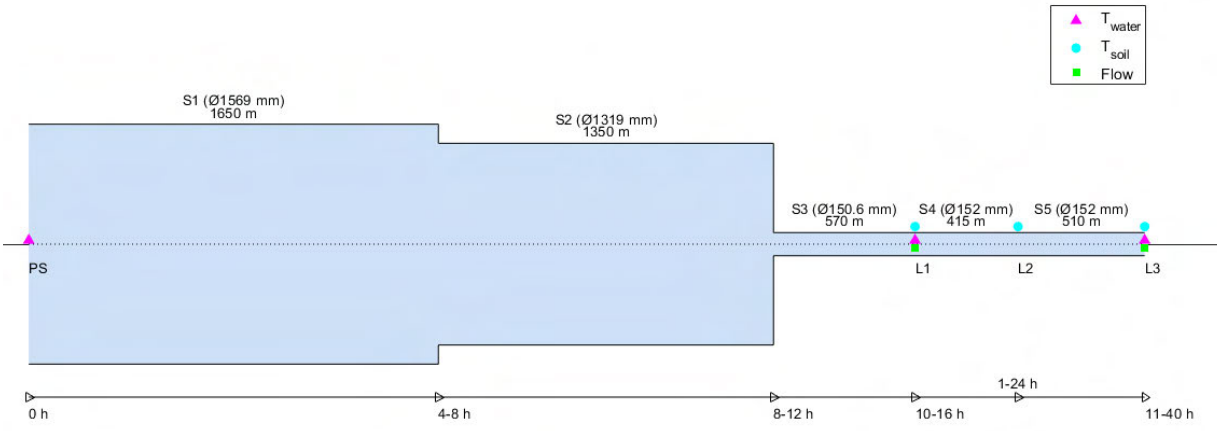


Figure 2. Measurement locations of case study 1. In green flow meter locations, in magenta and cyan drinking water temperature and soil temperature measurement locations. Stretch names with length and diameter are indicated above, and travel times are indicated below. PS is pumping station, L1, L2 and L3 are measurement locations.

The following was measured, with a logging frequency of once per 15 minutes:

- T_{water} was measured at the PS and at L1 and L3 (Figure 3). We may assume that T_{water} at the start of S3 is equal to T_{water} from the PS as S1 and S2 have large diameters and short residence times. However, T_{water} at the end of S3 (location L1) is not equal to T_{water} from the PS (Figure 3).
- T_{soil} (temperature of the soil) was measured at L1, L2 and L3, at various distances from the pipe.
- The flow was measured at L1 and L3. During the measurements there was limited demand from customers along the pipe, as there is only one residential customer, and some sports facilities that were closed during the Covid-19 pandemic. The demand at the customer location was measured, and it had a negligible effect on the flow at L1. The residence time between L1 and L3 was calculated from the pipe length, pipe diameter and flow at L3.

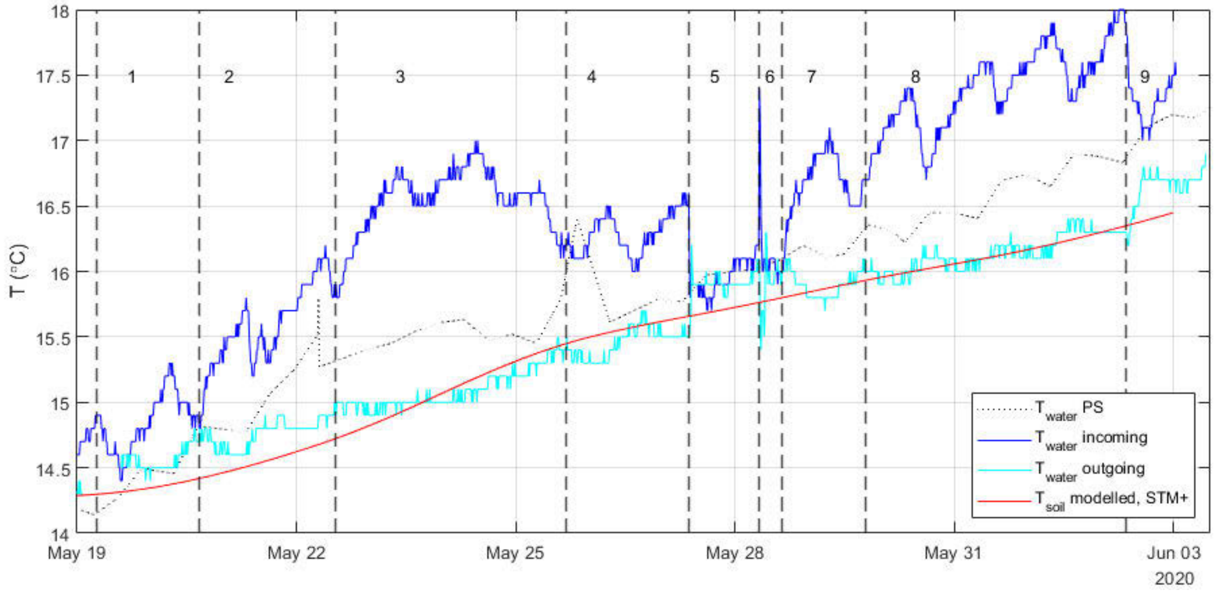


Figure 3 Measured drinking water temperatures and modelled soil temperature for the single pipe case study. Indicated are the starting times of the 9 tests (Table 1).

3.2 Case study 2: Drinking Water Distribution Network

Drinking water temperature measurements were done in the DWDN of Almere in the Netherlands. The Almere DWDN is ca. 700 km long, and consists mainly of PVC pipes, with nominal diameters of Ø63 (8%), Ø110 (51 %), Ø160 (15%), Ø200 (5%), Ø310 (6%) and > Ø400 mm (9%), [5]. The DWDN is fed by two feeding reservoirs, in the north and in the south part of the system that will be referred to in the paper as F1 and F2.

There is a variation of backfill for the pipes where the older network parts (built between 1980 and 2000) have a mixture of sand and clay and the newer parts where the backfill is sand. Also, there is a variation in installation depth with the distribution mains in the older parts at -1.3 m and in the newer parts at -1.1 m. The transport mains are installed at ca. -1.5 m. The western part of Almere has a district heating network (DHN) that may influence the Drinking Water Distribution Network (DWDN). The effect of the DHN parallel to the DWN was calculated with STM+ to be negligible as the distance between the two is more than 2.5 m. Where DWDN and DHN pipes cross, the distance between the two is much closer, but for a limited length. So, crossings may lead to an increased drinking water temperature. Roughly speaking, the area with a DHN is the older part of the DWDN with larger pipe diameters and a looped system, where the newer DWDN is designed as a self-cleaning network with a more branched structure and smaller pipe diameters [6].

Two areas (A and B) were selected to take temperature measurements. Area A and B are comparable in residence time from the sources, they have a similar year of installation and thus a similar design philosophy (with similar pipe diameters, pipe materials and DWDN layout, in this case self-cleaning network design) and a similar number of residents. The individual measurement locations differ in experienced soil temperatures and in residence times from the source. The soil in area A experiences a slower heat transfer compared to area B, because there is a mixture of clay and sand in A and only sand in B; the pipes in area A are installed a little bit deeper than in area B; and area A has a DHN installed next to the DWDN, and in area B there is no DHN. This means that the expected soil temperatures (based on the STM) in area A without a DHN are lower than for area B (18.7 °C instead of 20.5 °C), but with a DHN the soil temperature in area A is comparable to that in area B (20.5 °C). The hydraulic model of Almere (provided by Vitens)

cc BY NC SA

was used to calculate the flows at the measurement day (the demand pattern of 31 August 2020 was used). The hydraulic network model shows a residence time between 5 to 24 hours at the measurement locations in area A and between 10 and 15 hours in area B.

Drinking water temperature measurements were done by Vitens employees on 31 August 2020 at 35 hydrants in area A and B at two moments of the day (one in the morning, between 8:00 and 12:30, and one in the afternoon, between 12:45 and 16:30) leading to a total of 70 values. Measurements were taken at hydrants, in order to avoid an influence of the premise plumbing system, and during the Covid-19 pandemic to not have to enter people's homes. The hydrants were opened only with a small flow of water, in order not to influence travel time of the water too much. The drinking water temperature at the sources F1 and F2 are $T_{water,F1}$ = 13.7 °C and $T_{water,F2}$ =13.3 °C on 31 Aug 2020, respectively (see Figure 4).

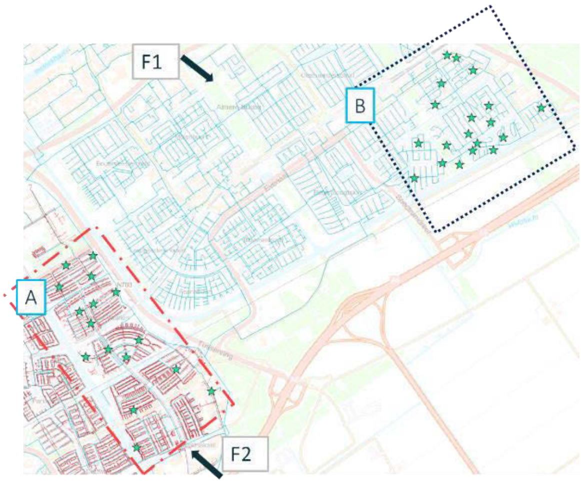


Figure 4 Almere DWDN in blue, and DHN in red. The measurement locations are indicated in green.

4. DATA ANALYSIS

4.1 Case study 1: Single pipe system

As during the measurement period (19 May 2020 – 03 June 2020) the boundary conditions ($T_{water,0}$ and $T_{boundary}$ in Eq. (1)) were not constant, we decided to introduce a normalized parameter (with values between 0 and 1) to be able to compare all measurements in a single graph. This leads to:

$$\Delta TN(\tau,t) = 1 - exp(-k(t)\tau(t)) = \frac{T_{water}(\tau,t) - T_{water,0}(t)}{T_{boundary}(t) - T_{water,0}(t)} \qquad (3)$$

The parameters of Eq. (3) are determined as follows (and shown in Figure 3):

- $\tau(t)$: The residence time follows from the measured flows, and is listed in Table 1. The accuracy of the flow meter is 0.004 m^3/h (1 litre with a log frequency of 15 minutes). During the test phases the flow is kept more or less constant, the accuracy and variability lead to a ΔQ equal to 0.012 m^3/h. This leads with Eq. (4) and the flow rates of Table 1 to an uncertainty in the residence time of less than 2% (less than 1 minute for the short residence times of tests 5 and 6 and almost 30 minutes for the longest residence time of test 3), and therefor can be neglected. In this Eq. V is the pipe volume (16.8 m^3 for a 152 mm, 925 m pipe):

$$\frac{\Delta\tau}{\tau} = \frac{\frac{V}{Q-\Delta Q}-\frac{V}{Q}}{\frac{V}{Q}} = \frac{\Delta Q}{Q-\Delta Q} \approx \frac{\Delta Q}{Q} \tag{4}$$

- $T_{water,0}(t)$: drinking water temperature measured at location L1.
- $T_{water}(\tau, t)$: drinking water temperature measured at location L3, time-shifter to adjust for residence time τ. For example for test 4, in which the flow Q = 1.4 m³/h results in τ = 5 hours (L1 to L2) + 7 hours (L2 to L3). Therefore, $T_{water}(\tau,t) = T_{water}(t+12)$, where the data during the transition time of 12 hours are discarded (adjustment for residence time). Table 1 shows the residence times needed for the adjustment, and the amount of datapoints left for the analysis after discarding the transition period. Figure 5 shows the time series after adjustment for residence time.
- $T_{boundary}(t)$: $T_{boundary}$ was estimated by using the STM+ [7] results at L3. The STM+ was validated with this particular case study, and then the STM+ was rerun to calculate the so called undisturbed soil temperature ($T_{boundary}$). This modelled temperature ($T_{\text{soil, modelled}}$) is also shown in Figure 3. For the tests we assume that during the residence time the $T_{boundary}$ that is experienced by the flowing water is the average of the soil temperature during this residence time: $T_{boundary}(t) = mean(T_{soil,modelled}(t{:}\, t+\tau))$

The normalized temperature difference ΔTN (Eq. (3)) is then calculated for every datapoint, i.e. for every 15 minutes. Also, the uncertainty in ΔTN is considered. This uncertainty is related to the accuracy of the drinking water temperature measurements. The measurement accuracy in $T_{water,0}$ and $T_{water}(t)$ is ± 0.05 °C and the uncertainty for $T_{boundary}$ is assumed to be small compared to the measurement uncertainty. With ε_w = 0.05 °C, the uncertainty in ΔTN is determined as (see appendix) $\bar{E} = \frac{\varepsilon_w(2T_{boundary}-T_{water,0}-T_{water})}{(T_{boundary}-T_{water,0})^2-\varepsilon_w^2}$. This means that when $(T_{boundary} - T_{water,0})$ is small, the uncertainty in ΔTN is large. As ΔTN is between 0 and 1, we will discard datapoints where $\bar{E}$ > 0.25. Figure 5 shows the time series after adjustment for residence time and without the datapoints with $\bar{E}$> 0.25.

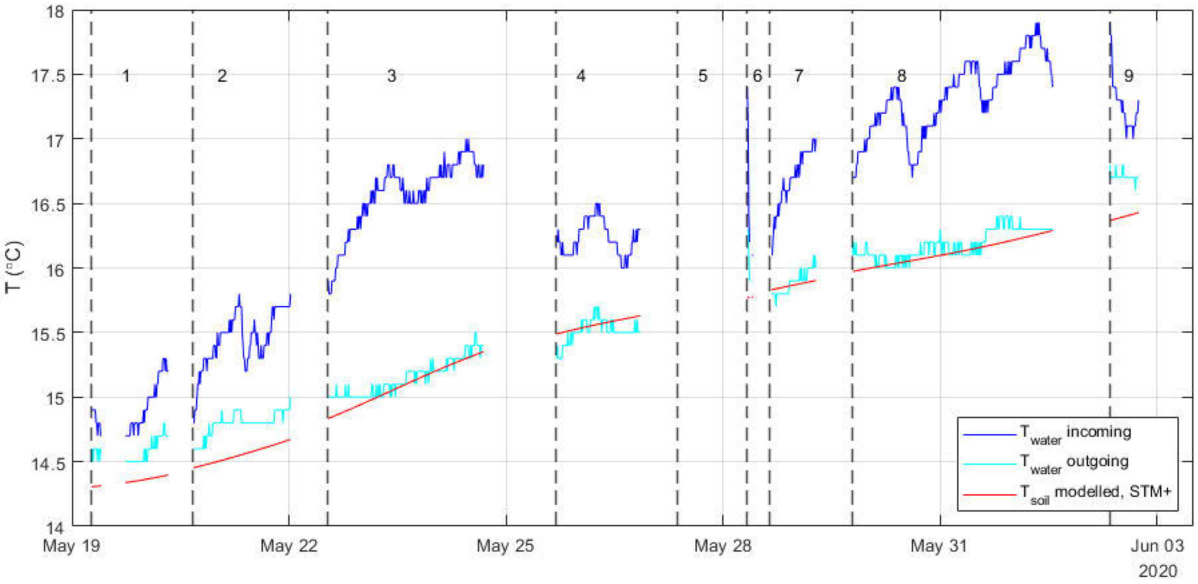


Figure 5. As Figure 3, but here T_{water} outgoing is adjusted for residence time. And datapoints where $\bar{E}$ > 0.25 are discarded.

CC BY NC SA

Table 1 Overview of data in the drinking water temperature measurements for various residence times. Changing the hydrant flow took 15-30 minutes. Test 1 started on 18 May, but some measurements were lost. Therefor we used data from 19 May onwards. Test 6 was started 15 minutes after test 5 ended. However, a passer-by closed the hydrant. The next morning during the check test 6 was started again. The measurements at L1 were stopped at 03-June 0:45. Between brackets the number of datapoints after removing high uncertainty measurements.

ID	Starting time	End time	Amount of datapoints after adjusting for residence time	Flow rate [m^3/h]	Residence time [h]
1	19-May-2020 6:30 (18-May-2020 12:00)	20-May-2020 16:00	103 (73)	2.1	8.0
2	20-May-2020 16:15	22-May-2020 12:30	130 (130)	1.4	12.0
3	22-May-2020 13:00	25-May-2020 16:30	207 (207)	0.7	24.0
4	25-May-2020 16:45	27-May-2020 08:30	112 (112)	1.4	12.0
5	27-May-2020 09:15	27-May-2020 20:00	41 (1)	16.7	1.0
6	28-May-2020 08:00	28-May-2020 15:00	21 (9)	8.5	2.0
7	28-May-2020 15:30	29-May-2020 18:45	62 (60)	1.4	12.0
8	29-May-2020 19:00	02-Jun-2020 07:45	266 (266)	0.9	18.6
9	02-Jun-2020 08:15	03-Jun-2020 10:00	39 (39)	2.8	6.0

Case study 2: Drinking Water Distribution Network

The parameters of Eq. (1) are determined as follows:

- τ: The residence time follows from the hydraulic network model. Here, the demand patterns of 31 August 2020 were applied, but the model was not calibrated for this particular day, so there may be some valve positions that are incorrect which may lead to errors in travel time.
- $T_{boundary}$: $T_{boundary}$was estimated by using weather data, processed with the STM [2, 3] and STM+ [4]. There were no soil temperature measurements available. Five different boundary conditions were suggested:
    - T_{TM} is the soil temperature around transport mains. T_{TM} = 18.0 °C. This is the STM calculated temperature at -1.7 m (Figure 6), in peri-urban area (clay/sand under grass);
    - $T_{\mathrm{TM\_DHN}}$ is the soil temperature around transport mains with crossing DHN. $T_{\mathrm{TM\_DHN}}$ = 20.1 °C. This is the STM calculated temperature at -1.7 m, in peri-urban area (clay/sand under grass) + 2.1 °C from the primary network DHN as from the STM+ [4];
    - T_{TM} is the soil temperature around distribution mains in the older part of Almere. T_{DM} = 18.7 °C. This is the STM calculated temperature at -1.35 m (Figure 6), in peri-urban area (clay/sand under grass);
    - T_{TM} is the soil temperature around distribution mains with crossing DHN, in the older part of Almere. $T_{\mathrm{DM\_DHN}}$ = 19.5 °C. This is the STM calculated temperature at -1.35 m, in peri-urban area (clay/sand under grass) + 0.8 °C from the secondary network DHN as from the STM+ [4];
    - $T_{\mathrm{TM\_B}}$ is the soil temperature around distribution mains in the newer part of Almere. $T_{\mathrm{DM\_B}}$ = 20.5 °C. This is the maximum drinking water temperature that was measured in area B (Figure 7), and is the average of the STM calculated temperatures at -1.15 m (Figure 6) for peri-urban (clay/sand under grass) and urban (sand, under tiles with various shade conditions).

- $T_{water,0}$: drinking water temperature measured at locations F1 and F2 (13.7 and 13.3 °C respectively).
- $T_{water}(\tau)$: drinking water temperature measured at hydrants (Figure 7).

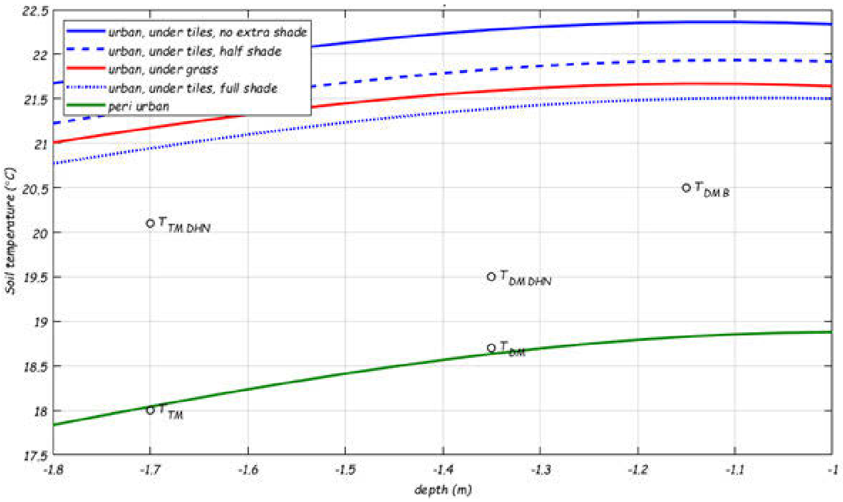


Figure 6. Modelled soil temperature (STM) at various depths for 31 August 2020, for KNMI data of Schiphol airport and circumstances around Almere DWDN. T_{TM}: Temperature around transport mains, $T_{TM\ DHN}$: Temperature around transport mains with crossing of DHN, T_{DM}: Temperature around distribution mains (area A) , $T_{DM\ DHN}$: Temperature around distribution mains with crossing of DHN (area A) , $T_{DM\ B}$: Temperature around distribution mains (area B).

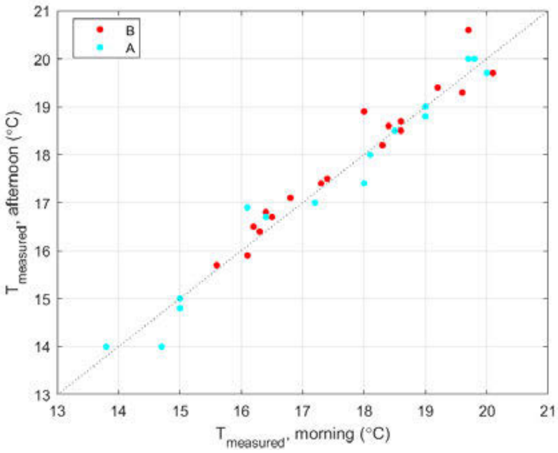


Figure 7. Measured drinking water temperatures in the morning and afternoon in areas A and B.

5. RESULTS OF MEASUREMENTS

5.1 Case study 1: Single pipe system

Figure 3 and Figure 5 show the measured and modelled temperatures. There are some remarkable results:

1) In Figure 3 temperatures are shown simultaneously, without taking into account the travel time between the locations. This makes it difficult to visually compare the temperatures. It looks like the outgoing water temperature is often lower than the modelled soil

BY NC SA

temperature, but after correction for the travel time, this is hardly ever the case (considering the 0.1 °C accuracy) as is shown in Figure 5.

2) Typically the drinking water temperature increases from PS to L1, most likely in stretch S3, and then decreases between L1 and L3. It would be worthwhile to also model the upstream drinking water temperature changes. The difference between drinking water temperature at L1 and L3 was < 2.0 °C.
3) During the test with only one and two hours residence time (tests 5 and 6) there is no increase or decrease of the drinking water temperatures between PS and L1, nor between L1 and L3.
4) The difference between the drinking water temperature at the source and the modelled soil temperatures is small (but be aware of the travel time that needs to be considered when comparing the two). The fact that there was a temperature increase between PS and location L1 lead to better testing circumstances. If the soil temperatures between PS and the beginning of the test side (L1) would have been the same as the soil temperatures around the case study pipe (between L1 and L3), than the test would probably not have given any useful results.
5) The temperature of the incoming water at location L1 (dark blue line) shows an influence of the time of day, most probably due to the change in flow (high demands, high flows during the morning and little flow during the night). The temperature of the outgoing water at location L3 (cyan line) has a much more constant value over the day. This is due to the fact that during the tests the flow between L1 and L3 was constant and for the longer residence times the drinking water temperature was almost in equilibrium with the soil temperature. For the shorter residence times, the measurements were not long enough to show a diurnal pattern.
6) At the start of tests 5, 6 and 9 it can be seen that there is a very quick change in the incoming temperature at location L1 (dark blue line); this is not the case for the other tests. It is suggested that these tests have a quicker heat transfer, involving convective heat transfer due to turbulent flows. For this case study a travel time of 7 hours means a Reynolds number of 5,000. It is assumed that for $Re > 5{,}000$ the flow is turbulent, where $Nu = 0.027 \times Pr^{0.33} \times Re^{0.8}$, while for smaller Re (laminar flows), $Nu = 3.66$. This means that k in Eq. (1) is different for laminar and turbulent flows.

7) Table 1 shows that the results of test 5 and 6 have led to a very limited dataset. Partly due to the fact that the tests were only short (and test 6 was shorter than intended because a passer-by closed the hydrant), and partly due to the fact that the temperature differences between incoming temperature and the soil temperature are small, and therefor datapoints with very high uncertainty needed to be discarded.

5.2 Case study 2: Drinking Water Distribution Network

The temperature measurements in the DWDN showed a wide range between 14 and 21 °C (Figure 7). There is no significant difference between the measured temperatures in areas A and B. The maximum measured temperature in area B is higher than in area A (Figure 8), which can be a coincidence, or due to the fact that the soil temperatures around the pipes in area B ($T_{\mathrm{DM\_B}}$) are higher than in area A (T_{DM}). This suggests that the influence of the DHN (in area A) is limited. However, more analysis is required here. The differences between the morning and afternoon measurements were limited, less than 1.0 °C (Figure 7), and Figure 9 shows that the distribution of the results in the morning and afternoon is the same for area A; for area B the afternoon temperatures are slightly higher.

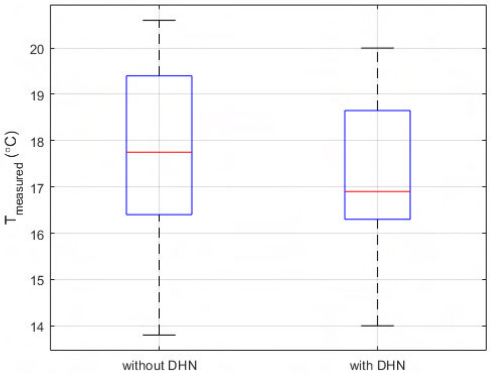


Figure 8. Range of measured drinking water temperatures in area without DHN (area B) and with DHN (area A).

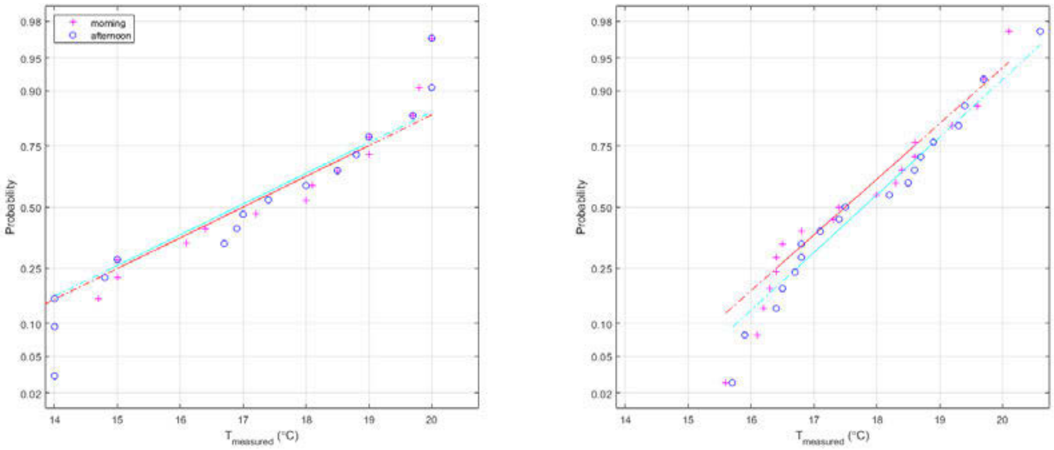


Figure 9. Normal probability plot of measured temperatures in area A (left) with DHN, and B (right) without DHN. In purple/red the morning data, in cyan/blue the afternoon data.

6. DISCUSSION

With respect to designing a good test, we present some lessons learned:

- Case study selection: It was a big challenge to find a good test site for the single pipe test, i.e. a pipe of sufficient length where travel time could be manipulated between 1 and 24 hours and where the soil temperature would not vary much over the pipe length. The DWDN test can be performed in any DWDN.
- Certainty of travel times: In the single pipe case it was possible to manipulate and exactly know the travel times. In the DWDN the hydraulic model was needed to estimate the travel times, which introduced some uncertainty. The measurement locations in the DWDN were such that there was a range of travel times, leading to a wide range of temperatures. More or less by accident there were hardly any measurement locations with a larger travel time beyond the maximum heating time, i.e. there were no measurement locations beyond where the influence of travel time could be assessed. Differences between WTM+ and measurements are potentially due to inaccuracies in the modelled flows (maybe due to inaccurate valve statuses), where the inaccuracies are not easy to estimate.
- Certainty of soil temperatures: In the single pipe test setup it was possible to measure soil temperatures, and soil heat capacity, and therefore validate the STM+ for the test location. This allowed to determine the $T_{boundary}$ with higher certainty than for the DWDN test site.
- Certainty of drinking water temperatures: For the single pipe the difference between drinking water temperature at the beginning of the pipe and the soil temperature was <

2.0 °C, with a resolution of 0.1 °C. The DWDN showed a larger range of temperatures, between 14 and 21 °C. Testing the same locations both in the morning and in the afternoon increased the reliability of the test results.

- Test duration: The single pipe test took a few weeks, so each test would lead to a good amount of datapoints. The test in the DWDN took just one day, and a limited number of datapoints were thus collected. Because the temperature differences in the single pipe system were relatively small and drinking water and soil temperatures varied a lot over time, there was a need for a high number of datapoints (and the test duration for the 1 and 2 hour travel times turned out to be too short). Because the temperature differences in the DWDN were relatively large, and were found in a broad range, the limited number of datapoints is acceptable.
- WTM+ validation: In the single pipe test, there was only a single pipe diameter, and the soil temperatures around the pipe were more or less the same over the entire pipe length. This ensured that the WTM+ for this single pipe was relatively simple, and validation should be straight forward. However, the drinking water and soil temperatures were not stable over time and some extra data preparation is required before the WTM+ can be validated. It would be worthwhile to also model the upstream drinking water changes. For the test in the DWDN multiple pipe diameters are involved, so the measurements need to be compared to a WTM+ over a trajectory of various pipe diameters and soil temperatures.
- Overall value of the tests: The case studies did prove to be suitable for validating the WTM+, including the effect of residence time (T.B.P.).
- How to measure: in the single pipe system, the drinking water temperature was measured by inserting a sensor in the pipe; in the DWDN at the hydrant. In both cases the influence of the service line and premise plumbing were not present.

Based on the experiences, we recommend the following:

- When selecting a test site we advise to use one with large temperature differences, i.e. a ground water source, and doing the measurements either in winter or summer, when soil temperatures are most different from the ground water temperatures. In this case a resolution of 0.1 °C will not be a problem.
- When taking measurements in a DWDN we advise to use a well calibrated hydraulic network model, select the measurement locations based on expected WTM+ results with a variety of travel times (in the required range of travel times), and at enough locations to be able to not be bothered by a few outliers.

7. ACKNOWLEDGEMENTS

This work is part of the TKI Project Engine

(https://www.tkiwatertechnologie.nl/projecten/engine-energie-en-drinkwater-in-balans/), which is paid for by the ten Dutch drinking water companies, Energie Nederland, Gasunie, Convenant Rotterdam and the Dutch government. We would like to thank Henk de Kater (Evides), and Ben Los (Vitens) for their efforts in the measurement campaigns in Kralingen and Almere.

8. REFERENCES

[1] Agudelo-Vera, C., S. Avvedimento, J. Boxall, E. Creaco, H. de Kater, A. Di Nardo, A. Djukic, I. Douterelo, K.E. Fish, P.L. Iglesias Rey, N. Jacimovic, H.E. Jacobs, Z. Kapelan, J. Martinez Solano, C. Montoya Pachongo, O. Piller, C. Quintiliani, J. Rucka, L. Tuhovcak, and M. Blokker, Drinking Water Temperature around the Globe: Understanding, Policies, Challenges and Opportunities. Water, 2020. 12(4): p. 1049.

[2] Blokker, E.J.M. and E.J. Pieterse-Quirijns, Modeling temperature in the drinking water distribution system. Journal - American Water Works Association, 2013. 105(1): p. E19-E29.

[3] Blokker, E.J.M. and E.J. Pieterse-Quirijns, Erratum - Blokker, E.J.M., Pieterse-Quirijns, E.J. Modeling temperature in the drinking water distribution system - JAWWA 105(2013)1, E19-E29. JAWWA, 2018. 110(10): p. 98.

[4] van Esch, J., BTM+ model development and validation and Expert tool computations. 2022, Delft: Deltares

[5] Blokker, E., A. van Osch, R. Hogeveen, and C. Mudde, Thermal energy from drinking water and cost benefit analysis for an entire city. Journal of Water and Climate Change, 2013. 4(1): p. 11-16.

[6] Vreeburg, J.H.G., E.J.M. Blokker, P. Horst, and J.C. van Dijk, Velocity based self cleaning residential drinking water distribution systems. Water Science & Technology, 2009. 9(6): p. 635-641.

[7] van Esch, J., J. van Summeren, H. de Kater, and E.J.M. Blokker, An enhanced soil temperature model, influenced by drinking water and district heating pipes. T.B.P.

Appendix

$$\Delta TN = \frac{T_{water} - T_{water,0}}{T_{boundary} - T_{water,0}} = \frac{T_w - T_{w,0}}{T_b - T_{w,0}} \tag{5}$$

$$\begin{aligned} E_1 &= \frac{T_w + \varepsilon_w - T_{w,0} + \varepsilon_w}{T_b + \varepsilon_s - T_{w,0} + \varepsilon_w} - \frac{T_w - T_{w,0}}{T_b - T_{w,0}} \\ &= \frac{(T_w - T_{w,0} + 2\varepsilon_w)(T_b - T_{w,0}) - (T_w - T_{w,0})\,(T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s))}{(T_b - T_{w,0})\,(T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s))} \\ &= \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{(T_b - T_{w,0})\,(T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s))} \end{aligned} \tag{6}$$

ε_w: error in water temperature measurement, ε_s: error in soil temperature measurement.

E_1: maximum error in ΔTN to plus side, E_2: maximum error in ΔTN to min side.

$$\begin{aligned} E_2 &= \frac{T_w - T_{w,0}}{T_b - T_{w,0}} - \frac{T_w - \varepsilon_w - T_{w,0} - \varepsilon_w}{T_b - \varepsilon_s - T_{w,0} - \varepsilon_w} \\ &= \frac{-(T_w - T_{w,0} - 2\varepsilon_w)(T_b - T_{w,0}) + (T_w - T_{w,0})\,(T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s))}{(T_b - T_{w,0})\,(T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s))} \\ &= \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{(T_b - T_{w,0})\,(T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s))} \end{aligned} \tag{7}$$

$$\bar{E} = \frac{E_1 + E_2}{2} \tag{8}$$

$$\begin{aligned} \bar{E} &= \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{2(T_b - T_{w,0})} \\ &\times \left(\frac{1}{T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s)} + \frac{1}{T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s)}\right) \end{aligned} \tag{9}$$

$$\begin{aligned} \bar{E} &= \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{2(T_b - T_{w,0})\,(T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s))\,(T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s))} \\ &\times (T_b - T_{w,0} - (\varepsilon_w + \varepsilon_s) + T_b - T_{w,0} + (\varepsilon_w + \varepsilon_s)) \end{aligned} \tag{10}$$

$$\bar{E} = \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{2(T_b - T_{w,0})\left((T_b - T_{w,0})^2 - (\varepsilon_w + \varepsilon_s)^2\right)} \times 2(T_b - T_{w,0}) \tag{11}$$

$$\bar{E} = \frac{2\varepsilon_w(T_b - T_{w,0}) - (\varepsilon_w + \varepsilon_s)(T_w - T_{w,0})}{(T_b - T_{w,0})^2 - (\varepsilon_w + \varepsilon_s)^2} \tag{12}$$

If ε_s small: (13)

$$\bar{E} = \frac{\varepsilon_w(2T_b - T_{w,0} - T_w)}{(T_b - T_{w,0})^2 - \varepsilon_w^2}$$

If $\varepsilon_s = \varepsilon_w$: (14)

$$\bar{E} = \frac{2\varepsilon_w(T_b - T_w)}{(T_b - T_{w,0})^2 - 4\varepsilon_w^2}$$

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14118

SMALL COMMUNITY WATER SUPPLIES IN THE ISIOLO COUNTY, KENYA

Paolo Vezza[1], Roberto Arnesano[2], Fulvio Boano[3], Andrea Bessone[4]

[1,2,3] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Italy
[4] LVIA – Associazione Internazionale Volontari Laici, Cuneo, Italy

[1] paolo.vezza@polito.it, [2] roberto.arnesano@outlook.it, [3] fulvio.boano@polito.it, [4] a.bessone@lvia.it.

Abstract

Lack of access to drinking water mostly affects the sub-Saharan Africa. In Kenya, since 2000, access to safe drinking water has increased by 12 per cent, whereas access to basic sanitation has fallen by five per cent. The Italian NGO LVIA – Lay Volunteers International Association – has been working in Kenya for more than 50 years to improve living conditions of the beneficiary populations, also with interventions in the water sector that contribute to the achievement of the No.6 Sustainable Development Goals (SDGs) which aims to ensure the availability and sustainable management of water and sanitation. Since 2012, several international cooperation projects have been carried out in collaboration with the pastoral communities of the rural areas of the Isiolo County. Isiolo County, located in Northern Kenya, has been plagued for years by severe droughts and floods that put a strain on water access and the already precarious water infrastructure. In this study, we present the rehabilitation of water supply systems in the villages of Duse, Bulesa/Godha, Sericho, Gafarsa, Oldonyiro, Kipsing, Kinna affected by floods at the end of 2019. In addition, the rehabilitation of Boji village water supply system will be described, as an example of good practise for implementing small community water supplies. In particular, the construction of pipelines, boreholes with solar pumping system, water kiosks, livestock troughs will be presented, as well as, the establishment of a water service pricing system, and the training of the water management committee.

Keywords
Rural areas, water supply systems, Sub-Saharan Africa.

1 INTRODUCTION

According to 2016 UNICEF/WHO Joint Monitoring Programme (JMP), the Kenyan population basic water supply level increased by only 12% in 15 years, between 2000 and 2015, whereas access to basic sanitation has fallen by five per cent. In particular, 46% of the population of Kenya (specifically 88% in urban areas and 36% in rural areas) had adequate access to water in 2000, which slightly increased to 58% in 2015 (specifically 83% in urban areas and 50% in rural areas) [1]. Extrapolating this trend, it can be assumed that this value is not greater than 62 % in 2020, with an annual growth rate of 0.8%. Currently, a big portion of the Kenyan population (about 40 %) continues to have no improved access to drinking water, that is, uses unsafe and discontinuous water supply sources such as rivers, dams, pans and ponds, with high risk to people health [2]. These issues mostly affect rural populations in arid and semi-arid areas (ASAL), where the humanitarian and non-governmental organizations (NGOs) investments are concentrated. Like many sub-Saharan countries, water access and sanitation in Kenya has not achieved the Millennium Development Goals (MDGs) targets set for 2015 [3].

Water availability is still quite irregular throughout the country and rainfall occurrence and intensity implies frequent droughts and severe floods due to climate change effects, worsening the risks associated with changes in the distribution and availability of water resources [4]. The alternation of severe droughts and heavy rainfall with flood occurrence is putting a strain on water distribution systems and the relative operating and maintenance costs.

CC BY NC SA

In the present study, we present the rehabilitation of water supply systems in several villages of the Isiolo County affected by floods at the end of 2019. In addition, we describe the rehabilitation of the water supply system in the Boji village, as an example of good practise for implementing small community water supplies.

2 STUDY AREA

Isiolo County is located in a central part of Kenya where the land is mostly flat (Figure 1). The altitude varies gradually within its area, from about 200 m a.s.l. at Lorian swamp (Habaswein) to 300 m a.s.l. at Merti Plateau and finally to approximately 1100 m a.s.l. at Isiolo town [5]. The capital, Isiolo, is the largest town, which is 285 km north from Nairobi, the capital city of Kenya. The total area covered by Isiolo County is 25,336.7 km^2 [6], which corresponds at 4.26% of the total area of Kenya (580,367 km^2 [7]). The county climate is hot and dry for most of the year. The county receives annual rainfall ranging between 100-750 mm [5] and records mean annual temperatures ranging from 24°C and 30°C [8].

Generally, the warmest period is between February-March, while the coolest is July-August, although seasonal variations in temperature are contained. The temperature and rainfall vary according to the prevailing winds that affect the Country, which determine two seasons: from October to March hot and dry winds from Arabia prevail, while from April to September the coolest and wettest winds from the Indian Ocean prevail. At the beginning of the two seasons, the two rainy periods are determined: the long rain season from March to May with the peak in April, and the less intense short rain season from October to December with the peak in November. The rains are generally showers or thunderstorms occurring mainly in the afternoon or in the evening.

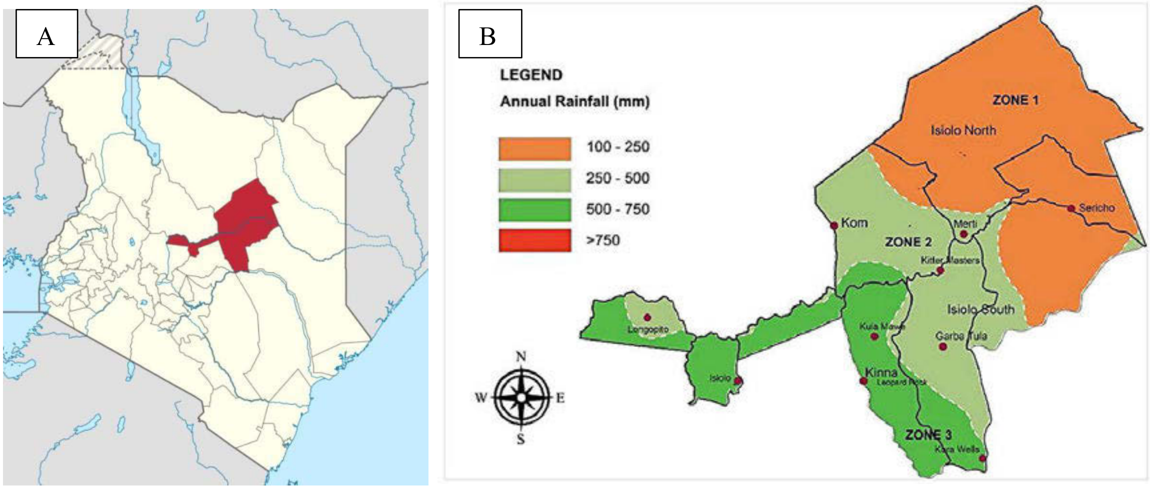


Figure 1: A) Location of the Isiolo County in Kenya. B) Annual rainfall distribution in the Isiolo County

3 PROJECT #1: IMPROVING ACCESS TO SAFE WATER AND HYGIENE PRACTICE

The project "*Improving access to safe water and safe hygiene practice to flood Emergency affected people in Isiolo County*" was funded by United Nations Children's Fund (UNICEF) and implemented in partnership with Lay Volunteers International Association (LVIA). This emergency project aimed to cope with the floods disastrous consequences that had affected many villages in Isiolo County at the end of 2019. As reported by LVIA technicians, after almost a year of drought, starting from the second week of October 2019 torrential rains occurred, exacerbating the water access and water supply issues experienced during the rest of the year.

The worst conditions were found in the Eresaboru and Gafarsa villages where boreholes were submerged and cut off. In Gafarsa, Duse, Bulesa Kipsing, Oldonyiro and Sericho villages, 2.8 km of the aqueduct pipeline were damaged, disrupting safe water access to people. Water pumping equipments for Oldonyiro and Kipsing villages were also damaged. In total the flood affected a population of about 24,000 people in the study area [9]. The emergency project was aimed to repair and desilt the intakes of boreholes, replace the damaged pipelines and pumping systems, and install other ancillary structures to ensure access to safe water for the affected population. In addition, it was proposed to install inline chlorine dosers for Kipsing, Oldonyiro and Kinna villages which rely on surface water abstraction from sand dams and natural springs (Figure 2).

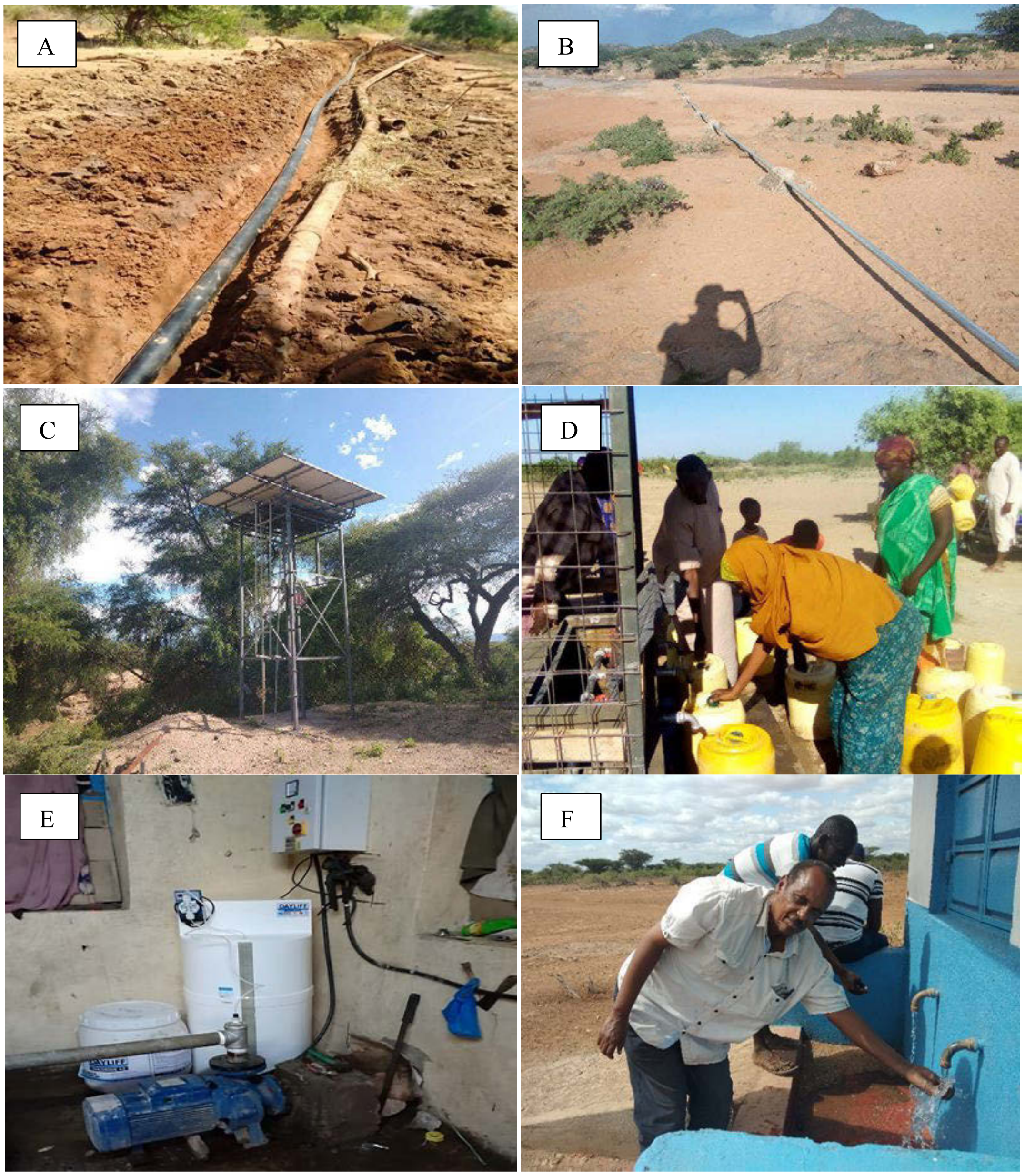


Figure 2. A) New installed HDPE pipes in the Bulesa village B) New pipeline in the Kipsing village. C) Borehole in the Kipsing Village with added solar panels for the new pumping system. D) Women fetching water at a community water point in Sericho village. E) Inline chlorine dosing equipment installed for Kinna village water supply system. F) The director of the water service during the launch of the new water kiosk in the Gafarsa village

LVIA coordinated the project activities in collaboration with the Isiolo County Departments of Water Services and Public Health and other local agencies and stakeholders. Community groups and beneficiaries actively participated in the implementation of the project. The programme was implemented in the wake of the COVID-19 pandemic. Therefore, some of the activities were aligned to the COVID-19 response, mainly to support prevention and control awareness through hygiene promotion activities.

The results of the project are listed below, partially redefined due to the COVID-19 pandemic:

Result 1: by 30th April 2020, 17,384 persons (8,969 male, 8,415 females of which 6,954 children) had permanent access to safe water at 7.5 to 15 litres/person/day from repaired water sources and water supply systems in the following villages: Duse, Gafarsa and Seiche in Garbatulla subcounty, Bulesa in Merti Sub-County and Oldonyiro and Kipsing in Isiolo Sub-County.

Result 2: by 30th April 2020, 1,354 school children (730 girls and 624 boys) had access to safe water at 1-2 litres/child/day from repaired water supply systems in Oldonyiro and Kipsing Primary and secondary schools.

4 PROJECT #2: REHABILITATION OF BOJI VILLAGE WATER SUPPLY SYSTEM

The project "*Rehabilitation of Boji Village Water Supply System*" was funded by Italian Government and United Nations Children's Fund (UNICEF) and implemented by Lay Volunteers International Association (LVIA) in partnership with National Drought Management Authority (NDMA) and in collaboration with the Isiolo County Departments of Water Services and Livestock.

This activity was part of a larger project – "*Improving access to water and the resilience of pastoral communities in Isiolo County, Kenya*", funded by Italian International Cooperation Programme – that had the general aim of improving the living conditions of pastoral communities of Isiolo County. The specific purpose was to develop climate change resilience and mitigation mechanisms through water supply protection and rehabilitation and, at the same time, through increasing capacity building of the local community.

The water supply protection and rehabilitation were completed within 18 months. The project achieved 100% of the targets set at the baseline which were measured through key indicators. This implies that the project improved the availability of safe water, through the rehabilitation of two boreholes and the construction of new pipelines and four water kiosks (Figure 3). This allowed to separate water for human use and livestock use, to provide water allocation also to schools, health dispensaries and mosques. Solar pumping systems, coupled with a diesel generator backup system, were installed to reduce water pumping operational costs.

The project was implemented using a participatory approach with active community engagement throughout the project lifetime. In addition, a proper water management structure was put in place to insure the sustainability of the system. This implied to create a water management committee in charge or insuring public access to water, collect water fees and manage funds for the system maintenance and management.

The direct beneficiaries reached through the project implementation were 2,175 people (1,110 males, 1,065 females), out of which 236 were school students. Indirect beneficiaries (i.e. people living within the zone of influence of the project that may have access to water) were 4,544 (2,113 males, 2,431 females). Livestock population, benefiting of watering system, were 15,000 shoats, 2,600 cattle and 4,500 camels.

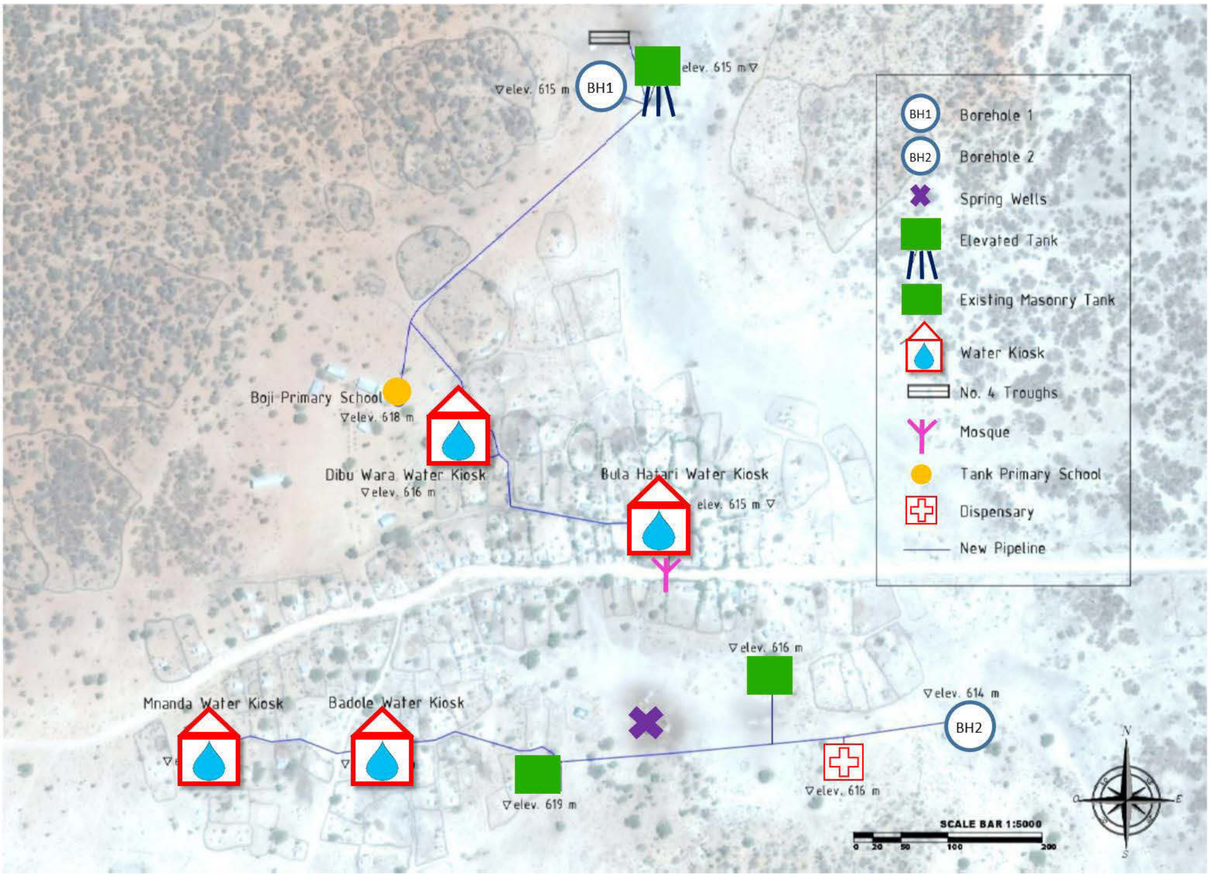

▽elev. 615 m
BH1
elev. 615 m▽
Boji Primary School
▽elev. 618 m
Dibu Wara Water Kiosk
▽elev. 616 m
Bula Hafari Water Kiosk
elev. 615 m ▽
▽elev. 616 m
▽elev. 614 m
BH2
Mnanda Water Kiosk
Badole Water Kiosk
▽elev. 619 m
▽elev. 616 m
BH1 Borehole 1
BH2 Borehole 2
Spring Wells
Elevated Tank
Existing Masonry Tank
Water Kiosk
No. 4 Troughs
Mosque
Tank Primary School
Dispensary
New Pipeline
N
O
E
S
SCALE BAR 1:5000
0 20 50 100 200

Figure 3: New water supply system at the Boji village.

5 DISCUSSION

Although technical notions and technologies for water distribution and allocation are roughly the same in all circumstances, their implementation is strongly affected by the social context. From a project management point of view, the construction or rehabilitation of water supply systems in rural areas of Kenya may consider a strong involvement of beneficiaries in the project activities. A management committee should be put in place to ensure the maintenance of the water supply system. The water management committee should be trained with technical and business skills, and it is important that committee members are salaried. Each water committee should develop a business plan in order to anticipate future investments, equipment replacement and maintenance. In addition, water committees should report revenue, costs and activities to communities for transparency.

From a technical point of view, pumping water from boreholes can be carried out using solar energy. The implementation of this technology is feasible in rural areas of Kenya and can be considered convenient both from an economic and management point of view. To insure continuous water distribution a backup system (e.g. diesel generator set) should be considered. Safe water access can be insured by the installation of water kiosks in strategic places, supported by appropriate water storage facilities that guarantee water availability in case of power failure or interruptions of pumping systems.

6 REFERENCES

[1] M. Mwangi, B. Gichangi, B. Kinuthia, J. Samo-Project Manager, and I. Githu-Project Director, “Evaluation of the Sustainability of Solar Powered Water Supply Systems in Kenya” 2018. [Online]. Available at: www.solarzzy.comwww.eedadvisory.com

[2] WHO and UNICEF, “Progress on Drinking Water, Sanitation and Hygiene - Joint Monitoring Programme 2017 Update and SDG Baselines,” 2017. doi: 10.1111 / tmi.12329.

[3] USAID, “USAID water and development country plan for Haiti,” 2016. [Online]. Available at: https://dec.usaid.gov/dec/content/Detail.aspx?vID=47&ctID=ODVhZjk4NWQtM2YyMi00YjRmLTkxNjktZTcxMjM2NDBmY2Uy&rID=MjM1MTM5

[4] R. Connor, The United Nations World Water Development Report 2015: Water for a Sustainable World, vol. 4, no. 2. 2016.

[5] R. of Kenya and I. C. Government, “Isiolo County Integrated Development Plan, CIDP, 2018-2022,” 2018.

[6] I. C. Government, “County Government of Isiolo,” 2020. https://isiolo.go.ke/.

[7] United Nations Statistics Division, “Demographic Yearbook—Table 3: Population by sex, rate of population increase, surface area and density” pp. 1–13, [Online]. Available at: https://unstats.un.org/unsd/demographic/products/dyb/dyb2005.htm.

[8] B. Mati, J. Muchiri, K. Njenga, F. P. de Vries, and D. J. Merrey, “Assessing water availability under pastoral livestock systems in drought-prone Isiolo District, Kenya,” in Water Management, 2006, p. 13.

[9] National Drought Management Authority, “Isiolo Flood Risk Assessment Report Dec 2019” 2019.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14127

ADVANCING TOWARDS SEMI-AUTOMATIC LABELING OF GPR IMAGES TO IMPROVE VISUALIZATIONS OF PIPES AND LEAKS IN WATER DISTRIBUTION NETWORKS USING MULTI-AGENT SYSTEMS AND MACHINE LEARNING TECHNIQUES

Gemma Stanton[1] and David Ayala-Cabrera[2]

[1,2] CWRR-School of Civil Engineering, University College Dublin, Dublin (Ireland)

[1] *gemma.stanton@ucdconnect.ie*, [2] *david.ayala-cabrera@ucd.ie*

Abstract

Critical infrastructures such as water distribution networks (WDNs) require reliable and affordable information at a reasonable cost to address challenges that can negatively affect their operation. Inadequate knowledge about WDN assets and their state of health presents challenges for essential activities such as network modeling, operation, assessment, and maintenance. This work seeks to increase the availability of WDN asset data through improved interpretability of GPR images. The semi-automatic labeling approach presented here expands upon existing multi-agent image-cleaning methods and feature characterization techniques. The division of a pre-processed image, in the form of a matrix, into a grid of smaller blocks allowed the identification of relevant features using density of nonzero values in the blocks; this approach, conducted manually in this proof of concept, can provide a basis for training an intelligent system (e.g., a convolutional neural network) to extract the families of interest and eliminate noise. Thus, this research expands this methodology to advance towards automatic detection of pipes and leaks and easily visualize the data. In this paper, 3D visualizations of WDN assets have been created to demonstrate the usefulness of this semi-automatic process in delivering easily-interpretable GPR data for managers and operators of WDNs.

Keywords

Water distribution networks, GPR image interpretation, 3D pipe models, Non-destructive testing methods, Water leakage, Intelligent data analysis, Multi-agent systems, Semi-automatic labeling of GPR images.

1 INTRODUCTION

Critical infrastructure such as water distribution networks (WDNs) present many challenges in their construction, operation, and maintenance. Effective management of these infrastructures requires economical, accessible, and reliable information [1] for decision-making processes. Information on the health of WDNs supports general operation activities such as modeling, operation, assessment, and maintenance [2] as well as resilient response to climate change [3], increasing digitization [4], and other such advances. WDNs and other buried assets may be difficult to assess and even locate, since manual inspection requires intensive labor and often necessitates excavation and significant social and economic disruption and even safety risks to carry out [5]. Minimizing the negative impacts of disrupting the WDNs and related infrastructure (*e.g.*, roads) is imperative [6] and calls for the incorporation of non-destructive testing methods to gather compile data on WDN health. Thus, non-destructive testing methods have been incorporated into the surveys to address the need for non-invasive procedures that provide relevant information on buried assets, including WDNs, gas pipelines, and more [7]. Non-destructive testing methods such as ground penetrating radar (GPR) are easy to deploy, but they present several difficulties (*e.g.*, absence of consistent GPR data interpretation protocols,

BY NC SA

excessive noise, and heterogeneity of environmental conditions and equipment settings [8]) in the data interpretation process.

GPR is a non-destructive testing method which is deployable in a variety of environmental conditions (*e.g.,* urban environments [9]) and does not require disruption in the operation of the critical infrastructure inspected; *e.g*., water supplies in WDN [1]. GPR is examined in this research as an alternative to destructive testing to obtain relevant information regarding buried assets of WDN due to its potential to detect components (*e.g.,* pipes) and their integrity with great detail (*e.g.,* pipe with water leak [10], clogged pipe [11], deteriorated pipe [12]). Among the main advantages of the GPR as a non-destructing method, its inspection ability for successfully inspect different pipe materials has been evidenced in several studies (see [11, 12, 13]). GPR signals are capable of detecting materials, both metallic and non-metallic [14] (*e.g.*, PVC, asbestos cement, polyethylene), which are commonly used in assets of WDNs. GPR has the sensitivity to detect leaks in the media surrounding pipes [8], and this non-destructive method also offers accurate depth estimation and has an advantage of high resolution of up to a centimeter [15]. In the field of leak detection, other non-destructive methods such as thermography and tracer gas are limited by temperature, leak depth, and disruption to WDNs [16]. Furthermore, GPR is an inexpensive and non-destructive method, and it is easy for unskilled personnel to operate. However, the raw images produced by current GPR method require personnel with expertise to interpret them, increasing the cost and time that this method demands for analysis, and causes human interpretation errors [17]. While current GPR analysis requires expert interpretation, innovative processing methods can facilitate overcome this barrier.

Reliable data about WDNs can help their managers make responsible asset management decisions on when and how to make repairs and maintenance. Additionally, leak detection through GPR image analysis has the potential to prevent (*i.e.* by favoring the detection of water leaks at early stages) the waste of water, energy, and infrastructure [18] in those leaks that can only be identified when damage to a road surface occurs (i.e. when leaked water is visible [19). A raw GPR image may be used to detect and characterize a variety of subsurface assets (e.g. pipes of drinking water [20], gas [21], among others), but it is not easily interpretable, particularly for unskilled personnel as mentioned above. Although the raw GPR images are difficult to interpret, they contain a wealth of data that can be processed to extract useful information [22]. The work presented in this paper will serve managers and operators of WDNs who need accurate information about their networks. Basic data such as size and layout of pipes as well as material, interior build-up, leakage, and illegal connections can be detected more easily with this image cleaning technology.

The generation of 3D models of buried assets is an important final step in GPR data analysis since a key motivation of this work is to enhance site safety and deliver interpretable information for any operator without expertise [23, 24]. The 3D models generated from the information obtained after processing the GPR images can be used to feed augmented reality visualizations that can facilitate the assets assessment process [25] in a dynamic manner. The literature has examples of 3D model generation ranging from simple raw data plotting [16] to contours created from a series of pre-processed cross sections [8, 26]. These contours created from pre-processed cross sections are a significant step towards interpretable GPR images; however, they rely on comparison of initial and final states of leaking pipes to extract the relevant features and construct a 3D image. In this work, this comparison is eliminated from the process using semi-automatic extraction of relevant features and reconstruction of cross-sectional profiles in 3D space. The results displayed here will demonstrate that it is possible to reconstruct 3D representations of buried pipes from a single set of GPR data collected at a the same time period.

2 METHODS

This section presents the treatment of GPR data proposed in this paper both for the semi-automatic labeling of the information and its previsualization in a 3D model. The generation of 3D models from GPR images can consist of 6 steps of which the current paper focuses particularly on the semi-automatic labeling, through a density boundary extraction, and its subsequent 3D model previsualization (see *Figure 1*). Each step of this process is described in the subsections below.

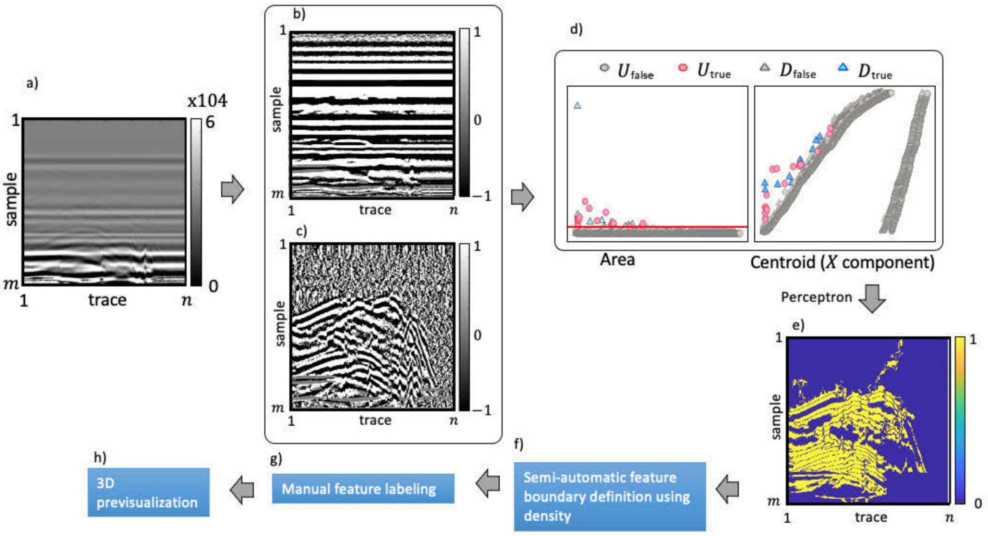


Figure 1. Proposed semi-automatic labeling of GPR images and 3D model previsualization. a) Raw GPR image, b) multi-agent pre-processing (vertical direction), c) multi-agent pre-processing (horizontal direction) d) semi-automatic feature extraction, e) semi-automatic feature extraction (visualization), f) semi-automatic labeling: density boundary, g) manual feature labeling, and f) 3D previsualization.

2.1 Raw GPR Images

GPR functions by propagating electromagnetic waves from a transmitter antenna below the ground at a specific velocity which is related to the medium through which it travels [7]. Parts of these waves return to a receiver antenna whenever an interface between two different media is reached; the wave is then partly propagated to deeper layers [17]. The reflected electromagnetic wave that is received by the receiver antenna has an amplitude proportional to the dielectric constant of the media through which it travels; a reflection at an interface between two materials correspond to a change in wave amplitude [22]. Thus, a raw GPR image, or radargram, contains a record of all the wave amplitudes recorded by a receiver in a matrix, A, of size m by n, where m represents the total two-way travel time of all the received signal of each trace ($i = 1, \dots, m$) and n corresponds to the total of traces captured ($j = 1, \dots, n$) with the survey (see Figure 1). A naked eye is unlikely to detect a hyperbola in a radargram, making these images difficult to interpret without processing.

2.2 Pre-Processing

[27] developed a multi-agent algorithm called "agent race" based on game theory to reduce the dimensionality of matrix A and prepare it for feature detection. This multi-agent algorithm pre-processes the radargram based on wave amplitudes intensity, as shown in Figure 2. The result of this algorithm is an $m1 \times n$ matrix (see Figure 2, Output 1), where the key features are registered

CC BY NC SA

as movements of the agents in each j trace, and m1 represents the maximum total of movements obtained by the agent winner of the competition.

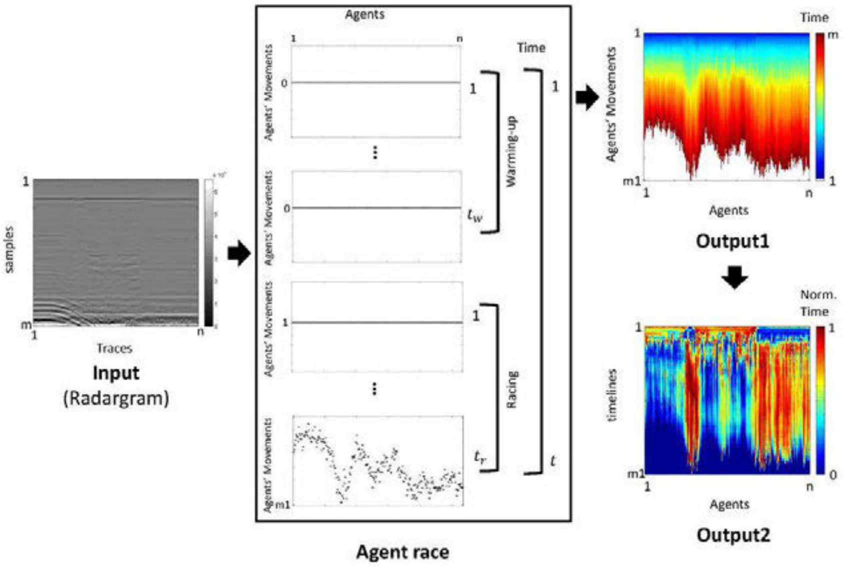


Figure 2. Agent race algorithm process. Adapted from [10].

The uses of the multi-agent pre-processing described above can generate groups (families), that contain features (in the form of functions) that specifically represent noise, horizontal lines, and objects (Figure 1, insets b and c). According to [28], agents run their moves on the rows of the A matrix (see the output in Figure 1c) and not on the columns (as it originally designed, see the output in Figure 1b) the removal of families that contain horizontal lines from these GPR images can be obtained. This removal was completed in [28] and [29] with the incorporation of noise removal through the use of perceptron neural network. In these studies, the geometric properties of the generated groups were used both to label (Area) and as a key feature for classification (X component of the centroid) of them (see Figure 1d).

Separating the noise and horizontal lines from the raw GPR images allows further analysis to detect features of interest in the objects group without the hindrance of noise and horizontal lines. This can be observed by comparing the raw GPR image (Figure 1a) with the resulting preprocessed image, A' matrix (Figure 1e). The $m \times n$ matrix A' (pre-processed image), consists of 0 and 1 values, where nonzero values indicate a feature of the target objects.

2.3 Semi-Automatic Labeling and Refinement

Labeling is an essential task in machine learning, in particular when the classification is conducted via supervised learning. This pre-preprocessing activity is conducted on various occasions manually as a preamble for the classification (via machine learning methods) of the embedded objects into the GPR images (*e.g.*, subsoil background [20], metallic and no- metallic pipes [31], among others). In order to reduce the dependency on personnel with high experience in interpreting GPR images and minimize the human errors that the manual labeling can generate, the analysis of densities is proposed as an alternative in this paper. In this sense, the GPR images cleaned by both the agents and the perceptron (see Section 2.2.), that had horizontal lines and noise removed, were used in this section as the starting point for the semi-automatic labeling proposal and to generate images that can be more easily interpreted.

Families of functions that represent embedded objects (*e.g.*, pipe objects) were preserved in A' matrix. These families contain additional information about the objects beyond the initial hyperbola identification that is commonly used to label the objects embedded into the GPR images [10, 32]. However, many of these families exist in the images and further analysis is necessary to extract and classify those families of interest. In order to locate the features of interest, the proposed density analysis can be formulated as in equation (1) [19].

$$B_{ir,jr} = \sum_{iy_{ir}}^{iy_{ir+1}} \sum_{jx_{jr}}^{jx_{jr+1}} \left(\frac{a'_{(iy_{ir},iy_{ir+1}],(jx_{jr},jx_{jr+1}]}}{(iy_{ir+1} - iy_{ir} + 1)(jx_{jr+1} - jx_{jr} + 1)} \right) \tag{1}$$

In there, the pre-processed images in A' were further processed by dividing them into a grid of blocks ($a'_{iy,jx}$) with size $iy = \{1: \frac{m-1}{ly} : m\}$ (the step in y-axis) by $jx = \{1: \frac{n-1}{lx} : n\}$ (the step x-axis). Where ly and lx represents the dimension of the new matrix, B, generated for the blocks. Each block in B consists of $ir = \{1, \dots, ly\} \in \mathbb{N} = \{1,2,3 \dots\}$ and $jr = \{1, \dots, lx\} \in \mathbb{N} = \{1,2,3 \dots\}$. In essence, B represents the density calculation of nonzero values, from A', in each grid block. Figure *3* presents an example of a grid of 75 × 75 blocks which represents the subdivision of matrix A' (inset a) and the density of each block in a color scale (inset b).

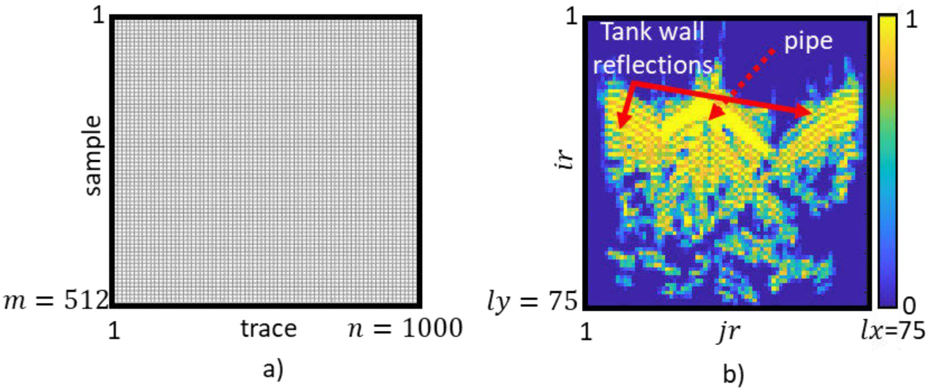


Figure 3. Generation of the density matrix B. a) Grid of 75×75 *blocks (*$ly = lx = 75$*) for A' matrix and, b) the respective densities; matrix B.*

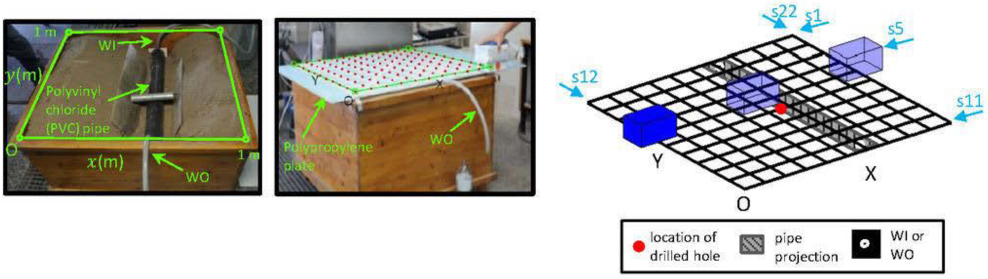
The example presented in Figure *3* corresponds to a GPR image of a pipe buried in a wooden tank to emulate the phenomena of water leaks. More detailed information is presented in Section 3. Figure *3*b the hyperbola in the center of the image corresponds to the pipe itself, and is the most dense part of the figure. This detection of the hyperbola by the density method proposed herein, in addition to the highlighting of the tank reflections on the sides, shows the use of density as a useful tool for extracting features from GPR images.

Matrix B can be rescaled to the dimension of the original matrix A' (*i.e.*, from size of $ly \times lx$ to size of $m \times n$). Based on the observation of the densities generated by the object in the resulting matrix B (resized), it is proposed to iteratively determine a density threshold for each object. This thresholding generates groups that are captured by using the Matlab's `bwboundaries` function. In this work, we have selected the boundary of the group with the largest area as a representative element of the desired object to be extracted.

This extraction process can be used to label certain and particular objects of the GPR image in order to feed more robust classification processes. Likewise, these labels can be used to refine the pre-processing process presented above. This will allow the inclusion of new parameters to be considered, among other advantages. In addition, it is already possible to preview these images in 3D models as a preamble to automatic visualizations and it is in this sense that we will discuss about this process in Section 3.

3 CASE STUDY

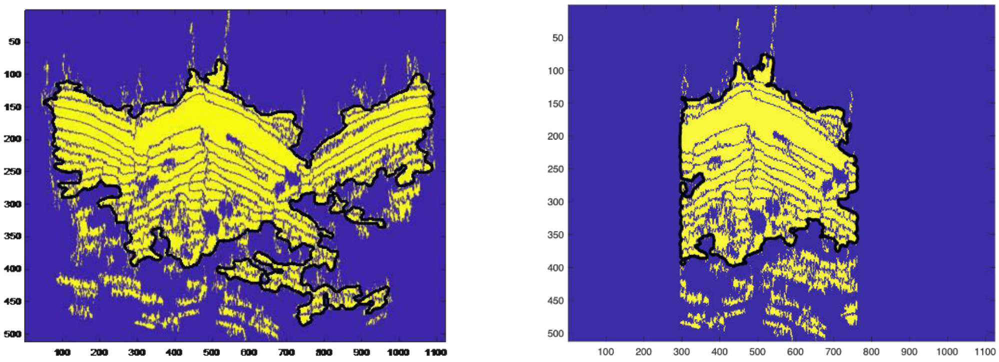
In this section, we present both the application of the proposed GPR image labeling procedure and its results through 3D reconstructions. To do this, we have used as a case study the work presented in [10], where GPR images of a pipe leaking water were collected (at different stages of maturity of the phenomena) in a laboratory set up, depicted in Figure 4. In this configuration, a PVC pipe with diameter 100 mm and length of 0.95 m with one hole drilled to mimic a leak in the pipe and two points water input (WI) and water output (WO) was buried in dry soil in a wood tank (with size of 1.0 m × 1.0 m × 0.70 m). The tank was covered in a polypropylene plate with eleven paths parallel to the X-axis (transversal paths) and eleven paths parallel to the Y-axis (longitudinal paths), each path spaced 0.10 m from the next. A GPR (with a central frequency of 1.5 GHz) was run over each path as shown in Figure 4c, s5. Two samples of each of the 22 profiles were taken, one set without water in the setup and one set with water in the pipe and leakage around it. These two sets of samples are referred to as without water and with water, respectively.



a) b) c)

Figure 4. Tank configuration (from [10]).

Density-boundary analysis. By using the density boundary procedure described in Section 2.3 and adjusting the number of blocks (both ly and lx) and density threshold, the pipe reflections were removed from transverse images to capture the boundary of the relevant feature (pipe, leak or tank reflection). Figure 5a shows a transversal profile, as an example, in which the boundary includes both the pipe and the tank reflections. Figure 5b shows how removing the tank reflections on the periphery of the central hyperbola allowed the pipe boundary to be captured.



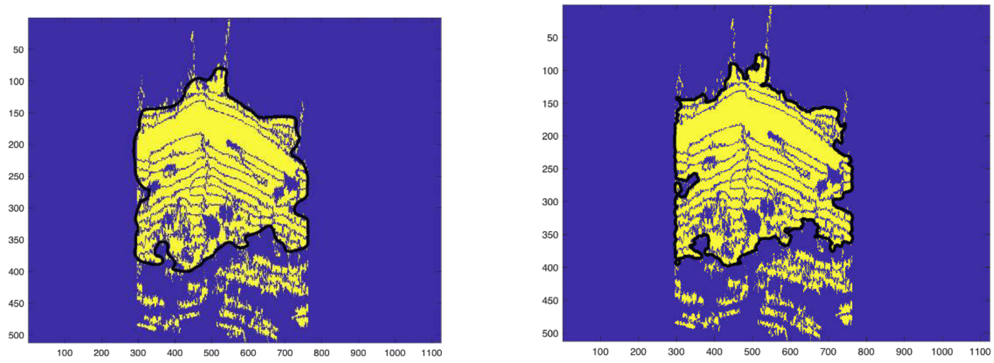
a) b)

Figure 5. Example of the labelling of the pipe in Profile 11 (s11) with the use of density-boundary process. a) before peripheral tank reflection removal and, b) after peripheral tank reflection removal.

The step and the density affect the resolution with which the boundaries of the features will be traced. The y blocks (ly) and x blocks (lx) the required density to locate the relevant features in each profile were adjusted manually in order to determine the optimal values for each parameter. The adjustments to these parameters ensured that the boundary of the relevant feature would be

CC BY NC SA

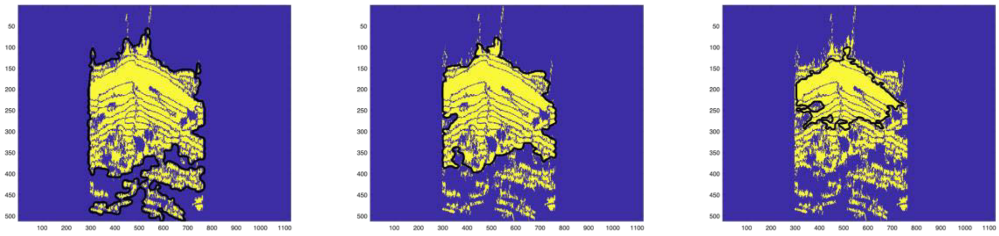
as accurate as possible. For example, Figure 6 shows how increasing the x and y blocks from 40 (Figure 6a) to 100 (Figure 6b) increases the precision of the boundary.



a) b)

Figure 6. Profile 11, a) $ly = lx = 40$ *and, b)* $ly = lx = 100$*.*

Some images contained spots of noise that would be included in the boundary if the required density of the feature was low (close to a value of 0); increasing the density would skip over this noise. In addition, a very high density (close to a value of 1) would skip over relevant parts of the desired boundary. Figure 7 illustrates the results for three density values.



a) b) c)

Figure 7. Profile 11, a) density threshold=0.1 b) density threshold=0.4 and, c) density threshold=0.8.

For the profiles taken without water, the only boundaries identified were the tank reflection and the pipe.

Preliminary classification. The three main classes of objects in the profiles are tank reflection, pipe, and leak. It should be noted that in longitudinal profiles (s12-s22), pipes are represented as horizontal lines, and thus are removed in the pre-processing stage. As a result, no pipe classes exist in the longitudinal profiles. These classes were identified for each boundary in order to produce a clear 3D image. For the profiles taken with water into the system, boundaries obtained fell into five classes:

- Pipe,
- Leak,
- Pipe+leak (this due to difficulty in separating the two in one of the profiles),
- Tank reflection,
- Tank reflection+ leak (this due to difficulty in separating the two in one of the profiles).

The process of identifying boundaries based on determining an appropriate grid size and density threshold was conducted on transversal profiles (s1-s11). For the case of the system with water, the boundaries obtained for s1-s11 are displayed in Figure 8.

CC BY NC SA

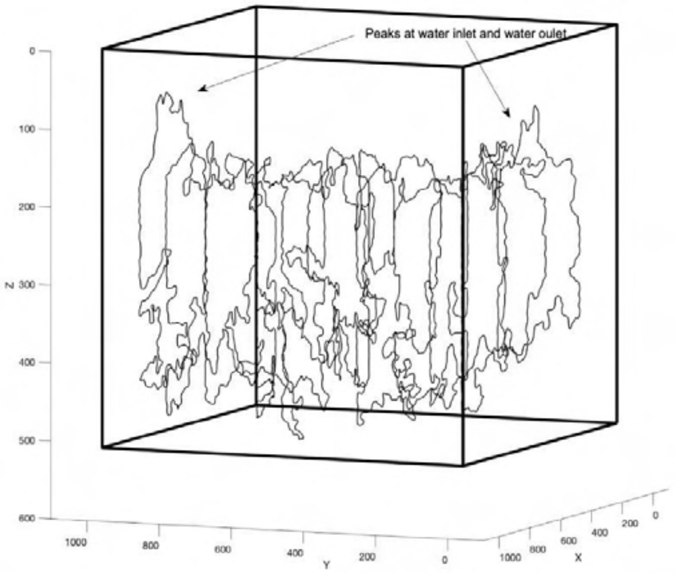


Figure 8. Boundary of pipe/pipe+leak for each transversal profile.

In Figure 8, the inlet and outlet for the system with water can be seen as peaks in the first and last profile. In [10], these peaks at the inlet (Figure 9a) and outlet (Figure 9k) are also observed in the images obtained through initial/final (without/with water into the system) image contrast. The capacity to identify the boundaries of the objects embedded into the GPR images without the requirement of a reference image to contrast, is one of the main virtues of the system proposed in this paper.

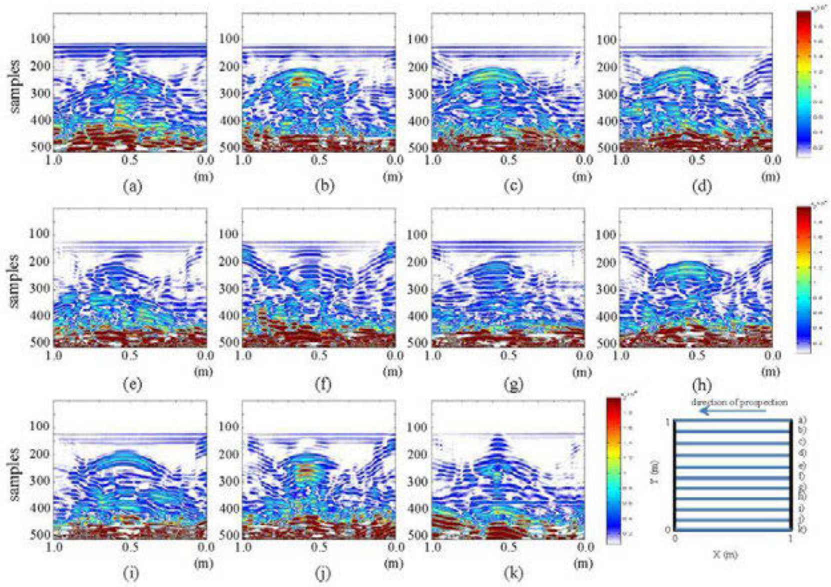
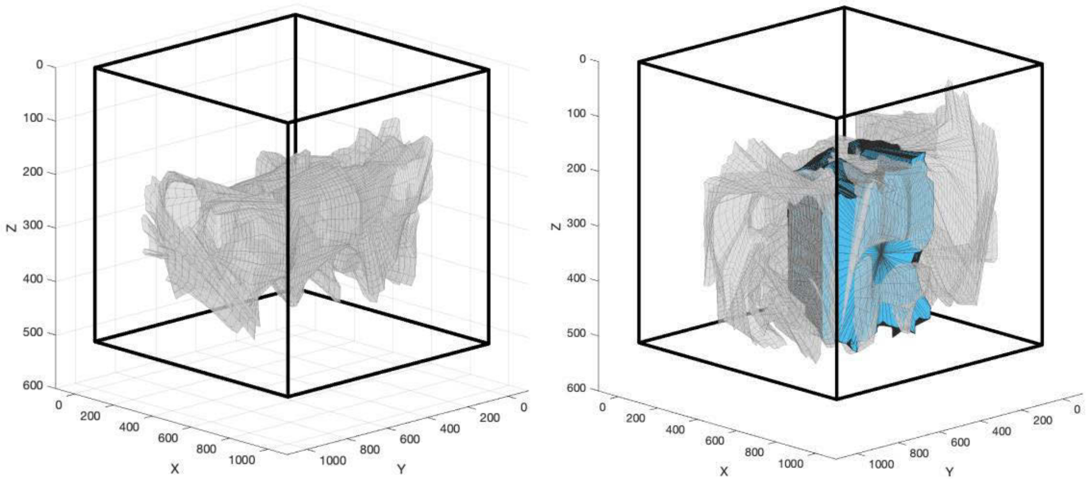


Figure 9. Images obtained using contrasts between initial and final states for transversal images (from [10]).

3D Image Construction. After the boundaries of these features were obtained, they were plotted in 3D space. The `interp2` function in Matlab was used to interpolate between 2D profiles. The known distance from the tank edge was inserted as X (for longitudinal) and Y (for transversal) coordinates. These distances are based on the GPR operator's selection of surface lines along which to take profile data. The interpolation between the boundaries of each profile without and with leakage in the transversal direction yielded the results in Figure 10(insets a and b; respectively).



a) b)

Figure 10. 3D reconstruction of pipe and leak for a) without water and b) with water into the system.

The pipe and leakage are clearly visible in the 3D image. The pipe/leakage representation in the transversal direction (gray) extends laterally past the leakage from the longitudinal direction (blue) in a few points as a result of the leakage from the longitudinal direction being difficult to capture in several profiles due to comprising disconnected shapes. For example, Figure 11 displays spots that represent the leak, that were not dense enough to be captured as one unit.

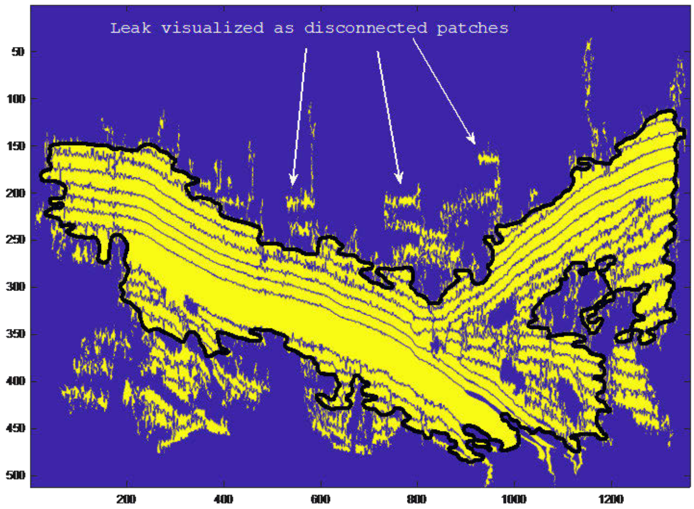


Figure 11. Profile 20, with tank reflection outlined by the boundary and leaks faintly represented.

Figure 12 presents the success in outlining the tank reflections in the longitudinal direction by means of the proposed semi-automatic labeling process. Outlining these reflections was useful to characterize their shape in the pre-processed GPR images, and may allow conversion to their real shape (a box) in the future.

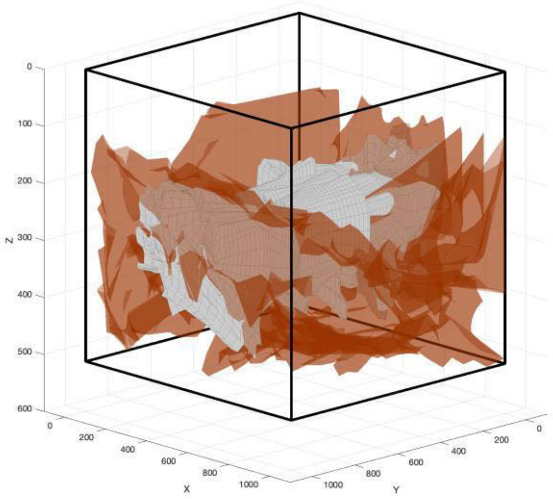


Figure 12. Representation of pipe and tank reflections for the system without water.

A few difficulties arose in the reconstruction of the 3D images from GPR data. For example, the sizes of the radargram matrix A for each profile had the same n dimension and varying m dimensions. This variation was due to the speed at which the GPR was run along the profile, taking a 120 traces/second, regardless of the total distance covered. This resulted in the first 3D reconstructions being askew. This problem was remedied by incorporating the estimated spatial coordinates for each trace in the analysis. However, the interpolation used in this reconstruction resulted in some inaccuracies due to the number of points in each cross section. For example, the interpolation between profiles 4 and 5 resulted in the constriction of the diameter of the profile as shown in Figure 13. This discrepancy in the interpolation may be resolved through the use of another interpolation function that will be tested in future work.

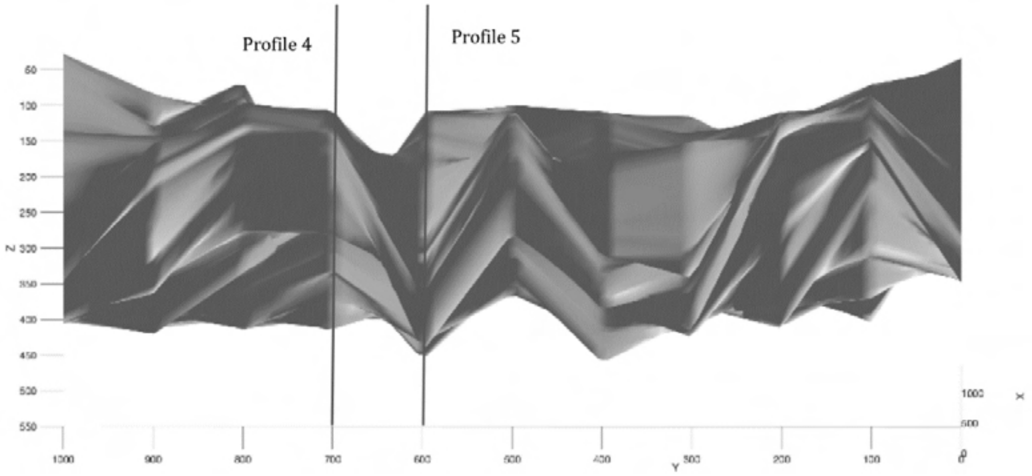


Figure 13. Transversal 3D image in Y-Z plane.

4 CONCLUSIONS

This approach to 3D image construction from GPR data began with a radargram matrix of the raw GPR data that was pre-processed using multi-agent techniques. The semi-automatic labeling of relevant features was conducted manually on the pre-processed images using the numerical approach of assessing the density of nonzero values in the pre-processed matrix. By extracting desired boundaries manually and plotting these, the need for initial/final state comparison was eliminated. Basic identification of classes was also possible due to the manual extraction and classification of boundaries. This semi-automatic approach to labeling through density of grid blocks, as explained, offers a potential pathway to machine learning.

5 REFERENCES

[1] A. Aijazi, L. Malaterre, L. Trassoudaine, T. Chateau, and P. Checchin, "Automatic Detection and Modeling of Underground Pipes Using a Portable 3D LiDAR System," Sensors, vol. 19, no. 24, Dec. 2019, pp. 1-21.

[2] J. Liang, Q. Liu, H. Zhang, X. Li, Z. Qian, M. Lei, X. Li, Y. Peng, S. Li, and G. Zeng, "Interactive Effects of Climate Variability and Human Activities on Blue and Green Water Scarcity in Rapidly Developing Watershed," Journal of Cleaner Production, vol. 265, Aug. 2020, pp. 1-13.

[3] M. Karl, E. Culbertson, and J. Abrera, "Optimize Water Distribution Pipes and Water Loss with Digital Solutions," Journal AWWA, vol. 113, no. 7, Sept. 2021, pp. 69-73.

[4] P. Lau, B. W. Y. Cheung, W. W. Lai, and J. Sham, "Characterizing Pipe Leakage with a Combination of GPR Wave Velocity Algorithms," Tunnelling and Underground Space Technology, vol. 109, no. 1, Mar. 2021, pp. 1-12.

[5] T. Hao, C.D.F. Rogers, N. Metje, D.N. Chapman, J.M. Muggleton, J.Y. Foo, P. Wang, S.R. Pennock, P.R. Atkins, S.G. Swingler, J. Parker, S.B. Constello, M.P.N. Burrow, J.H. Anspach, R.J. Armitage, A.G. Cohn, K. Goddard, P.L. Lewin, G. Orlando, M. Redfern, A.C.D. Royal, and A.J. Saul, "Condition Assessment of the Buried Utility Service Infrastructure," Tunnelling and Underground Space Technology, vol. 28, Mar. 2012, pp. 331-344.

[6] S. Kerwin and B. Adey, "Optimal Intervention Planning: a Bottom-Up Approach to Renewing Aging Water Infrastructure," Journal of Water Resources Planning and Management, vol. 146 no. 7, Jul. 2020, pp. 1-16.

[7] M. Gupta, M. A. Khan, R. Butola, and R. M. Singari, "Advances in Applications of Non-Destructive Testing (NDT): a Review," Advances in Materials and Processing Technologies, Apr. 2021, pp. 1-22.

[8] D. Ayala-Cabrera, J. Izquierdo, S.J. Ocaña-Levario, and R. Pérez-García, "3D Model Construction of Water Supply System Pipes Based on GPR Images," in Proc. 7th International Congress on Environmental Modelling and Software, San Diego, C.A., USA: iEMSs, 2014.

[9] P. Song, Z. Pu, B. Ren, X. Tang, and J. Wang, "Application of Transient Electromagnetic Method in Municipal Buried Gas Pipelines Detection," in Proc. ASME 2020 Pressure Vessels and Piping Conference, Aug. 2020.

[10] D. Ayala-Cabrera, M. Herrera, J. Izquierdo, S.J. Ocaña-Levario, and R. Pérez-García, "GPR-Based Water Leak Models in Water Distributions Systems," Sensors, vol. 13, no. 12, Nov. 2013, pp. 15912-15936.

[11] S. Zhang, Y. Li, G. Fu, W. He, D. Hu, and X. Cai, "Wavelet Packet Analysis of Ground-Penetrating Radar Simulated Signal for Tunnel Cavity Fillings," Journal of Engineering Science and Technology Review, vol. 11, no. 6, Dec. 2018, pp. 62-69.

[12] C. Lin, X. Wang, Y. Li, F. Zhang, and Z.H. Xi, "Forward Modelling and GPR Imaging in Leakage Detection and Grouting Evaluation in Tunnel Lining," KSCE Journal of Civil Engineering, vol. 24, no. 1, Dec. 2019, pp. 278-294.

[13] R. Ahmadi, N. Fathianpour, and G.H. Norouzi, "Detecting Physical and Geometrical Parameters of Some Common Geotechnical Targets through their Effects on GPR Responses," Arabian Journal of Geosciences, vol. 8, Jul. 2014, pp. 4834-4854.

[14] A. S. Rao, M. Radanovic, Y. Liu, S. Hu, Y. Fang, K. Khoshelham, M. Palaniswami, and T. Ngo, "Real-Time Monitoring of Construction Sites: Sensors, Methods, and Applications," Automation in Construction, vol. 136, April 2022, pp. 1-22.

[15] S. Li, H. Cai, and V. Kamat, "Uncertainty-Aware Geospatial System for Mapping and Visualizing Underground Utilities," Automation in Construction, vol. 53, May. 2015, pp. 105-119.

[16] A. De Coster, J.L Pérez Medina, M. Nottebaere, K. Alkhalifeh, X. Neyt, J. Vanderdonckt, and S. Lambot, "Towards and Improvement of GPR-Based Detection of Pipes and Leaks in Water Distribution Networks," Journal of Applied Geophysics, vol. 162, Mar. 2019, pp. 139-151.

[17] N. Kim, K. Kim, Y.K. An, H.J. Lee, J.J. "Deep learning-based underground object detection for urban road pavement," International Journal of Pavement Engineering, vol. 21, no. 13, Jan. 2020, pp. 1638-1650.

[18] S. Ali, M. A. Hawwa, and U. Baroudi, "Effect of Leak Geometry on Water Characteristics Inside Pipes," Sustainability, vol. 14, no. 9, April. 2022. pp. 1-21.

[19] B. Wong, and J. A. McCann, "Failure Detection Methods for Pipeline Networks: From Acoustic Sensing to Cyber-Physical Systems," Sensors, vol. 21, no. 15, Jul. 2021. pp. 1-83.

[20] D. Ayala-Cabrera, "Characterization of Components of Water Supply Systems from GPR Images and Tools of Intelligent Data Analysis," Ph.D. thesis, Universitat Politècnica de València, Valencia (Spain), 2015.

[21] A. Akbarpour, S. Chamaani, J. Sachs, G.D. Galdo "Clutter Removal of Near-Field UWB SAR Imaging for Pipeline Penetrating Radar," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, April. 2020, pp. 1527-1539.

[22] M. Rasol, J. Pais, V. Pérez-Gracia, M. Solla, F. Fernandes, S. Fontul, D. Ayala-Cabrera, F. Schmidt, H. Assadollahi, "GPR Monitoring for Road Transport Instracture, A Systematic Review and Machine Learning Insights," Construction and Building Materials, vol. 324, Mar. 2022, pp. 1-21.

[23] A. Shekargoftar, H. Taghaddos, A. Azodi, A. N. Tak, and K. Ghorab, "An Integrated Framework for Operation and Maintenance of Gas Utility Pipeline Using BIM, GIS, and AR," Journal of Performatnce of Constructed Facilities, vol. 36, no. 3, Mar. 2022.

[24] S. Li, H. Cai, and V. Kamat, "Uncertainty-Aware Geospatial System for Mapping and Visualizing Underground Utilities," Automation in Construction, vol. 53, May 2015, pp. 105-119.

[25] M. Zahlan Abdul Muthalif, D. Shojaei and K. Khoshelham, "A Review of Augmented Reality Visualization Methods for Subsurface Utilities," Advanced Engineering Informatics, vol. 51, Jan. 2022, pp. 1-18.

[26] S. Ocaña-Levario, E. Carreño-Alvarado, D. Ayala-Cabrera, J. Izquierdo, "GPR Image Analysis to Locate Water Leaks from Buried Pipes by Applying Variance Filters," Journal of Applied Geophysics, vol. 152, May. 2018, pp. 236-247.

[27] D. Ayala-Cabrera, J. Izquierdo, I. Montalvo, and R. Pérez-García, "Water supply system component evaluation from GPR radargrams using a multi-agent approach," Mathematical and Computer Modelling, vol. 57, no. 7-8, April. 2013, pp. 1927-1932.

[28] D. Ayala-Cabrera, M. Herrera, J. Izquierdo, R. Perez-Garcia, O. Piller, "A New GPR Image Analysis Proposal Based on a Multi-agent Approach and Properties of Groups – Towards Automatic Interpretations," in Proc. 8th International Congress on Environmental Modelling and Software, Toulouse, France: iEMSs, 2016, pp. 678-685.

[29] D. Ayala-Cabrera and J. Izquierdo, "GPR Image Interpretation Advancement for Smarter Technical Management of Water Leakage in Urban Water Infrastructures," in Proc. Earth Resources and Environmental Remote Sensing/GIS Applications XII, Madrid, Spain: SPIE 118630S, 2021.

[30] N. Kim, K. Kim, Y.K. An, H.J. Lee, J.J. Lee, "Deep Learning-Based Underground Object Detection for Urban Road Pavement," International Journal of Pavement Engineering, vol. 21, no. 13, 2020, pp. 1638-1650.

[31] Z. Zong, C. Chen, X. Mi, W. Sun, Y. Song, J. Li, Z. Dong, R. Huang, B. Yang, "A Deep Learning Approach for Urban Underground Objects Detection from Vehicle-Borne Ground Penetrating Radar Data in Real-Time," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., XLII-2/W16, Sep. 2019, pp. 293-299.

[32] W. Lei, J. Luo, F. Hou, L. Xu, R. Wang, and X. Jiang, "Underground Cylindrical Objects Detection and Diameter Identification in GPR B-Scans via the CNN-LSTM Framework," Electronics, vol. 9, no. 11, pp. 1804, Oct. 2020, pp. 1-16.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14132

THE USE OF RESILIENCE METRICS TO SUPPORT DECISION MAKING IN DRINKING WATER SYSTEMS

Joana Carneiro[1], Dália Loureiro[2] and Dídia Covas[3]

[1,3]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisboa, Portugal
[2]Urban Water Unit, Hydraulics and Environment Department, Laboratório Nacional de Engenharia Civil, Av. Do Brasil 101, 1700-066 Lisboa, Portugal

[1]*joana.carneiro@tecnico.ulisboa.pt,*
[2]*dloureiro@lnec.pt,* [3]*didia.covas@tecnico.ulisboa.pt*

Abstract

Performance assessment is essential for effectively managing drinking water systems. It allows to understand the system's behaviour, identify critical components and subsystems, and help with the decision analysis of measures to improve economic, infrastructural and water and energy resources. The current paper proposes an updated methodology for diagnosing drinking water networks, considering different perspectives: water and energy efficiency, infrastructural sustainability, and the quality of service provided, including resilience and redundancy concepts. Several performance indicators and indexes, including Resilience index and Entropy, are recommended to describe each perspective. The methodology is applied to a real-life network, and the attained results are discussed. The status quo situation for the network area is assessed, and the main problems are identified (i.e., high friction losses, inefficient pump operation, old pump equipment). Different improvement solutions are considered (e.g., pipe replacement with higher diameter, pump operation improvement, pump replacement). These interventions are considered individually or combined. Using the resilience-based perspective with other performance assessment criteria has provided a broader assessment. The resilience index has proven sensitive to the different alternatives, endorsing the system's efficiency. Smaller values of the resilience index indicate that a small amount of energy in excess is available as surplus energy in the consumption nodes, being dissipated, either as pumping inefficiencies, water losses, or friction and singular head losses. The Entropy metric is sensitive to the network layout and is helpful in alternatives that consider flow path alterations to prevent the impact of pipe failure. The best improvement alternative achieves a considerable enhancement in energy efficiency, maintains a good quality of service, improves the system's infrastructural sustainability, and corresponds to the highest resilience index value.

Keywords
Drinking water systems, resilience, performance metrics, system diagnosis, alternative comparison.

1 INTRODUCTION

Drinking water systems are crucial infrastructures worldwide and are managed by water utilities that have the mission to continuously and efficiently deliver water in the desired amount and with good quality to all consumers. Therefore, it is of the utmost importance that drinking water systems perform well and respond adequately to future challenges. As water utilities look to incorporate sustainable and improvement measures in their daily practices, performance assessment has become essential for managing drinking water systems. It is a tool capable of identifying critical components and prioritising subsystems within the water system.

In the last couple of decades, different approaches were developed to assess different performance aspects of water systems. A comprehensive effort to analyse and quantify the technical performance of water networks was developed by [1]. Based on the quantified evaluation of system behaviour as a function of pre-defined objectives, this approach can be applied to assess hydraulic and water quality performance (i.e. pressure, water age).

Performance assessment frameworks have been developed using performance indicators to monitor the effectiveness and efficiency of water utilities in different aspects, such as water losses [2], infrastructure asset management [3,4], energy efficiency [5,6], among others. These frameworks were developed to evaluate a specific aspect of the system, neglecting a broader vision. Additionally, most of these indicators are not dependent on the hydraulic and water quality network model, lacking a more detailed analysis that allows the identification of specific problems.

In recent years, resilience has become a more noticeable topic in the water sector. A broader definition describes resilience as "the ability to absorb local failures, quickly recover and maintain the essential service functions, and adapt to long-term changes in the environment and uncertainty disturbances" [7]. Several metrics have been used to assess resilience in the literature, classified into two main groups: hydraulic and topological [8]. Hydraulic resilience measures the hydraulic capacity to maintain supply under failure or uncertain demand conditions, being the resilience index [9] one of the most commonly used metrics. Topological reliability considers the graphical linkage among nodes in water distribution systems, most used graph-theory metrics [10] and flow-entropy measures [11,12]. In Portugal, a Security and Resilience index was proposed in the 4th edition of the "Guide for the Assessment of the Quality of Service in Water and Waste Services" [13]. The index assess the water utility by its concern and measures planned in terms of water security, safety, draught and flood contingency and emergency management and response. To the authors' knowledge, resilience metrics have not yet been used to support diagnosis and decision analysis in drinking water systems.

The current paper proposes an updated methodology for diagnosing drinking water systems and assessing improvement measures considering different perspectives. The paper's main novelty is considering resilience metrics together with hydraulic, energy efficiency, water losses, infrastructural condition, and water quality in the performance assessment system. The methodology is applied to a real-life network, and the attained results are discussed, mainly the importance of attending to different dimensions of drinking water networks and considering resilience.

2 METHODOLOGY

2.1 Approach Overview

A methodology to assess the performance of a drinking water system, attending different dimensions of analysis, is described herein. The energy efficiency performance assessment [5], in the scope of the Portuguese peer-to-peer innovation project "Assessment of energy efficiency and sustainability in urban water system – Avaler+", and the decision-making approach developed by [4] was improved in this study.

Avaler+ project first developed a performance assessment framework, including indicators capable of assessing the system's energy efficiency and effectiveness [5]. This approach calculates the indicators using historical information gathered annually by the water utility, without hydraulic modelling. Energy efficiency was assessed for drinking water systems by calculating Standardised energy consumption and Energy in excess per unit of authorised consumption. Non-revenue water and Infrastructure value index were selected to assess the system's effectiveness. The diagnosis phase is carried out for a reference year and allows the prioritization of subsystems for further analysis and intervention to improve energy efficiency.

In the present methodology, other indicators and assessment methods that require the modelling of the system are introduced for more comprehensive performance analysis. The setting of the assessment framework is composed of three main stages: (i) identifying the relevant dimensions for performance analysis in drinking water systems, including resilience, (ii) selecting a set of metrics to assess each dimension and (iii) identification of reference values for each metric. This framework is used to assess the priority subsystem for the reference year and applied to each alternative identified as improvement measures. This study assumes that measures are implemented in the year following the reference year, and the improvement impact is immediately reflected. Further analysis is recommended to compare long-term alternatives and consider different scenarios.

2.2 Framework for Performance Assessment

The first dimension under scope is Energy efficiency using the same metrics those used in Avaler+ project [5]. Standardized energy consumption [13] allows assessing pumping energy efficiency. Energy in excess per unit of authorised consumption, calculated through the energy balance [6], represents the theoretical potential for energy reduction per volume of authorised consumption since energy in excess is calculated as the total input energy into the system minus the minimum energy required to supply the consumers. The energy balance can be calculated by a top-bottom approach requiring minimum data, has in the Avaler+ project, or by a bottom-up approach requiring a calibrated hydraulic model of the network providing a detailed assessment of energy consumption in every component of the balance, used in the present work, according to work developed in [6].

The second dimension is Infrastructural sustainability. From the Avaler+ project [5], Infrastructure value index [4] assess the infrastructure's age. Non-revenue water is replaced by Real water losses [13] in this study, as the later reflects the network leaks that directly impact energy consumption, and the former also considers apparent losses due to theft or metering inaccuracies. Infrastructural sustainability can be in jeopardize if a system has elevated pressures, significant pressure fluctuations and high flow velocities in pipes. Therefore, performance assessment includes Maximum pressure, Pressure fluctuation and Maximum velocity, following the methodology developed by [1] and considering the modifications in the performance scale proposed by [14]. This methodology is based on applying a penalty function over the results of the variable under analysis. The penalty function relates the variable value and a scale of performance for each network element. A generalising function, weighted average, is used to extend the element-level calculation to the network. The network performance is assessed for 24h simulations, and the results are in the form of extended-period performance graphs showing the system performance and 25% percentiles bands of the network elements. In the present study, penalty and generalising functions for Maximum pressure, Pressure fluctuation and Maximum velocity are the same as [14]. The system's performance for a specific aspect is calculated as the average of the system performance over the 24h. As an example, Pressure fluctuation performance is optimal if a node is always kept at constant pressure and starts dropping linearly until it reaches the threshold of nodal head fluctuation of 30 m.c.a. (normal water utility establishment), corresponding to a performance of 25 %, meaning that it has poor performance with no service interruption. The penalty function for pressure fluctuation is presented in Figure 1. Once the penalty function is applied to node results, nodal performance is generalised by the weighted average in terms of demand, to obtain the system global performance.

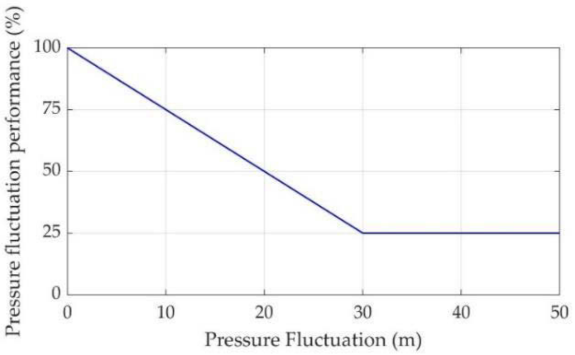


Figure 1. Penalty curve for pressure fluctuation (based on [14]*).*

The third dimension under analysis is the service provided, in both quantity and quality. In this dimension, the capacity of the system to provide water is assessed through the Minimum pressure performance and the Todini's Resilience index [9]. Resilience index requires hydraulic simulation and measures the amount of energy available at each end-use point, evaluating the capacity of the system to continue the provision of water, ranging from 0 to 1. Comparing the Resilience index equation with the energy in excess (obtained through the energy balance [6]), this resilience metric is a measure of surplus energy over energy in excess. Surplus energy refers to the energy supplied to the end-use points minus the minimum energy required to supply the consumers.

Water quality is assessed through Minimum velocity and Water age performance, as these are related to the time of water inside the system. Minimum pressure performance and water quality metrics are calculated with the same procedure as [1] and exemplified for Pressure fluctuation performance. For Minimum pressure and Minimum velocity, penalty and generalising functions are the same as in [14] while for Water age are the same as in [1].

Another metric evaluated is the Entropy of the system [12]. This metric measures flow uncertainty inside the network and considers flow path redundancy. This metric is important as it accounts network layout and the respective capacity for successfully surpassing any link failure in the network.

The metrics selected to assess each dimension are summarised in Table 1, along with a brief description and the respective reference values that allows the classification of the metrics as good (represented by ●), fair (represented by ●) or poor (represented by ●). The metrics requiring hydraulic simulations were calculated in Matlab via Epanet-Matlab-toolkit [15], except entropy calculated through the Water Network Tool for Resilience (WNTR)[16]. As the entropy and resilience metrics do not have reference values, the performance assessment for these metrics is performed by a comparison between the results of the diagnosis and the assessment of alternatives, as improve (↗), stay the same (→) and decrease (↘).

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14132

Table 1. Performance indicators for diagnosis and improvement alternative assessment, description and reference values

Dimension	Metric	Description	Reference values 🔴 (poor), 🟡 (fair), 🟢 (good)
Energy Efficiency	Standardised energy consumption [kWh/(m³.100m)] [13]	Energy consumption for pumping/Sum of the volume elevated multiplied by the pump head/100	🟢 [0.27; 0.40] 🟡]0.40; 0.54] 🔴]0.54; 5.00]
	Energy in excess per unit of authorised consumption [kWh/m³] [4]	(Energy supplied to the system - Minimum energy necessary)/Volume of authorised consumption	🟢]0.0; 0.15] 🟡]0.15; 0.30] 🔴]0.30; +∞[
Infrastructural Sustainability	Real water losses [l/(connection·day)] [13]	Average daily volume lost/Number of connections	🟢 [0.0; 100] 🟡]100; 150] 🔴]150; +∞[
	Infrastructure value index [-] [4, 13] *	Current value of the infrastructure/Replacement cost of the infrastructure	🟢]0.6; 1.0] 🟡 [0.4; 0.6] 🔴]0.0; 0.4[
	Maximum pressure performance in network nodes [%] [1, 14] **	Average of the 24h system maximum pressure performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
	Pressure fluctuation performance in network nodes[%] [1, 14] **	Average of the 24h system pressure fluctuation performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
	Maximum velocity performance in network links [%] [1, 14] **	Average of the 24h system maximum velocity performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
Service Provided	Minimum pressure performance in network nodes [%] [1, 14] **	Average of the 24h system minimum pressure performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
	Resilience index [-] [9]	Surplus energy / energy in excess	***
	Water Age performance [%] [1]	Average of the 24h system water age velocity performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
	Minimum velocity performance in network links [%] [1,14] ***	Average of the 24h system minimum velocity performance	🟢]75; 100] 🟡]50; 75] 🔴 [0; 50]
	Entropy [-] [12]	Uncertainty in the flow of the network pipes	***

Note: *Reference values obtained in [13]; ** Reference values obtained in [14]; ***Reference values not found in literature

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14132

3 CASE STUDY

The case study is a drinking water system located in the Algarve region (Portugal) and is one of the systems analysed in Avaler+ project. The drinking water system comprises five subsystems (S1 to S5) and receives water from three inlet delivery point provided by Águas do Algarve. It comprises five storage tanks, eight pumping stations, approximately 130 km of pipes and 5730 service connections. In the reference year of 2018, the system supplied ca. 4 Mm^3 of water to the consumers. Figure 2 presents the drinking water network, identifying subsystems S1 to S5.

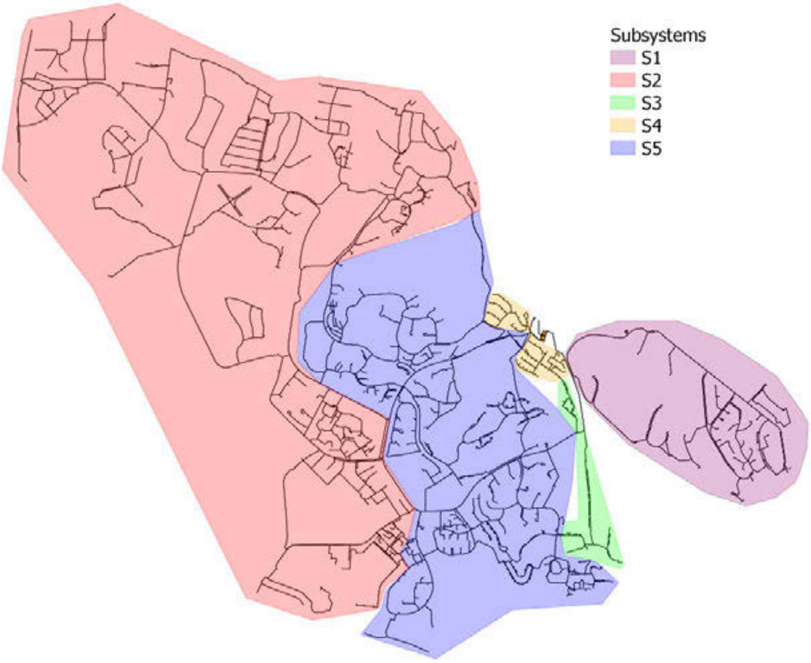


Figure 2. Subsystems of the drinking water distribution system

The updated performance assessment was implemented in the subsystem identified as the priority (S3) as explained in section 4.1. This subsystem is composed of approximately 3 km of pipes in asbestos cement (78 %) and polyvinyl chloride (22 %), with nominal diameters ranging from 80 to 110 mm, 276 service connections and 15 % of annual water losses. The subsystem is supplied by a storage tank with 53 m of water level and has a pumping station downstream needed to pump water to the highest elevation area (ca. 43 m), assuring the minimum required pressure-head (20 m). This pumping station has four pump groups installed in parallel, each with a nominal flow rate of 16 m^3/h and a nominal head of 58.4 m, with a total hydraulic power of 7.63 kW.

The water utility provided two hydraulic models of the network developed in EPANET [17], set to 24h simulation period, corresponding to the summer and winter seasons with a total water supply of 775 947 l/day and 109 433 l/day, respectively. The performance assessment falls on the summer model, as it requires higher hydraulic capacity from the network due to higher demands (demand is five times higher in the summer season). Water losses were simulated as emitters in EPANET and calculated as the pressure-driven method described in [18], with leak coefficient constant c=1.2×10^{-5} l/(s.m.m$^{1.18}$) and pressure exponent of 1.18. Data was provided for the reference year 2018, and the alternatives are implemented and assessed in year 2019.

4 METHODOLOGY APPLICATION AND RESULTS

4.1 Subsystem #3 diagnosis

Avaler+ performance assessment framework [5] requires data gathered by the water utility that was not provided. A simpler version of the assessment was carried out for the diagnosis phase considering Energy consumption, Standardised energy consumption, Energy in excess per unit of authorised consumption, Non-revenue water and Infrastructure value index. The diagnosis results carried out during the Avaler+ project are presented in Table 2, and subsystem S3 was identified as a priority to implement energy efficiency improvement measures.

Table 2. Avaler+ diagnosis results for subsystem prioritisation (reference year)

Performance Indicator	**Subsystems**				
	S1	**S2**	**S3**	**S4**	**S5**
Energy consumption (%)	19	8	15	8	49
Standardised energy consumption [kWh/(m^3.100m)]	0.48 🟡	0.84 🔴	1.13 🔴	0.53 🟡	0.41 🟡
Energy in excess per unit of authorised consumption [kWh/m^3]	0.15 🟢	0.12 🟢	0.40 🔴	0.21 🟡	0.16 🟡
Non-revenue water [%]*	2.4 🟢	15.0 🟢	15.0 🟢	15.0 🟢	15.8 🟢
Infrastructure value index [-]	0.15 🔴	0.12 🔴	0.15 🔴	0.00 🔴	0.00 🔴

Note: *Reference values obtained in [13].

Subsystem S3 represents 15% of the water utility's energy consumption in drinking water transport and distribution. The subsystem has poor performance in both energy efficiency performance indicators, indicating low efficiency of the pumping equipment. Additionally, the Infrastructure value index has poor performance, indicating assets with low residual life. In terms of Non-revenue water, all subsystems have good performance.

Identified as the subsystem with higher priority for intervention, subsystem S3 was selected for a more detailed diagnosis in this paper, and the results are presented in Table 3. Figure 3 presents the extended-period performance simulation graphs to understand the dynamic of subsystem S3 in 24h.

Table 3. S3 performance metrics results for diagnosis (reference year). Complementary metrics are shaded

Dimension	**Performance Metrics**	**Value**
Energy efficiency	Standardised energy consumption [kWh/(m^3.100m)]	1.13 🔴
	Energy in excess per unit of authorised consumption [kWh/m^3]	0.39 🔴
Infrastructural sustainability	Real water losses [l/(connection·day)]	227 🔴
	Infrastructure value index [-]	0.15 🔴
	Maximum pressure performance [%]	90.6 🟡
	Pressure fluctuation performance [%]	77.0 🟢
	Maximum velocity performance [%]	96.4 🟢
Service provided	Minimum pressure performance [%]	100.0 🟢
	Resilience Index [-]	0.16
	Minimum velocity performance [%]	63.8 🟡
	Water age performance [%]	100 🟢
	Entropy [-]	2.65

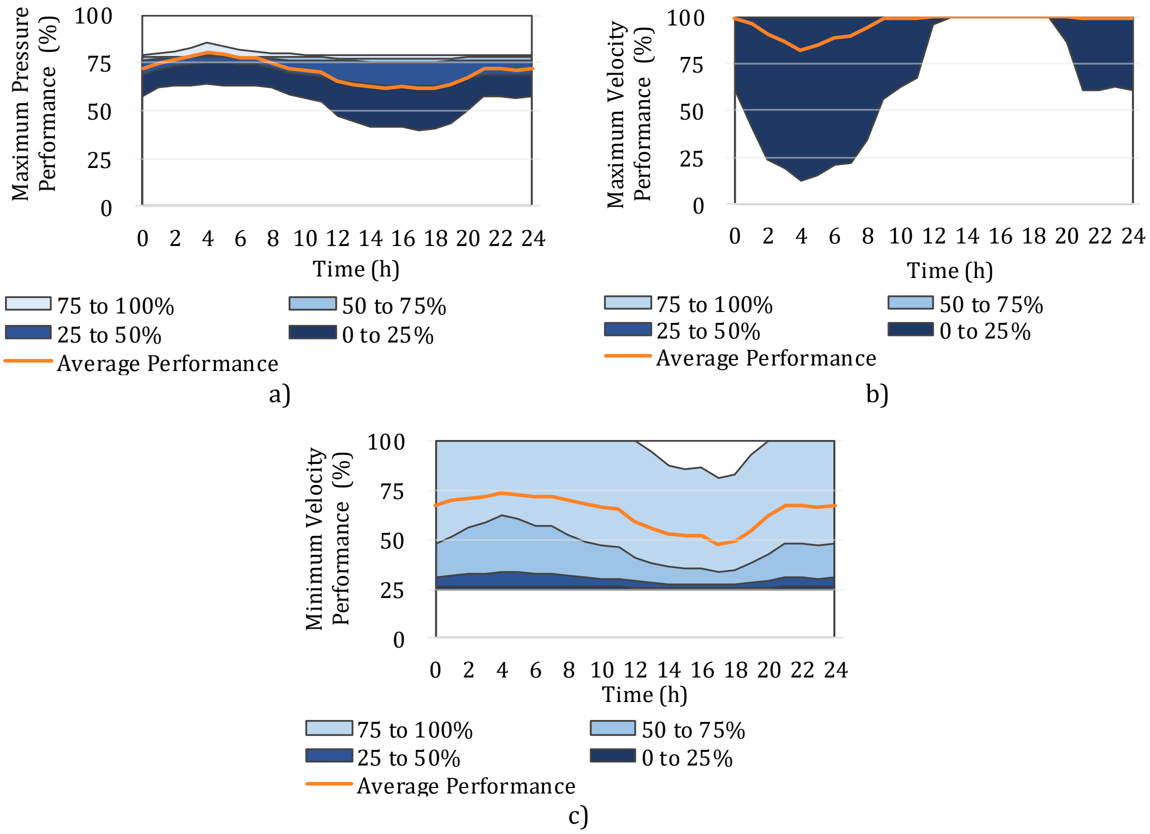


Figure 3. Subsystem S3 extended-period performance assessment for: a) maximum pressure, b) maximum velocity, c) minimum velocity

In addition to the poor performance in energy efficiency, infrastructural sustainability needs improvement since the system has poor performance in Real water losses and Infrastructural value index and fair performance in Maximum pressure. Looking at the extended-period performance simulation graphs for Maximum pressure (Figure 3a), in the higher pressure period (from 12h to 20h) the system has fair performance (average maximum pressure between 50% and 75%). Moreover, less than 25% of the network nodes have poor performance (25th percentile of maximum pressure below 50%). At 4h time instant, Maximum pressure performance increases while Maximum velocity performance decreases (Figure 3b). This instant corresponds to the peak demand, with nodes' pressure decreasing and pipes' velocity increasing. Note that in the Maximum velocity performance (Figure 3b), for less than 25% of the network's pipes, velocity significantly increases with the increase in the demand, suggesting the existence of pipes with high head losses, identified as critical assets according with EPANET simulation.

In terms of quality of the provided service, Minimum velocity has fair performance (Table 3). In the extended-period performance simulation graph for minimum velocity (Figure 3c), over 50% of the network pipes have poor performance throughout the 24h, potentially leading to high water retention time. However, Water age has good performance staying in the range of 100% for all nodes, emphasising the importance of looking at different dimensions of analysis for diagnosis. From the results obtained, it can be inferred that subsystem S3 has a good performance in terms of water quality.

The minimum pressure required by the utility is always met, and the performance is 100% in terms of quantity. The resilience index result of 0.15 is close to 0 and indicates that the denominator is considerably higher than the numerator, which means that from the energy in excess present in the system, only a small part is available at nodes as surplus energy. On the other side, the performance is poor in terms of Energy in excess per unit of authorised consumption, Standardised energy consumption and real water losses (Table 3). Therefore, we can infer that energy in excess, is mainly due to pumping inefficiencies, water losses and possible high head losses in some pipes.

4.2 Alternative Assessment

From the diagnosis phase, it was possible to identify priority assets that require measures to improve the system's performance in terms of energy efficiency. For that, along with the prediction of the *status quo* situation, several improvement measures were selected as follows:

- Alternative A0: *Status quo*, not considering interventions;
- Alternative A1: Replacement of the pump groups with the same rated characteristics (like-for-like replacement);
- Alternative A2: Replacement of the pump groups properly designed, considering the opportunity to improve the operating points of the pump groups;
- Alternative A3: Replacement of ca 900 m of water pipes properly designed, considering appropriate nominal diameters;
- Alternative A4: a combination of alternatives A1 and A3;
- Alternative A5: a combination of alternatives A2 and A3.

The capital costs, obtained as in [4], for each alternative are presented in Table 4.

Table 4. Capital cost of each analysed alternative

Alternatives	A0	A1	A2	A3	A4	A5
Investment (€)	- €	13 700 €	12 896 €	38 835 €	52 535 €	51 731 €

Like-for-like replacement of pump group alternatives (A1 and A4) are slightly more expensive than adequately designed new pump groups (A2 and A5), because the installed pumping groups are overdesigned, with the operating point far from the maximum efficiency point. The new, properly designed pump groups were selected to have 5 m over the minimum pressure required by the water utility (20 m). The detailed performance assessment results for each alternative are presented in Table 5.

Alternatives involving pump replacement, either like-for-like (A1 and A4) or properly designed (A2 and A5), promote an improvement in metrics that depend on pump efficiency, such as Standardised energy consumption, Energy in excess per unit of authorised consumption and Resilience index metric (Table 5). The replacement of properly designed pump groups (A2 and A5) directly impacts metrics that depend on pressure results, particularly on the Maximum pressure performance and Real water losses. The replacement of water pipes properly designed (A3, A4 and A5) decrease the head loss in the pipes, promoting a better Pressure fluctuation performance, Maximum velocity performance and Resilience index. It can also be noted that the replacement of water pipes has a smaller impact in the Resilience index than the pump replacement.

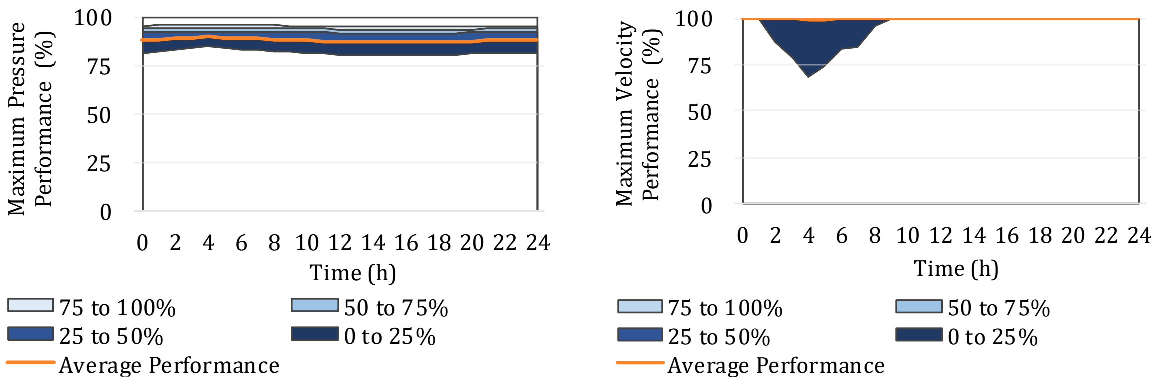
The proposed alternatives do not contemplate the improvement of network redundancy (i.e., installing new pipes to create new paths to support the supply of the branched part of the network). As such, the flow paths remain the same for all alternatives and the Entropy index, which is why the Entropy index is not shown in Table 5. A more exploratory analysis of this metric is presented in the next section.

Overall, alternative A5 has a higher performance in the different dimensions and is the one that should be implemented by the water utility, even though it requires the second-highest investment (Table 4). This alternative combines different measures that improve pump efficiency, decrease high head losses in critical pipes and decrease the pressure input into the system, maintaining the necessary pressure to provide a good service to the consumers.

Table 5. S3 performance metrics results for alternatives (implementation year). Complementary metrics are shaded.

Dim.	Performance Indicator	Analysed alternatives					
		A0	A1	A2	A3	A4	A5
Energy efficiency	Standardised energy consumption [kWh/(m³.100m)]	1.18 🔴	0.54 🔴	0.48 🟡	1.13 🔴	0.54 🔴	0.48 🟡
	Energy in excess per unit of authorised consumption [kWh/m³]	0.41 🔴	0.18 🟡	0.09 🟢	0.39 🔴	0.19 🟡	0.07 🟢
Infrastructural sustainability	Real water losses [l/(connection·day)]	227 🔴	227 🔴	153 🔴	239 🔴	239 🔴	147 🟡
	Infrastructure value index [-]	0.1 🔴	0.3 🔴	0.3 🔴	0.5 🟡	0.7 🟢	0.7 🟢
	Maximum pressure performance [%]	70.6 🟡	70.6 🟡	87.5 🟢	64.4 🟡	64.4 🟡	88.0 🟢
	Pressure fluctuation performance [%]	77.0 🟢	77.0 🟢	77.2 🟢	91.4 🟢	91.4 🟢	91.7 🟢
	Maximum velocity performance [%]	96.4 🟢	96.4 🟢	96.6 🟢	99.9 🟢	99.9 🟢	99.9 🟢
Service provided	Minimum pressure performance [%]	100 🟢	100 🟢	100 🟢	100 🟢	100 🟢	100 🟢
	Resilience Index [-]	0.16	0.34 ↗	0.32 ↗	0.19 ↗	0.38 ↗	0.39 ↗
	Minimum velocity performance [%]	63.8 🟡	63.8 🟡	63.1 🟡	62.7 🟡	62.7 🟡	61.8 🟡
	Water age performance [%]	100 🟢	100 🟢	100 🟢	100 🟢	100 🟢	100 🟢

Comparing the extended-period performance simulation graph for alternative A5 (Figure 4) and the diagnosis (Figure 3), it can be concluded that the maximum pressure performance has improved to good performance in all network nodes. The replacement of the under-designed pipes decreases unit head loss from 15 m/km in diagnosis to 5 m/km in alternative A5, corresponding to a reduction of 32%. This replacement promotes a significant improvement of the subsystem pressure fluctuation performance (Table 5) and is also responsible for the improvement shown in alternative A5 of maximum velocity performance (Figure 4b), where no pipes have poor performance, as opposed to the diagnosis assessment (Figure 3b).



cc BY NC SA

a) b)

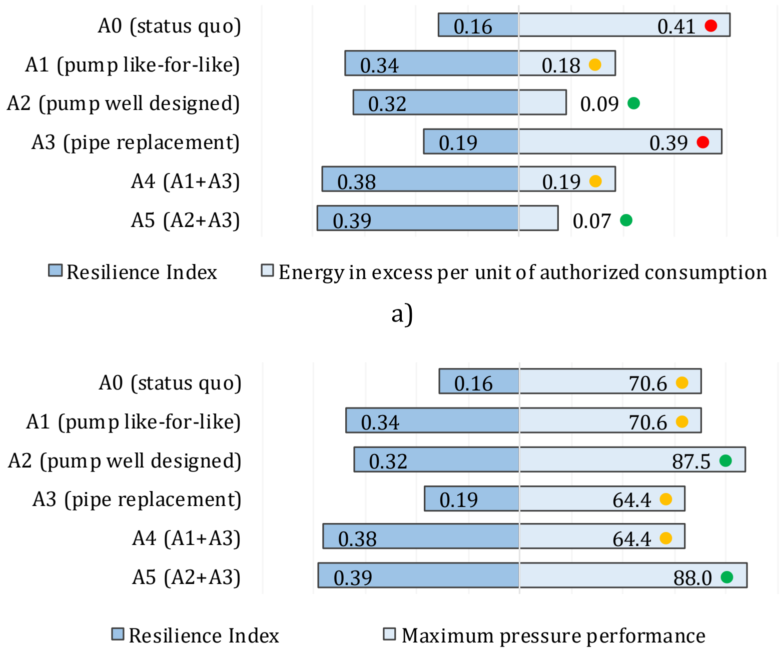
Figure 4. Alternative 5 extended-period performance simulation for: a) maximum pressure, b) maximum velocity.

The best alternative (A5) demonstrates a good compromise between energy efficiency and maintaining a good quality of service, providing the required quantity of water with sufficient pressure and good quality while improving the infrastructural sustainability of the system.

4.3 Resilience metrics exploratory analysis

The introduction of resilience indexes brought new information to the assessment of the system. Comparing Resilience index and Energy in excess per unit of authorised consumption (Figure 5a), both increase their performance with the implementation of improvement measures and are more sensitive to pump replacement (A1, A2, A4 and A5) than under-designed pipes replacement (A3). For alternatives considering pump replacement, like-for-like alternatives (A1 and A4) have similar Resilience index results as properly designed alternatives (A2 and A5) while Energy in excess per unit of authorised consumption have better results for properly designed alternatives (A2 and A5) than like-for-like alternatives (A1 and A4). This infers that the Resilience index is not very sensitive to implementing adequately dimensioned pumping groups as opposed to Energy in excess per unit of authorised consumption.

Comparing the equivalent alternatives without and with pipe replacement (A0 and A3, A1 and A4, A2 and A5), it is observed that the replacement of under-designed pipes has a more positively impact in Resilience index than in the Energy in excess per unit of authorised consumption. Both the Maximum pressure performance and the Resilience Index provide different information (Figure 5b): the Maximum pressure performance is not sensitive to the change of efficiency in the pump groups (like-for-like pump group replacement) but highly sensitive to the correct design of pump groups, as opposed to the Resilience index.



a)

b)

Figure 5. Bi-directional bar chart to compare alternatives' Resilience index and: a) Energy in excess per unit of authorized consumption, b) Maximum pressure performance

BY NC SA

Todini's Resilience index reflects the percentage of the energy in excess provided to the system that is available at nodes as surplus energy. From the energy balance scheme [6], energy in excess can be decomposed into three main components: (i) pumping inefficiencies, (ii) energy associated with water losses, (iii) energy associated with friction and singular head losses and surplus energy. Water systems with a Resilience index closer to 1 are more robust to demand increase, and since the numerator is the surplus energy, this metric infers that the energy in excess main component should be available as surplus energy. The system should have low pumping inefficiencies, water losses and energy dissipated due to friction and singular head losses.

The alternatives proposed to improve energy efficiency of subsystem S3 and do not contemplate improvement of Network redundancy. The flow paths remain the same for the alternatives attested along with the Entropy index. To explore this metric, a couple of exploratory alternative measures to change the original layout of the northern looped part of the network were assessed: (A0') northern looped area *status quo*, (A1') closure of valve in pipe 1 of the loop, (A2') closure of valve in pipe 2 of the loop, and (A3') alternative supply from the north connection to subsystem S4. The scheme and illustration of these exploratory alternatives are shown in Figure 6.

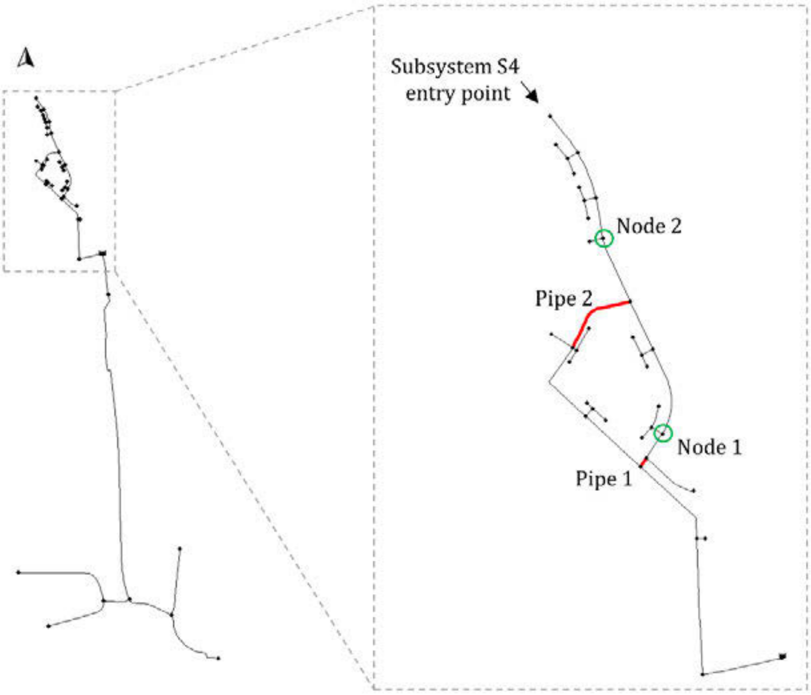
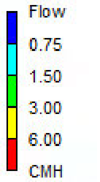
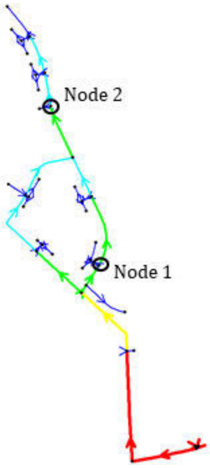
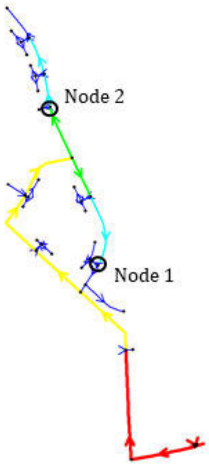
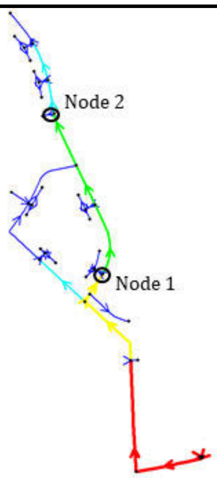
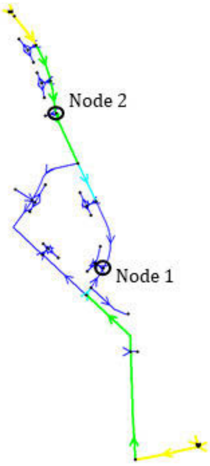


Figure 6. Illustration of explorative alternatives in the network northern looped area for entropy analysis

The present work calculates Entropy according to [12] and depends on link flows and network topology. Entropy accounts for redundancy among the network nodes and the dependence of paths from the source to any demand node by introducing parameter a_{ij} a correction factor to reduce the number of alternate paths, if some are dependent, to the network entropy measure. Entropy results are shown in Table 6, along with flow distribution, the number of paths (independent or dependent) and a_{ij} for two example nodes (Node 1 and Node 2 identified in Figure 6).

Table 6. Results of Entropy, number of paths and parameter a_{ij} to nodes 1 and 2 for the alternatives.

	A0'		**A1'**		**A2'**		**A3'**	
Flow 								
Number of paths	Node 1	1	Node 1	1	Node 1	1	Node 1	2
	Node 2	2	Node 2	1	Node 2	1	Node 2	1
a_{ij}	Node 1	1	Node 1	1.0	Node 1	1.0	Node 1	[1;1]
	Node 2	1.37	Node 2	1.0	Node 2	1.0	Node 2	1
Entropy	2.75		2.68		2.66		2.94	

To understand the parameter a_{ij}, looking at Node 2 of alternative A0', there are two possible paths, one from each side of the loop. However, for both paths the flow must pass through the same links before and after the loop, demonstrating the dependence on these paths. Though there is only one path, it has redundancy and, therefore, the correction factor a_{ij} has a value higher than 1. When looking at the alternative with two water sources (A3'), Node 1 receives water from two different paths, and as the source of each path is different, it is considered two independent paths. The closure of pipes in the network loop (A1' and A2') transforms the looped network to a branched network with independent paths from the source to all consumption nodes, leaving the network with no redundancy and, consequently, with lower entropy values. In this case, the influence in entropy measure is mainly due to the flows where alternative A1' has slightly more equally flow paths than alternative A2'.

5 CONCLUSIONS

The methodology presented herein promotes a broader and more complete assessment of drinking water systems, including the resilience dimension. It allows assessing different dimensions providing different information capable of complementing each other. The best alternative to improve the priority subsystem identified in the diagnosis phase is to improve energy efficiency without compromising the quality of service provided and improve the system's infrastructural sustainability. The introduction of resilience metrics provides new information in the alternative assessment, particularly in assessing the capacity of the system to overcome increasing demands and pipe failures. The resilience index endorses that the energy in excess should be available as surplus energy at the consumption nodes. Systems with smaller values of resilience index indicate that from the total energy in excess provided to the system, a small amount is available as surplus energy in the consumption nodes and was dissipated in the system, either as pumping inefficiencies, water losses or friction losses. The entropy metric depends on

BY NC SA

the variability of network flow inside the system and considers the network's redundancy with the accountability of alternative paths to each node. As the case study alternative assessment did not promote a variation in this component, this metric did not bring new information to the performance assessment in terms of energy efficiency. However, this metric is helpful in alternatives that contemplate flow path alterations and should be assessed in studies where network redundancy is under assessment since none of the other metrics considers alternative paths in case of pipe failure.

The present methodology is the first approach to developing a complete water system resilience-based approach performance assessment. This performance assessment system should evaluate other and more complex case studies to improve the methodology and validate the conclusions drawn.

Funding: This work was supported by the Fundação para a Ciência e a Tecnologia through Ph.D. Research Studentship of Joana Carneiro (grant number PD/BD/150694/2020) and the WISDom project (grant number DSAIPA/DS/0089/2018).

Acknowledgements: The authors would like to thank the Fundação para a Ciência e Tecnologia (FCT) through funding UIDB/04625/2020 from the research unit CERIS.

6 REFERENCES

[1] S. T. Coelho, Performance in Water Distribution - a systems approach. Series Water Engineering and management systems, Research Studies Press, John Wiley and Sons, Lisboa, 1997.

[2] A. O. Lambert, T. G. Brown, M. Takizawa, and D. Weimer, “A review of performance indicators for real losses from water supply systems,” J. Water Supply Res. Technol., vol. 48, no. 6, pp. 227–237, Sep. 1999, doi: 10.2166/aqua.1999.0025.

[3] H. Alegre, S. T. Coelho, D. I. C. Covas, M. Do Céu Almeida, and A. Cardoso, “A utility-tailored methodology for integrated asset management of urban water infrastructure,” Water Sci. Technol. Water Supply, vol. 13, no. 6, pp. 1444–1451, 2013, doi: 10.2166/ws.2013.108.

[4] M. Cabral, “Towards a better urban water infrastructure asset management by condition assessment and cost modelling,” Ph.D dissertation, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 2021.

[5] D. Loureiro, C. Silva, M. A. Cardoso, A. Mamade, H. Alegre, and M. J. Rosa, “The development of a framework for assessing the energy efficiency in urban water systems and its demonstration in the portuguese water sector,” Water (Switzerland), vol. 12, no. 1, 2020, doi: 10.3390/w12010134.

[6] A. Mamade, D. Loureiro, H. Alegre, and D. Covas, “A comprehensive and well tested energy balance for water supply systems,” Urban Water J., vol. 14, no. 8, pp. 853–861, Sep. 2017, doi: 10.1080/1573062X.2017.1279189.

[7] Q. Shuang, H. J. Liu, and E. Porse, “Review of the quantitative resilience methods in water distribution networks,” Water (Switzerland), vol. 11, no. 6, pp. 1–27, 2019, doi: 10.3390/w11061189.

[8] D. Jung and J. H. Kim, “Water Distribution System Design to Minimize Costs and Maximize Topological and Hydraulic Reliability,” J. Water Resour. Plan. Manag., vol. 144, no. 9, p. 06018005, 2018, doi: 10.1061/(asce)wr.1943-5452.0000975.

[9] E. Todini, “Looped water distribution networks design using a resilience index based heuristic approach,” Urban Water, vol. 2, no. 2, pp. 115–122, Jun. 2000, doi: 10.1016/S1462-0758(00)00049-2.

[10] A. Yazdani, R. A. Otoo, and P. Jeffrey, “Resilience enhancing expansion strategies for water distribution systems: A network theory approach,” Environ. Model. Softw., vol. 26, no. 12, pp. 1574–1582, 2011, doi: 10.1016/j.envsoft.2011.07.016.

[11] H. Liu, D. Savić, Z. Kapelan, M. Zhao, Y. Yuan, and H. Zhao, “A diameter-sensitive flow entropy method for reliability consideration in water distribution system design,” Water Resour. Res., vol. 50, no. 7, pp. 5597–5610, Jul. 2014, doi: 10.1002/2013WR014882.

[12] K. Awumah, I. Goulter, and S. K. Bhatt, “Entropy-Based Redundancy Measures in Water-Distribution Networks,” J. Hydraul. Eng., vol. 117, no. 5, pp. 595–614, 1991, doi: 10.1061/(asce)0733-9429(1991)117:5(595).

[13] ERSAR, LNEC, and NOVA, “Guide for the Assessment of the Quality of Service in Water and Waste Services,” 4rd ed.; Technical guide 27; Entidade Reguladora dos Serviços de Águas e Resíduos, Laboratório Nacional de Engenharia Civil, Nova School of Science and Technology. Lisbon, Portugal, 2021.

[14] A. C. Jacob, “Avaliação de perdas em sistemas de distribuição de água: o caso de estudo da ZMC 320 da EPAL,” M.S. Thesis, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 2006.

[15] D. G. Eliades, M. Kyriakou, S. Vrachimis, and M. . Polycarpou, “EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB,” 2016, doi: 10.5281/zenodo.831493.

[16] K. A. Klise, M. Bynum, D. Moriarty, and R. Murray, “A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study,” Environ. Model. Softw., vol. 95, pp. 420–431, Sep. 2017, doi: 10.1016/j.envsoft.2017.06.022.

[17] L. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, EPANET 2.2.0. Cincinnati, USA: U. S. Environmental Protection Agency, 2020.

[18] A. Mamade, “Enhancement of water and energy efficiency in water supply systems,” Ph.D dissertation, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 2019.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14133

AQUARELLUS: A NUMERICAL TOOL TO CALCULATE ACCUMULATION OF PARTICULATE MATTER IN DRINKING WATER DISTRIBUTION SYSTEMS

Joost van Summeren[1], Amitosh Dash[2], Mark Morley[3], Luuk de Waal[4], Jip van Steen[5]

[1,2,3,4,5]KWR Water Research Institute, Groningenhaven 7, 3433 PE, Nieuwegein (The Netherlands)
[5]Currently at: Het Waterschapshuis, Stationsplein 89, 3818 LE, Amersfoort (The Netherlands)

[1] Joost.van.Summeren@kwrwater.nl, [2] Amitosh.Dash@kwrwater.nl, [3] Mark.Morkey@kwrwater.nl,
[4] Luuk.de.Waal@kwrwater.nl, [5] Jip.van.Steen@kwrwater.nl.

Abstract

Despite preventive measures, turbid (vernacular: “discolored”) distributed drinking water is still a common cause for customer complaints across the world. Discoloration events are caused by the accumulation of particulate matter in drinking water distribution systems (DWDSs) and subsequent remobilization during hydraulic events, although uncertainties remain concerning the specific accumulation and transport processes. For Dutch DWDSs, it is plausible that microscopic particles originating at treatment plants contribute substantially to the particulate matter that resides in DWDSs, and that physical processes *within* the distribution network are cardinal in the subsequent transport during distribution.

Aquarellus is a predictive numerical tool that has been developed to predict the location and amount of accumulated particulate material in DWDSs. It integrates hydraulic calculations using the EPANET toolbox with a particle transport module that is based on a description of gravitational settling, particle stagnation, bed load transport, and resuspension of particles in distribution pipes, depending on the shear stress near the pipe wall. Despite the particle transport calculations being very computationally intensive, parallelization allows for simulating distribution network sizes that are common to Dutch water utilities (100s of km total pipe length). The user can assign the injection of multiple particle species corresponding to temporal patterns at multiple source locations. A graphical user interface handles user IO and the visualization of geographical maps as well as time-dependent build-up of particulate material across the distribution network and within individual pipes.

To characterize particle properties (critical input parameters) encountered in DWDSs, we performed lab experiments on nine samples from one Flemish (chlorinated drinking water) and two Dutch (non-chlorinated) DWDSs to determine particle size distributions, mass density, mobility thresholds, and a measure for gravitational settling. Using the outcomes of these lab experiments, a sensitivity test with a range of input parameters was performed in *Aquarellus*. This helped determine how the variation in the relevant input parameters influences the calculated spatial patterns of accumulated particulate matter– a measure for the discoloration risk. We compared the modeling results to turbidity measurements from systematic cleaning actions in a real-life Dutch distribution network. Finally, we discuss the potential for applying the tool to assist the planning of cleaning actions and monitoring programs.

Keywords
Discoloration, water quality modelling, drinking water distribution system.

1 INTRODUCTION

Despite preventive measures, turbid (vernacular: “discolored”) distributed drinking water is still a common cause for customer complaints across the world. Discoloration events are caused by the accumulation of particulate matter in drinking water distribution systems (DWDSs) and subsequent remobilization during hydraulic events [1], although uncertainties remain concerning the specific accumulation and transport processes. For Dutch DWDSs, it is suspected that microscopic particles originating from treatment plants contribute substantially to the particulate matter that resides in DWDSs, and that physical processes *within* the distribution network are cardinal in the subsequent transport during distribution.

Observational evidence suggest that the accumulation of particulate material in DWDSs often occurs in a repeatable spatial and temporal patterns and diminishes above a threshold hydraulic vigor (e.g. [2, 3]). This formed the basis for design and application of self-cleaning networks in The Netherlands that have been applied in The Netherlands since the 1990 and that have been shown to reduce the number of customer complaints [4]. These observations also suggests that can be described with a predictive model .

Several models for discoloration in DWDSs have been developed previously. Richardt et al. (2009) [5] developed a model to optimize flushing intervals to prevent deposit growth in pipes. These predictions are based on maximum daily velocities from a hydraulic simulation and deposit measurements for all pipes in the study area. The Particle Sediment Model [6] includes the processes of particle settling, resuspension, and wall deposition. However, the potentially important process of bed-load transport is absent from this model. The semi-empirical Variable Condition Discoloration Model (VCDM) [7] builds on the Prediction and Control of Discolouration in Distribution Systems (PODDS), [8]. VCDM provides a mathematical formulation for the simultaneous removal (by means of erosion) and build-up (by means of adhesion to pipe walls) of particulate material, depending on the material strength in relation to the actual shear stress. VCDM has been validated for UK transport mains. However, theoretical considerations [9] suggest that conditions and mechanisms may differ from DWDSs with different source qualities due to other treatment processes or non-chlorinated Dutch systems that include self-cleaning network layouts. It is emphasized that the location and amount of particle accumulation in DWDSs cannot be reliably predicted on the basis of actual hydraulic conditions of single pipe segments, or even its hydraulic history. This is because the evolution of sediments relies on the accumulation *history* that includes bed-load transport and recurrent resuspension-settling that can result in complex transport of material through the distribution network.

This paper presents the development of *Aquarellus*, a numerical tool to predict the accumulation of particulate material in DWDSs. It integrates hydraulic calculations using the EPANET toolbox with a particle transport module (PTM) that is based on a description of gravitational settling, particle stagnation, bed load transport, and resuspension of particles in distribution pipes, depending on the shear stress near the pipe wall [10]. To characterize particle properties (critical input parameters for *Aquarellus*) encountered in Dutch DWDSs, we performed lab experiments on nine field samples collected by three water utilities to determine particle size distributions, mass density, on which the particle settling velocity depends. Based on the outcomes of these lab experiments, a sensitivity analysis with a range of input parameters was performed to determine the influence on calculated spatial patterns of accumulated particulate matter– a measure for the discoloration risk. The potential for applying Aquarellus in assisting in the planning of cleaning actions and network monitoring is discussed.

2 METHOD

2.1 Conceptual framework and model assumptions

The theoretical framework of *Aquarellus* was formulated in a previous study that investigated the leading processes governing particle transport and discoloration events in DWDSs [10]. This study assessed which particle transport mechanisms are dominant for hydraulic conditions and particle properties that are typical of Dutch distribution pipelines in DWDSs. The theoretical framework identifies three modes of particle mobility that are discriminated based on the shear stress exerted by the turbulent flow near the pipe wall, τ_b (Figure 1). The shear stress exerted is closely related to the flow velocity of the water (u_f). The dimensionless Shields number (θ) expresses the ratio between the driving (shear stress) and resisting forces to particle mobility, in which the and resisting forces depend on the particle properties:

$$\theta = \frac{\tau_b}{(\rho_p - \rho_f) g d_p} \tag{1}$$

$$\tau_b = \frac{\rho_f u_f^2 C_f}{8} \tag{2}$$

with ρ_p and ρ_f the mass densities of the particulate material and drinking water, respectively, g the gravitational acceleration, d_p the particle diameter, and C_f the Darcy friction factor. Eq. (1) and (2) can be combined as:

$$\theta = \frac{u_f^2 C_f}{8 s g d_p} \tag{3}$$

where s is the excess mass density of particles relative to that of water: $s \equiv \frac{(\rho_p - \rho_f)}{\rho_f}$ and g the gravitational acceleration. For the settling velocity, u_s, we used Stokes' Law which assumes free settling of spherical and incompressible particles in a fluid:

$$u_s = \frac{g s d_p^2}{18 \upsilon_f} \tag{4}$$

With υ_f kinematic viscosity of the fluid. Constant values were assumed for ρ_f ($1 \cdot 10^3$ kg m-3), $\upsilon_f = 1 \cdot 10^{-6}$ m^2 s^{-1} , and C_f=0.02.

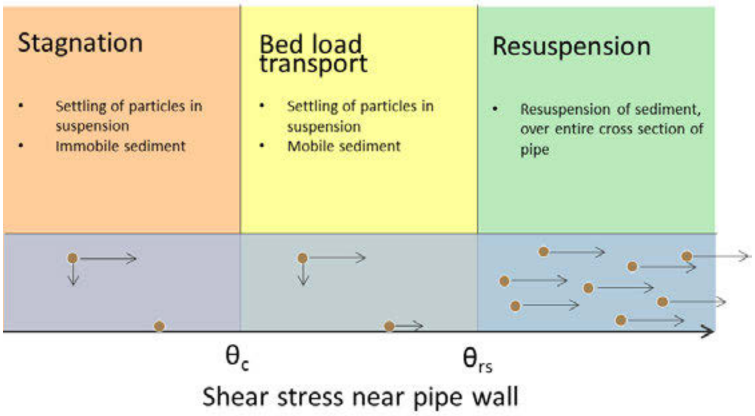

Stagnation
Settling of particles in suspension
Immobile sediment
Bed load transport
Settling of particles in suspension
Mobile sediment
Resuspension
Resuspension of sediment, over entire cross section of pipe
θc
θrs
Shear stress near pipe wall

Figure 1. Conceptual representation of the particle transport model (modified from [10]).

The modes of particle mobility illustrated in *Figure 1* are:

i. *Particle settling under the influence of gravity followed by stagnation*

At low velocities, when the Shields number θ is below the threshold value for incipient motion ($\theta<\theta_c$, red area in *Figure 1*), particles in suspension settle according to Stokes' Law to the invert of the pipe under the influence of gravity and, once settled, remain immobile.

ii. *Particle settling under the influence of gravity followed by bed-load transport*

At shear stress conditions between the critical Shields number for incipient motion, θ_c, and for particle entrainment by resuspension, θ_{rs} (i.e. $\theta_c<\theta<\theta_{rs}$ yellow area in *Figure 1*.), particles in suspension settle, but particles at the pipe's invert move by means of bed-load transport. *Aquarellus* assumes the bed-load velocity to increase linearly from 0 to uu_f when θ increases from θ_c to θ_{rs}.

iii. *Particle resuspension by hydraulic forces*

At shear stresses above the critical Shields number for resuspension ($\theta>\theta_{rs}$, green area in *Figure 1*), it is assumed that all particles resuspend instantly. The particles are redistributed uniformly over the cross section of the pipe and move with velocity u_f.

The above-mentioned processes were identified as the dominant physical processed for particle diameters and flow conditions common to distribution pipes with fully developed pipe flow in Dutch DWDSs. Other mechanisms related to turbulent motion-particle interaction are likely relevant, although probably not dominant, including turbulent diffusion, turbulent dispersion, turbophoresis, gradual rather that abrupt transitioning to bed-load transport and resuspension (see [10] for a more extensive discussion). Irregular pipe structure (due to e.g. culverts, bends, and appurtances) may further complicate the local hydraulic conditions and the associated particle transport. Furthermore, complex interactions with microbiological (e.g. biofilm formation) and/or chemical processes (e.g. coagulation) may also influence the mobility and accumulation of particulate material but are difficult to describe and were kept out of the scope of the current study.

2.2 Architecture of *Aquarellus*

The overall architecture of *Aquarellus* is depicted in Figure 2. A graphical user interface (GUI) is coupled to a computational core that combines:

(i) the functionality of EPANET [11] to calculate hydraulic transport in the DWDSs and

(ii) the particle transport module (PTM) calculates the transport of particulate matter across distribution pipes, according to the theoretical framework discussed in Section 2.1.

The bed load transport velocity of particles can have any velocity between zero and the fluid velocity, depending on the actual Shields number. This requires a fine spatial resolution of particle transport in the streamwise direction. Consequently, the PTM computations are generally more expensive than the hydraulic computations.

From the concentration of particulate material at the source nodes, *Aquarellus* calculates the corresponding number of particles that enter the DWDS. To keep a practically feasible memory load, the particles are combined into packages that move collectively. This particle collation number is user-defined and 1 million by default. Particle packages are injected and subsequently transported through the DWDS at two user-specified time steps that can be set independently between 1 and 60 s. The accuracy of the results of *Aquarellus* were verified using a series of benchmarks of small (9-node) network models for a model time of 30 days. The particle masses were accurate within ~1% and ~6%, respectivelyfor injection and transport time steps of 1 and 60 seconds, respectively.

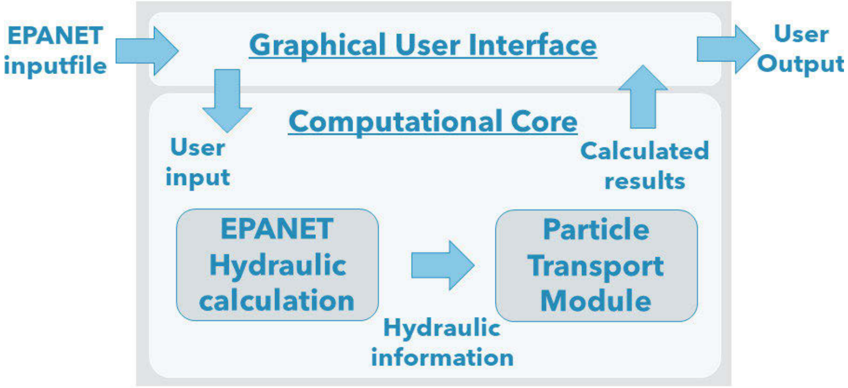


Figure 2. Architecture of the numerical tool.

The memory use and calculation time of *Aquarellus* primarily depend on the total network length, the amount of particle packages in the network model, and the numerical resolution settings. Multi-threading and code optimization were performed to enable calculation of full-scale hydraulic network models on a commonly used laptop quadcore (Intel Core i7, 16 GB). Specifically, a memory of 16 GB was sufficient to calculate a typical hydraulic network model with a total pipe length 500 km in 48 hours.

In practical terms, calculations with *Aquarellus* are performed by providing a modified EPANET input file in which the user defines multiple network junctions as source nodes. For each source node, the incoming mass concentration of particulate material is provided either as a constant baseline level or time-dependent pattern. It is also possible to assign sources of distinct particle material properties at a single node, thereby creating a “multi-species” source, and interpret the distinct species separately. Within the GUI, the user can assign the material properties of each material species, and settings for the numerical resolution and graphical representation of the simulation results that includes exportable tables, graphs, and maps of accumulated masses in the DWDS. The maps make use of geographical information from OpenStreetMap [12]. An impression of the GUI of *Aquarellus* is shown in Figure 3.

3 DETERMINING MATERIAL PROPERTIES OF PARTICULATE MATERIAL

3.1 Motivation

The material properties of particles affect the particle transport and the resulting spatial configuration of stagnant and bed-load particulate material that forms in the DWDS (hereinafter referred to as “sedimentation configuration). In the calculations of *Aquarellus*, the relevant material properties are: the particle mass density (ρ_p), particle diameter (d_p), and the critical shear stresses at which particles mobilize as bed-load transport (θ_c) and resuspend (θ_{rs}). Within the framework of *Aquarellus* an increase in particle size (ρ_p) and density (ρ_p) result in: (i) a faster particle settling velocity, u_s (through Stokes’ Law) and (ii) a lower dimensionless shear stress (Shields number θ) which, in turn, results in a lower propensity to particle mobility, i.e. a larger hydraulic disturbance is required for the transition from stagnation to bed-load transport and from bed-load transport to resuspension.

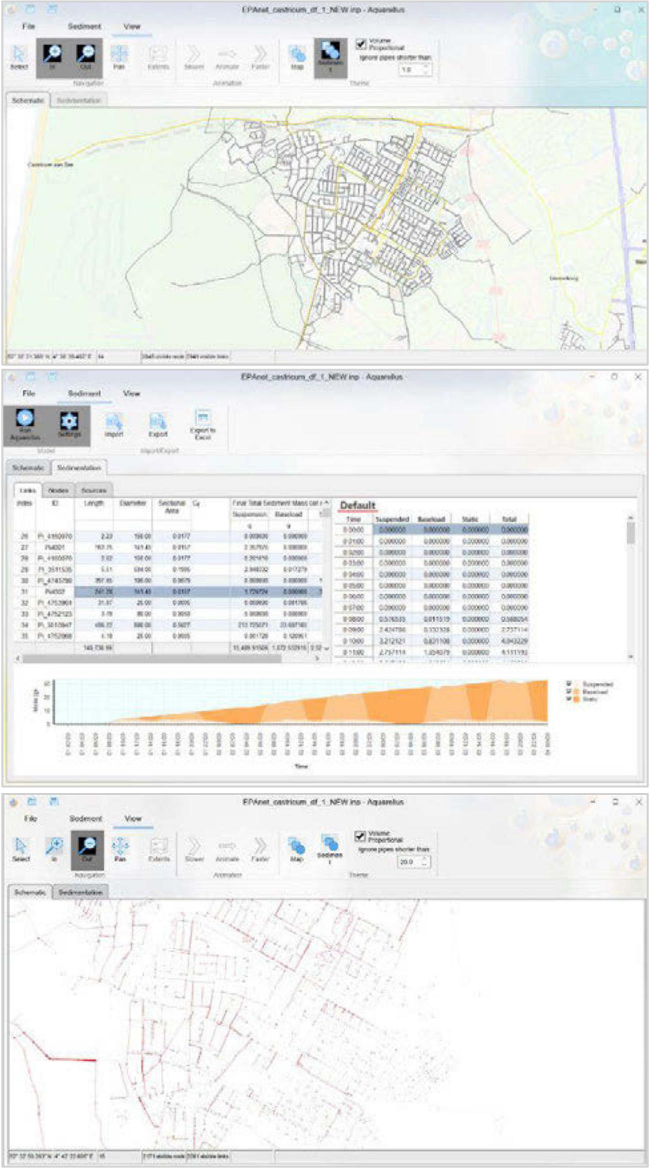


Figure 3. Impression of *Aquarellus'* Graphical User Interface showing examples of (top) map of imported hydraulic network model projected on top of geographical map, (middle) modelling result showing masses of suspended, bed-load, and stagnant particle masses for each pipe (left table), time series of particle masses (right table and graph at the bottom), and (bottom) map of sediment masses per pipe.

Previous measurements have given some information on d_p and ρ_p and lab experiments provide insight into the mobility thresholds θ_c and θ_{rs} [10]. To determine the range of input parameters that can be expected in Dutch and Flemish DWDSs, a series of nine field samples were collected and analyzed in a series of lab experiments. The nine samples are taken in 3 distribution areas at three different locations, aimed at gaining insight into the variation between and within distribution areas. The results form the basis of the sensitivity analysis described in Section -.

3.2 Sampling of drinking water using disruptive actions

Three water companies conducted disturbance tests in the summer of 2020 at three locations in Sint Jansklooster (Vitens), Spijkenisse (Evides), and Neerpelt (De Watergroep). The disturbance test were performed following a protocol similar to that of flushing actions. Samples of 20 L were

BY NC SA

collected from hydrants installed at PVC distribution pipes (80 to 110 mm). The areas were selected for a suspected high risk of particulate accumulation. Thus, turbidity values are not indicative of the drinking water quality that these water utilities distribute. To avoid misportrayal, the areas have been anonymized as Areas *A*, *B*, and *C*. In each area, three sample locations were selected: 'near', 'mid', and 'far' from the supply location and within a pipeline length of 2 to 5 km. The samples locations were positioned on the same flow path, meaning that the same water flow passed through the three locations 'near', 'mid', and 'far' for each area. Locations 'near' and 'far' were separated by a minimum pipe length of 2 to 5 km, depending on the area .

3.3 Laboratory tests to characterize particles in samples of drinking water

Particle settling test

Particle settlement tests were performed for all nine samples using a beaker setup. From each sample, a homogenized subsample of 2 L was taken at 7 cm above the bottom of the beaker glass.

The beaker glass was filled to a height of 14 cm and was subsequently mixed with a stirrer for 10 minutes at 150 r.p.m. after which turbidity was measured for 24 h using a Hach 2100 IQ turbidity meter. The rate of turbidity decrease was quantified by calculating the half-life time ($\tau_{1/2}$) (from an exponential fit of the measured values). A short $\tau_{1/2}$ is taken as an indicator value for a high settling velocity, u_s. It must be noted that this is a global measure of a mixture of particles types with different individual settling velocities.

Particle size determination

The particle size distribution was determined for the 9 samples using a PAMAS particle counter in particle size intervals of 1-3 μm; 3-5 μm; 5-10 μm; 10-20 μm; 20-35 μm; 35-60 μm; and 60-100 μm. The size was interpreted as particle diameter (d_p). It was verified that the particle numbers of samples were within the interval for accurate measurements by diluting the original samples (5 times and 10 times dilution) and verifying the corresponding dilution factors in the results.

Particle density estimation

All nine samples were analyzed in the lab to estimatethe particle density, ρ_p, following these steps:

- Determine the volume of the wet sediment, V_{wet}. In this context, "wet sediment" refers to the turbid part of the sample after 72 h of settling in a measuring cone beaker.
- Determine the dry weight, m_{dry}, after heating at 105 °C of the 0.45 μm filtrated residue of 1 L of homogenized sample.
- The particle density, ρ_p, was estimated as the density of the "wet" sediment:

$$\rho_p = \frac{1}{V_{wet}}\left(\rho_w \left[V_{wet} - \frac{m_{dry}}{\rho_{dry}}\right] + m_{dry}\right) \tag{4}$$

The term between []-brackets expresses the volume of the *water* fraction in the wet sediment. The parameter ρ_{dry} is not readily measured and remains undetermined in this study. This introduces an uncertainty in ρ_p, which converges from its minimum value, $\rho_{dry} = \rho_w$ at $\rho_{dry} = 0$, to its maximum value at $\rho_{dry=\infty}$. It can be shown that the convergence (in terms of the relative excess density) depends on ρ_{dry} as: $s(\rho = \rho_{dry})/\,s(\rho = \infty)$ =1 - ρ_w/ρ_{dry}. In the remainder of this paper we assume a negligible dry sediment volume by assuming an infinite ρ_{dry}.

3.4 Results of laboratory experiments

The main results of the laboratory experiments are shown in Table 1. The main findings are:

- The turbidity measured at the beginning of the settling experiment (when the sample is fully mixed) is hereinafter referred to as the "initial turbidity". It shows a strong variation both *across* and *within* the three distribution areas, ranging from 60 to 150 NTU (Area A), 0 NTU (Area B), and 0 to 40 NTU (Area C). The initial turbidity and the dry weight increase with an increasing number of particles in the sample, as expected.
- A fraction of 40% to 70% of the particles are in the 1 – 3 μm range (relative to particle count in the 1- 100 μm range). The number of particles decreases approximately logarithmically with particle size and with mildly varying fractions at location 1, 2, and 3 (not shown).
- The half-life time $\tau_{1/2}$ shows a wider variation between the three areas (averages of 5.5, 11.7, and 6.0 h for Area A, B, and C, respectively) than within each area (ranges of 4.1 to 6.8 h, 9.4 to 13.3 h, and 4.4 to 9.2 h, respectively). Although $\tau_{1/2}$ cannot be readily translated to the settling velocity, it is useful to try to put these in perspective. A 10 μm particle of ρ_p = 1100 kg·m^{-3} has a settling velocity of 5.45·10^{-5} m·s^{-1} and would sink across the height of the water column in the beaker (14 cm) during 7.1 h which of the same order as the measured values.

Table 1. Lab experiment results of properties of particles in drinking water sampled during disturbance tests. (*): Could not be estimated accurately because of the small dry weight.

	Settling test		Particle count		Lab analysis	
Sample location	Initial turbidity (NTU)	Half-life time turbidity, $\tau_{1/2}$ (h)	Log$_{10}$(N) (number of particles)	Fraction of 1-3 μm particles (% of total)	Dry weight (mg·L^{-1})	Estimated particle density (kg·m^{-3})
A_{near}	149	4.1	6.33	68.7	100	1052.6
A_{mid}	82.3	5.5	6.19	59.4	64	1045.7
A_{far}	56.1	6.8	5.91	64.1	42	1084.0
B_{near}	3.32	12.4	4.62	68.7	3.2	*
B_{mid}	3.35	9.4	4.25	59.4	4.6	*
B_{far}	1.85	13.3	4.11	64.1	2.1	*
C_{near}	3.94	9.2	4.58	66.5	5.7	*
C_{mid}	42.3	4.5	5.54	47.0	63	1018.0
C_{far}	20.8	4.4	4.95	40.0	37	1018.5

- For each of the three areas (but not for the nine samples combined), particle settling is faster (shorter $\tau_{1/2}$) for more turbid samples. A correlation based on chance cannot be ruled out. Another possible explanation is that fast-settling particles concentrate more readily at locations than slow-sinking particles, which at the same flow rate would settle over a larger area.

- For Area A, a lower initial turbidity and longer $\tau_{1/2}$ are measured at locations further away from the pumping station (locations 'near', 'mid', and 'far', respectively). This would be in line with preferential settling close to the source location, especially for fast-settling particles. However, Area B and C do not show this relation, suggesting that other mechanisms are important.
- For five samples, sufficient material was present to derive densities of the wet sediment, calculated according to Eq. 4 and assuming an infinite ρ_{dry}. The estimated particle densities range from 1018 to 1084 kg/m^3.

4 SENSITIVITY OF CALCULATED SEDIMENTATION CONFIGURATIONS TO PARTICLE PROPERTIES

4.1 Scenarios of the Castricum distribution network

Next, we investigated how the sedimentation configurations calculated with *Aquarellus* are influenced by variations in the particle diameter (d_p), particle density (ρ_p), and the shear stress mobility thresholds for critical motion (θ_c) and resuspension (θ_{rs}). To this end, a series of scenarios was run of which a selection of seven will be discussed in this study. The input parameters used in the reference (RFS) and sensitivity scenarios (S1 to S6) are shown in Table 2. In the present study, it is assumed that the ratio between these two Shields numbers is constant at 1:10.

The range of values for d_p and ρ_p are based on the lab test results presented in Chapter 3. The reference model values of d_p and ρ_p are slightly larger than the average measured values; this was done to focus on results with sufficient accumulation in the DWDS models of 30 days (i.e. relatively fast settling velocities). The critical Shields numbers for incipient particle motion and resuspension are based on empirical relationships for rectangular channels documented in the sediment transport literature (see [10] for details and references).

Table 2. Input parameters of *Aquarellus* for seven model scenarios used in the sensitivity analysis. RFS refers to the reference scenario. S1 to S6 refer to scenarios in which parameters indicated in boldface differ from the RFS.

Quantity	Symbol	RFS	S1	S2	S3	S4	S5	S6
Particle mass density (kg ·m$^{-3)}$	ρ_p	1100	**1050**	**1150**	1100	1100	1100	1100
Dimensionless mass density (-)	S	0.1	**0.05**	**0.15**	0.1	0.1	0.1	0.1
Particle diameter (μm)	d_p	10	10	10	**8.8**	**11.2**	10	10
Settling velocity (Stokes) (μm·s^{-1})	u_s	54.5	**22.3**	**82.0**	**42.2**	**68.3**	54.5	54.5
Critical Shields number for incipient motion (-)	θ_c	1	1	1	1	1	**0.5**	**2**
Critical Shields number for resuspension (-)	θ_{rs}	10	10	10	10	10	**5**	**20**

The sensitivity analysis was applied to the 141 km hydraulic network model of Castricum, a municipality of ca. 35,000 inhabitants in the province of North-Holland, The Netherlands. Its network is part of the distribution area of PWN water utility and has a relatively simple supply structure (single water supply; no water tanks; no substantial water exchange with connecting areas).

The hydraulic network model is shown in Figure 4 (pipe diameters and velocity during the morning demand peak in frame (a) and (b), respectively). The supply is located in the Southwest of the area and feeds the Castricum municipality via two transport mains as indicated. In the Northern half of the area, the main flow is in the clockwise direction around Castricum municipality and supplies this area at several locations. All scenarios were calculated for a model period of 30 days. Initially the network contains no sediment, and the particles are introduced at the source at a concentration of 0.023 mg/L. A demand pattern for a standard weekday was deliberately repeated throughout the model period, which makes the results easier to interpret.

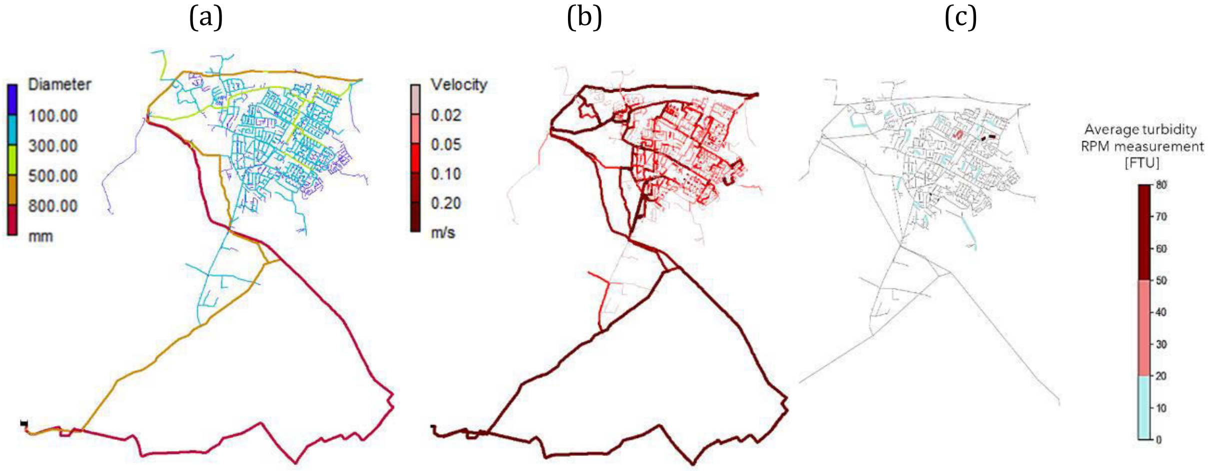

(a) (b) (c)

Figure 4. DWDS of Castricum (PWN) with (a) pipe diameter, (b) flow velocity at 8AM, and (c) RPM field measurements of turbidity. No measurements were available for the pipes indicated in grey in frame (c).

Figure 4c shows an overview of 22 turbidity measurements from the period 2018-2020. The measurements were performed by PWN following the Resuspension Potential Method (RPM) [13], a measuring protocol aimed at measuring the presence and mobility of particulate material in pipe segments with diameters in the range of 50-200 mm. The RPM consists of a controlled and reproducible increase of 0.35 m/s in addition to the actual velocity at the time of measuring. This hydraulic disturbance was created by opening a hydrant on an isolated pipe segment to create a unidirectional flushing. PWN maintained the disturbance for a duration of 15 minutes and the turbidity was monitored every 3 minutes of which the average was used in this study. All the RPM measurements used in this study had a preceding RPM measurements three years earlier, thus creating similar RPM disturbance histories. The area did not experience major disruptions (from flushing actions, major construction works, or network extensions) in the evaluation period, although minor disruptions did occur (such as minor construction works and operational actions). These are beneficial properties when comparing calculated sediment masses to field measurements.

4.2 Sedimentation configurations of model scenarios

The sedimentation configurations after 30 days of simulation at 12 AM (midnight) for the reference scenario (see Table 2) is shown in Figure 5. It shows that distribution pipes with and without substantial sediment masses can be in close proximity; and most distribution pipes with high accumulation levels are relatively close to where the distribution area branches off from the transport mains.

(a) (b)

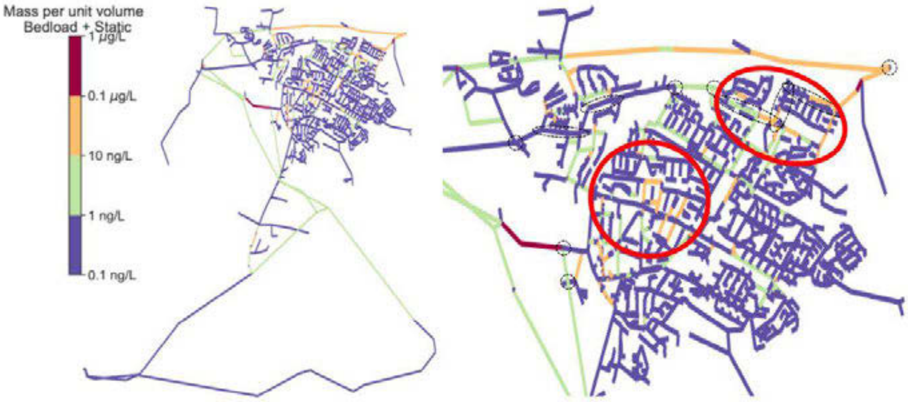


Figure 5. (a) Calculated sedimentation configuration for the reference scenario (RFS) of Castricum after 30 days, and (b) zoom-in of the same scenario. The encircled red areas show distribution pipe clusters of anomalously large sediment masses. The black dotted circles and ellipses show area where >300 mm transport mains connect to distribution pipes smaller than 300 mm.

The sensitivity of sedimentation configurations to input parameters is demonstrated with a selection of results in Figure 6 (cf. to the reference scenario in Figure 5). The sensitivity analysis captures the progressive increase of d_p (from 0.88 to 1 to 1.12 µm), ρ_p (from 1050 to 1100 to 1150 $kg\cdot m^{-3}$) and the mobility thresholds for incipient motion and resuspension (from $\theta_c = 0.5$; $\theta_{rs} = 5$ to $\theta_c = 1$; $\theta_{rs} = 10$ to $\theta_c = 2$; $\theta_{rs} = 20$).

For each of the three tests, an increase in the parameter resulted primarily in (i) a larger amount of sediment is present in the DWDS after 30 days and, (ii) a larger fraction of sediments closer to where the transport mains supply Castricum municipality (i.e. the dotted circles and ellipses in Figure 5b. These two phenomena are stronger for the variation of ρ_p, than for θ_c and θ_{rs} ,than for d_p over the parameter ranges tested.

These results reflect that bigger and heavier particles are associated with a faster settling velocity (linear dependence $u_s \propto s\cdot d_p^2$) and a lower propensity to particle mobilization (linear dependence $\theta \propto s\cdot d_p$), while a simultaneous increase in θ_c and θ_{rs} results in a lower propensity to particle mobilization. Although these outcomes may seem obvious from *Aquarellus*' theoretical framework, it is emphasized here that the calculated sediment mass cannot be reliably predicted on the basis of actual hydraulic conditions of single pipe segments, or even its hydraulic history. This is because the evolution of sediments relies on the sedimentation history that includes bed-load transport and recurrent resuspension-settling that can result in complex transport of material through the distribution network.

Based on a visual interpretation of Figure 5 and Figure 6, two neighborhoods in the Castricum DWDS stand out in expected to have above-average sediment masses (the transport mains that feed the municipality are not taken into account in this analysis). These sediment 'hotspots' are indicated in Figure 5b with the red ellipses. Although their spatial extent varies, these locations are also present as hotspots in most of the scenarios of the sensitivity test analysis. The results also show that within these hotspot locations, the sediment masses are spatially irregular, i.e. clean pipes can coexist adjacent to contaminated pipes, even in hotspot locations. A more quantitative analysis is required to substantiate this visual interpretation.

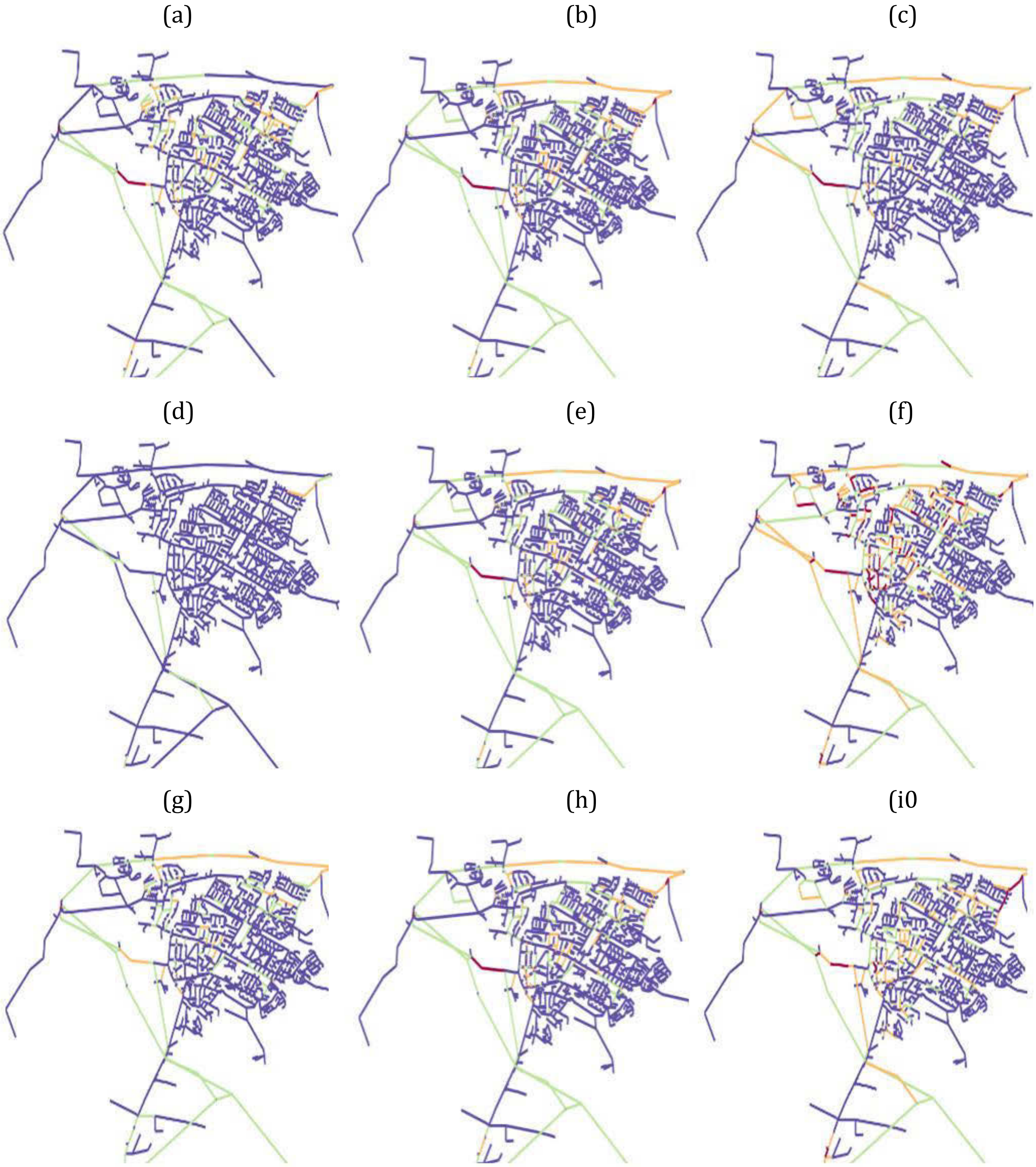


Figure 6. Calculated sediment configurations after 30 days at 12 AM for Castricum DWDS. The reference model is shown in frame (b), (e), and (h) and was assigned input parameters $\rho_p = 1100$ kg·m^{-3}, $d_p = 1 \cdot 10^{-6}$ m, $\theta_c = 1$ and; $\theta_{rs} = 10$. Input parameters that deviate from the reference model are (a) model S1 ($\rho_p = 1050$ kg·m^{-3}), (b)S2 ($\rho_p = 1150$ kg·m^{-3}), (c) S3 ($d_p = 0.88 \cdot 10^{-6}$ m), (d) S4 ($d_p = 1.12 \cdot 10^{-6}$ m), (e)S5 ($\theta_c = 0.5$; $\theta_{rs} = 5$), and (f) S6 ($\theta_c = 2$; $\theta_{rs} = 20$). Color legend is identical to Figure 5.

4.3 Comparing results to turbidity measurements

A proper validation of the modelling results based on the RPM measurements in Castricum was difficult to make for a number of reasons.

Hydraulic uncertainty

Although it was a deliberate choice to perform the numerical calculations with recurring 24 h demand patterns, this also simplified the hydraulic variations that are expected in a real-life DWDS. Those variation originate from daily and seasonal total demand variations, the stochastic nature of customer demand, and the occurrence of small network interventions. Despite the area having been relatively “quiet” in terms of large network operations and cleaning actions (see Section 4.1), smaller interventions are known to have occurred throughout the evaluation period.

In general, this level of detail of the hydraulic variations is not available for DWDSs. Nevertheless improved knowledge on the uncertainty imposed on the results can be obtained, e.g. by extending the sensitivity analysis with stochastic demand modelling and hydraulic variations related to daily and seasonal demand variations and network interventions.

Limited monitoring resolution of turbidity measurements

Despite the fact that the area being monitored in a systematic way over more than a decade, an even finer resolution of measurements would be required to make a convincing comparison between the contaminations calculated with *Aquarellus* and those inferred from RPM measurements.

This level of detail would require an even more extensive measuring program. While this is feasible for specific areas, large-scale implementation would involve big investments and take many years. To address this issue, a comparison at a larger spatial scale could be facilitated by aggregating both the results of *Aquarellus* and the available measurements on the scale of neighborhoods.

Sediment mass vs. RPM turbidity

The RPM is devised to monitor the *mobile fraction* of particulate material, rather than *all* material in a pipe segment. However, in the sensitivity tests the presence of *all* sediment mass was analysed. A more precise comparison is possible, but this require the simulation of multi-species sediments that would respond selectively to RPM shear stresses.

It is known that sediment mass relates to turbidity in an approximately linear fashion for a particular distribution area, but that the slope of this dependence can vary strongly between areas. This makes it difficult to interpret the absolute levels of sediment masses predicted by *Aquarellus*. Consequently, the predicted sedimentation configurations are best compared to field measurements in a relative sense (e.g. by dividing the sediment masses by the average sediment mass in the study area).

With the above issues in mind, it is not possible to draw strong conclusions from the comparative analysis. Of the two identified hotspots in the results, the Northeastern hotspot coincides with the two locations where the highest turbidity values were measured (cf. Figure 5, Figure 4c, and Figure 5b). The Southwestern hotspot coincides with few measurements, and no high-turbidity measurements in its vicinity.

5 CONCLUDING REMARKS

This paper presents the numerical tool *Aquarellus* that was developed to predict the accumulation of particulate material from source locations in a DWDS via particle settling, bed-load transport and resuspension. The code allows for calculations of hydraulic network models of several hundreds of km on a standard laptop computer. The user can specify the properties of particulate material (diameter d_p, mass density ρ_p, and shear stress thresholds to incipient motion θ_c and resuspension θ_{rs}). This material can enter the network from multiple (user-specified) source locations according to temporal concentration patterns.

The lab experiments presented in this paper demonstrate the variation of particle properties (d_p, estimated ρ_p, and half-life time $\tau_{1/2}$ of the turbidity in a settling experiment, which was used as a proxy for the setting velocity u_s) found across and within three Dutch and Flemish DWDSs. The results of a sensitivity analysis show the influence of d_p, ρ_p, θ_c and θ_{rs} on the absolute levels of sedimentation configurations in the model of the real-life DWDS of Castricum. The resultsfrom *Aquarellus* suggest the presence of two hotspot areas with a higher propensity to particle accumulation than their surroundings.

Although an attempt was made to validate the predictions of *Aquarellus* with field measurements of a full-scale DWDS, it was difficult to derive strong conclusions from this comparative study. The reason lies mainly in (i) a limited knowledge of hydraulic conditions at the level of individual pipes, related to stochastic customer demand, daily and seasonal demand variations, as well as operational anomalies, and (ii) a limited monitoring resolution of turbidity measurements; although thoroughly monitored (pilot) areas exist, this limitation will remain innate to most DWDSs for the foreseeable future. A more thorough validation with field data is required to gain confidence in the predictive power of *Aquarellus* (and, by extension, a deterministic approach).

Several future modelling steps can address some of these limitations. A more extensive sensitivity analysis could investigate the influence of plausible demand fluctuation (stochastic, seasonal, etc.) on sedimentation configurations. In light of the (limited) available turbidity measurements, the aggregation of calculated sediment masses and measured turbidity values (e.g. at scale of neighborhoods), would allow for a more significant comparison between predictions and observations (albeit on a lower spatial resolution). From the measurement perspective, it is recommended to monitor turbidity over long periods in a standardized manner to gain further insight in contamination pattern of areas that can be used for validation purposes.

The potential benefit for system operations of *Aquarellus* lies in the prioritization of subregions of faster than average accumulation of particulate material though prioritization. As such, *Aquarellus* can help the planning of cleaning actions and measuring programs. Ultimately, with enough confidence in its predictive power, *Aquarellus* could also be used to further optimize the self-cleaning capacity of DWDSs [1, 2, 14]. Although more thorough validation is desired, *Aquarellus* may already be used as a prioritization aid to plan cleaning actions or measuring campaigns.

6 ACKNOWLEDGEMENTS

The research presented in this paper was conducted within a project of the Joint Research Program (BTO) of the Dutch Water Companies and De Watergroep. We greatly appreciated the help of Evi Loozen (De Watergroep), Henk de Kater (Evides), and Jan Pot (Vitens) in organising the field sampling campaigns and Matthijs Rietveld (PWN) in making available the Castricum hydraulic model and measurements.

7 REFERENCES

[1] J. Vreeburg, “Discolouration in drinking water systems: a particular approach”, Ph.D. Thesis, Delft University of Technology, 2007.

[2] E. Blokker, J. Vreeburg, P. Schaap, and J. van Dijk, “The self-cleaning velocity in practice”. 12th Annual Conference on Water Distribution Systems Analysis (WDSA). 2011.

[3] S. Mounce, E. Blokker, S. Husband, W. Furnass, P. Schaap, and J. Boxall. “Multivariate data mining for estimating the rate of discolouration material accumulation in drinking water distribution systems. IWA J. Hydroinformatics, 18, pp.96-114, 2016.

[4] E. Blokker, P. Schaap, and Vreeburg, “Self-cleaning networks put to the test”. IWA Leading Edge conference on Strategic Asset Management (LESAM 2007), Lisbon, 2009.

[5] S. Richardt, A. Korth, and B. Wricke, ”Model for the calculation of optimized flushing concepts”, TECHNEAU: Safe Drinking Water from Source to Tap, Ed. By T. van den Hove, and C. Kazner., IWA Publishing, London, 2009.

[6] G. Ryan, P. Mathes, G. Haylock, A. Jayaratne, J. Wu, N. Noui-Mehidi, C. Grainger, and B. Nguyen, “Particles in Water Distribution Systems.” Rpt. No. 33. The Cooperative Research Centre for Water Quality and Treatment, 2008.

[7] W. Furnass, R. Collins, S. Husband, R. Sharpe, S. Mounce, and J. Boxall, “Modelling both the continual and erosion and regeneration of discolouration material in drinking water distribution systems”. Water Sci. Technol. 14, pp. 81-90, 2014.

[8] J. Boxall, A. Saul, J. Gunstead, and N. Dewis, "A novel approach to modelling sediment movement in distribution mains based on particle characteristics”. Water Software Systems, Ch. 1: Theory and Applications, Eds.: B. Ulnacki B. Coulbeck, J. Rance. Hertfordshire, UK, 2001.

[9] P. van Thienen, J. Vreeburg, and E. Blokker, “Radial transport processes as a precursor to particle deposition in drinking water distribution systems". Water Res., 45, pp. 1807-1817, 2011.

[10] J. van Summeren & M. Blokker, “Modeling particle transport and discoloration risk in drinking water distribution networks”, Drink, Water Eng. Sci. 10, 2017, pp.99-107, 2017.

[11] L. Rossman, “EPANET 2 Users Manual”, U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-00/057, 2000.

[12] OpenStreetMap. https://www.openstreetmap.org/ (2022).

[13] J. Vreeburg, P. Schaap, and J. van Dijk, “Measuring discoloration risk: resuspension potential method,”in M. van Loosdrecht & J. Clement (Eds.), 2nd IWA Leading-Edge Conference on Water and Wastewater Treatment Technologies, Prague, Czech Republic. 2005, pp. 123-132.

[14] E. Abraham, E. Blokker, I. Stoianov, “Decreasing the discolouration risk of drinking water distribution systems through optimized topological changes and optimal flow velocity control”. J. Wat. Res. Planning and Management. 144 (2), 2018.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14135

ADVANCED FIRE FLOW RISK ANALYSIS USING EPANET

Alexander Sinske[1], Altus de Klerk[2] and Adrian van Heerden[3]

[1,2,3] GLS Consulting, Stellenbosch, WC, South Africa

[1]Alex.Sinske@gls.co.za, [2]Altus.deKlerk@gls.co.za , [3]Adrian.vanHeerden@gls.co.za

Abstract

A water reticulation system is key infrastructure that enables hydraulic water services. It is therefore a critical component in providing the full level of water service to a city's consumers. Extended water outages and below minimal pressure is a risk for the water network and especially for fire flows.

GLS has developed and implemented a multi-threaded client-server software application based on the latest open-source EPANET 2.2 hydraulic analysis engine which now enables city-wide fire flow risk analyses on a property-by-property basis in reasonable time.

Previously it has been virtually impossible to assess the risk on a city-wide basis, for each and every property, and to consider the improvement of such risk in the master plan (MP) of the water distribution system. Hence, the focus of the MP has been the provision of flows and pressures for the peak hour demand scenario in the network. Consequently, many township developments, densifications, and land use re-zonings require a separate and focussed specific fire risk analysis to ensure the existing system is capable of providing the requisite fire flow and pressure. If not, specific additional MP items related to the fire requirements of the specific property have to be investigated and considered for implementation.

The fire flow risk analysis produces a GIS-based heat map displaying a Fire Risk Compliance Score (FRCS) of areas and properties where the current system is inadequate to deliver the fire flows and pressures. Such a heat map provides valuable information that can be used to improve and prioritise the MP and minimise future additional ad-hoc analyses for specific properties or developments.

The fire flow risk analysis also allows the identification of pressure management zones where adjustments to the pressure regime are required in order to ensure requisite fire flows/pressures are achieved.

Various methodologies based on Pressure Driven Analysis and Demand Driven Analysis have been evaluated and tested on models from South African cities. Great care has been take to optimise the multi-threaded communication of the application with the EPANET engine to streamline performance and support concurrent hydraulic analyses.

In addition, a concept of automatically creating unique fire events to reduce the number of analyses, has been introduced for large models.

This has resulted in smaller cities that can be analysed on modern PCs with few processor cores in a few minutes and large cities that can be analysed in reasonable time on high-core cloud-computing platforms.

Visualisation in GIS-based software greatly helps to control the analyses and interpret results visually. Critical areas can be identified on a broader scale and allows for a rational approach to decide where to focus on network augmentation, or alternatively to provide on-site fire fighting capabilities.

Keywords
Fire flow simulation, EPANET, PDA, Wadiso.

1 INTRODUCTION

The water reticulation system is key infrastructure that enables hydraulic water services. It is therefore a critical component in providing the full level of water service to a city or town's consumers. Extended water outages and below minimal pressure is a risk for the water network and especially for fire flows.

GLS has developed new functionality in their *Wadiso* [1] software which enables system-wide fire flow risk analysis on a property-by-property basis.

Whereas it has previously been possible to assess the risk of providing the requisite fire flow and pressure at a specific property, up until this development, it has been virtually impossible to assess the risk on a city-wide basis, for each and every property, and to consider the improvement of such risk in the master plan (MP) of the water distribution system. Hence, the focus of the MP has been the provision of flows and pressures for the peak hour demand scenario in the network. Consequently, many township developments, densifications, and land use re-zonings require a separate and focussed specific fire risk analysis to ensure the existing system, in combination with the implementation of certain (peak hour demand) MP items, is capable of providing the requisite fire flow and pressure. If not, specific additional MP items related to the fire requirements of the specific property have to be investigated and considered for implementation.

A fire risk analysis provides a heat map of areas and properties where the current system is unable to deliver the fire flows and pressures. Such a heat map provides valuable information that can be used to improve and prioritise the MP and minimise future additional ad-hoc analyses for specific properties or developments.

As part of Water Conservation and Water Demand Management (WCWDM) initiatives pressure management zones are often proposed, requiring Pressure Reducing Valves (PRVs) to reduce pressures and hence reduce leakage and water consumption. This is commendable practice, but is often proposed without considering the potential negative effect it might have on fire flow risk. The fire flow risk analysis allows the identification of such pressure management zones where certain adjustments to the pressure regime are required in order to ensure requisite fire flows and pressures are achieved.

This paper covers the Software Development, Model Preparation and Result Presentation.

2 SOFTWARE DEVELOPMENT

2.1 Using the EPANET 2.2 Toolkit

EPANET [2] is a software application and programmers toolkit that was originally developed by the US EPA for modelling the flow of drinking water and constituents within distribution systems. The most common release is 2.00.12 released in 2008. More recently EPANET 2.2 was released in December 2019 and is managed as an Open Source initiative by OpenWaterAnalytics [3] to maintain and extend EPANET. Two key improvements of EPANET 2.2 toolkit are the ability to analyse multiple projects in parallel in a thread-safe manner and the ability to use pressure dependent demands in hydraulic analyses [4]. Both these improvements have been utilised in the *Wadiso FireFlowServer* application. The documentation of the OWA-EPANET Toolkit [5] provides detail on the usage.

2.2 Choice between Demand Driven & Pressure Driven Analysis

Pipe network hydraulic analyses tools and engines have historically only supported Demand Driven Analysis (DDA), where fixed demands are given at all nodes in the model. This type of analysis will often result in large negative pressure heads calculated at nodes as a result of excessive frictional head losses in pipes. In reality this is not possible, as the demand at nodes will drop.

Modern hydraulic analysis software such as EPANET 2.2 also provide the option of performing a Pressure Driven Analysis (PDA), where the demand is defined by a power function of pressure, up to the point where the full demand is met.

The *Wadiso FireFlowServer* has been developed to support both analysis options, DDA and PDA.

2.3 Formulation of DDA fire flow analysis

Before the analysis, each property is assigned a specific fire risk category which states the requisite fire demand and pressure as well as the number of hydrants required within a prescribed distance to enable fire fighting. The closest nodes to the hydrants are retrieved for each property, and together with the proportionally distributed fire demand, is then assigned to a node set which constitutes a fire event. Each fire event will be analysed in a separate analysis run/processing thread.

A base demand DDA simulation is performed, i.e. before any fire demand is superimposed. This could be the peak hour or peak day demand scenario, as per the relevant city's fire flow criteria.

For each fire event the fire demand is superimposed on the nodes, and a DDA analysis is performed. If however the resultant pressure for the base analysis is below the requisite pressure as defined by the fire event's risk category, then a Fire Risk Compliance Score (FRCS) of 0% will be assigned to the fire event and the fire event will not be analysed.

If the requisite fire pressure is achieved at all nodes in the node set for the fire event, then the FRCS is 100%. If the resultant pressure is below the required pressure at any of the nodes, then a heuristic calculation is employed to calculate the fire flow demand that could be achieved at the prescribed pressure, by utilising the base analysis and the fire event results. Both a linear and power function headloss curve is considered, and the averaged flow at the prescribed pressure between the two curves is reported as the achievable fire demand considering the prescribed fire flow pressure value. The FRCS is then the ratio (%) of the fire flow that can be achieved at the requisite pressure, e.g. the required fire flow is 50 L/s @ 20m, but at 20m a fire flow of only 30 L/s is achievable, then the FRCS is 30/50x100 = 60(%).

The FRCS score can be further expanded to include other criteria. For example, the distance to the nearest fire station. This, however, is beyond the scope of this paper.

2.4 Formulation of PDA fire flow analysis

The fire event creation phase is identical to that of the DDA, with the addition of providing the PDA analysis settings required for the EPANET solver.

PDA assumes the demand delivered is a function of nodal pressure as follows [6]:

$$d = D\left[\frac{p - P_{min}}{P_{req} - P_{min}}\right]^{P_{exp}} \tag{1}$$

where d is the delivered demand, D is the full required demand, P_{min} is the pressure below which demand is zero, P_{req} is the pressure required to deliver the full required demand (before reduction occurs) and P_{exp} is an exponent (typically 0.5). When $p < P_{min}$ demand is 0 and when $p > P_{req}$ demand is equal to D.

As with the DDA, a base demand PDA simulation is performed, i.e. before any fire demand is superimposed. If the achieved demand for the base analysis is below the base demand then a FRCS of 0% will be assigned to the fire event and the fire event will not be analysed.

For each fire event the fire demand is superimposed on the nodes, and a PDA analysis is performed. The FRCS is then the ratio (%) of the achieved demand compared to the fire event demand (base demand + superimposed fire demand).

2.5 Developing a multi-threaded analysis server software

The software is currently designed to run on a server accessible from the Local Area Computer Network. The server-side software has been developed in the latest edition of Delphi programming language. A folder is monitored for input files (*.ffinput) to be processed in a queue. The typical format for the input file is:

```
NodeSet,NodeCode,FireDemand
```

NodeSets are numbered and represent a set of nodes that should have a *FireDemand* fire flow demand superimposed to a base demand, for every *NodeCode* node. A single period steady state analysis is then performed on the hydraulic model for every *NodeSet*.

A matching EPANET text-based input file (*.inp) represents the hydraulic model. Coordinates of nodes are not required in this file. Optionally a meta file with additional parameters controlling the analysis (e.g. the choice between DDA & PDA analyses and convergence parameters) can be provided. Additional control files can be placed in the folder to stop or restart a running process. Once the process is complete a (*.ffdone) marker file is written, together with an output file (*.ffoutput). The output file typically has this format:

```
NodeSet,NodeCode,FireDemand,ResStatus,ResPressure,ResDemand,ResRatio
```

where in addition to the input fields that are repeated for readability, *ResStatus* stores the EPANET balanced status information and other critical error conditions, *ResPressure* stores the resulting pressure head, *ResDemand* the resulting demand, and *ResRatio* the FRCS for the specific node.

MaxHeadloss & *MaxVelocity* for all the links in the model is also exported in a separate file for additional verification by the engineer.

2.6 Performance comparison of multi-threaded server software

For larger cities, such as the City of Cape Town in South Africa, the number of modelled nodes can exceed 100,000. In addition, the City of Cape Town has in excess of 500,000 stands.

Performing 100,000+ analyses on a 100,000 node sized hydraulic model can take a considerable amount of time. The development of a multi-threaded and highly optimised analysis engine capable of running on fast multi-core servers was therefore essential.

Using multi-threading code of Delphi, in a 64 bit compiler and ensuring thread-safety is maintained at all times, a performance optimised application has been developed. One thread is used for a complete hydraulic analysis at a time, through the EPANET 2.2 Toolkit.

The application has been tested on various Windows environments. Figure 1 shows result where the maximum number of Concurrent Threads to be used and selected in the software is displayed on the x-axis and the number of completed hydraulic analysis per minute, of the 138,073 node CoCT sample model, is shown on the y-axis.

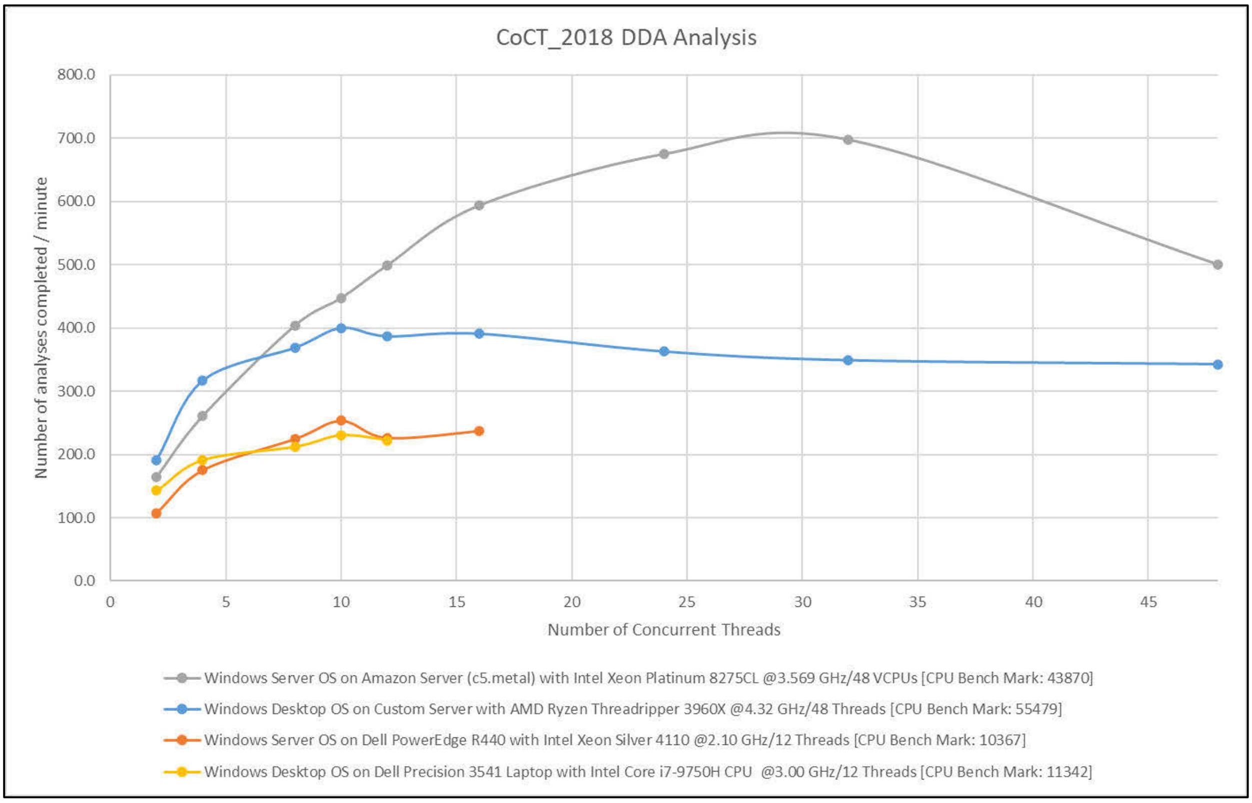


Figure 1. Performance of the Wadiso FireFlowServer.

The following observations are made:

Even a high-end laptop is capable of running up to 250 analyses per minute using 10 of the 12 computing threads available in the computer (yellow line). An older 16 computing thread server performs similar (red line).

A custom server using a mid-range 48 computing thread processor can reach 400 analyses per minute, but interesting also only at around 10 computing threads. This implies that other hardware in the computer such as the speed of the SSD hard disk, speed of memory, size of CPU cache and the bus architecture play a role (blue line) in performance, not allowing scaling to all computing threads with this specific model.

The superior server architecture of a cloud computing server is clear. Although at under 8 concurrent threads the faster CPU of the custom server outperforms the cloud computing server, the latter scales to about 32 concurrent threads before other overheads catch up. A maximum of about 700 analyses per minute was recorded (grey line).

2.7 Visualisation in GIS-based hydraulic modelling client software

The GIS-based visualisation of input and output data from the Wa*diso FireFlowServer* is critical for the engineer to ensure the input is correct and the results are meaningful. This was best accomplished using the *Wadiso* hydraulic modelling software, which is also based on the EPANET 2.2 engine. Individual analyses can be verified in the software. For comparison a single steady state analysis of the CoCT model within the GIS-based environment, with updating of the result map display (e.g. pressure head) takes normally about 50 seconds on the laptop mentioned before. The industry standard non-GIS EPANET 2.0 takes about 20 seconds for comparison. The *Wadiso FireFlowServer* on the cloud computing server is therefore 233 faster than attempting to perform the analysis manually in the EPANET 2.0 UI and 583 times faster than attempting to use the GIS-based hydraulic analysis software package.

Figure 2 shows the extent of the CoCT_2018 hydraulic model for reference.

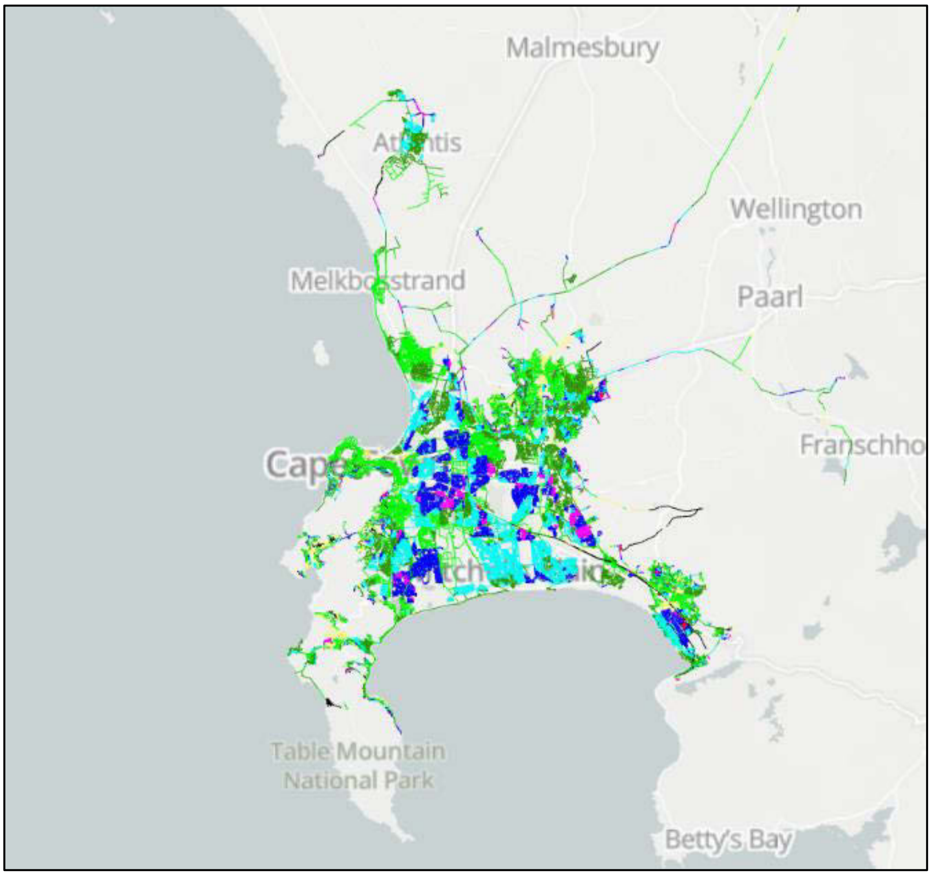


Figure 2. Extent of the CoCT_2018 hydraulic model as visualised in Wadiso.

3 MODEL PREPARATION

3.1 Model Creation

The model needs to be created and verified. The model can be populated with theoretical demands per land use or in our case study the Average Annual Daily Demand (AADD). A model should generally be well calibrated before performing a fire flow study. An extended period time simulation could be run, but the instantaneous nature of the fire event allows for a steady-state analysis. In a fire flow scenario tank levels are not looked at in detail.

Another important factor for using a steady-state analysis rather than a time simulation, is time. When analysing a system for example with 60 000 stands. Every second of computation time adds more than 16 hours to the analysis time if running on a single thread. This can be mitigated by using machines with multiple cores (as discussed earlier), breaking models into smaller discrete zones, or finding unique fire events to reduce the number of simulations required (discussed below).

3.2 Stand Categorisation

In our case study, stands are categorised according to local standards, such as the South African Bureau of Standards (SANS) code [7]. The code defines the fire flow requirements to be delivered to a stand in the event of a fire. This will vary according to each nation's or region's own standards.

The requirements in the SANS code are generally the minimum number of hydrants to be within a distance from the fire, the minimum flow that needs to be delivered and the minimum allowable pressure at the hydrant. Normally the base analysis for a fire event is simulated at two times the Annual Average Daily Demand to simulate a relative high demand of water in the area.

Below is an extract of the SANS codes for fire flow events.

Table 1. SANS Codes [7]

Code	Hydrant count	Distance (m)	Flow (L/min)	Min. Pressure (m)	Description
SANS D1	2	300	1 900	15	SANS 10090 Category D1: Houses > 30 m apart
SANS D2	2	200	2 850	15	SANS 10090 Category D2: Houses 10,1 to 30 m apart
SANS D3	3	200	3 800	15	SANS 10090 Category D3: Houses 3 – 10 m apart
SANS D4	3	200	5 700	15	SANS 10090 Category D4: Houses < 3 m apart
SANS C	3	200	6 000	15	SANS 10090 Category C: Non-residential buildings with divisions not greater than 1250 m^2
SANS B	5	120	9 000	15	SANS 10090 Category B: Non-residential buildings having divisions not greater than 2 500 m^2
SANS A	7	85	13 000	15	SANS 10090 Category A: Non-residential buildings with divisions not greater than 5000 m^2

3.3 Hydrants

Hydrants can either form part of the model by flagging nodes as hydrants or they can be placed as points on a different model layer, known as model appurtenances. The advantage of having the hydrants physically part of the model is accuracy. On larger systems it might be preferable to have fewer nodes and rather allocate the hydrant's demand to the nearest node. This approach would also be acceptable when looking at the bulk system's capacity for handling fires, while having detail up to the node level would be preferable when looking at the local system's ability to deliver adequate water during a fire.

This process can also be done in two different ways. The first is to find the closest hydrant "as the crow flies". This is the easiest but can often lead to a hydrant being chosen that would not necessarily be used for fire fighting. The second option would be to find the closest pipe to the fire and then traversing the network to find the closest hydrants along the route. This often leads to more accurate results as pipes tend to follow roads and as such the path is generally the shortest road path to the hydrants. A more accurate method would be to get actual road layouts, but this information is not always available and would not necessarily increase the accuracy by much.

3.4 Finding Unique Fires

When analysing larger systems it becomes advantageous to recognise unique fire events. A number of stands in the same area often share the same land use category as well. This means that they have the same fire flow requirements. If the closest hydrants also happen to be the same hydrants, then the fire event can be classified as the same event. As a result a number of stands can be analysed only once.

For example on the CoCT system with 587 000 stands, the number of unique fires can be reduced to only 82 112 individual events that need to be analysed. Using the metrics from previous sections this will reduce the analysis time on a laptop doing 250 analyses per minute from 39 hours to less than 6 hours. When using the cloud computing server doing 700 analyses per minute the time will decrease from 14 hours to under 2 hours.

Figure 3 shows the distribution of unique fire events in a residential area where hydrants were selected using the basic "as the crow flies" method.

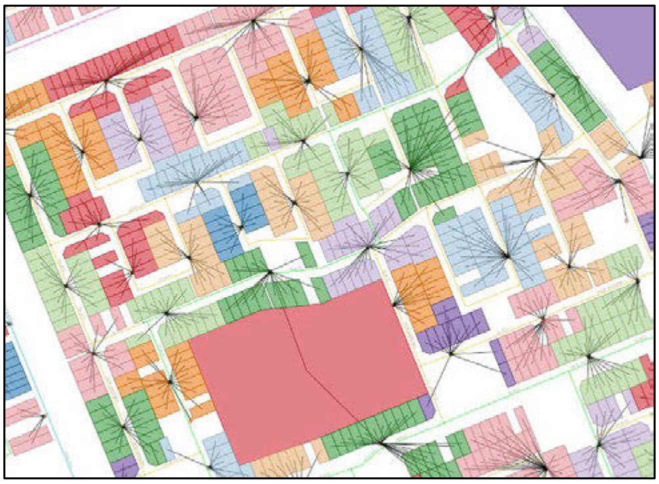


Figure 3. Example of unique fire events

4 RESULT PRESENTATION

4.1 Post Analysis

For result presentation a smaller coastal municipality near the City of Cape Town was used as a test case mainly due the larger variation of property land use, and thus risk categorisation, as well as greater elevation changes in close proximity, to better illustrate differing FRCS results.

Analysis results are imported for each fire event and applied to the relevant property. This allows for a spatial visualisation for levels of compliance in terms of the categorisation table. Critical areas can be identified where intervention will be required.

BY NC SA

Table 2 shows a small extract example of the result table after results are imported and assigned to the properties that are linked to each unique fire event. A status message is provided that indicates which requirements have not been met in cases where the FRCS is less than 100% (imported as a ratio). A comment is provided in cases where an insufficient number of hydrants (hydrant shortage) are within the required distance from the fire event.

Table 2 Example of Result table

ID	Status	FRCS	Comments	Category	Description
HIND 02362	Not achieved: Fire demand	0.71	The nearest hydrant is beyond the required distance away	BUS_COMM	SANS 10090 Category C
HMP 06183	All requirements achieved	1.00		RES[2000]	SANS 10090 Category D1
HSB 00250	All requirements achieved	1.00		RES[1500]	SANS 10090 Category D1
HSB 01239	Not achieved: Fire demand	0.97		RES[1000]	SANS 10090 Category D2
HSB 01378			Excluded from analysis	NO FIRE	
HZW 00658	Not achieved: Fire demand & minimum pressure at a node/hydrant	0.61		RES[500]	SANS 10090 Category D3

Hydrant compliance is important to ensure redundancy for fire fighting capability. Areas may have insufficient hydrant distribution regardless of network capabilities. Figure 4 provides and example where no hydrants are found within the required 200 m distance for a property classified as SANS 10090 Category D2: Houses 10.1 to 30 m apart.

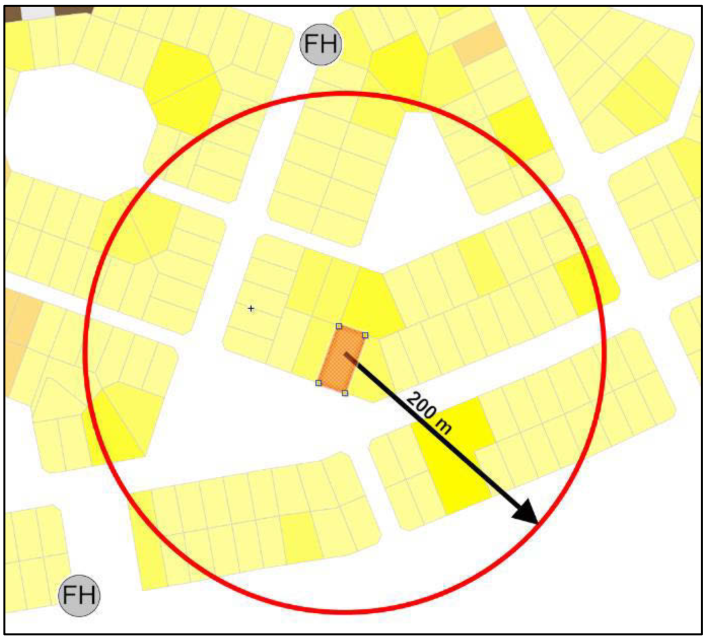


Figure 4. Example of hydrant shortage (non-compliance)

Only part of the larger test case area is displayed in Figures 5, 6 and 7 for clearer illustration of the analysis results. Properties that have a hydrant shortage based on their risk categorisation have a dark red border.

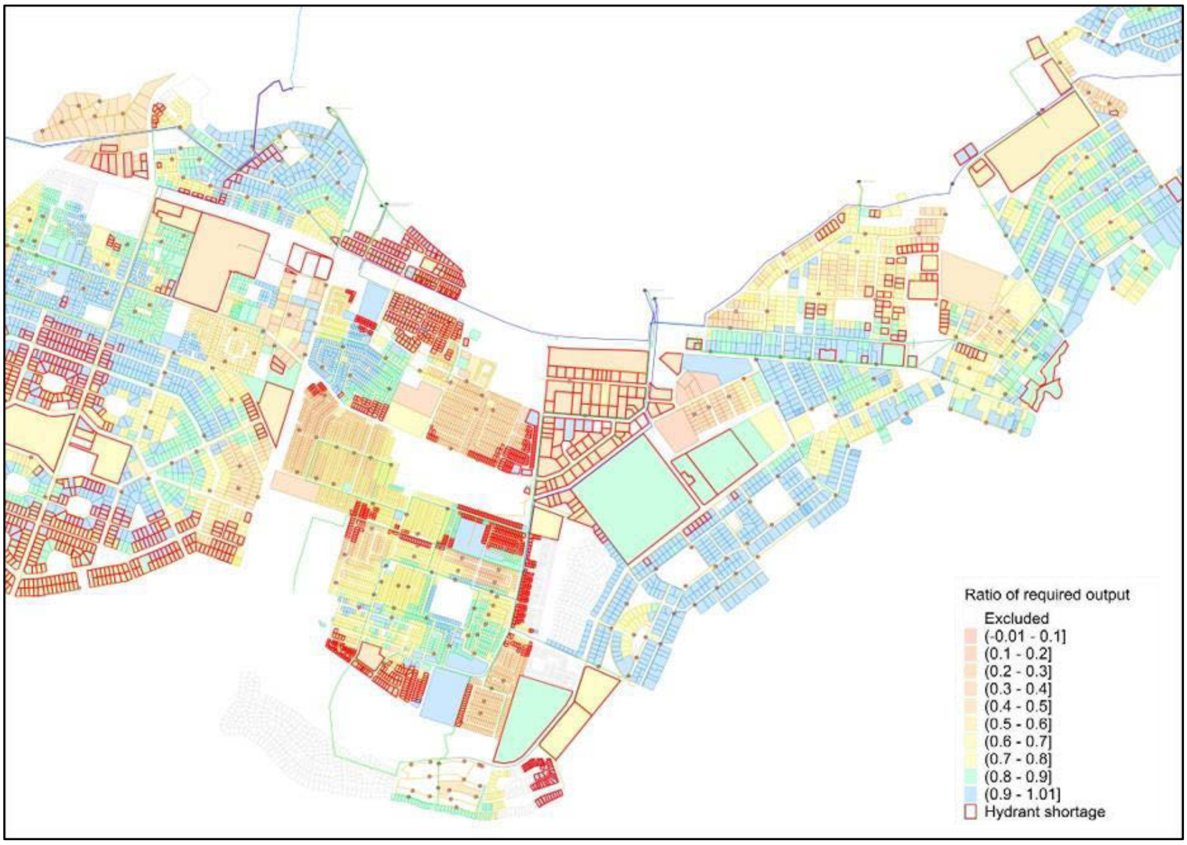

Ratio of required output
Excluded
(-0.01 - 0.1]
(0.1 - 0.2]
(0.2 - 0.3]
(0.3 - 0.4]
(0.4 - 0.5]
(0.5 - 0.6]
(0.6 - 0.7]
(0.7 - 0.8]
(0.8 - 0.9]
(0.9 - 1.01]
Hydrant shortage

Figure 5. Test Case fire compliance visualisation before network reinforcements

Interpretation of the flow and pressure results are dependent on the analysis method used. With a PDA, achieved fire flow is considered along with the pressure, as demands in the model are reduced based on simulation settings for the PDA analysis. With a DDA, the achievable fire flow at the prescribed pressure is provided where the FRCS is below 100%.

If the network cannot deliver the required fire flow at the prescribed pressure, then intervention is required to augment the network to allow for fire fighting compliance. The augmented model is then analysed to determine if adequate compliance is achieved. For quicker turn-around time a subset or even a single fire event can be analysed for the augmented model, to determine if adequate compliance is achieved when considering localised improvements/changes to the network.

Figure 6 shows the required network reinforcements required for fire fighting compliance, over and above standard MP upgrades that are required for sufficient nodal pressure and flow velocity limitation within the system, that would be required regardless of the fire flow analysis.

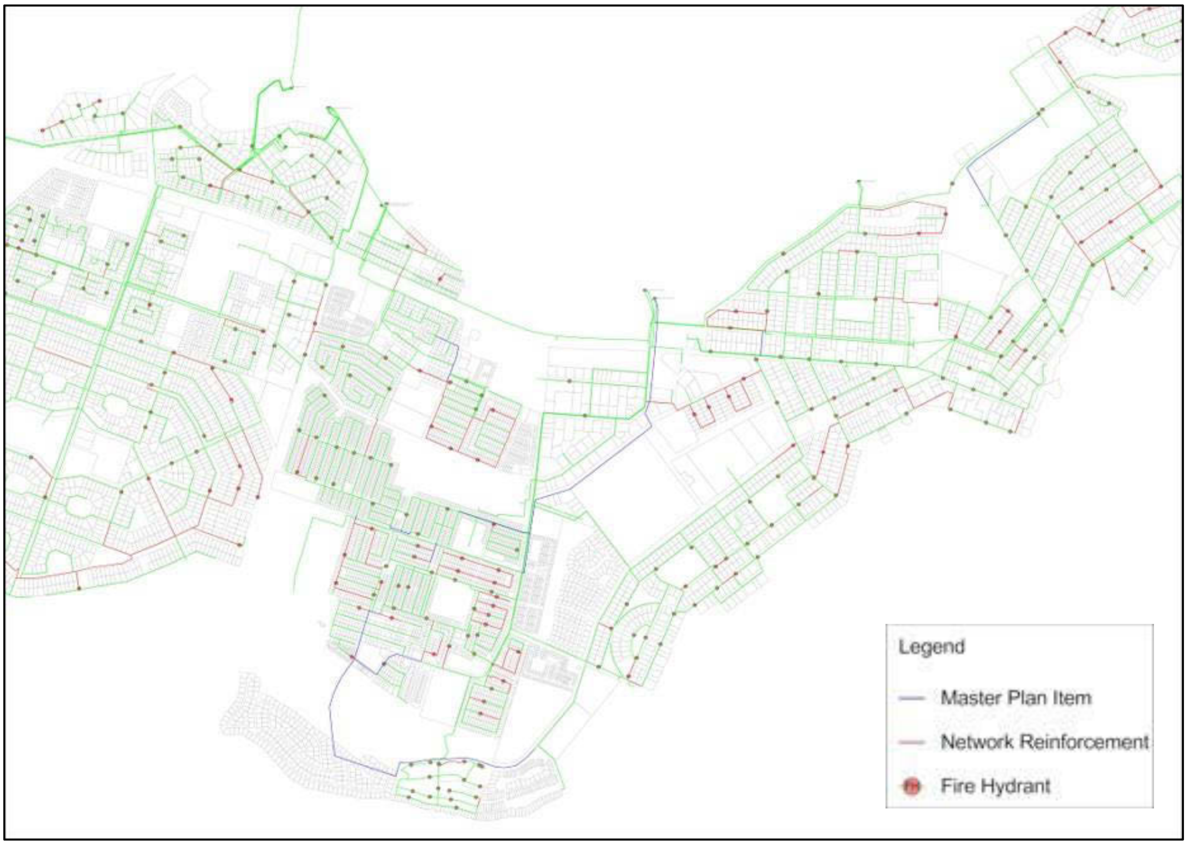


Figure 6. Test Case Network upgrades

After implementation of the proposed improvements the fire risk is dramatically improved as shown in Figure 7, with the majority of the properties now indicating full to only slight non-compliance, when compared to Figure 5.

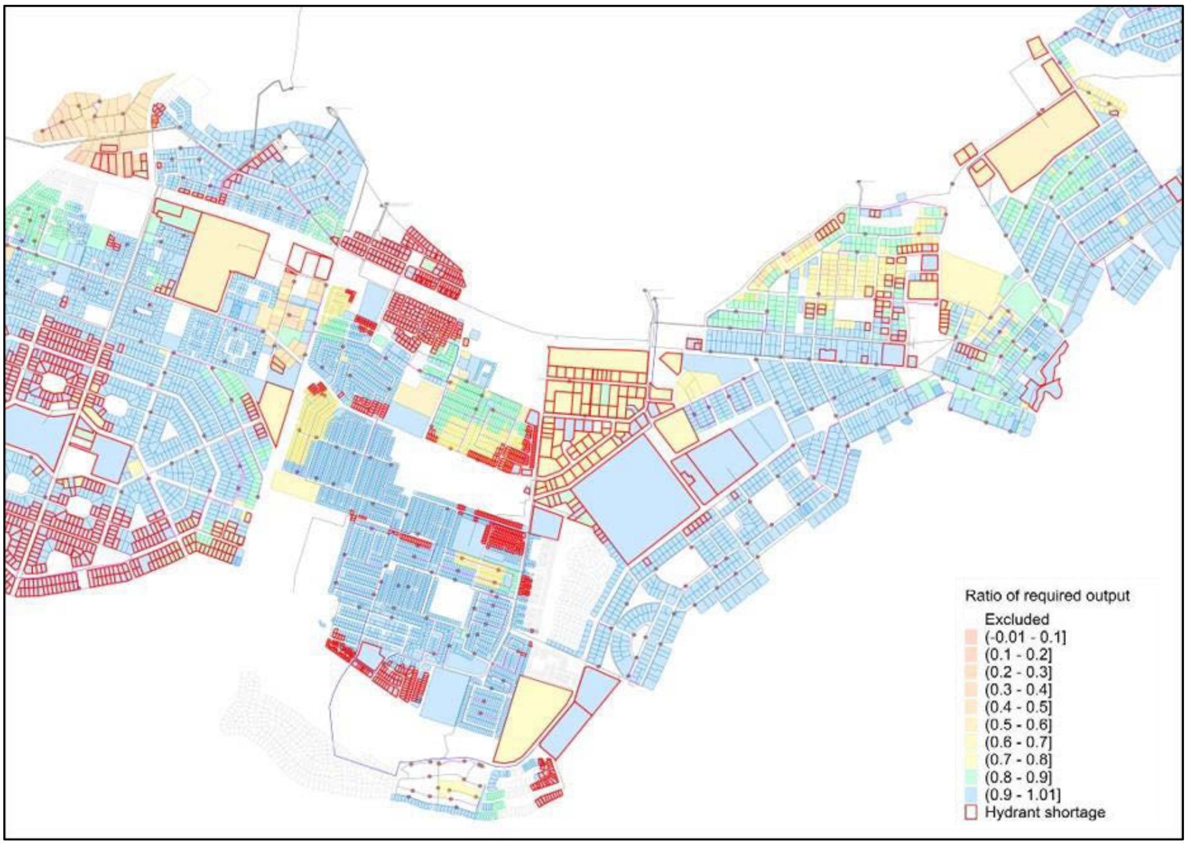

Ratio of required output
Excluded
(-0.01 - 0.1]
(0.1 - 0.2]
(0.2 - 0.3]
(0.3 - 0.4]
(0.4 - 0.5]
(0.5 - 0.6]
(0.6 - 0.7]
(0.7 - 0.8]
(0.8 - 0.9]
(0.9 - 1.01]
Hydrant shortage

Figure 7. Test Case fire compliance visualisation after network improvements

Besides proposing network reinforcements for pipes with high velocity and headloss, it is important to apply engineering judgement and consider existing pressure reduction zones, and investigate if any pressure reduction settings can be increased to improve the results of the fire flow analysis before other reinforcements are proposed. Consider Figures 8 and 9 below which shows the FRCS before and after a PRV bypass is implemented.

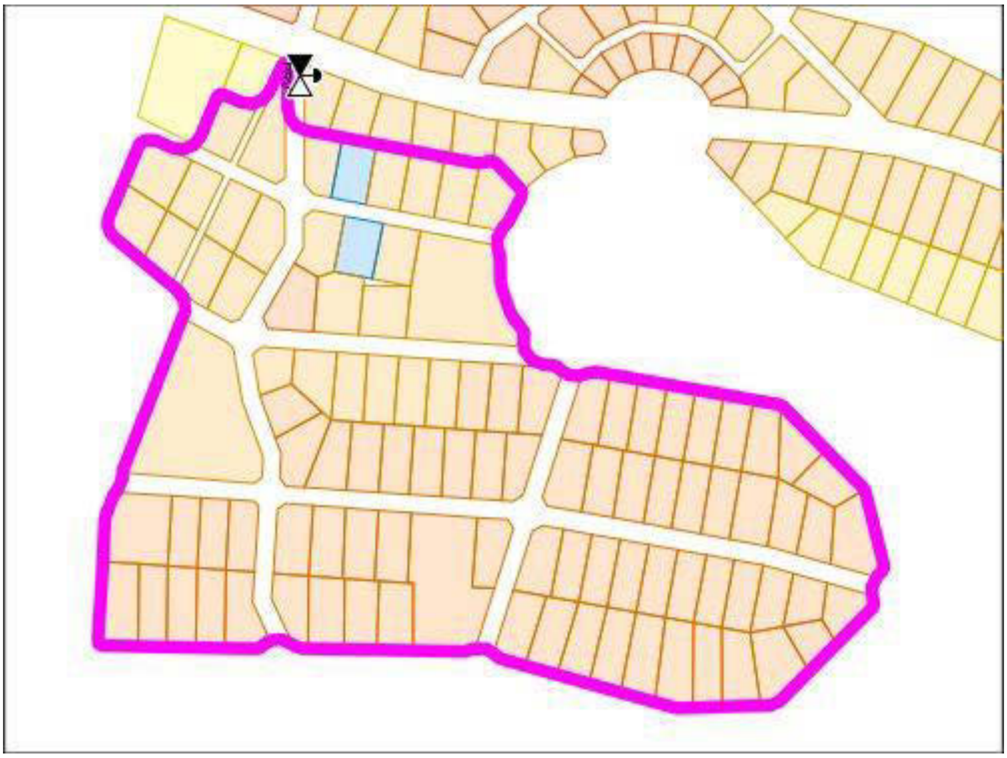


Figure 8. PRV zone FRCS before intervention

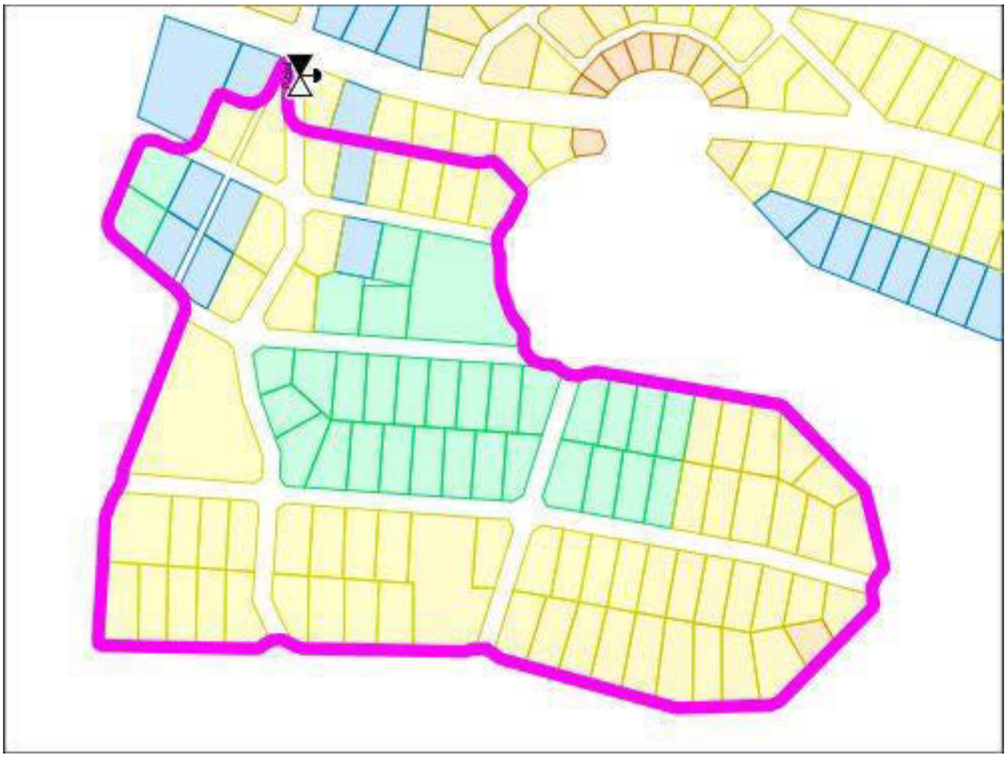


Figure 9. PRV zone FRCS after intervention (PRV bypassed)

5 CONCLUSIONS

The development of a multi-threaded client server software application, based on the latest open source EPANET 2.2 hydraulic analysis engine, in partnership with a GIS-based water analysis software enables city wide fire flow risk analysis on a property-by-property basis.

A Fire Risk Compliance Score is assigned for each property and produces a heat map of fire flow requirement compliance based on required flow, pressure and number of feeder hydrants within a minimum distance. This provides an invaluable view of the state of the water network to determine where the system is inadequate to deliver fire flows at the required pressure. Critical areas can be identified on a broader scale and allows for a rational approach to decide where to focus on network augmentation, or alternatively to provide on-site fire fighting capabilities.

The ease of setup via the *Wadiso* front-end allows for quick creation of various fire flow risk analysis scenarios to compare severity of fire category assignment.

Future work includes moving the server to the cloud to eliminate Local Area Computer Network dependence and to improve the display of results by adding more meta data to the node sets supporting filtering based on criteria.

6 REFERENCES

[1] GLS Consulting, “Wadiso – Water Distribution and System Optimization,” https://www.gls.co.za/software/products/wadiso.html, January 2022.

[2] United States Environmental Protection Agency, “EPANET: Application for Modeling Drinking Water Distribution Systems,” https://epa.gov/water-research/epanet, February 2022.

[3] OpenWaterAnalytics,“OWA-EPANET,” https://github.com/OpenWaterAnalytics/EPANET, April 2022.

[4] OpenWaterAnalytics, “OWA:EPANET 2.2,” https://github.com/OpenWaterAnalytics/EPANET/releases/tag/v2.2, Dec. 2019.

[5] OpenWaterAnaltyics, “OWA-EPANET Toolkit 2.2,” http://wateranalytics.org/EPANET/index.html, Dec. 2019.

[6] OpenWaterAnalytics, “Release Notes for EPANET 2.2: Pressure Dependent Demands,” https://github.com/OpenWaterAnalytics/EPANET/blob/dev/ReleaseNotes2_2.md, Dec. 2019.

[7] SANS 10090:2018: Community protection against fire, Edition 3.01, Pretoria: South African Bureau of Standards, 2018.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14136

DATA DRIVEN APPROACH FOR EQUITABLE SUPPLY IN WATER NETWORKS

Adhityan R[1], Varghese Kurian[2] and Sridharakumar Narasimhan[3]

[1,2]Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India),
[3]Dept. of Chemical Engineering and Robert Bosch Centre for Data Science and Artificial Intelligence,
Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India)

[1]*nsraadhi@gmail.com*, [2]*varghese@alumni.iitm.ac.in*,
[3] *sridharkrn@iitm.ac.in*

Abstract

In many localities with limitations on the availability of water, Water Distribution Networks (WDNs) supply water only for a few hours in a day. In such networks, inefficient operational policies can lead to inequitable supply of water. This work proposes a data-driven approach for scheduling the supply side of WDNs for equitable distribution of available water. We formulate and solve a scheduling problem that makes use of flow measurements from the real network rather than a hydraulic model of the WDN. This helps reduce the effort required for model development and the errors arising out of it. Further, to limit the number of additional measurements required, we propose a heuristic for choosing measurements that are informative. In each step of this iterative procedure, a schedule is generated using the available measurements and based on this, a new system state that has to be measured is identified. The iteration is completed when acceptable performance is achieved. We demonstrate the advantages of this technique through simulations of a real WDN and experiments performed on a lab-scale network.

Keywords

Water distribution networks (WDN), Equitable distribution, Scheduling algorithms, LabVIEW.

1 INTRODUCTION

Water Distribution Networks (WDNs) are used for transporting water from source to consumers. These systems are composed of reservoirs, pipes, pumps, and other instrumentation. The flow rates in these systems may not be proportional to the demands at the consumers' side. In such situations, utility providers control the valves and pumps in the system to distribute water efficiently and equitably. However, the mathematical program for scheduling valves and pumps in a WDN is a difficult problem to solve (1). Recently, in (2), the authors proposed an efficient technique for scheduling a class of rural WDNs. However, the technique requires a substantial amount of data from the system to develop the schedule. In the present work, we propose a technique for judiciously choosing experiments, leading to the generation of useful data, which in turn, can be used for scheduling the system efficiently.

Scheduling WDNs is a complex problem to solve and for the same reason, a variety of formulations and algorithms have been proposed for solving these efficiently (3; 4; 5). However, only a handful of these techniques address the drinking water networks widely prevalent in developing countries. Except for a few recent works, the majority of the research addresses urban water networks (6; 7; 1). Among these, the scheduling technique proposed by Amrutur et al. (6) is computationally expensive, the method of Bonvin et al. (1) applies only to systems with continuous control valves and the improved algorithm presented by Bonvin et al. (8) (also (1; 6)) requires the full specification of the WDN in consideration. Moreover, none of these techniques address the problem of equitable distribution, a key challenge in rural drinking water networks (9). While the technique proposed by Kurian et al. (2); (henceforth referred to as $\mathcal{A}1$) presents an efficient technique for equitable distribution of water, the technique is data intensive. That is, the number of measurements required for optimal scheduling increases exponentially with the number of valves in the system. Therefore, it is desirable to have a method for optimal experiment design that would generate informative data for the scheduling problem.

The literature addressing the estimation problems in water networks primarily focus on identifying the Hazen William's coefficient, system demands or the pump characteristics for model calibration (10; 11). These works identify the model parameters that are to be adjusted, choose the data for model calibration and finally present methods for identifying the parameters. Here, the objective is to identify a system model that is predictive of all characteristics of the system (pressures, flow rates at all locations) arising at any operating condition of the system. However, in the context of scheduling problems in drinking water networks, what we are really interested in are the *flow rates received by the demand locations under the control settings that the system is expected to be operated in*. A tailor-made design of experiments could potentially help us achieve the objective with lesser data. This is the key idea distinguishing our work from existing literature on identification in WDNs.

The rest of the paper is organised as follows. We present a brief review of the scheduling technique $\mathcal{A}1$ and highlight the problem of optimal experiment design that is addressed in the present work. Following this, we present a method for solving the problem along with a brief analysis of the algorithm. The implementation is demonstrated on simulated networks and a laboratory scale test network. We conclude with the key takeaways and recommendations for future work.

2 PROBLEM STATEMENT

We address the problem of equitable distribution in a class of rural drinking water networks. *Figure 1* shows the schematic of one such network, originally presented by Bhave and Gupta (12). In these systems, water is supplied from a single source to a set of downstream consumer locations (village level storage tanks/standpipes) using pumps or gravity. As there are storage facilities available at the consumer locations, water may be provided at any time of the day to meet the daily demand. Further, it is assumed that there is a valve upstream of every consumer location which can be turned ON or OFF to regulate the supply.

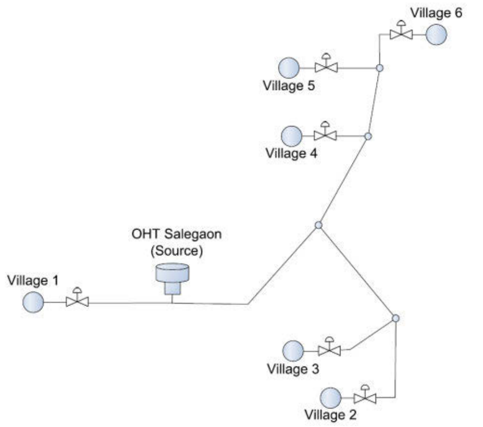


Figure 1. Schematic of the Osmanabad WDN (12)

The objective of the scheduling task is to identify the time points for operating the valves to distribute the available water equitably. For this, the objective function in the mathematical program minimises the deviation of the supplied water from the daily demand of the respective consumer location.

In algorithm $\mathcal{A}1$, the authors decoupled the system model from the optimisation problem. As the state space for the system is finite (due to the finite number of valve settings), the flow rates in the system under different valve configurations were measured *a priori*, and these measurements were used for solving the optimisation problem for identifying the schedule (*Figure 2*). This decoupling allows the transformation of the scheduling problem - which has to be solved on a daily basis - from a non-linear program (P) to a linear problem ($P1$), at the expense of a set of experiments (or model simulations) that have to be performed only on the first day. As the only measurements required are the flow rates at the final consumer locations, the technique could be implemented even without any prior knowledge of the network topology. However, the number of experiments required to explore the entire state space rises exponentially with the number of valves and this could become prohibitively high in large systems. This leads to the two key questions addressed in the present work:

1. *Can we implement the scheduling technique using the information on flow rates from only a subset of the total network configurations?*
2. *If yes, what is a good subset of network configurations in which the flow rates have to be measured?*

In the following sections, we present arguments substantiating that the scheduling technique could indeed be implemented with information from only a subset of the network configurations, and propose an algorithm for choosing a good subset of network configurations or system states.

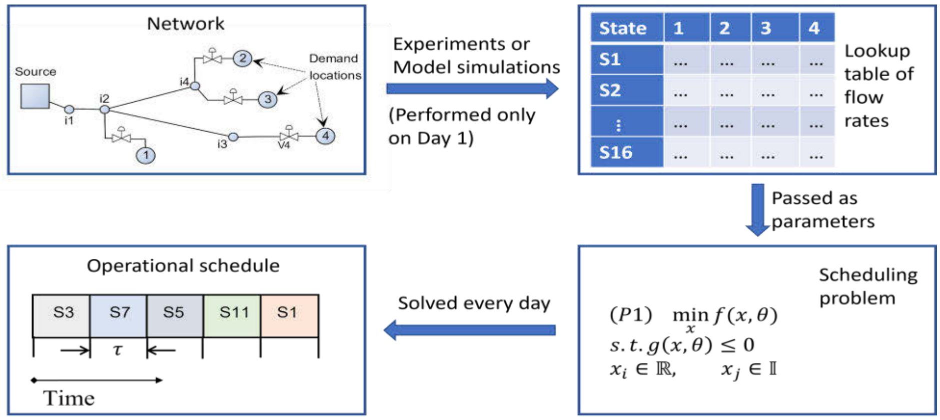

Network
Source
i1
i2
i3
i4
V4
Demand locations
1
2
3
4
Experiments or Model simulations
(Performed only on Day 1)
State
1
2
3
4
S1
S2
⋮
S16
...
Lookup table of flow rates
Passed as parameters
Scheduling problem
$(P1)\quad \min_x f(x,\theta)$
$s.t.\, g(x,\theta) \le 0$
$x_i \in \mathbb{R}, \quad x_j \in \mathbb{I}$
Solved every day
Operational schedule
S3
S7
S5
S11
S1
τ
Time

*Figure 2. Schematic describing the scheduling technique (**A1**) presented in* (2)*.*

3 A HEURISTIC FOR STATE DISCOVERY

3.1 Is a subset of network configurations enough?

In algorithm $\mathcal{A}1$, it is assumed that the lookup table (see *Figure 2*) contains information about all network states. This enables the scheduling problem to prepare an operational schedule that is optimal. In the event the table is incomplete, the program could still prepare a schedule using the network configurations for which data is available. In this case, the resultant schedule need not be optimal as the solution space is only a subset of the original feasible set. If one could ensure that the new search space still contains the optimal solution (or a near optimal solution), then the resultant schedule would still be (near) optimal. In other words, if the information available in the lookup table corresponds to network configurations that are likely to be present in the optimal schedule chosen by the original problem, any schedule prepared using this information can also be expected to be near optimal. In the following paragraphs, we propose a heuristic for identifying

BY NC SA

such *good* network configurations that could preferably be considered for inclusion in the lookup table.

3.2 Iterative scheme for discovery of states

The algorithm for discovery of states ($\mathcal{A}2$) begins with known flow rates from a few random network configurations. The number of states chosen in this manner is approximately equal to the total number of valves and pumps in the network. With these available states, the scheduling problem can be solved to determine a schedule. However, this schedule would be far from optimum. The next task would be to find a new state that can improve the schedule by the largest extent possible. The flow rates of this new state may be identified and added to the database to prepare the schedule once again. This process of scheduling and state selection may then be repeated until an acceptable schedule is obtained.

In the procedure described above, the challenge is in identifying the next state to be measured/simulated after every iteration. For improvement of the objective function, the new state that is chosen for simulation/measurement should preferentially supply water to the demand points that face the maximum deficiency in supply (negative deviation of supply from demand). This can be achieved by a state that provides water *only* to the demand points with insufficient supply. That is, in the current schedule, if two demand points receive insufficient supply, the newly chosen state should provide water to these two demand points alone. The reason behind the argument is that, in a network configuration supplying *only* to a few demand nodes, these nodes tend to receive a high flow rate, as the others are turned off. This new state can now be simulated/measured, and the flow rates added to the database. Thereafter, the procedure has to be repeated until the resultant schedule is acceptable.

The selection procedure described above could potentially lead to infinite cycles, e.g., the algorithm predicts a new state that has already been selected in a previous iteration. To avoid this, the selection process is made probabilistic. That is, the set of chosen nodes may not be the entire set of nodes receiving insufficient supply in the previous iteration. It may even be a subset of it with certain probability. To implement this, in each iteration, the demand points with insufficient supply have to be selected, their shortfall normalised and then sorted in descending order of the magnitude of the deficit in supply. Thereafter, we pick only a subset of these demand points for supply in the next state. The demand nodes on the top of the list have the maximum deficit in supply and need to be chosen for supply with a high probability while those lower down the list can be chosen with a lower probability, as they are in a much better position with respect to demand satisfaction. We start selecting from the top and go down the list until a certain 'stopping node' is reached. This 'stopping node' has to be decided probabilistically. For this, in each iteration of the heuristic, we generate a random number r between 0 and 1. Demand points are now chosen from the beginning of the sorted list until the sum of their normalised deficit in supply reaches r. If r is close to zero, only the first demand point, i.e., the one with the maximum deficit would be chosen. If r is one, all demand points with shortfall in supply would be chosen. For intermediate values of r, a subset of the nodes with inadequate supply would be chosen. It has to be noted that, for any chosen demand node, any other node with a higher deficit in supply would also be there in the set of chosen nodes. These chosen demand points would receive supply in the new state that has to be measured. The stochastic or probabilistic feature of the algorithm ensures that the algorithm does not get caught in infinite cycles.

The heuristic is further improved by ensuring that a completely new state is chosen randomly, albeit with a small probability. This is included because the rationale used here for choosing new states is ideal only for an unconstrained scheduling problem. If there are constraints in implementation (time or resources), one cannot expect the state selection procedure to be perfect. In the cases we tested, in each iteration, we chose the next state randomly with a probability of 0.05 and with the remaining probability, chose the next state based on the shortfall in supply. In the flow chart of the heuristic given in *Figure 3*, the block verifying the condition - $p >= 0.05$? - decides whether a state is chosen completely randomly or not. If FALSE (Probability = 0.05), a random state is chosen from the set of unexplored states. Otherwise (TRUE, Probability = 0.95), a state is chosen following the steps described in earlier paragraphs. This procedure is similar to the *ε-greedy exploration* described in (13)

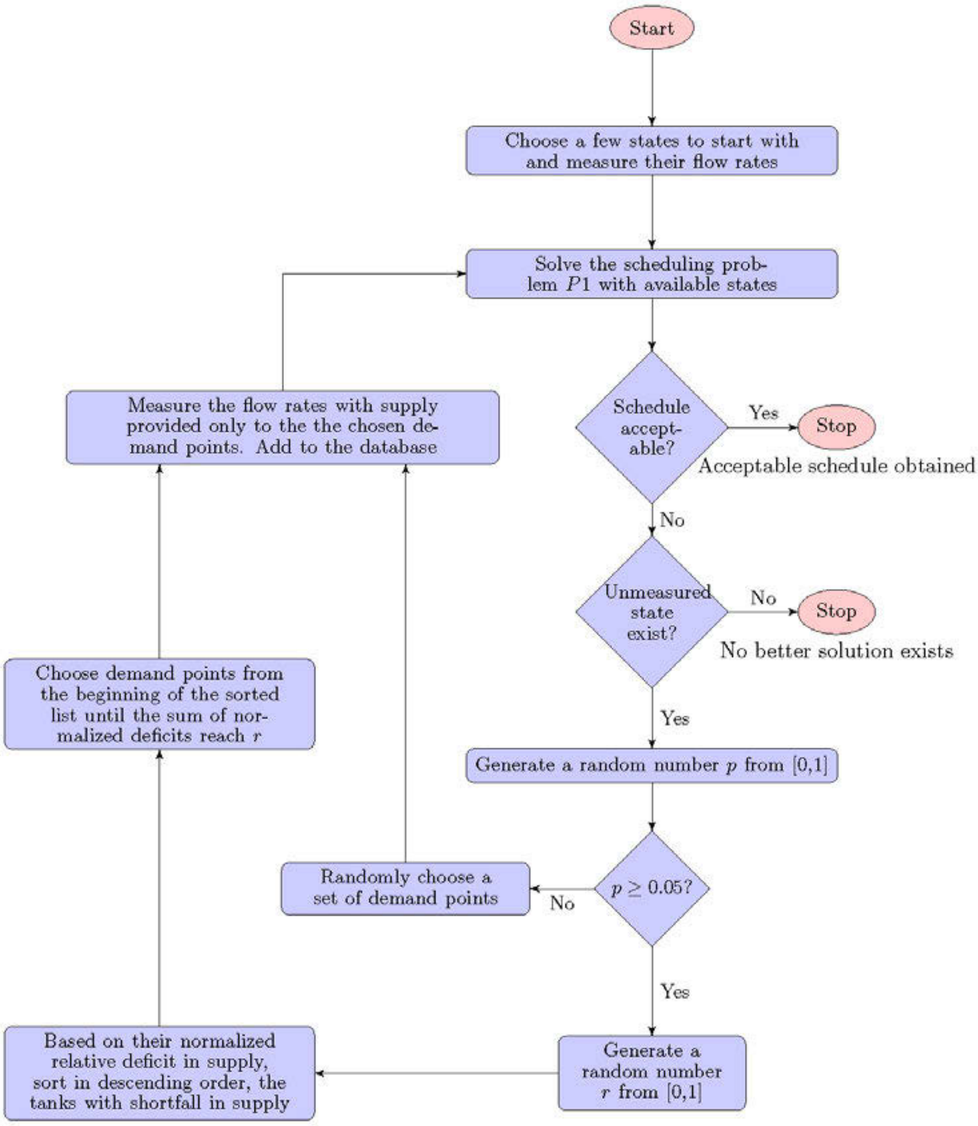

Start
Choose a few states to start with and measure their flow rates
Solve the scheduling problem P1 with available states
Schedule acceptable?
Yes
Stop
Acceptable schedule obtained
No
Unmeasured state exist?
No
Stop
No better solution exists
Yes
Generate a random number p from [0,1]
p ≥ 0.05?
No
Randomly choose a set of demand points
Yes
Generate a random number r from [0,1]
Based on their normalized relative deficit in supply, sort in descending order, the tanks with shortfall in supply
Choose demand points from the beginning of the sorted list until the sum of normalized deficits reach r
Measure the flow rates with supply provided only to the the chosen demand points. Add to the database

Figure 3 Flowchart describing the algorithm for discovery of states ($\mathcal{A}2$)

In the following sections, we show the implementation of the algorithm in both in-silico and real (laboratory scale) networks.

4 CASE STUDIES – MODEL SIMULATIONS

We tested the heuristic on the mathematical model of the Osmanabad network shown in *Figure 1*. The objective of the scheduling problem was to meet the demand given in *Table 1*. (Please see (2) for more details of the scheduling problem). To start with, it was required to have a set of states for which flow rates were known. Six randomly chosen states, S1 (all valves off), S2, S3, S26, S44 and S64 were used. The states and the flow rates corresponding to each of these are given in *Figure*

4. Thereafter, the heuristic was performed and after each iteration, the flow rates of one new state were identified. This was then added to the database and the procedure was repeated.

Table 1. Eight-hour demand of the villages in Osmanabad

Village	Demand (m^3)
1	81.01
2	30.00
3	30.00
4	52.80
5	75.00
6	50.00

In the first iteration, on preparing the schedule with the six states mentioned above, the sum squared relative deviation (value of the objective function in the scheduling problem) was 2.94. Villages 3 and 5 reported the highest relative deviation of 100% and the deviations corresponding to remaining villages are available in *Table 2*. A random number p was generated to decide whether to choose a state completely random or not. p was greater than 0.05 and hence the next state had to be chosen based on the relative deviations in supply. The random number generated for state selection was 0.8147 (r = 0.8147). As per the procedure described earlier, states had to be selected from the top of the list until the sum of normalised relative deviation reached r. The normalised relative deviation for the first two villages (3 and 5) added up to 0.662 ($< r$) and the first three villages (3, 5 and 6) added up to 0.978 ($\geq r$). Hence, the three villages at the top had to be supplied water in the new state to be explored. This corresponds to state 53 as shown in *Figure 4*. Following this, on preparing a schedule with the seven available states (six initial and the newly explored state S53), the objective function value dropped to 0.45 (from 2.94 in the previous iteration).

Table 2. Villages and their relative deviation on preparing a schedule with the initial six states (sorted on relative deviation)

Village	Relative deviation between supply and demand	Normalised shortfall in supply
1	-1	0.331
2	-1	0.331
3	-0.955	0.316
4	-0065	0.022
5	0.039	NA
6	0.131	NA

Repeating the procedure for five more iterations, states 51, 59, 33, 49 and 8 were explored and the objective function in the final iteration came down to an acceptable level of 0.016. The largest relative deviation for any village had also reduced to 0.09, this time for Village 2. All states chosen by the heuristic for preparing the schedule are shown in *Figure 4* and the variation in objective function value of the optimisation problem with the progression of the heuristic is given in *Figure 5*. It has to be noted that, out of the total 63 non-trivial states, the heuristic required measurements of only 11 states to prepare a reasonable schedule for supply.

	S2	S3	S8	S26	S33	S44	S49	S51	S53	S59	S64
1	18.36		18.36	18.36		18.36					18.36
2		17.23	17.23			15.59		16.60		15.59	15.59
3									11.30		0.00
4				51.58		40.20				47.08	40.20
5							18.81	16.15	17.11	0.00	0.00
6					8.91	5.32	7.32	7.03	7.13	5.54	5.32

Figure 4. The states of the Osmanabad network identified using the heuristic. In the table, each column corresponds to a state and each row corresponds to a village. A colored cell indicate that the valve leading to the village is open in the corresponding state. The numbers in the cell give the flow rate received by the village in m^3 / h.

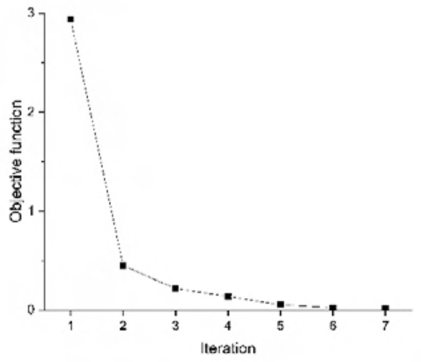


Figure 5. The change in the scheduling objective function (deviation between demand and supply) with the progression of the heuristic.

5 CASE STUDY – EXPERIMENTAL SYSTEM

In this section, we present the results obtained while implementing the method on a laboratory scale WDN.

5.1 Experimental Network:

The network comprises one Over the Head Tank (OHT) supplying water to nine Small Tanks (STs) located downstream. A schematic of the network is shown in *Figure 6* with the STs numbered T1 - T9. The network is an extended version of the setup described in [14] with nine STs instead of five. The pipe and tank dimensions remain unchanged. The control valve installed downstream of OHT was kept open throughout this study. The STs have solenoid valves (Burkert 6011), placed at the inlet and the outlet. The valves could switch ON and OFF the inflow and the outflow(drain) for each ST. Ultrasonic level transmitters (Baumer U500) are installed on top of all tanks to monitor the level. All transmitters and actuators are interfaced to a computer using DAQ card form National Instruments and they are programmed and controlled using LabVIEW. The OHT represent the source and STs represent the storage available at beneficiary villages. The objective

BY NC SA

is to equitably distribute water by scheduling the solenoid valves which regulates supply to STs in the setup. Two different scenarios are tested to verify the utility of the heuristic for state discovery.

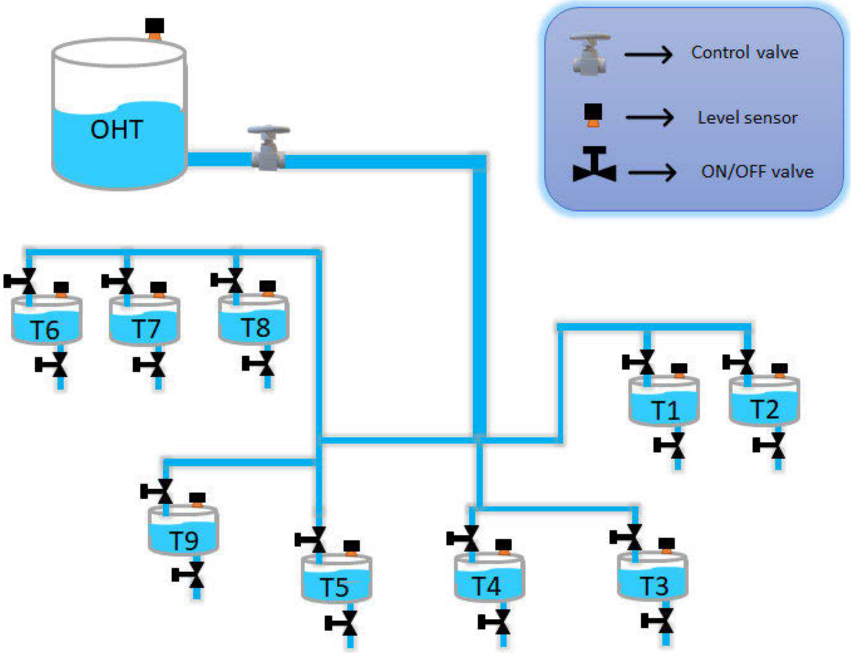


Figure 6. The Experimental Network

5.2 Data Acquisition:

The level of water in the OHT was maintained between 45 cm and 40 cm with a control loop in LabVIEW. The experiments were carried out to identify the inflow rates into STs under different valve combinations. The total number of valve combinations for the system with nine STs is 512(i.e., 2^9). Since it is difficult to measure flow rates for all 512 combinations., we use the heuristic for discovering useful states. We begin with flow rates of few states chosen random and then flow rates of the network configurations chosen by the heuristic was measured and added to database iteratively.

The Ultrasonic level sensor placed at top of the tank gives 4-20 mA analog output. This analog output is sampled at 5Hz frequency and then its filtered and converted to corresponding to level value in LabVIEW. The system was kept idle for 15s before and after the valve activation for every combination, for the level readings to stabilise, following which, level measurements were recorded. Each configuration is kept active for 60s and the difference in the level in each ST was used to calculate the inflow rates into respective ST. The solenoid valve at outflow of STs is kept closed while recording the data.

5.3 Scenario-1

A schedule has to be prepared to meet the demand of 7 litres for each ST. The total time available for water supply is 24 minutes which is discretised into 24 equal intervals of 60 s. The following restrictions are also imposed.

- No of time intervals allowed for the system is 24.
- No of switching allowed for each valve is 4.

We formulated scheduling problems and discovered operational states for the system following the heuristic described in Section 3. To start with, a set of 9 random states were used. Then, the

cc BY NC SA

heuristic ($\mathcal{A}2$) was carried out to discover new states and after each iteration within the heuristic, flow rates of one new state were identified.

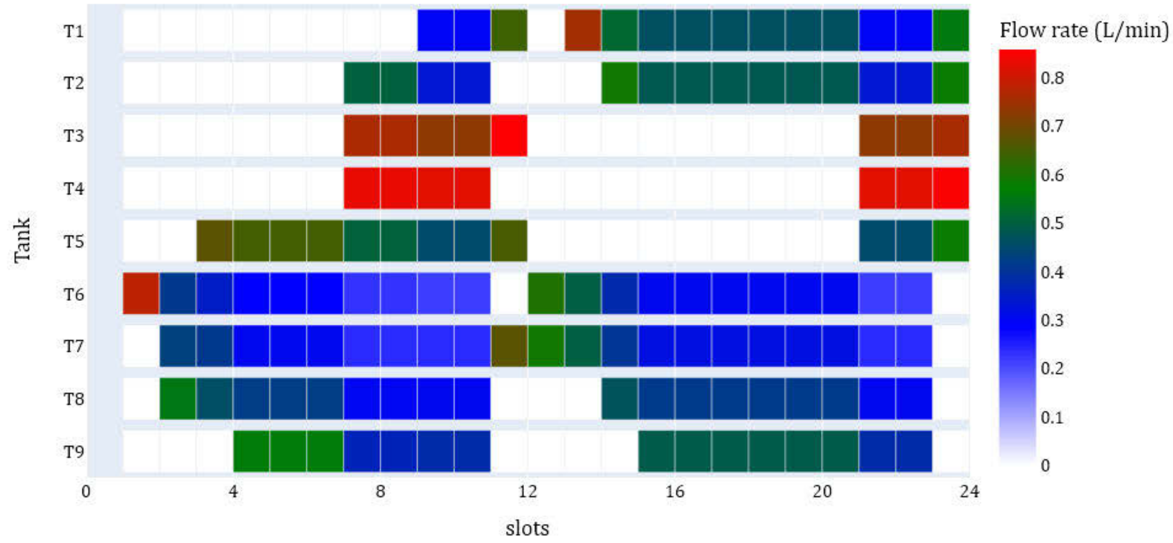


Figure 7. The Final schedule for scenario-1 identified using the heuristic

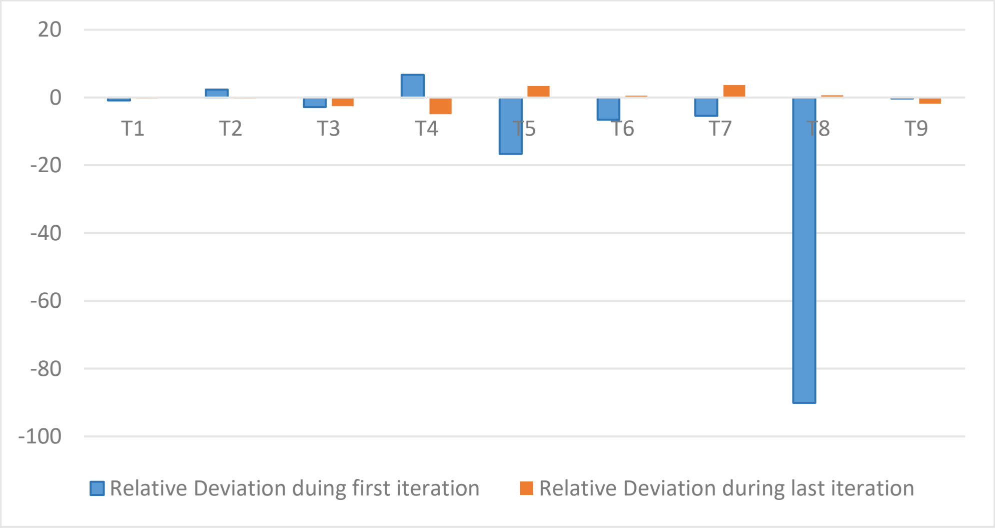


Figure 8. Relative Deviation for scenario-1 in percentage

This was added to the database and the procedure was repeated. The iteration is continued till the maximum relative deviation for every ST was less than 10%. After 7 iterations, a schedule with a maximum relative deviation of 3.6% for tank T7 was obtained. The final schedule and the supply obtained by different STs is shown in the *Figure 7*. In *Figure 7*, each row represents a ST and each column represent a state (network configuration). The coloured cell represents the flow rate leading to that ST is ON and the colour range indicates the flow rate range. The relative deviation in the final schedule is shown in the *Figure 8*.

5.4 Scenario-2

In the second scenario considered here, we added a more stringent condition on the obtained schedule - the total number of switches allowed for each valve was reduced to 2. The scheduler ran initially with 9 different random states and then new states were added with new iterations with the help of the heuristic ($\mathcal{A}2$). The final schedule was obtained after 16 iterations with a maximum relative deviation of 7.4% for T5. The final schedule is shown in the figure 11. On comparing *Figure 7*and *Figure 9*, it is evident that the number of valve operations did decrease substantially in the latter case.

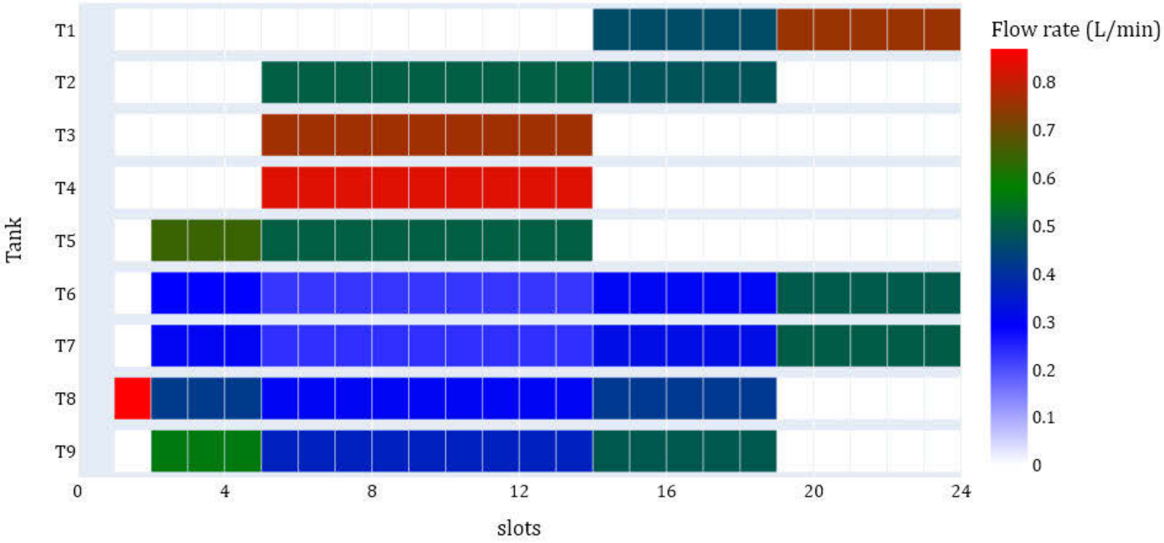


Figure 9. Final schedule for scenario-2 identified using the heuristic.

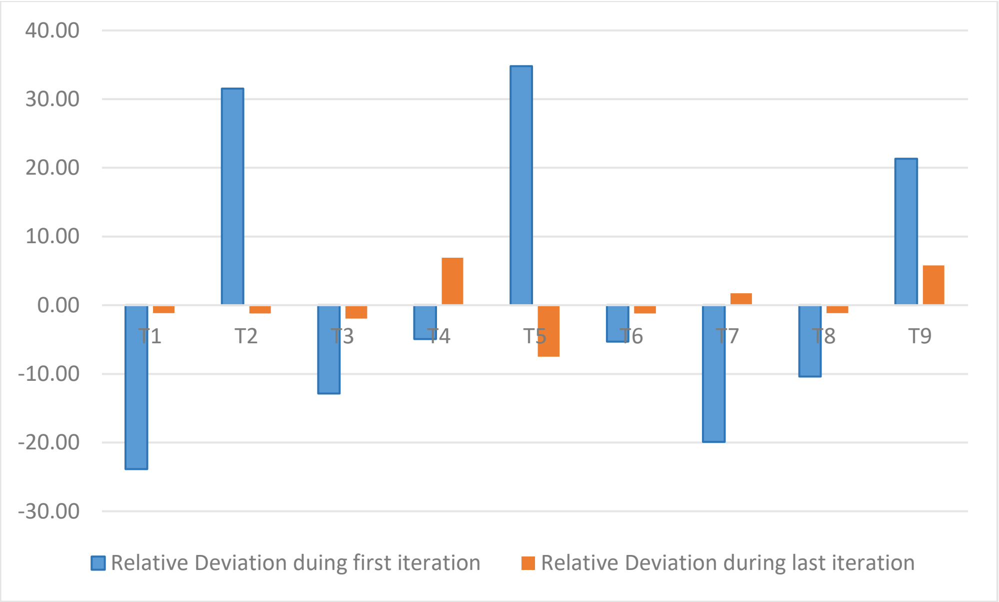


Figure 10. Relative Deviation for scenario-2 in percentage

Out of the total 512 non-trivial states, the heuristic required measurements of only 25 states to prepare a near optimal schedule for supply to STs. It may also be noted that the number of states required in the scenario with more stringent constraints (Scenario 2) was more than that of the case with less stringent constraints (Scenario 1). Through these experiments, we were able to establish that near optimal scheduling with less relative deviation can be obtained with a selective subset of network configurations identified with the help of the heuristic ($\mathcal{A}2$).

6 CONCLUSIONS

In this work, we proposed a heuristic for design of experiments to aid the scheduling in rural WDNs. To the best of our knowledge, this is the first attempt to develop a methodology for design of experiments in WDNs specifically for scheduling. The approach uses cues from a schedule prepared using the available measurements to identify the new network configuration that has to be explored. We demonstrated the applicability of the technique through model simulations and experiments performed on a lab scale WDN. The results show a clear advantage for identifying the states using the heuristic.

The approach presented here assumes that the states are identified prior to the network being commissioned for normal operation. That is, the experimentation does not affect the regular supply of water. The now popular reinforcement learning approaches may help exploring the network states while the network is operational and may also allow correcting the existing measurements for scaling in the pipes. We are also exploring strategies for identifying states that are robust to changes in demands.

7 ACKNOWLEDGEMENTS

This work was partially funded by the Department of Science and Technology, Govt. of India under the Water Technology Initiative (No: DST/TM/WTI/WIC/2K17/82(G) WATER-IC for SUTRAM of EASY WATER) Madras and by the Robert Bosch Center for Data Science and Artificial Intelligence, IIT Madras.

8 REFERENCES

[1] Bonvin, G., S. Demassey, C. Le Pape, N. Maïzi, V. Mazauric, and A. Samperio, "A convex mathematical program for pump scheduling in a class of branched water networks," Applied Energy, 2018, pp. 1702–1711.

[2] Kurian, V., P. Mohandoss, S. Chandrakesa, S. Chinnusamy, S. Narasimhan and S. Narasimhan, "Optimal scheduling of water distribution networks for equitable supply," In prep.

[3] Burgschweiger, J., B. Gnadig, and M. C. Steinbach, "Optimization models for operative planning in drinking water networks," Optimisation and Engineering, vol. 10, no. 1,2009, pp. 43-73.

[4] D'Ambrosio, C., A. Lodi, S. Wiese, and C. Bragalli, "Mathematical programming techniques in water network optimisation," European Journal of Operational Research, 2010, pp. 10-16.

[5] Mala-Jetmarova, H., N. Sultanova, and D. Savic, "Lost in optimisation of waterdistribution systems? a literature review of system operation," Environmental Modelling & Software, 2017, pp. 209-254.

[6] Amrutur, B., M. M. Kumar, K. S. Kumar, L. Patel, R. Sundaresan, and N. K. Vaidhiyan, Wateropt, "a method for checking near-feasibility of continuous water supply," In Proceedings of the 2nd International Workshop on Cyber-Physical Systems for Smart Water Networks. IEEE, 2016.

[7] Kurian, Varghese, Saravanan Chinnusamy, Ashok Natarajan, Sridharakumar Narasimhan, and Shankar Narasimhan. "Optimal operation of water distribution networks with intermediate storage facilities," Computers & Chemical Engineering,2018, pp. 215-227.

[8] Bonvin, Gratien, Sophie Demassey, and Andrea Lodi. "Pump scheduling in drinking water distribution networks with an LP/NLP-based branch and bound." Optimisation and Engineering vol.22, no. 3,2021, pp. 1275-1313.

[9] Prasad, Pooja and Sohoni, Milind. Sugave Water Scheme, Multi-village drinking water scheme analysis. 2011.

[10] Dini, Mehdi, and Massoud Tabesh. "A new method for simultaneous calibration of demand pattern and Hazen-William's coefficients in water distribution systems," Water Resources Management, vol. 28, no. 7, 2014, pp. 2021-2034.

[11] Ostfeld, Avi, Elad Salomons, Lindell Ormsbee, James G. Uber, Christopher M. Bros, Paul Kalungi, Richard Burd et al. "Battle of the water calibration networks," Journal of Water Resources Planning and Management, vol.138, no. 5, 2012, pp. 523-532.

[12] Bhave, Pramod R. and Gupta, Rajesh. Analysis of water distribution networks, 1st ed., Alpha Science Int'l Ltd,2006.

[13] Russo, Daniel J., Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen, "A tutorial on thompson sampling," Foundations and Trends® in Machine Learning, vol. 11, no.1, 2018, pp. 1-96

[14] Chinnusamy, Saravanan, Prasanna Mohandoss, Varghese Kurian, Sridharakumar Narasimhan, and Shankar Narasimhan, "Operation of intermittent water distribution systems: An experimental study," In Computer Aided Chemical Engineering, vol. 44, 2018, pp. 1975-1980.

BY NC SA

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14138

URBANLEMMA: A SERIOUS GAME TO SUPPORT THE ADOPTION OF SUSTAINABLE URBAN DRAINAGE SOLUTIONS

Aashna Mittal[1], Jessica Nguyen[2], Lisa Scholten[3], Zoran Kapelan[4]

[1,2,4] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, Delft 2628CN, The Netherlands
[3]Faculty of Technology, Policy, and Management, Delft University of Technology, Building 31, Jaffalaan 5, Delft 2628 BX, The Netherlands

[1] *a.mittal@tudelft.nl, [2] J.Nguyen@student.tudelft.nl, [3] l.scholten@tudelft.nl,[4] z.kapelan@tudelft.nl.*

Abstract

Long-term planning of urban drainage systems is required to prevent pluvial flooding and the effects of droughts. Sustainable urban drainage systems (SUDS), e.g., permeable pavements, detention ponds, and rainwater harvesting systems are increasingly used to manage water and build resilience in urban environments. However, the adoption of these solutions remains low due to various socio-institutional barriers. Improving awareness about SUDS and highlighting their multifunctional benefits and trade-offs through an engaging medium could help address these barriers. Serious games provide an immersive and engaging experience that can be used to motivate and impart knowledge or train skills to improve decision-making. The potential of serious games to support the adoption of SUDS has not been investigated so far. This paper presents the proof-of-concept of the serious game UrbanLemma, designed to improve awareness and decision-making of stakeholders about SUDS along with the additional goals of increased engagement and communication among stakeholders. Initial results from play-test sessions conducted with researchers at TU Delft are presented and lessons learned from the game development process are discussed.

Keywords
SUDS, Pluvial flooding, Urban water management, Serious games, Adoption, Decision-making.

1 INTRODUCTION

Urbanization has become a global phenomenon and it is expected that 60% of all population will be living in urban areas by the year 2050 [1]. A consequence of increasing urban densities is that cities across the world are facing stormwater management challenges. Pluvial flooding occurs when the underground sewage networks cannot drain the stormwater. Contrary to the assumption that pluvial flooding is a minor 'nuisance', it can lead to direct loss of life, contamination of surface bodies, disruption of transportation networks, and cause damage to property and critical infrastructure [2]. Furthermore, the drivers of climate change and increased urbanization are expected to increase the risk of flooding in the coming years [3], [4]. Extreme rainfall events may occur more frequently leading to more stormwater being processed by the urban water system and higher peak discharges due to decreased perviousness of the ground.

Piped drainage systems have been the traditional response to stormwater management. But this solution can no longer deal with the impacts of climate change and increasing urbanization [5], [6]. Given this situation, sustainable urban drainage systems (SUDS) have become a widely utilized solution to deal with the risk of urban flooding. The SUDS approach is to reduce the quantity of runoff water that enters the underground drainage system by harvesting, infiltrating, slowing, storing, conveying, or treating the runoff on-site [7]. By doing so, the downstream flood risk is managed and the risk of the runoff causing pollution of the receiving water body is also reduced. Typical examples of SUDS include green roofs, rainwater harvesting systems, permeable

BY NC SA

pavements, swales, bio-retention systems, pervious pavements, and wetlands among many others.

Although SUDS are widely considered as a core stormwater management strategy, they have not yet received widespread uptake and their adoption remains patchy [8], [9]. Several studies exploring the barriers behind the low adoption of SUDS highlight that the barriers are not technical but rather socio-institutional in nature [5], [9]–[11]. Key barriers to the adoption of SUDS include lack of awareness/knowledge, and lack of cooperation and coordination among stakeholders involved in decision-making [12]–[15]. To overcome these barriers, suggested strategies include engaging stakeholders and raising awareness around SUDS, and promoting their co-benefits [15], [16].

Serious games are a promising solution to overcome the above barriers. These are games that not only entertain but are developed to achieve a more serious purpose such as education, skills training, understanding people's behaviours, or inducing a behaviour change [17]. Serious games have been widely used to support many urban water management problems such as water distribution, flood prevention, and urban infrastructure management [18]–[20] but their application to improve awareness about SUDS has not been attempted so far. To do so, we present a proof-of-concept of a serious game called Urbanlemma in this paper.

2 BACKGROUND

2.1 Barriers to adoption of SUDS

SUDS are generally accepted as a more sustainable approach to managing stormwater. Although they cannot replace the underground sewage networks entirely, they have great potential to reduce peak runoffs and provide additional water storage that can provide benefits such as biodiversity, heat stress reduction, groundwater recharge, recreation, and improvement in aesthetics [21]. Even though the benefits of SUDS are abundant and their technical performance is established, their adoption remains low and patchy.

Many studies that have explored the barriers to adoption of SUDS establish that the adoption is low not due to technical but socio-institutional barriers. These studies recognize that lack of knowledge and awareness about SUDS and their functions is one of the key barriers to their adoption [9], [11], [13]–[15], [22]. For instance, in a review of 85 studies reporting knowledge, attitude, intentions, and behaviours towards SUDS, general awareness about different types of SUDS, their ecosystem services, and stormwater best management practices was found to be low [11]. Low public awareness of urban drainage issues can impede public participation in local water management issues and create resistance to the adoption of SUDS [22]. A study conducted on residents in Amsterdam, Netherlands further shows that although citizens are aware of the problem of urban flooding, it is not enough for residents to adopt SUDS [23]. In addition to educational barriers, another category of barriers that are frequently reported relates to stakeholders involved in the decision-making around stormwater management. Lack of communication/interaction and collaboration among stakeholders is frequently reported to impede the uptake of SUDS [9], [10], [24]. Therefore, more awareness of the solutions to overcome flooding is needed along with recognition of responsibilities and interdependencies between stakeholders.

2.2 Serious games

Urban water management presents a complex decision-making environment involving multiple stakeholders, uncertainties, and the complexity of the biophysical environment. Serious gaming (or simulation gaming) is a method that allows capturing this complexity by simplifying the real world [25]. Serious games are those in which the primary purpose is not to entertain the target audience but to attain a more serious purpose such as education or learning, skills training, or changing behaviours. In this paper, the game presented is an educational game designed to inform players about the consequences of drainage issues (traditional and SUDS), while providing a platform for actors to cooperate and understand their role and responsibilities in stormwater management.

3 DESIGN METHODOLOGY

The overall design process can be simplified into 5 distinct phases [26]–[28]. In this paper, we present the first 4 of the 5 design phases. Phase 1: *Design Specifications* starts by listing and understanding the requirements and objectives of the game. For research purposes, we executed this phase by looking at the literature on the adoption of SUDS, and shortlisting the barriers that the serious game should focus on. These were then translated into the game objectives, i.e., the game must increase awareness of different types of SUDS, and their performance and effects.

Phase 2: *Systems Analysis* involves analysis and representation of the part of the real world that is relevant to the problem being focused on in Phase 1. In this phase, important elements to be highlighted in the game, e.g. concepts, alternatives, relations, stakeholders, information, theoretical artefacts, responsibilities, etc. are captured in the form of a conceptual diagram. Information about these parameters was gathered through desk research with a focus on stormwater management in the Netherlands. The elements from systems analysis were then converted into game elements in Phase 3: *Detailed game design*. A matrix was created that mapped each system analysis element such as actors, uncertainties, SUDS, etc. into corresponding game elements such as player roles, event cards, and actions/choices in the game [29]. Furthermore, in this phase, the game concept and elements were worked out in detail on paper. Lastly, in Phase 4: *Game construction, validation, and testing*, the game was designed and made into a tangible product. This included constructing/arranging the game paraphernalia such as the game board, event cards, role cards, player pawns, SUDS, associated impact cards, etc. Once the game was translated into a playable version, a test session was conducted to debug the game and validate whether the intended message of the game is received by the players, and obtain initial feedback on aspects that can be improved in the game design.

4 GAME DESIGN

4.1 Game objective and target audience

The objective of the serious game is to raise awareness of the consequences of SUDS. It is important to acknowledge that decision-making around the adoption of SUDS involves multiple stakeholders – citizens, municipalities, commercial landowners, and housing developers who have the power to implement SUDS on public/private land.

4.2 Conceptual system map

The real-world problem to be tackled by the serious game revolves around increasing awareness about SUDS among stakeholders involved in making decisions about the adoption of SUDS. To analyse the real-world system relevant to the problem, we focused on 4 key components: a typical Dutch neighbourhood and built environment, the stakeholders who take decisions on adopting SUDS, the objectives/criteria they care about, and the external factors (or unknowns) under which they operate. A conceptual map was prepared. It includes the following system components.

1. *Built environment*: The built environment in a typical Dutch city can be divided into public or private land depending on ownership by an individual/company vs a government organization. Private land mostly comprises houses or apartments and industrial buildings, while public land consists of parks, pavements, roads, and other parcels of land that have not been built upon.

2. *Stakeholders*: We consider stakeholders who have the power to implement SUDS on public or private infrastructure as part of the system under consideration. These are municipalities, housing developers, homeowners, and commercial landowners. Many other actors are also involved in stormwater management but we consider them out of the system boundary as they only indirectly impact the decision to adopt SUDS.
3. *Criteria*: These are the objectives stakeholders in the system care about. We consider objectives spanning social, economic, and environmental goals and resilience to urban flooding. These categories are further divided into more specific criteria.
4. *External factors*: These are factors that are outside the control of the actors within the system, e.g., climate change, population growth, and regional hydrology. However, some of the external factors might be under the control of actors outside the system of interest. For instance, the local water boards or the national government may push for more housing development initiatives or change regulations around the acceptable return period of flooding.

4.3 Game set-up and mechanics

The game revolves around a hypothetical town containing a mix of impervious areas, e.g., houses and apartment buildings, schools, shops, roads, and pervious areas such as gardens, parks, undeveloped land pockets, and a river. The town is struggling to develop in a way that balances protection against flooding, environmental quality, and social standards while staying within available financial budgets. In the game, players adopt one of the roles – home-owner with garden, housing developer, commercial land-owner, and municipality. Each role comes with access to different developments. They aim to win against the game by cooperatively developing the town while achieving flood resilience, environmental quality, and high social standards with available budgets. On the way, they have to deal with uncertainties such as climate change, changing budgets, acceptable return periods, or increasing flood risk, among others.

A set of paraphernalia is used to visually represent the game elements and facilitate the dynamics of the game:

- Game board (see Figure 1): The game board spatially represents the neighbourhood in which the game is set. It also visually represents the game scores against the Environmental quality and Social standards criteria, and the sewer and flood capacity.
- Role cards (see Figure 2a): The 4 playable roles of the game have a role card that describes the role, what developments they have access to, any special actions, and their maintenance and income.
- Development cards (see figure 2b): Each development option has an associated development card that describes the development and its impacts.
- Event Cards: External influences on the system are represented through event cards to be revealed throughout the game.
- Flood damage cards: Cards explaining monetary costs associated with a level of flood damage.
- Weather forecast card: Card outlining the number of dice to be rolled per round.
- Development blocks: The options for developments to place on the board are represented by small blocks that fit onto the squares of the board.
- Water cubes: Volume of rainwater represented by water cubes.
- Protest tokens: physical representations of the limited protests a role can spend.
- Score maker: A transparent block to be placed on the score scale bar to represent the current score.
- Play money: Money to represent the economics of the system.

BY NC SA

- Dice: To roll for a weather event.
- Erasable pen: To update maintenance costs.

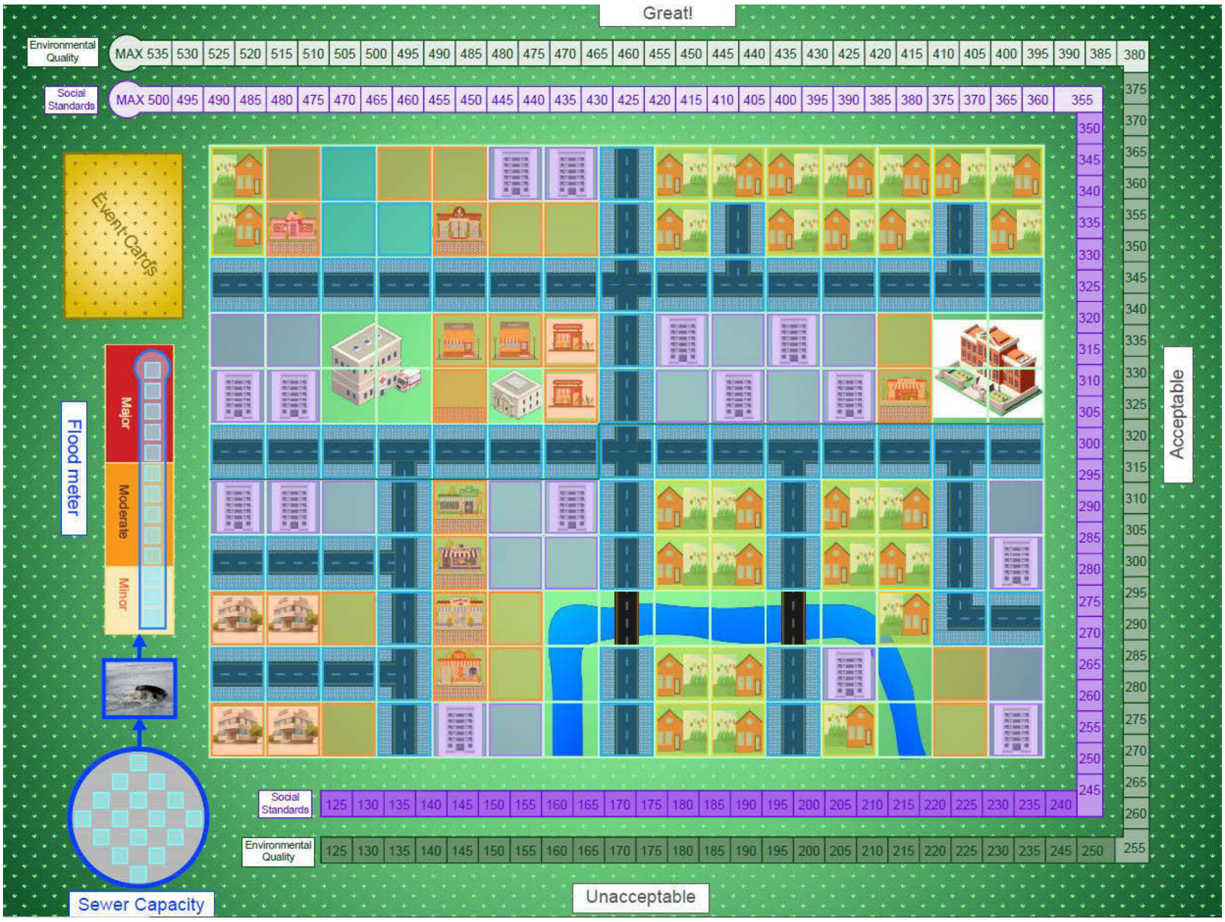
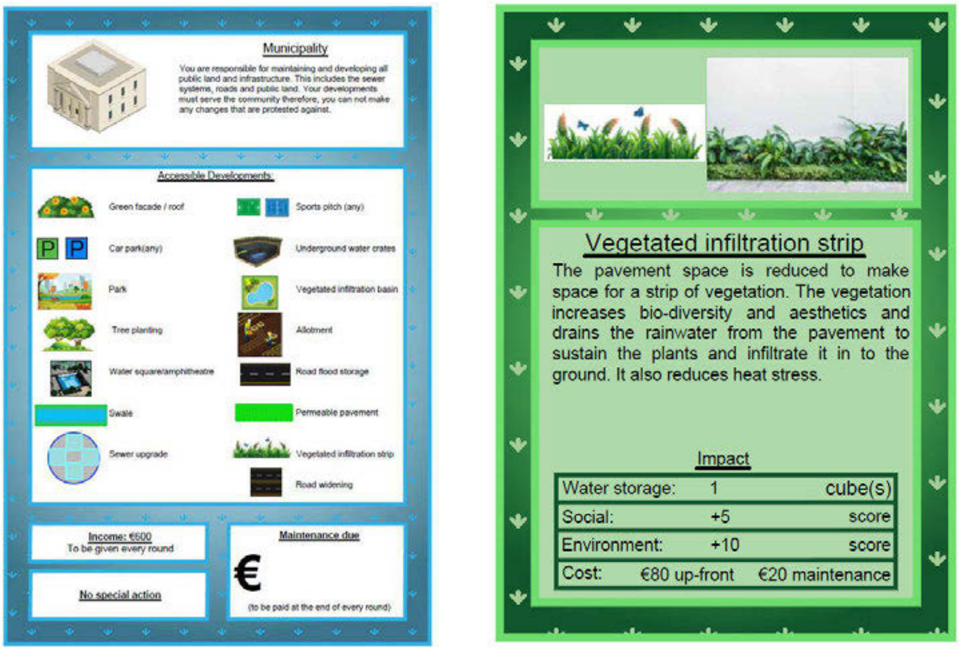

Great!
Environmental Quality
MAX 535 530 525 520 515 510 505 500 495 490 485 480 475 470 465 460 455 450 445 440 435 430 425 420 415 410 405 400 395 390 385 380
Social Standards
MAX 500 495 490 485 480 475 470 465 460 455 450 445 440 435 430 425 420 415 410 405 400 395 390 385 380 375 370 365 360 355
Event Cards
Flood meter
Major
Moderate
Minor
Acceptable
Social Standards
125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240
Environmental Quality
125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 245 250
Sewer Capacity
Unacceptable

Figure 1. Game board of UrbanLemma



Municipality

You are responsible for maintaining and developing all public land and infrastructure. This includes the sewer systems, roads and public land. Your developments must serve the community therefore, you can not make any changes that are protested against.

Accessible Developments:

Green facade / roof	Sports pitch (any)
Car park(any)	Underground water crates
Park	Vegetated infiltration basin
Tree planting	Allotment
Water square/amphitheatre	Road flood storage
Swale	Permeable pavement
Sewer upgrade	Vegetated infiltration strip
	Road widening

Income: €600
To be given every round

No special action

Maintenance due
€
(to be paid at the end of every round)

Vegetated infiltration strip

The pavement space is reduced to make space for a strip of vegetation. The vegetation increases bio-diversity and aesthetics and drains the rainwater from the pavement to sustain the plants and infiltrate it in to the ground. It also reduces heat stress.

Impact

Water storage:	1	cube(s)
Social:	+5	score
Environment:	+10	score
Cost:	€80 up-front	€20 maintenance

Figure 2a (left): Example of a role card, 2b (right): Example of a development card

BY NC SA

The game starts with a short briefing, followed by 10 game rounds, followed by a debriefing session. At the start, a facilitator explains the problem, storyline, objective, and rules of the game. The rules that players need to follow deal with the placement of infrastructure developments and water cubes on the board and managing money-related matters. A round consists of each player taking their turn during which they can take two out of four actions: (1) Remove a block, (2) Add a block, (3) Evoke special action, or (4) Pass. At the end of the round, players perform the following actions:

1. Drain the board: empty sewer, empty flood meter, remove 1 water cube from each SUD
2. Pay maintenance and receive income
3. Pick up an event card and follow the instructions
4. Roll dice for the weather forecast and place the required number of water cubes
5. Check for flooding and pay damages if necessary

The players play the game until they either win or lose. Players win the game if they achieve and maintain the maximum score on environmental and social standards for 2 rounds without any flooding or debt. They lose after 10 rounds, if they have 2 major floods in a row or if the players cannot collectively pay the flood damages. The game is concluded by a short survey and a discussion about aspects of the game that can be improved.

5 GAME VALIDATION

To test if the game designed in the previous section has the intended impact, a validation session was conducted in May 2022. This was done to get an overall sense of how the game is perceived by the players, whether the game mechanisms work as desired, and to debug potential errors. A gameplay session was organized with four Ph.D. researchers in the field of water management (see Figure 3). The session started with a brief explanation of the game setting and rules, followed by 9 rounds of gameplay, and a post-game survey followed by a short discussion on the players' game experience. The survey assessed 7 criteria scored on a 10-point Likert scale: realism, learning, fun, interaction among players, feeling of engagement throughout the game, cooperation among players, and clarity of rules. The session lasted for 2 hours. The game was facilitated by the second author while the first author observed the gameplay and took note of aspects for further improvement.

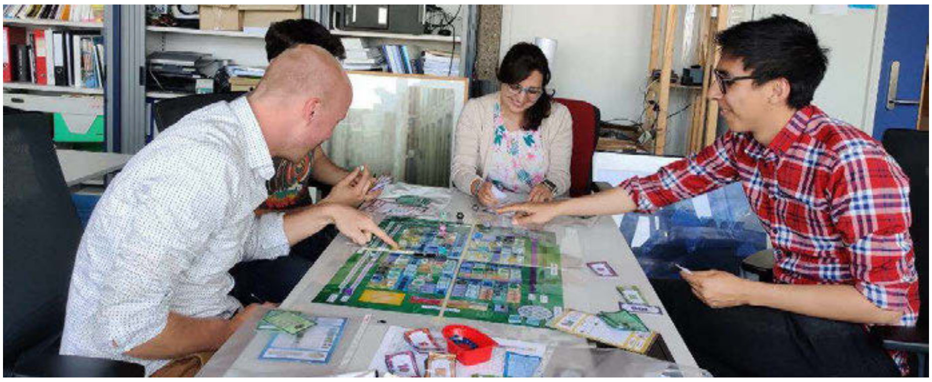


Figure 3: Validation session in progress with researchers at TU Delft

Overall, players found the game to be fun and engaging, and game rules were perceived as clear, reflected in scores > 8 (out of 10). Two out of four players liked the cooperative aspect of the gameplay, while others appreciated the overall design and realism of the game. On the other hand, aspects of realism, learning, and interaction among players could be improved. Concrete suggestions were mentioned, such as giving players more useful and realistic powers to influence the conditions of another role, or adding more status-quo developments to emphasize conflict of choice and trade-offs concerning flood prevention, cost, social and environmental quality impacts.

BY NC SA

Assessing the learning outcomes, players noted that they learned about the need for prevention and SUDS' ability to prevent flooding. Interestingly, all players identified a different purpose: "demonstration of the importance of installing SUDS to avoid flooding", "cooperation", "prevention of floods", and "raise awareness on climate change in urban environments". This implies that although the game increases awareness of flood prevention more broadly, the focus on SUDS and their consequences is still missing. In terms of the level of challenge, players found the game to be moderately challenging, as it was easy to achieve the maximum environmental and social score in the game. To remediate this, it was found that the use of event cards could be exploited more, e.g. by having to draw a positive and negative event card in each round. Although players appreciated the cooperative nature of the game, they suggested that personal competition could be introduced between roles as it is more realistic that different stakeholders have their own, conflicting objectives. This could motivate the players further and increase game difficulty. Lastly, the game structure could be adapted for different numbers of players by reducing the number of actions or increasing the winning criteria.

6 DISCUSSION AND CONCLUSION

Although SUDS are widely recognized as a sustainable and future-proof approach for stormwater management, their adoption remains patchy. Many educational and stakeholder-related barriers have been identified that hinder the adoption of SUDS including lack of awareness and lack of interaction and cooperation among stakeholders. This paper presents a prototype of the serious game UrbanLemma, aimed at increasing awareness of urban flooding prevention needs and SUDS' ability to respond to these while enabling players to understand the benefits of cooperation between roles and responsibilities of different stakeholders in stormwater management.

The game was developed by following common design phases proposed by [28]. Several jumps and iterations between phases were required during the design process. It is argued that the design specification cannot be fully addressed without input from the phase that follows it, the system analysis. This is particularly apparent when addressing the aspects of realism that should be contained in the game. In addition, it is difficult to fully construct and respond to the matrix of system components and gaming elements without considering the game format first. If considered at a preliminary stage, the game format can inform the gaming elements that feature in the matrix. As the game design is a rather creative process, different other creativity techniques could be used, such as writing a report of the game concept [28], or concept and mind mapping to visualize game elements and their interactions.

The prototype was overall well-received by the small number of participants in the validation session. Several recommendations to improve realism and interaction between players were made. Systematic testing and evaluation are required to understand whether the game has the intended impact.

7 REFERENCES

[1] United Nations Department of Economic and Social Affairs, "World Urbanization Prospects 2018: highlights," 2018. Accessed: Sep. 14, 2020. [Online]. Available: https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf.

[2] B. R. Rosenzweig et al., "Pluvial flood risk and opportunities for resilience," Wiley Interdiscip. Rev. Water, vol. 5, no. 6, p. e1302, Nov. 2018, doi: 10.1002/WAT2.1302.

[3] E. C. O'Donnell and C. R. Thorne, "Drivers of future urban flood risk," Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 378, no. 2168, Apr. 2020, doi: 10.1098/RSTA.2019.0216.

[4] P. S. Kaspersen, N. H. Ravn, K. Arnbjerg-Nielsen, H. Madsen, and M. Drews, "Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding," Hydrol. Earth Syst. Sci, vol. 21, pp. 4131–4147, 2017, doi: 10.5194/hess-21-4131-2017.

[5] Q. Zhou, "A review of sustainable urban drainage systems considering the climate change and urbanization impacts," Water (Switzerland), vol. 6, no. 4, pp. 976–992, 2014, doi: 10.3390/w6040976.

[6] M. Davis and S. Naumann, "Making the Case for Sustainable Urban Drainage Systems as a Nature-Based Solution to Urban Flooding," in Nature-Based Solutions to Climate Change Adaptation in Urban Areas, 2017, pp. 123–137.

[7] CIRIA, "The SuDS manual (No. C753)," London, UK, 2015. Accessed: Jul. 05, 2021. [Online]. Available: https://www.ciria.org/Memberships/The_SuDs_Manual_C753_Chapters.aspx.

[8] R. E. De Graaf, R. J. Dahm, J. Icke, R. W. Goetgeluk, S. J. T. Jansen, and F. H. M. Van De Ven, "Receptivity to transformative change in the Dutch urban water management sector," Water Sci. Technol., vol. 60, no. 2, pp. 311–320, Jul. 2009, doi: 10.2166/WST.2009.179.

[9] R. R. Brown and M. A. Farrelly, "Delivering sustainable urban water management: A review of the hurdles we face," Water Sci. Technol., vol. 59, no. 5, pp. 839–846, 2009, doi: 10.2166/wst.2009.028.

[10] X. J. Qiao, A. Kristoffersson, and T. B. Randrup, "Challenges to implementing urban sustainable stormwater management from a governance perspective: A literature review," J. Clean. Prod., vol. 196, pp. 943–952, 2018, doi: 10.1016/j.jclepro.2018.06.049.

[11] V. Venkataramanan et al., "Knowledge, attitudes, intentions, and behavior related to green infrastructure for flood management: A systematic literature review," Sci. Total Environ., vol. 720, p. 137606, Jun. 2020, doi: 10.1016/J.SCITOTENV.2020.137606.

[12] A. K. Sharma, S. Cook, G. Tjandraatmadja, and A. Gregory, "Impediments and constraints in the uptake of water sensitive urban design measures in greenfield and infill developments," Water Sci. Technol., vol. 65, no. 2, pp. 340–352, 2012, doi: 10.2166/WST.2012.858.

[13] I. Winz, S. Trowsdale, and G. Brierley, "Understanding barrier interactions to support the implementation of sustainable urban water management," Urban Water J., vol. 11, no. 6, pp. 497–505, 2014, doi: 10.1080/1573062X.2013.832777.

[14] L. Li, A. M. Collins, A. Cheshmehzangi, and F. K. S. Chan, "Identifying enablers and barriers to the implementation of the Green Infrastructure for urban flood management: A comparative analysis of the UK and China," Urban For. Urban Green., vol. 54, no. June, p. 126770, 2020, doi: 10.1016/j.ufug.2020.126770.

[15] E. C. O'Donnell, J. E. Lamond, and C. R. Thorne, "Recognising barriers to implementation of Blue-Green Infrastructure: a Newcastle case study," Urban Water J., vol. 14, no. 9, pp. 964–971, 2017, doi: 10.1080/1573062X.2017.1279190.

[16] A. H. Roy et al., "Impediments and solutions to sustainable, watershed-scale urban stormwater management: Lessons from Australia and the United States," Environ. Manage., vol. 42, no. 2, pp. 344–359, 2008, doi: 10.1007/s00267-008-9119-1.

[17] D. Michael and S. Chen, Serious games: Games that educate, train, and inform. Thomson Course Technology PTR, 2006.

[18] A. H. Aubert, R. Bauer, and J. Lienert, "A review of water-related serious games to specify use in environmental Multi-Criteria Decision Analysis," Environ. Model. Softw., vol. 105, pp. 64–78, 2018, doi: 10.1016/j.envsoft.2018.03.023.

[19] D. A. Savić, M. S. Morley, and M. Khoury, "Serious gaming for water systems planning and management," Water (Switzerland), vol. 8, no. 10, pp. 1–17, 2016, doi: 10.3390/w8100456.

[20] A. Mittal, L. Scholten, and Z. Kapelan, “A review of serious games for urban water management decisions: current gaps and future research directions,” Water Res., vol. 215, p. 118217, May 2022, doi: 10.1016/J.WATRES.2022.118217.

[21] C. Choi, P. Berry, and A. Smith, “The climate benefits, co-benefits, and trade-offs of green infrastructure: A systematic literature review,” J. Environ. Manage., vol. 291, p. 112583, Aug. 2021, doi: 10.1016/J.JENVMAN.2021.112583.

[22] A. H. Roy et al., “Impediments and solutions to sustainable, watershed-scale urban stormwater management: Lessons from Australia and the United States,” Environ. Manage., vol. 42, no. 2, pp. 344–359, 2008, doi: 10.1007/s00267-008-9119-1.

[23] K. Krijnen, “Increasing the resilience of urban areas to extreme precipitation: Are the residents ready?,” Delft University of Technology, Delft, 2020.

[24] A. F. Vasconcelos and A. P. Barbassa, “Sustainable urban stormwater management in developing countries: integrating strategies to overcome Brazilian barriers,” Urban Water J., 2021, doi: 10.1080/1573062X.2021.1969415/SUPPL_FILE/NURW_A_1969415_SM9898.PDF.

[25] H. K. Lukosch, G. Bekebrede, S. Kurapati, and S. G. Lukosch, “A Scientific Foundation of Simulation Games for the Analysis and Design of Complex Systems,” Simul. Gaming, vol. 49, no. 3, pp. 279–314, 2018, doi: 10.1177/1046878118768858.

[26] R. D. Duke, “The Game Design Process,” in Gaming Simulation: Rationale Design and Applications, C. S. Greenblat and R. D. Duke, Eds. John Wiley & Sons, Inc., 1975, pp. 99–168.

[27] R. D. Duke and J. L. A. Geurts, Policy games for strategic management: Pathways into the unknown. Amsterdam: Dutch University Press, 2004.

[28] V. Peters and M. Westelaken, Simulation games - A concise introduction to game design. Nijmegen, The Netherlands: Samenspraak Advies, 2014.

[29] A. Pendleton, “Introducing the Game Design Matrix: A Step-by-Step Process for Creating Serious Games,” Air Force Institute of Technology, 2020.

MODELLING WATER MIXING AND RENEWAL IN STORAGE TANKS

**Dídia Covas[1], Alexandre Pinheiro[2], Sofia Vaz[3],
Laura Monteiro[4], Nuno Martins[5] and Ana Ricardo[6]**

[1]CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[1] *didia.covas@tecnico.ulisboa.pt*, [2] *alexandre.pinheiro@tecnico.ulisboa.pt*,
[3]*sofia.anibal.vaz@tecnico.ulisboa.pt*, [4] *laura.monteiro@a4f.pt*,
[5] *nunomiguelmartins@tecnico.ulisboa.pt*,
[6] *ana.ricardo@tecnico.ulisboa.pt*

Abstract

The current paper aims at presenting the main developments and achievement attained in research project IMiST which aimed at a better understanding of flow dynamics inside water supply systems' storage tanks to find practicable solutions to improve the design and the rehabilitation and to support operation of the existing tanks. The project comprised the development of an extensive experimental programme in small-scale models, full-scale testing in a real storage tank and advanced numerical modelling. Small-scale tests were carried out in the Laboratory of Hydraulics and Water Resources of Instituto Superior Técnico, Portugal, for different tank configurations and operating conditions. Two cross-section tanks (circular and rectangular) were tested with and without interior structures (baffles), with the inlet and outlet pipes at different locations and for constant and variable water level. Three sets of experimental tests were carried out using different instrumentation to collect complementary data, namely traditional tracer tests, dye tracer tests and velocity field measurements by Particle Image Velocimetry (PIV). A methodology for assessing mixing conditions in full-scale storage tanks was developed and tested in a real tank with a rectangular cross-section and 3500 m3 capacity with inner baffles. Advanced numerical modelling using Computational Fluid Dynamics (CFD) was carried out for the submerged and plunging get in a circular tank for betted understanding the velocity fields and the flow patterns. This complementary experimental and numerical modelling have allowed drawing conclusions concerning improving mixing measures, by readjusting the tank configuration (e.g., location, number and diameter of inlet/outlet pipes, jet inflow, fluctuating stored water volume and geometry) and changing operating conditions (e.g., extreme tank levels). New knowledge on water storage tank hydrodynamics has been created and recommendations for improving water mixing in existing tanks are briefly summarised herein.

Keywords
Storage tanks, water mixing, water renewal, water quality, experimental tests, CFD, PIV, improvement measures.

1 INTRODUCTION

Water storage tanks are essential components of water distribution systems. These infrastructures are designed and operated to meet the variation of water demands of the distribution systems, while equalizing pressure and providing emergency storage. Nonetheless, they are also frequent sources of deterioration of drinking water quality, owing to inadequate tank design, operation and maintenance [1-2]

Typically, tank design does not account for water mixing and renewal. Tanks with inlet/outlet via single or adjacent pipes are common solutions in engineering practice, thus with bulky dead zones with minimal or null renewal [1, 3-4]. As large volumes are stored to respond to fire fighting and emergency demands, tanks are generally operated with high water levels with small amplitude draining/filling cycles. In addition, owing to temperature differences between inflowing and stored water, thermal stratification can be established, which can lead to ineffective mixing and stable stratified conditions [5-7].

Short-circuiting, zonal recirculation and thermal stratification allow for excessive residence times, particularly high in low or no flow zones with increased chlorine decay, biofilm development and sediment accumulation [1, 8-11]. In these circumstances, formation of potentially carcinogenic toxic disinfection by-products occurs, as increased amounts of chlorine are consumed, including those additionally added upon necessary re-chlorination [12-19]. Concomitantly, the existence of biofilm and sediments in a low or null chlorine concentration environment is a prone situation for microbial regrowth and associated risks, as well as for appearance of microbial taste and odour compounds [2]. In addition to the microbial regrowth upon re-suspension, accumulated sediments increase risks of discolouration in downstream pipes [9-11, 20-25]. Therefore, deficient mixing also leads to increases in frequency of required tank cleaning and disinfection [2, 13, 16-18, 26]. Improving mixing and reducing heterogeneous ageing and stagnation in storage tanks is, thus, crucial for the quality and safety of drinking water in supply systems.

Storage tanks have various cross-section shapes (e.g., rectangular, circular), sizes and inlet/outlet configurations (e.g., one single pipe for inlet and outlet, separated pipes located near each other or in extremities). Water flow paths and velocity and, thus, the prevalence of mixed or plug-flow regimes within the tanks, will depend on the tank configuration and operating conditions. Plug flow regimes are to avoid as tanks operated under these conditions tend to lose more disinfectant [27] and entailed mixing is minimal comparatively to mixed flow [2, 27]. On the other hand, the inclusion of baffles does not provide proper mixing and, generally, worsens storage tanks water quality [16, 27-30]. Water thermal stratification can even change the flow regime from mixed to plug-flow [31]. Tanks are normally operated without any active mixing devices, such as turbines or impellers [32], although mechanical mixing have been prescribed in particular situations [2]. Hence, mixing in storage tanks depends on water movement during the filling/draw cycle, i.e., on the kinetic energy of the inflowing water jet [4, 27]. Thus, effectiveness of mixing is mainly dependent on the inlet flow momentum (determined by inlet diameter and flow rate) and inlet location and orientation [33-36], as well as on tank geometry and size [27], though mixing dependence on various geometric parameters is not fully known [36]. Hence, tank water mixing and age, and, thus, quality and safety, do not depend exclusively on storage tanks characteristics and condition. They are also influenced by the fluctuations of inlet water flow and levels [37], as determined by volumes stored for emergency and demand management, and draining/filling times and frequency according to serviced areas consumption profiles. Nevertheless, water level in tanks and filling frequency are most frequently a function of electricity tariff schemes, as there is an emerging trend to minimize energy costs by scheduling pumps' operation [38-39]. Such operational decisions are made disregarding potential impacts on water mixing and turnover associated matters.

Mathematical and physical models of mixing and water quality dynamics have been used to determine how alternative designs or operational policies affect the quality of water within the tanks [3, 8, 32, 40]. In this context, Computational Fluid Dynamics (CFD) modelling is a proven tool for describing flow fields in tanks, thus allowing to simulate different designs, configurations and operational conditions, in order to optimize mixing and residence times distribution [41-45]. However, due to the large variety of existing tanks sizes, shapes and configurations, only a small number of conditions have been studied. Most studies focused on a single factor (e.g., inlet flow location; tank shape) not integrating other critical and overlooked variables (e.g., varying stored volumes, tank filling/discharging flows; operating rules), nor testing single or combined mixing improvement measures. In addition, most CFD models lack experimental calibration and validation.

The current paper aims at presenting the main developments and achievement attained in research project IMiST – Improving mixing in Storage Tanks for safer water supply, funded by Fundação para a Ciência e Tecnologia, carried out between 2018 and 2022. This project aims at understanding flow dynamics inside storage tanks to find practicable solutions to improve the design and rehabilitation and to support operation of the existing storage tanks. The project, comprising advanced numerical modelling and lab and full-scale testing, innovatively includes velocity fields' measurement in laboratory tanks of different configurations and operating conditions by Particle Image Velocimetry (PIV). Likewise, mathematical modelling hydrodynamics are carried out by using Computational Fluid Dynamics (CFD). Improving mixing measures, by readjusting tank configuration (e.g., location and number of inlets/outlets, jet inflow, fluctuating stored water volume and geometry) and changing operating conditions (e.g., extreme tank levels, pump scheduling) are evaluated both experimentally and numerically for their effectiveness. Complementarily, the effect of implementing multiple or moving inlets on mixing at several filling levels is analysed. A methodology for assessing mixing conditions in full scale storage tanks is developed and tested. In addition to new knowledge on storage tanks hydrodynamics, outcomes include recommendations for upgrading tank design, rehabilitation and operation with measures for the improvement/rehabilitation of existing storage, as integrated on the normal or adjusted operation and configuration of the overall supply systems.

IMiST project is organized in seven main tanks: T1 - Field survey and database assembly; T2 - Lab scale studies; T3 - Numerical modelling; T4 - Full scale studies; T5 - Mixing improvement measures; T6 - Modelling of mixing in hydraulic simulators; T7 - Recommendations design, operation and rehabilitation. Results have given rise to several publications [46-51]. Relevant results attained in tasks T2-T6 are briefly presented herein.

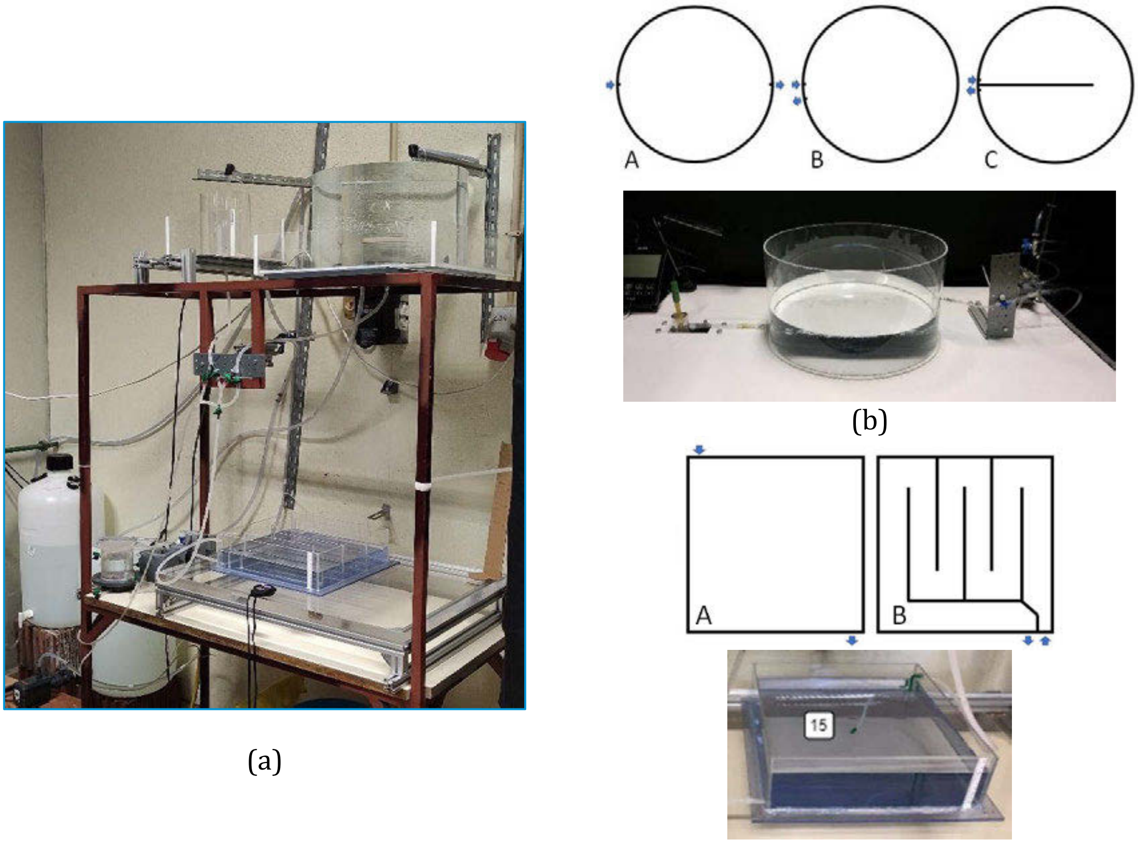
2 SMALL-SCALE TESTING: TRACER TESTS

Small-scale testing aims at analysing the effect of the tank configuration and the operating mode on water mixing and renewal processes in two cross-sections (circular and rectangular) storage tanks of typical configurations in Portugal, operated at steady-state and at variable water-level. Experiments were carried out in a laboratory facility, assembled at the Laboratory of Hydraulics and Water Resources of Instituto Superior Técnico, Portugal. This facility was designed to develop tracer tests (Figure 1a) and determine the water residence time distribution (RTD). The tank was supplied by gravity fed with a tracer solution from two secondary tanks. Step input tracer tests consisted of continuously injecting a NaCl solution (0.05 gL^{-1}) at the inlet of the tank and monitoring the tracer concentration at the tank outlet. A control valve located immediately upstream of the tank regulated the inflow rate.

The small-scale tanks used are in 1:100 circular cross-section tanks of acrylic with 392 mm diameter and a maximum water-depth-to-tank-diameter ratio of 0.15; and 1:77 rectangular tanks square cross-section (350 x 350 mm^2), with a maximum water depth of 70 mm (corresponding to

BY NC SA

a maximum depth-width ratio of 0.2). Three different inlet/outlet pipe configurations were tested for the circular tanks (Figure 1b) and two configurations for the rectangular tanks (Figure 1c). Different inlet flowrates were tested for a constant water level (i.e., Q1=5 l/h, Q2=7 l/h, Q3=9 l/h and Q4=11 l/h) and for a variable water level (i.e., constant inlet flow 9 l/h and three outlet flow rates Q1=1.5 l/h, Q2=3.0 l/h and Q3=4.5 l/h) to simulate the fill-and-draw cycles in real tanks.



(a)

(b)

(c)

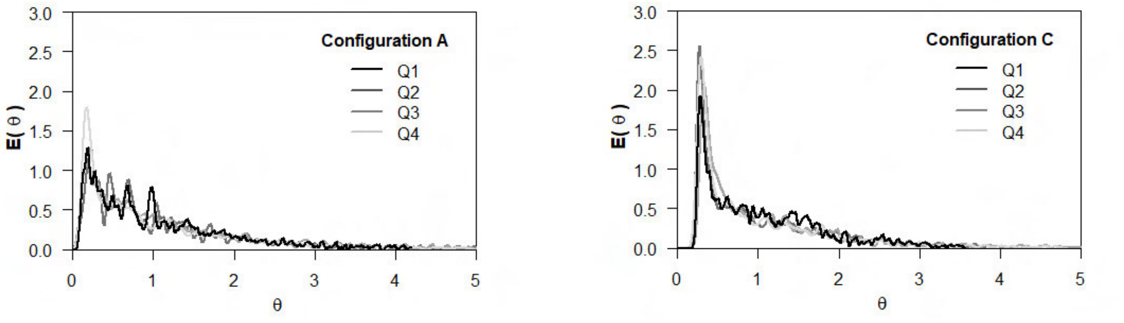
Figure 1. Small-scale test facility: (a) general view: (b) circular cross-section tanks; and (c) rectangular cross-section tanks.

The testing procedure was the following: i) the tracer concentration in the outlet pipe is measured along time, *C(t)*, and ii) measured values are normalized to obtain the non-dimensional cumulative distribution function *F(t)= C(t)/C₀*, where C_0 is the tracer concentration in the inlet pipe; iii) residence time distribution (RTD) function, *E(t)*, is calculated by differentiating *F(t),* representing the time spent by the different fluid elements inside the tank; iv) several hydraulic indices are calculated, namely Short-circuiting time (t_{10}), Morrill index (Mo = t_{90}/t_{10}) and the turnover/renewal time (t_{95}).

Circular cross-section tanks

The effect of inflow momentum flux on mixing in circular cross-section tanks is assessed by comparing the distribution curves (Figure 2) and the mean residence times (t') for flow rates *Q1* to *Q4*. The normalized RTD functions at different inlet flow rates present similar profiles, showing that the tested momentum flux range did not significantly change mixing within the tanks, except for Configuration A where the first peak increases with the flow rate. Increasing the flow rates makes the mean residence time (t') to approach the theoretical value (τ) because of diminished dispersion effects. In contrast, in Configuration B, a large part of the fluid entering the tank leaves almost immediately, which results in a mean residence time lower than τ. In the baffled tank (Configuration C), the mean and the theoretical values are very similar.

CC BY NC SA



a) b)

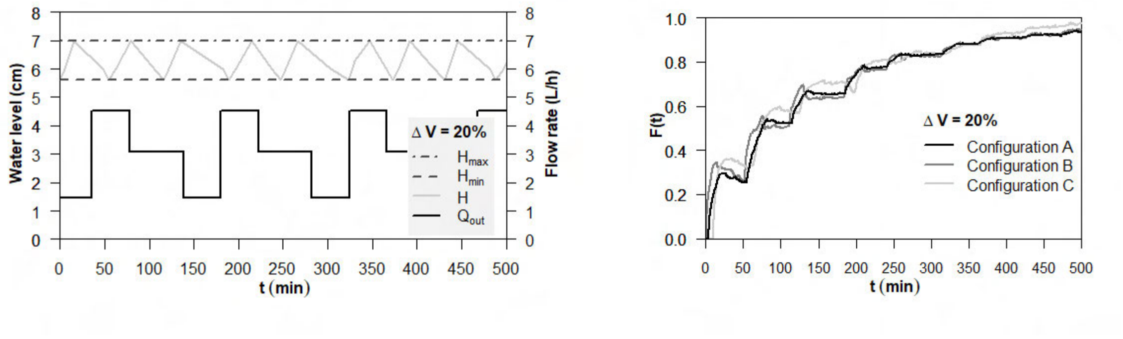
Figure 2. Exemples of normalized residence time distribution for constant water level and for different flow rates: (a) configuration circular A; b) Configuration circular C.

Short-circuiting and mixing indexes for the three tanks at steady-state conditions are presented in Table 1. Configuration B shows the lowest t_{10} values. 10% of the injected tracer reaches the outlet pipe very quickly, in less time than needed for the same amount of tracer to leave a CSTR, denoting short-circuiting. For the higher flow rates (9.2 and 11.6 Lh^{-1}), t_{10} is lower than half of a completely mixed tank. For Configuration A, t_{10} is higher than for the CSTR, because of recirculation within the tank. This index presents the highest values for the baffled tank, much higher than those for the completely mixed tank. In Configuration B, $\bar{\sigma}$ is also the highest observed, which indicates that flow paths within the tank are quite diverse; yet, other indexes indicate that mixing is far from perfect. Configuration A shows high dispersion indexes, reaching values close to 1 and, thus, close to perfect mixing. *Mo* values range from 11 to 18, which is close to the ideal value of 22. Configuration C presents the lowest $\bar{\sigma}$ and, thus, the lowest degree of mixing. The Morrill index also shows that a lower degree of mixing is achieved in the tank with the baffle (Conf. C) than in the one without the baffle (Conf. A), as *Mo* are lower for tank C.

Table 1. Hydraulic indexes for the three circular cross-section configurations in steady-state conditions.

Config.	Q (L/h)	τ (min)	t' (min)	t_{10} (min) [θ_{10}]	t_{95} (min)	$\bar{\sigma}$ (-)	Mo (-)	t_{10} CSTR (min)	t_{95} CSTR (min)
A	5.1	85.2	87.7	16.5 [0.19]	264	0.7	12	9.0	255
	7.4	58.7	61.1	12.3 [0.21]	180	0.8	11	6.2	176
	9.4	46.2	47.3	8.5 [0.18]	147	0.8	18	4.9	138
	11.6	37.5	37.7	6.1 [0.16]	116	0.9	15	4.0	112
B	5.4	80.5	76.4	8.0 [0.10]	328	0.9	27	8.5	241
	7.4	58.7	54.7	4.4 [0.08]	258	1.0	36	6.2	176
	9.2	47.2	44.1	2.6 [0.06]	211	1.1	46	5.0	141
	11.6	37.5	33.9	1.4 [0.04]	140	1.1	70	4.0	112
C	5.2	83.6	83.4	24.7 [0.30]	240	0.5	8	8.8	250
	7.3	59.5	59.0	16.2 [0.27]	162	0.6	8	6.3	178
	9.1	47.7	48.8	15.3 [0.32]	132	0.7	7	5.0	143
	11.6	37.5	36.2	10.3 [0.27]	96	0.8	7	4.0	112

Three levels of variation in tank volumes were tested (20%, 50% and 80%) and an outflow pattern was used (Figure 3) to simulate real water storage tanks conditions. The inflow rate was either 9.0 Lh-1, during the filling period, or null, while the water level was decreasing.



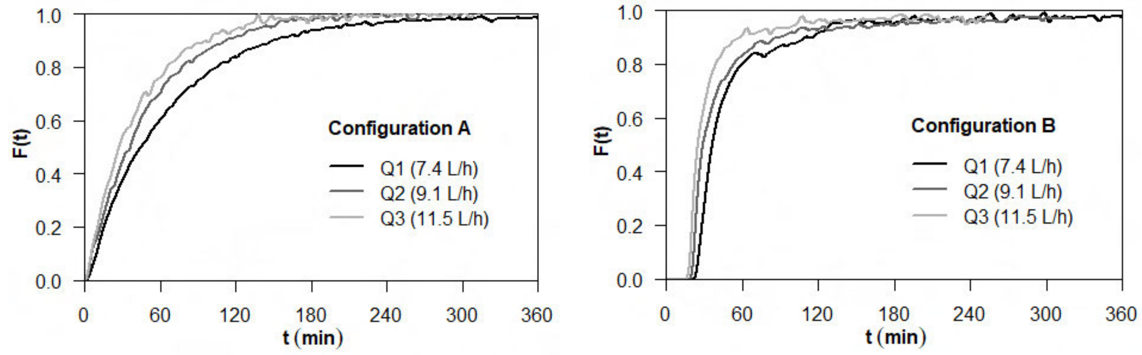
a) b)

Figure 3. Tracer tests at variable water level for circular cross-section tanks for 20% volume variation.

In general, stronger short-circuiting effects occur in the storage tanks operated at variable water levels than in steady-state conditions. Under fill-and-draw mode, the establishment of water pockets of higher residence time is promoted in all tank's configurations, particularly if small water level variation occurs in each cycle. Even when the filling periods are longer than the theoretical mixing time, mixing in the tanks operated with small volume variation in each cycle is far from complete. Consequently, despite results from CFD and tracer small-scale tests at steady-state conditions providing good insights on the hydrodynamics in the tanks, care must be taken when inferring results for full-scale tanks operated at variable water levels.

Rectangular cross-section tanks

A similar analysis was carried out for two configurations (A and B, see Figure 1c) of rectangular cross-section tanks. The cumulative distribution *F(t)* determined for the two configurations A and B with mode 1 and three flow rates is presented in Figure 4. The *F(t)* increases in the beginning of the test in Configuration A (fully open) (Figure 4a), highlighting the existence of short-circuiting phenomena. Configuration B (baffled tank) presents a delayed and sharp increase in *F(t)* at initial times, showing the effect similar to a plug flow regime (Figure 4b). The fact is that baffles prevent the inflowing jet to mix with older water and push older water to the outlet pipe.



a) b)

Figure 4. F(t) for the two configurations at steady-state conditions and for three flow rates: (a) Configuration A; (b) Configuration B.

Short-circuiting and mixing indexes are presented in Table 2 Results are coherent with those obtained for circular cross-section tanks. In rectangular tanks operated with a relatively constant water level, higher flow rates increase the water renewal in open tanks (Configuration A). However, the opposite effect is observed in tanks with interior walls (baffles structures) (Configuration B). When operating the rectangular tanks at variable water levels, the higher the water volume variation in each fill-and-draw cycle, the faster the renewal of stored water becomes, with slightly better results in Configuration B despite the worse mixing conditions. For

BY NC SA

high-volume variations (50% to 80%), the renewal times are very similar in the two configurations; for a 20% volume variation, Configuration B promotes a faster water renewal.

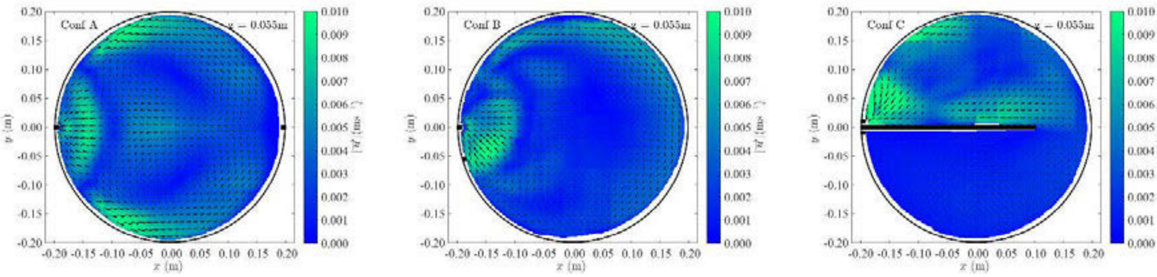
Table 2. Hydraulic indexes for the two rectangular cross-section configurations at steady-state conditions.

Config.	Q (L/h)	τ (min)	t' (min)	t_{10} (min)	t_{95} (min) [1st, last]	$\bar{\sigma}$ (-)	Mo (-)
A	7.4	59.6	61.5	9.6	204	0.86	15.6
	9.1	48.5	49.9	5.7	145	0.92	19.5
	11.5	38.3	38.9	5.4	127	0.70	18.5
B	7.4	59.6	49.1	26.1	130, 180	0.60	4.3
	9.1	48.5	40.5	22.0	175, 183	0.73	4.1
	11.5	38.3	33.7	19.2	103, 238	0.83	3.0

3 SMALL-SCALE TESTING: PIV MEASUREMENTS

Particle image velocimetry (PIV) technology has been used to study flow dynamics in the three configurations of circular-cross section tanks. The experimental facility was adapted to run these tests. A stirring mechanism is installed in the feed tank to mix the PIV seeding particles and to homogenize the inflow water. The experimental PIV system is composed of a laser head and lens, emitting a pulsed laser beam optically transformed into a 2-mm-thick laser sheet, a power supply or laser beam generator, a digital camera, a timing unit and an acquisition and control software. Polyamid Seeding Particles (PSP-50 9080A5011 Dantec Dynamics®, Denmark) with 1.03 of density and a mean particle diameter of 50 µm are used as seeding material in the current tests.

The 2D-PIV can only measure the projection of the velocity into the plane of the light sheet. The PIV system is operated between 0.5 and 15 Hz sample frequency. In each measuring gap, horizontal PIV velocity maps are acquired at three water heights, namely, at 0.010, 0.030 and 0.055 m from the tank bottom. Each plane was composed of the 9 or 12 small sections with 9.5 cm long and 12.5 cm wide. For each acquisition, approximately, 15 minutes of consecutive data are collected. The time-averaged velocity fields, obtained based on instantaneous velocity measurements, are presented in Figure 5 for each tank configuration (A, B and C) and for the horizontal z=0.055 m.



a) b) c)

Figure 5. Time-averaged velocity maps at each configuration and horizontal plane at z=0.055 from the bottom for Configurations a) A, b) B and c) C

The three configurations have very different flow patterns. Configuration A develops two symmetrical vortices in which the water moves onwards along the tank walls and returns in the central zone. Configuration B, given the proximity of the inlet and outlet pipes, enhances the short-circuiting flow path with poor mixing in most of the tank, since part of the incoming fluid is directed to the outlet without mixing with the stored water. Also, the region with lower velocities is located in the central zone of the tank and with higher expression near the tank bottom. In Configuration C, the baffle reduces the short-circuiting and promotes de recirculation and mixing in the first half of the tank, delaying the new water from reaching the outlet pipe.

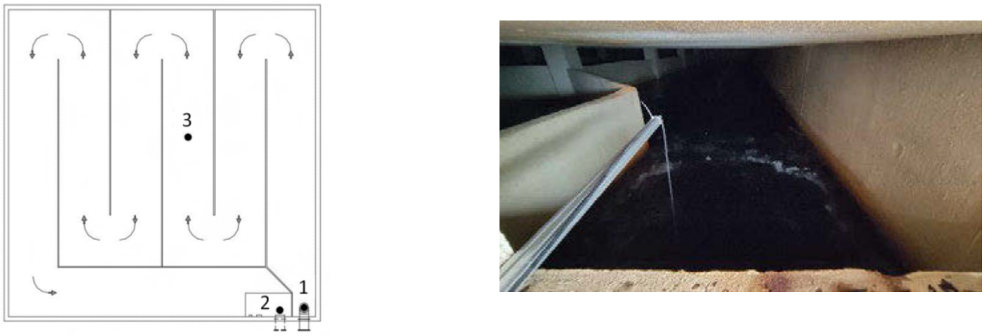
BY NC SA

4 FULL-SCALE TESTING

The full-scale tank selected to undertake the tests is an aboveground square cross-section structure made of concrete belonging to Águas do Algarve, the water utility responsible for the water supply and wastewater transport in the South of Portugal. The tank has a capacity of about 3 450 m^3, with an inner side with of 26.6 m, and corridors 5.5 m wide. An inlet pipe, with a 0.8 m diameter, enters the tank near the bottom of the structure, and an elbow with an upward angle directs the flow to the opening at 4.5 m height. The outlet pipe, also with 0.8 m diameter, is located inside a ditch, 1.5 m below the tank floor. This tank features internal baffles that force the flow through a series of 90º and 180º turns between inlet and outlet, and the total length of the water path is approximately 160 m (Figure 6).

Conductivity was monitored at three different locations (Figure 6a): 1) at 1 m inside the inlet pipe; 2) inside the ditch and 1 m below the floor of the tank next to the outlet pipe; and 3) in the middle of the third corridor and 2.8 m from the floor of the tank. Three conductivity probes were installed at each sampling site (two YSI Pro 30 and one YSI 556 MPS) and used to monitor the conductivity and the temperature during the test at 10-minutes intervals.

The test comprised the application of the step tracer method, in which the inflow water with a higher conductivity enters the tank filled with water with lower conductivity. The test started when the intended conductivity was achieved in the WTP (about 3 km from the tank). Inflow and outflow rates respond to the needs of the distribution systems, allowing for a minimum volume variation (ca. 20%). The test lasted until the conductivity between inflowing water and the outflowing differed by less than 5%.



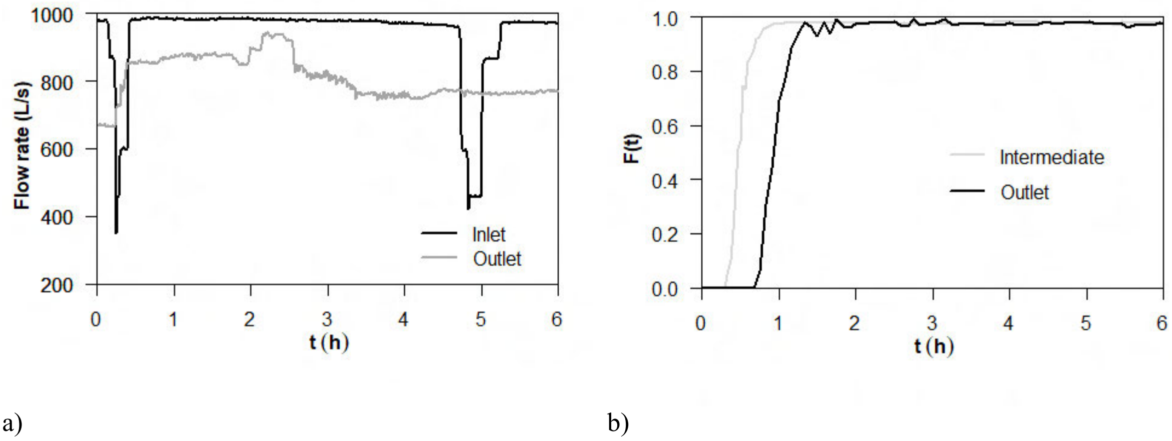
a) b)

Figure 6. Full-scale tank: a) aerial photo; b) inner view.

Collected data during 6 hours are shown in Figure 7a. During the first six hours, the inflow was practically constant (1 000 ls^{-1}); the outflow varied between 665 and 945 ls^{-1} and the water level varied between 3.6 and 4.5 m, which corresponds to a volume variation of 20%.

The electric conductivity evolution at the three monitoring sites has shown that, even though the water was leaving the WTP with a constant conductivity (ca. 402 μScm^{-1}), it still took approximately 30 minutes for the inflow conductivity to stabilize. Cumulative distribution curves, $F(t)$, show a rapid increase even in the outlet pipe (Figure 7b). Only after 40 minutes it is possible to start noticing the newer water reaching the outlet pipe. After 46 min, the 10% inflow conductivity was reached at the outlet point (t_{10}) and the complete renewal was attained at the intermediate monitoring point (t_{95}); after 1 h 41 min, the water in the tank was completely renewed. Even with the plug flow created by the existing baffle structure, the high flow rates allowed a rapid renewal of the stored water. Some older water pockets come out later and, after 6 hours, the renovation is almost complete. Calculated indexes at the outlet pipe are t_{10}=46, t_{95}=75 and 101 min (first and last) and Mo=1.5

CC BY NC SA



a) b)

Figure 7. Full-scale tests: a) flow rate data; b) cumulative distributions curves.

5 CFD MODELLING

Computational Fluid Dynamics (CFD) modelling has been carried out for a better understanding of the flow hydrodynamics (flow pattern) in circular storage tanks for two inflow configurations: the inlet pipe above the water level (referred to as a plunging jet) and the inlet pipe below the water level (referred as submerged jet).

A 3D CFD model of the small-scale circular cross-section tank of Configuration A was developed using OpenFOAM, a free and high-level language CFD library. The final mesh is refined near the tank walls, inlet and outlet regions and in the air-water interface, which are areas with high-velocity gradients region. The resulting mesh is characterized by 1.4 million mainly hexahedral cells with 1.5 million points of calculus. Less refined meshes settings in the wall were analysed however, the flow pattern obtained by the CFD did not follow the one observed by recorder images of the movement of the dye tracer solution in the laboratory [52]. For the flow pattern analysis, a multiphase solver has been used. For each phase, the solver multiphaseInterFoam captures the interfaces between two (water and air) and includes surface tension and contact angle effects. It uses the finite volume along with the volume of fluid method.

Results for the two configurations are presented in Figure 8 in terms of velocity fields at mid water height. Main differences are related to the direction of the flow paths inside the tank. In a plunging jet, the water enters the tank and sinks downwards; due to the 'Coanda effect' in the tank wall, the water from the jet progresses clung to the tank walls, symmetrically, in the left and right directions, and returns backwards through the tank center. In the submerged jet, the water is directed directly to the outlet pipe and returns backwards along the tank walls.

This research demonstrates that numerical models can capture the 3D behaviour of flow dynamics in water storage tanks. CFD results allow assessing the influence of the nozzle position on the hydrodynamics conditions inside the tank. Each jet generates a preferential flow path or distinct flow pattern and velocity field in the water storage tank. The plunging jet breaks into droplets that fall onto the water surface generating a void responsible for the progression of the jet clung to the tank wall and, then produces a symmetrical flow clung to the walls and a flow in the centre of the tank. The submerged jet generates a long entrainment path with high velocities through the center of the tank, promoting an enhanced water mixture by the turbulence created and a flow towards the inlet along the walls.

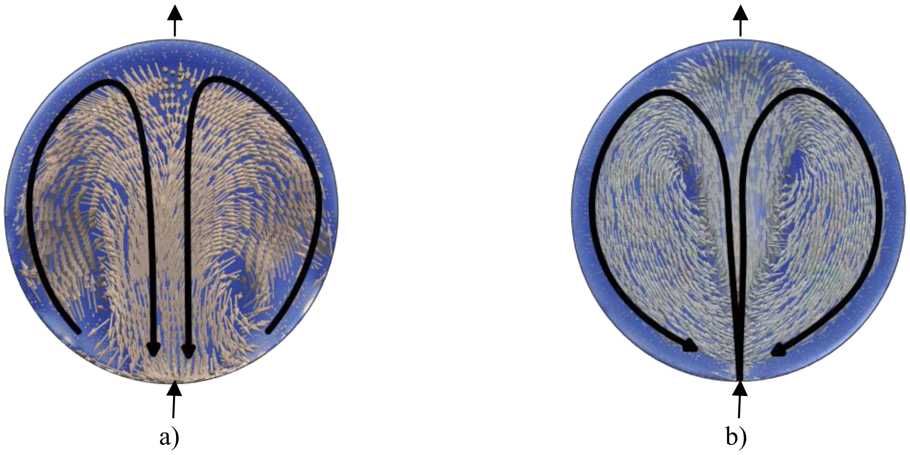


Figure 8. CFD results at mid water height for circular tanks: a) plunging jet and b) submerged jet.

6 MIXING IMPROVEMENT MEASURES

The effect of different types of measures on water mixing in circular and rectangular cross-section tanks is investigated. Several structural improvement measures are analysed: i) the inlet pipe diameter effect (two diameters were tested, 2 and 4 mm, in the circular cross-section tank); ii) the use of multiple nozzles effect (three nozzles were tested in circular cross-section tank and 10 nozzle configurations in rectangular cross-section tanks); iii) the effect of baffles with different lengths (0%, 50% and 75% of the diameter of the circular cross-section tank); iv) the use of baffles with holes (in rectangular cross-section tanks); and v) one operational improvement measure which corresponds to operating the tank with variable water level, that is with fill-and-draw cycles with several amplitudes of the volume variation (20%, 50% and 80%) in the circular cross-section tanks (Configurations A, B and C).

Tracer tests with NaCl have been carried out for a more qualitative analysis and dye tracer tests for a more qualitative analysis. Figure 9 shows examples of dye tracer tests carried out in the rectangular tanks of configuration A with different number, orientation and location of nozzles (referred as S2, S3, S5, S6 and S7). Different flow paths are established depending on the location and orientation of the nozzles within the tank.

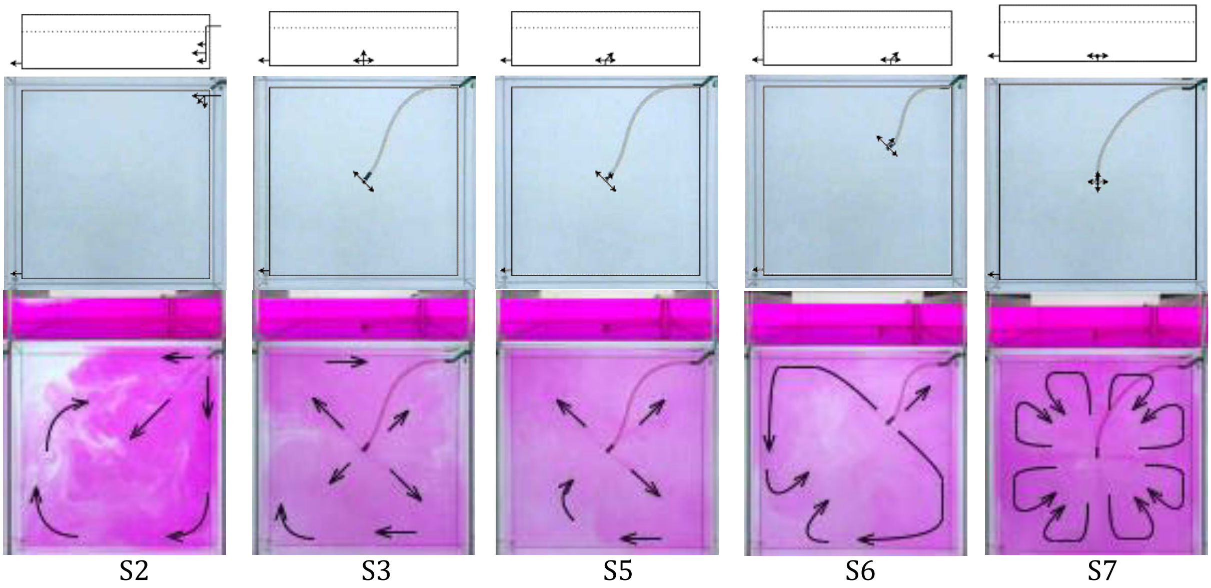


Figure 9. Rectangular tank configuration A: dye tracer tests for several locations and orientations of nozzles.

CC BY NC SA

The hydraulic indexes calculated for the circular cross-section tank for constnat water level and for the flow rate Q3 are shown in Table 3. Similar conclusions can be drawn from the tests in retangular cross-section tanks.

Table 3. Hydraulic indexes for the circular cross-section tanks for constant water level and for Q3.

Configu-ration	**Improvement measure**	**Inlet pipe** (mm)	**Morrill index** *Mo*	**Short-circuiting index** θ_{10}	Variation	**Turnover time** θ_{95}	Variation
Circular A	Initial conditions	4	12	0.19		3.10	
	Inlet diameter reduct.	2	23 🟢	0.11	– 39 % 🔴	3.67	+ 15 % 🔴
	Nozzles	6 × 1	18 🟡	0.13	– 30 % 🔴	3.18	0 % 🟡
		4 × 1	20 🟢	0.12	– 36 % 🔴	3.36	+ 6 % 🔴
		6 × 1	22 🟢	0.10	– 44 % 🔴	3.10	– 2 % 🟢
Circular B	Initial conditions	4	46	0.06		4.78	
	Inlet diameter reduct.	2	18 🟢	0.13	+ 117 % 🟢	3.13	– 34 % 🟢
	Nozzles	5 × 1	15 🟢	0.17	+ 183 % 🟢	3.27	– 32 % 🟢
		7 × 1	19 🟢	0.14	+ 133 % 🟢	3.32	– 31 % 🟢
		3 × 1	26 🟢	0.10	+ 67 % 🟢	3.12	– 35 % 🟢
	Baffle 50%	4	9 🔴	0.26	+ 550 % 🟢	2.87	– 31 % 🟢
	Baffle 75%	4	7 🔴	0.32	+ 433 % 🟢	2.76	– 42 % 🟢
Circular C	Initial conditions	4	7	0.32		2.76	
	Inlet diameter reduct.	2	11 🟢	0.21	– 33 % 🔴	3.35	+ 21 % 🔴

Note: 🟢 positive effect; 🟡 identical effect; 🔴 negative effect

Operating the tanks with fill-and-draw cycles, with large water volume variation, globally improves the water mixing conditions, reduces the short-circuiting effect and reduces the renewal time for all circular cross-section tanks and for rectangular tanks without baffles. The higher the water volume variation is, the most effective this measure becomes on improving mixing and renewal conditions. Still, this measure increases slightly the short-circuiting in rectangular tanks with baffles. Operating the tanks with fill-and-draw cycles is, indeed, the simplest and the most cost-effective water mixing and renewal measure. The only disadvantage is that operating the tank with not-full storage reduces the supply system reliability in case an abnormal situation occurs at upstream the tanks that interrupts the supply during several hours.

Both the reduction of the inlet pipe diameter and the use of multiple nozzles near the tank bottom far from the inlet pipe are the most efficient structural measures to improve mixing; though the short-circuiting effect can increase in circular tanks configurations A and C. Tests in circular cross section tanks have shown that the nozzles should be more than three directed to the horizontal or making a 45º angle with the horizontal. Tests in rectangular tanks carried out for 10 different nozzles locations, sizes and number have shown that: the most efficient solutions those with nozzles near the tank inlet at different water depths and pointing to the tank lateral wall, or nozzles (≥4) in the tank centre with horizontal or 45º angle jets. Few nozzles located in the centre of the tank with vertical jets should be avoided.

The use of baffles promotes the plug-flow and, consequently, worsens the water mixing conditions; however, baffles reduce the short-circuiting effect and reduce the renewal time. These structures are very important in tanks of water transmission systems, namely at downstream water treatment works or at intermediate locations in the system, far from distribution networks; that is in tanks in locations in which the water passes and stays very few time. Conversely, in tanks located immediately at upstream distribution systems, with large volumes, with high retention times and with tendency to have lower disinfectant concentrations, it is recommendable to increase inlet water momentum by reducing the inlet pipe diameter and by using nozzles.

The use of smaller-size baffles (i.e., 50% of the diameter, instead of 75%) is more efficient for improving mixing conditions in tanks with close inlet/outlet pipes. The use of small holes in baffle structures (as tested in rectangular tanks) have shown hardly any effect, which may have been due to the small-size of the tested holes (1-2 mm). Further research is necessary to assess the effectiveness of this measure.

7 CONCLUSIONS AND FUTURE WORK

The main findings of this research are that the most effective operational measure is operating tanks with fill-and-draw cycles, however, when the tank is operated nearly full to assure the maximum reliability of the system, structural measures are recommendable. Reducing the inlet pipe diameter and installing nozzles near the tank bottom improve the mixing conditions being advisable in large storage tanks located upstream distribution network to maintain disinfectant concentration levels. The use of baffles is recommendable in large tanks located at intermediate locations of the transmission system and with the inlet and outlet pipe located very close.

In terms of future work, other tank configurations (e.g., inlet and outlet pipe locations; different inlet nozzles configurations; with other H/D ratio) and operating conditions (e.g., different demand profiles) should be tested. PIV tests should be carried out for vertical planes for studying the 3D nature of the flow and for different inlet pipe configurations (e.g., for the submerged jet and for different nozzles). CFD simulations should be extended to find the optimal design of nozzles in terms of their location (above the water surface, near the tank bottom, or elsewhere), the ideal number of nozzles and the direction of the water jets.

8 ACKNOWLEDGEMENTS

The authors would like to thank the Fundação para a Ciência e Tecnologia (FCT) for funding: UIDB/04625/2020 from the research unit CERIS; the research project PTDC/ECI-EGC/32102/2017 (IMIST – Improving Mixing in Storage Tanks for safer water supply) and the H2DOC doctoral programme for the PhD grant PD/BD/135214/2017.

9 REFERENCES

[1] Clark, R.M., F. Abdesaken, P.F. Boulos, and R.E. Mau (1996). Mixing in Distribution System Storage Tanks: Its Effect on Water Quality. Journal of Environmental Engineering-ASCE. **122**(9): p. 814-821.

[2] USEPA (2022). Finished Water Storage Facilities (internet). Washington D.C., USA: United States Environmental Protection Agency; 2002 p. 24. Available from: https://www.epa.gov/sites/production/files/2015-09/documents/2007_05_18_disinfection_tcr_whitepaper_tcr_storage.pdf.

[3] Grayman, W.M., R.A. Deininger, A. Green, P.F. Boulos, R.W. Bowcock, and C.C. Godwin (1996). Water Quality and Mixing Models for Tanks and Reservoirs. Journal American Water Works Association. **88**(7): p. 60-73.

[4] Mahmood, F., J.G. Pimblett, N.O. Grace, and W.M. Grayman (2005). Evaluation of water mixing characteristics in distribution system storage tanks. AWWA. **97**(3).

[5] Fisher, I., A. Sathasivan, P. Chuo, and G. Kastl (2009). Effects of stratification on chloramine decay in distribution system service reservoirs. Water Res. **43**(5): p. 1403-13.

[6] Gwerder, C., L. Lötscher, J. Podhradsky, M. Kaufmann, A. Huggenberger, S. Boller, B. Meier, I. Mojic, and M.Y. Haller (2016). Horizontal Inlets of Water Storage Tanks With Low Disturbance of Stratification. Journal of Solar Energy Engineering. **138**(5).

[7] Moncho-Esteve, I.J., M. Gasque, P. González-Altozano, and G. Palau-Salvador (2017). Simple inlet devices and their influence on thermal stratification in a hot water storage tank. Energy and Buildings. **150**: p. 625-638.

[8] Boulos, P.F., W.M. Grayman, R.W. Bowcock, J.W. Clapp, L.A. Rossman, R.M. Clark, R.A. Deininger, and A.K. Dhinga (1996). Hydraulic Mixing and Free Chlorine Residual in Reservoirs. Journal American Water Works Association.

[9] Ho, C.K., J.M. Christian, E.J. Ching, J. Slavin, J. Ortega, R. Murray, and L.A. Rossman (2016). Sediment Resuspension and Transport in Water Distribution Storage Tanks. Journal American Water Works Association. **108**: p. E349-E361.

[10] Huang, C.-C., J.-S. Lai, F.-Z. Lee, and Y.-C. Tan (2018). Physical Model-Based Investigation of Reservoir Sedimentation Processes. Water. **10**(4).

[11] Zahabi, H., M. Torabi, E. Alamatian, M. Bahiraei, and M. Goodarzi (2018). Effects of Geometry and Hydraulic Characteristics of Shallow Reservoirs on Sediment Entrapment. Water. **10**.

[12] Angeloudis, A., T. Stoesser, and R.A. Falconer. Disinfection Kinetics in CFD Modelling of Solute Transport in Contact Tanks. in 3rd IAHR Europe Congress. 2014. Porto, Portugal.

[13] Xu, Q. (2010). Internal Hydraulics of Baffled Disinfection Contact Tanks Using Computational Fluid Dynamics.

[14] Angeloudis, A., T. Stoesser, D. Kim, and R.A. Falconer (2014). Modelling of flow, transport and disinfection kinetics in contact tanks. Proceedings of the Institution of Civil Engineers - Water Management. **167**(9): p. 532-546.

[15] Angeloudis, A. (2014). Numerical And Experimental Modelling of Flow and Kinetic Processes in Serpentine Disinfection Tanks, in Division of Civil Engineering. Cardiff School of Engineering: Cardiff University. p. 193.

[16] Bruno, P., G. Di Bella, and M. De Marchis (2020). Perforated Baffles for the Optimization of Disinfection Treatment. Water. **12**(12).

[17] Angeloudis, A., T. Stoesser, and R.A. Falconer (2014). Predicting the disinfection efficiency range in chlorine contact tanks through a CFD-based approach. Water Research. **60**: p. 118-129.

[18] Kennedy, A., L. Flint, A. Aligata, C. Hoffman, and M. Arias-Paic (2021). Regulated disinfection byproduct formation over long residence times. Water Res. **188**: p. 116523.

[19] Wang, H. and R.A. Falconer (1998). Simulating disinfection processes in chlorine contact tanks using various turbulence models and high-order accurate difference schemes. Water Research. **32**(5): p. 1529-1543.

[20] Hadi, G.A. and J. Kris (2009). A CFD methodology for the design of rectangular sedimentation tanks in potable water treatment plants. Journal of Water Supply Research and Technology. **58**(3): p. 212-220.

[21] Goula, A.M., M. Kostoglou, T.D. Karapantsios, and A.I. Zouboulis (2008). A CFD methodology for the design of sedimentation tanks in potable water treatment - Case study: The influence of a feed flow control baffle. Chemical Engineering Journal. **140**(1-3): p. 110-121.

[22] Samaras, K., A. Zouboulis, T. Karapantsios, and M. Kostoglou (2010). A CFD-based simulation study of a large scale flocculation tank for potable water treatment. Chemical Engineering Journal. **162**(1): p. 208-216.

[23] Oehy, C. (2003). Effects of obstacles and jets on reservoir sedimentation due to turbidity currents.

[24] Gao, H. and M.K. Stenstrom (2018). Evaluation of three turbulence models in predicting the steady state hydrodynamics of a secondary sedimentation tank. Water Research. **143**: p. 445-456.

[25] Dufresne, M., J. Vazquez, A. Terfous, A. Ghenaim, and J.-B. Poulet (2009). Experimental investigation and CFD modelling of flow, sedimentation, and solids separation in a combined sewer detention tank. Computers & Fluids. **38**(5): p. 1042-1049.

[26] Angeloudis, A., T. Stoesser, R.A. Falconer, and D. Kim (2015). Flow, transport and disinfection performance in small- and full-scale contact tanks. Journal of Hydro-environment Research. **9**(1): p. 15-27.

[27] Grayman, W.M., L.A. Rossman, R.A. Deininger, C.D. Smith, C.N. Arnold, and S. JF. (2004). Mixing and ageing of water in distribution system storage facilities. Journal of the American Water Works Association. **96(9):70–80.** .

[28] Kizilaslan, M.A., N. Nasyrlayev, A.T. Kurumus, H. Savas, E. Demirel, and M.M. Aral (2020). Experimental and Numerical Evaluation of a Porous Baffle Design for Contact Tanks. Journal of Environmental Engineering. **146**(7).

[29] Nasyrlayev, N., M.A. Kizilaslan, A.T. Kurumus, E. Demirel, and M.M. Aral (2020). A Perforated Baffle Design to Improve Mixing in Contact Tanks. Water. **12**(4).

[30] Samet, A., M.A. Ben Souf, T. Fakhfakh, and M. Haddar (2019). Numerical investigation of the baffle plates effect on the solar water storage tank efficiency. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. **42**(16): p. 2034-2048.

[31] Zhang, J.-M., B.C. Khoo, H.P. Lee, C.P. Teo, N. Haja, and K.Q. Peng (2012). Effects of Baffle Configurations on the Performance of a Potable Water Service Reservoir. Journal of Environmental Engineering. **138**(5): p. 578-587.

[32] Rossman, L.A. and W.M. Grayman (1999). Scale-Model Studies of Mixing in Drinking Water Storage Tanks. Journal of Environmental Engineering-ASCE. **125**(8): p. 755-761.

[33] Xavier, M.L.M. and J.G. Janzen (2016). Effects of inlet momentum and orientation on the hydraulic performance of water storage tanks. Applied Water Science. **7**(5): p. 2545-2557.

[34] Tian, X. and P.J. Roberts (2008). Mixing in Water Storage Tanks. II: With Buoyancy Effects. Journal of Environmental Engineering-ASCE. **134**(12): p. 986-995.

[35] Tian, X. and P.J. Roberts (2008). Mixing in Water Storage Tanks. I: No Buoyancy Effects. Journal of Environmental Engineering-ASCE. **134**(12): p. 974-985.

[36] Nordblom, O. and L. Bergdahl (2004). Initiation of Stagnation in Drinking Water Storage Tanks. Journal of Hydraulic Engineering. **130**(1): p. 49-57.

[37] Zhang, J.-M., B.C. Khoo, C.P. Teo, N. Haja, T.K. Tham, L. Zhong, and H.P. Lee (2013). Passive and Active Methods for Enhancing Water Quality of Service Reservoir. Journal of Hydraulic Engineering. **139**(7): p. 745-753.

[38] Hashemi, S.S., M. Tabesh, and B. Ataeekia (2014). Ant-colony optimization of pumping schedule to minimize the energy cost using variable-speed pumps in water distribution networks. . Urban Water Journal. **11 (5),** : p. 335–347.

[39] Menke, R., E. Abraham, P. Parpas, and I Stoianov (2015). Approximation of system components for pump scheduling optimisation. Procedia Engineering, 119, 1059–1068. Procedia Engineering. **119**: p. 1059–1068.

[40] Boulos, P.F., Walter M. Grayman, J. Robert W. Bowcock, o.W. Clapp, L.A. Rossman, R.M. Clark, R.A. Deininger, and A.K. Dhingra (1996). Hydraulic mixing and free chlorine in reservoirs. AWWA.

[41] Bumrungthaichaichan, E. and S. Wattananusorn (2019). CFD modelling of pump-around jet mixing tanks: a reliable model for overall mixing time prediction. Journal of the Chinese Institute of Engineers. **42**(5): p. 428-437.

[42] Ganjare, A.V. and A.W. Patwardhan (2019). CFD simulations of single-phase flow in settling tanks: comparison of turbulence models. Indian Chemical Engineer: p. 1-14.

[43] Bernal-Colio, V.R., J. Gómez-Goñi, and J.L. Cercos-Pita (2020). CFD computation of the hydrodynamic torque due to free-surface antiroll tanks with 3D dynamics. Ships and Offshore Structures: p. 1-13.

[44] Alizadeh Fard, M., A. Baruah, and B.D. Barkdoll (2021). CFD modeling of stagnation reduction in drinking water storage tanks through internal piping. Urban Water Journal: p. 1-9.

[45] Martins, N.M.C. and D.I.C. Covas (2022). Induced Circulation by Plunging and Submerged Jets in Circular Water Storage Tanks Using CFD. Water. **14**(8).

[46] Monteiro, L., A. Pinheiro, J. Carneiro, and D. Covas (2021). Characterization of drinking water storage tanks in Portugal. Ingeniería del agua. **25**(1).

[47] Brentan, B., L. Monteiro, J. Carneiro, and D. Covas (2021). Improving Water Age in Distribution Systems by Optimal Valve Operation. Journal of Water Resources Planning and Management. **147**(8).

[48] Monteiro, L., R. Algarvio, and D. Covas (2021). Enhanced Water Age Performance Assessment in Distribution Networks. Water. **13**: p. 2574.

[49] Pinheiro, A., L. Monteiro, J. Carneiro, M. do Céu Almeida, and D. Covas (2021). Water Mixing and Renewal in Circular Cross-Section Storage Tanks as Influenced by Configuration and Operational Conditions. Journal of Hydraulic Engineering. **147**(12).

[50] Monteiro, L., A. Pinheiro, J. Carneiro, and D.I.C. Covas (2021). Characterization of drinking water storage tanks in Portugal (in Portuguese). Ingeniería del agua. **25**(1).

[51] Monteiro, L., J. Carneiro, and D.I.C. Covas (2020). Modelling chlorine wall decay in a full-scale water supply system. Urban Water Journal. **17**(8): p. 754-762.

[52] Pinheiro, A., J. Carneiro, L. Monteiro, M.d.C. Almeida, and D.I.C. Covas (2021). Water Mixing and Renewal in Circular Cross-section Storage Tanks as Influenced by Configuration and Operational Conditions. Journal of Hydraulic Engineering-ASCE.

[53] Tran, T., H. de Maleprade, C. Sun, and D. Lohse (2013). Air entrainment during impact of droplets on liquid surfaces. Journal of Fluid Mechanics. 726.

BY NC SA

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14141

DISCOVERING DIFFERENCES IN IRON AND MANGANESE BEHAVIOUR IN SERVICE RESERVOIRS

Anastasia Doronina[1], Stewart Husband[2], Joby Boxall[3], and Vanessa Speight[4]

[1,2,3,4] University of Sheffield, Department of Civil and Structural Engineering, Mappin St, Sheffield S1 3JD

[1] *anastasia.doronina@pantonmcleod.co.uk*, [2] *s.husband@sheffield.ac.uk*
[3] *j.b.boxall@sheffield.ac.uk*, [4] *v.speight@sheffield.ac.uk*

Abstract

Service reservoirs (SRs) are crucial components in drinking water distribution systems (DWDSs). In the UK, regulatory monitoring is conducted infrequently at SR outlets and for disinfectant residual and bacteriological indicators only, providing limited information on SR performance and its effect on drinking water quality. In this research, long term monitoring at the inlet and outlet of multiple SRs across the UK for various water quality parameters, and analysis of accumulated material collected from SRs during maintenance activities, provided valuable information on iron and manganese behaviour in these assets. Results showed that a lower proportion of iron appears to be retained within SRs in comparison to manganese. These findings challenge the understanding of how water quality risks posed by iron and manganese (e.g., discolouration) manifest in SRs and this knowledge can be used to help inform a targeted approach specific to SRs for the proactive management of those risks.

Keywords
Service, Reservoirs, Distribution, Water, Quality, Iron, Manganese, Monitoring, Deposits, Discolouration.

1 INTRODUCTION

Service reservoirs (SRs) are treated water storage tanks, used to balance water supply and maintain diurnal demand in drinking water distribution systems (DWDSs), making them crucial and integral components of the system. Despite their importance, there is a limited understanding surrounding SR performance and its effect on drinking water quality, in comparison to other stages of the DWDS, such as water treatment works (WTW) and pipes [1]. A key reason for this gap is the current UK regulatory SR monitoring programme, which although considered good practice in comparison to other countries where there is no regulatory requirement for routine sampling at SRs at all, is fundamentally limited [2]. For instance, there is no regulatory requirement to monitor water quality at SR inlets, making it impossible to track water quality changes with travel through these assets, or to attribute any observed changes in water quality directly to the SR [2,3]. Additionally, disinfectant residual and bacteriological indicators are the only regulated parameters monitored at SR outlets, unlike other key indicator water quality parameters (e.g., metals, turbidity, pH, and conductivity), which are regulated at other stages of the DWDS [4]. The lack of inlet sampling and monitoring for a greater range of water quality parameters substantially limits the amount of information that could be obtained on SR performance and on water chemistry and microbiology within SRs, making it difficult to proactively manage these assets [2,5].

All inorganic compounds in the DWDS influence water quality but much of the research focus has been on regulated metals, primarily iron and manganese. Both metals are found in every DWDS, but concentrations vary based on source water type, the efficacy of treatment, and system materials [5]. Upon contact with oxygen or disinfectant, both metals may be oxidised to their soluble states, facilitating accumulation as particulate matter in the DWDS; both have been shown to be the most common metals found in DWDS accumulated material [6,7,8]. Material accumulation occurs throughout DWDSs, especially in low flow zones like SRs where the low velocities in these assets allow material entering storage to settle on interior surfaces [5,9]. The remobilisation of iron and manganese containing material (triggered by changes to normal operating flow regimes) can cause discolouration events, which makes the water appear unattractive and can lead to customer complaints and penalties for water utilities [6,10]. Controlling the accumulation of material in DWDSs and mitigating discolouration is therefore important and water utilities take measures to remove network material through procedures like flushing pipes and cleaning SRs [5,11], typically as a reactive measure [10]. It is only from thorough understanding of entire system processes and mechanisms that lead to discolouration events that water utilities will be able to foresee such incidents and proactively reduce risks [10]. Such comprehensive understanding is not currently possible as, despite the risks associated with the presence of iron and manganese and the knowledge that material accumulation is a continuous process in SRs, metals are not monitored at this stage of the DWDS, nor is SR accumulated material collected and analysed for its inorganic constituents.

This paper reports results from an extensive water quality monitoring programme conducted at the inlet and outlet of multiple SRs from four UK water utilities, and analysis of accumulated material collected from SRs during maintenance activities, on the behaviour of iron and manganese in these critical assets in comparison to the rest of the DWDS. Results provide a first-time insight into the differences in the accumulation and remobilisation behaviour of these metals in SRs, challenging current literature on the composition of DWDS accumulated material and adding to the understanding of discolouration risk posed by iron and manganese at this stage of the network.

2 METHODS

Field site selection

Field study sites were selected from each of the four different UK water utilities (termed A, B, C and D) who were part of a larger study on assessing SR performance and its effect on drinking water quality. SR site selection requirements included: (1) having a consistent source water and treatment; (2) being completely post treatment; (3) either having a pre-existing accessible inlet and outlet sampling line and tap or having the scope for installation; (4) having separate inlet and outlet pipes; and (5) having scope to carry out downstream/upstream sampling in the surrounding DWDS.

Water quality monitoring programme

As inlet monitoring at SRs is not a regulatory requirement for water utilities in the UK, there were no available functioning inlet sampling points at the four project SRs. To enable investigation of SR performance using incoming and outgoing water quality, inlet sample lines and taps were either installed or recommissioned in accordance with UK Water Supply Regulation 8 and made from materials complying with BS 6920 [12,13]. A water quality monitoring programme was then undertaken at both the inlet and outlet of project SRs for a range of water quality parameters including total iron and manganese, conducted on a weekly basis across 2018-2020. All samples were collected and analysed in accordance with standard procedures for regulatory compliance monitoring ISO/IEC 17025, ISO/IEC 17024, and the Drinking Water Testing Specification, accredited by the United Kingdom Accreditation Service [14,15].

Water quality monitoring results presented in this paper are for SR D4; an underground, rectangular, twin-compartment, 9.1 ML reinforced concrete (with a brick compartment-dividing wall) SR with an average retention time of 28 hours. It is fed by a 0.22 km ductile iron cement-lined trunk main directly from the WTW. Treated water is reservoir abstracted with free chlorine as the secondary disinfectant and further UV dosing at the inlet of the SR. Table 1 below presents summary site information for D4 and other studied SRs discussed in this paper.

Collection and analysis of SR accumulated material samples

SR accumulated material was collected from five SRs (B6, C7, C8, C9, and D4; note, D4 was also monitored for water quality as described above) during scheduled structural inspections and cleans. The criteria for site selection was to collect material from a variety of different SRs with respect to geometry, retention time, distance from WTW, and secondary disinfectant type (see Table 1 below for summary site details). As no standardised existing procedure was found, material was collected using a protocol developed in this research, built on experience from SR cleaning operations and based upon elements of existing literature on material collection from SRs for microbiological analysis and material collected as loose deposits from raw waterbodies [16,17,18]. Sampling guidelines stipulated that material needed to be collected in a way that represented the entire floor area of the SR. This included, where feasible, collection from each corner, the middle, and around any other features such as inlet and outlet pipes. All samples were collected immediately after the SRs were drained and before any further work or cleaning had commenced, except in one case where cleaning was inadvertently started prior to collection of one sample (C7 SR - middle of compartment 2). Samples were collected using sterile plastic spatulas, kept in sterile plastic sample bottles, and refrigerated as soon as possible using cooler boxes and fridges. Material was then analysed for its organic and inorganic composition in a set of two experiments, one investigating the effect of material resuspension on the surrounding water quality (material from SRs B6 and D4) and the other on the influence of sampling location within SRs on material composition (material from SRs C7, C8, and C9).

For the resuspension experiment, different masses of dry material were placed into deionised water at concentrations of 60 mg/l (low), 120 mg/l (medium), and 240 mg/l (high). The bottles were placed on a shaker table and shaken for a period of 48 hours. The table was stopped, and samples collected at 2 and 48 hours. The shaking process was done to recreate the effect of immediate resuspension of material for the 2-hour samples (e.g., following a sudden event like a pipe burst) and prolonged resuspension for the 48-hour samples (e.g., from a more permanent change in hydraulic operation of a SR). 6 of the 18 samples were removed from the table entirely at the 24-hour mark and left to settle for an additional 24 hours, after which a sample for analysis was taken. The purpose of removing the samples from the shaker table was to observe whether the re-settling of material begins to reverse any potential effects that resuspension might have on the surrounding water quality.

Samples were analysed for a range of metals and elements, focussing on those often found in DWDSs including aluminium, arsenic, cadmium, chromium, cobalt, copper, iron, lead, magnesium, manganese, nickel, potassium, uranium, and zinc. The analysis was done using an Agilent 7900 ICP-MS with Ultra High Matrix Introduction following manufacturers' guidelines [19]. One sample, 'C7 SR - middle of compartment 2', required a five-factor dilution (using deionised water) as the total organic carbon content of the sample was over 20 mg/l and could have interfered with the ICP-MS.

Table 1. – Summary information for characteristics of project SRs.

Service reservoir name	Geometry and Material	Volume (ML)	Height (m)	Average retention time (hours)	Inlet pipe (direct upstream main) length, diameter (inches), and material	Source water	Secondary and on-site disinfectant (where applicable)
D4	Underground Rectangular Twin compartment Reinforced concrete with a brick dividing wall	9.1	5.4	28.1	Length: 0.223 km 23.6” DI cement lined	Surface (reservoir) Soft water	Secondary: Chlorine On site: Ultraviolet irradiation (dosed at the inlet)
B6	Underground Rectangular Twin compartment Reinforced concrete	9.4	4.8	NA	Length: ~52 km 22.1” HPPE down to 17.7” DI	Surface (river and reservoir) Slightly hard water	Secondary: Chlorine
C7	Underground Twin compartment Rectangular Reinforced concrete	5.2	6.7	63.5	Length: 0.51 km 15.7” DI	Surface (reservoir) Soft water	Secondary: Chlorine
C8	Underground Twin compartment Rectangular Reinforced concrete	0.8	4.8	82.4	Length: 1.1 km 12” SI	Surface (reservoir) Soft water	Secondary: Chloramine
C9	Underground (part exposed) One compartment Circular Concrete	0.4	3.3	71.6	Length: primary main from WTW 11.5 km, but immediate main tapping off this is 0.02 km 9” primary main and 6” immediate main, both AC	Surface (reservoir) Soft water	Secondary: Chloramine

3 RESULTS

Total iron and manganese in bulk water pre and post SR clean

D4 SR was inspected and cleaned during the water quality monitoring programme conducted in this project. A variety of water quality parameters were monitored pre, during, and post clean, enabling the investigation on the efficacy of SR maintenance on asset performance. Figures 1 and 2 show the water quality monitoring results for SR D4 over 4 time periods: (1) prior to the clean; (2) when the first compartment was taken out of service, so the entire flow was travelling through a single, uncleaned compartment; (3) when the second compartment was taken out of service, so the entire flow was travelling through a single, cleaned compartment; and (4) after both compartments had been cleaned. When looking at discrete sampling data for total iron and manganese, results show that iron concentration does not change significantly at the SR outlet pre, during, or after the clean, suggesting that the clean did not affect SR performance with respect to iron levels (Figure 1). Results for manganese sampling show that pre clean manganese levels were slightly higher at the SR outlet in comparison to SR inlet and that upon the commencement of the clean, and post clean, manganese levels became lower at the SR outlet in comparison to SR inlet, irrespective of the concentration of incoming manganese (Figure 2). Based on these results, it appears that cleaning the D4 SR improved its performance in terms of it removing, instead of contributing manganese, into the water supply. With a reduction of manganese entering the downstream network, it can be proposed that this reduces the rate of accumulation on pipe surfaces and thus alleviates future mobilisation and hence reduces discolouration risk posed by this metal.

BY NC SA

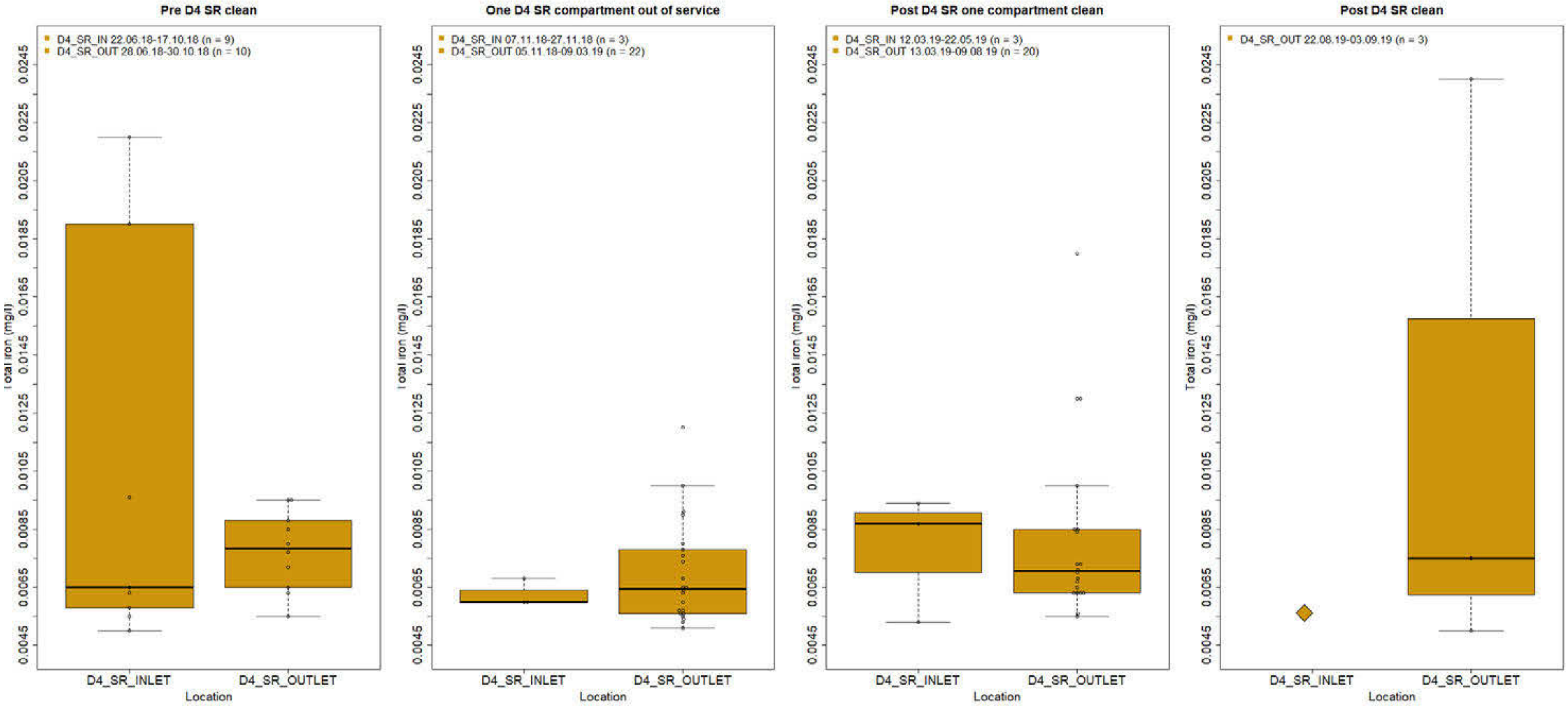


Figure 1. Discrete total iron concentrations (mg/l) at the D4 SR across four time periods; pre SR clean, with one compartment out of service, post one compartment clean, and post the entire SR clean. In the fourth time period, all SR inlet values (n = 25) were below the minimum detection limit of 0.011 mg/l and are represented here as limit of detection/2 (0.0055 mg/l).

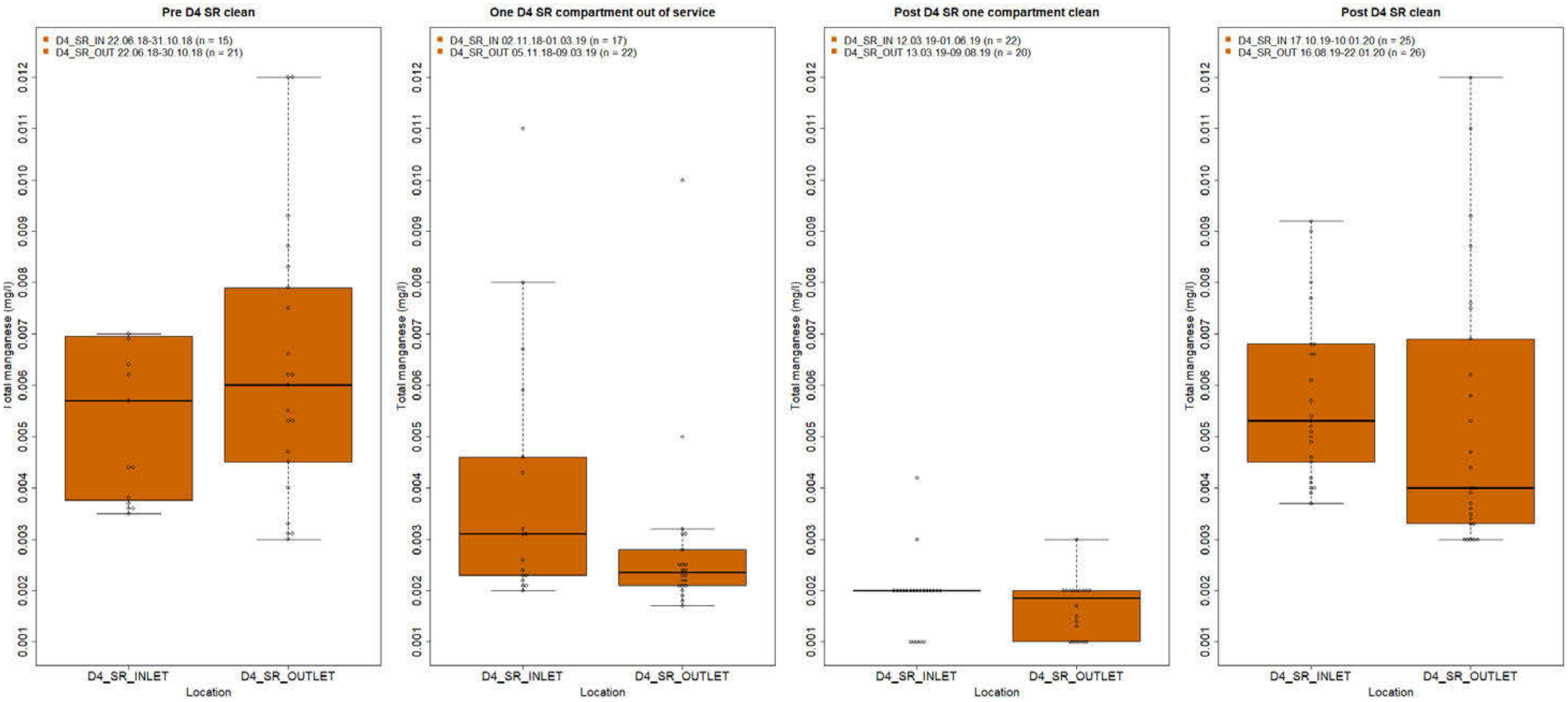


Figure 2. Discrete total manganese concentrations (mg/l) at the D4 SR across four time periods; pre SR clean, with one compartment out of service, post one compartment clean, and post the entire SR clean.

cc BY NC SA

Composition of metals in SR accumulated material

Figure 3 provides the results of the resuspension experiments for D4 and B6 SR material samples. Results are shown for the different concentrations of material (low 60 mg/l, medium 120 mg/l, and high 240 mg/l) and different times in resuspension (2 hours, 48 hours, and settled). Total aluminium was also present in all D4 and B6 SR samples but is not presented in Figure 3; values ranged from 0.0038-0.0526 mg/l in samples from D4 SR and 0.0024-0.0053 mg/l in samples from B6 SR.

Figure 4 presents the dominant metals identified in the accumulated material collected from C7, C8, and C9 SRs. Results from sample 'C7 SR - middle of compartment 2' were much higher than in material samples from other SRs. This was also the only sample where arsenic (0.0010 mg/l), lead (0.0007 mg/l), and uranium (0.0001 mg/l) were identified.

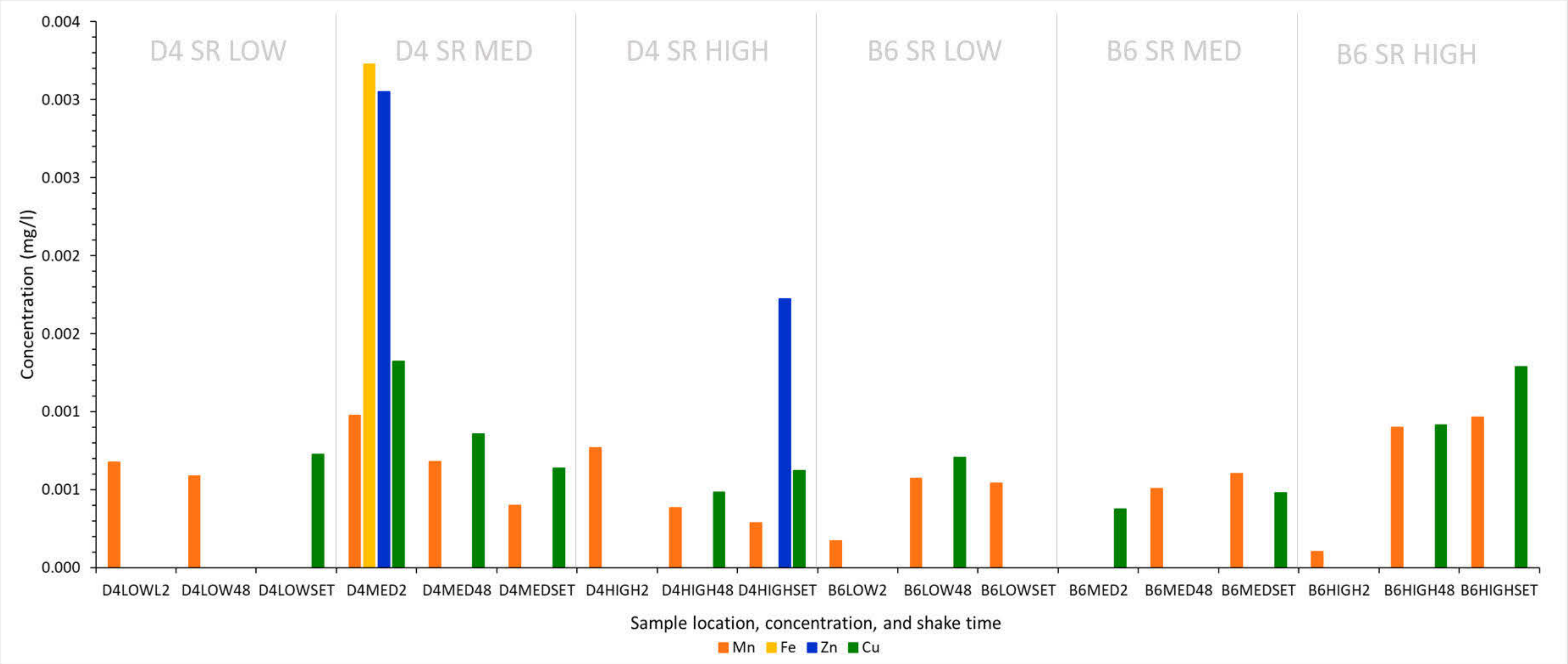


Figure 3. Total metals (mg/l) in resuspension analysis of material collected from D4 and B6 SRs.

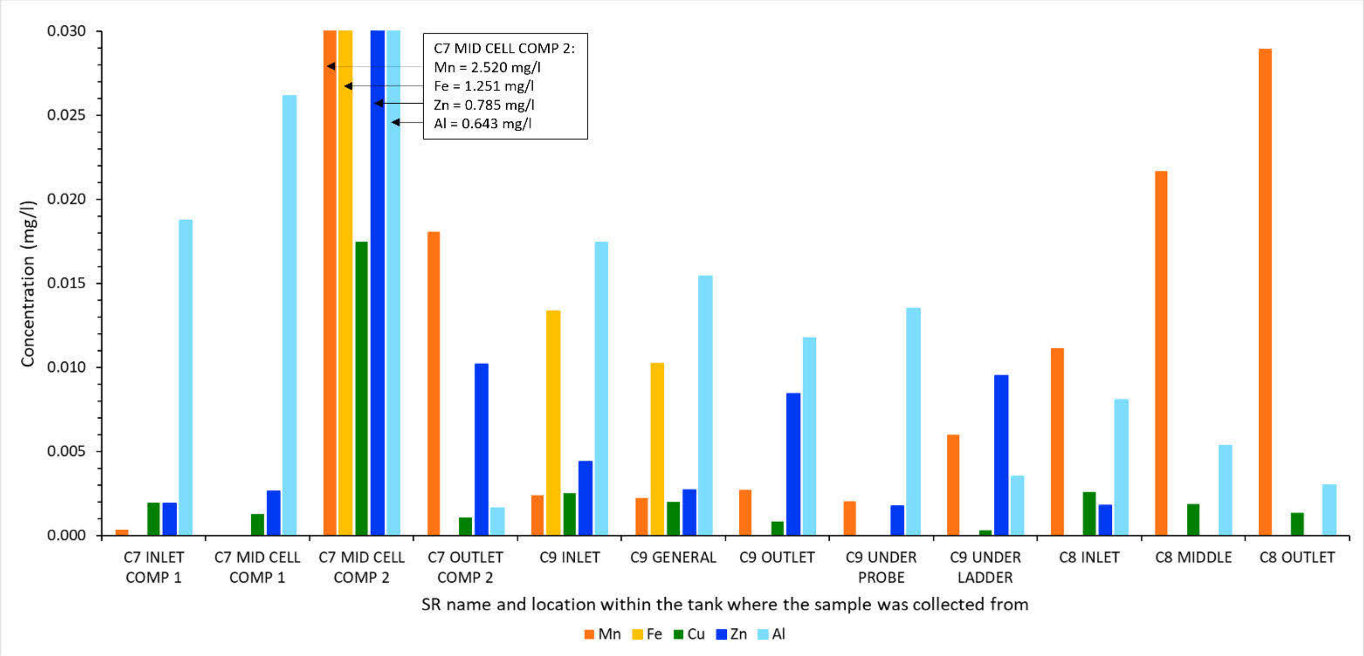


Figure 4. Total metals (mg/l) in material samples collected from different locations within C7, C8, and C9 SRs.

Results from Figures 3 and 4 show that aluminium was the most dominant metal, present in 100% of the samples, followed by manganese in 90%, copper in 73%, zinc in 40%, and iron in just 13% of the samples. The results also demonstrate that the accumulated material is not homogenous, with variation in results from different locations and from different subsamples within a single material sample. For example, different metals were detected in the different subsample concentrations prepared from the same larger material sample in the resuspension experiments (Figure 3). An exception was the sample collected from the middle of compartment 2 in C7 SR, an outlier sample where iron was present at levels that breached UK prescribed concentration value (PCV) in drinking water, set at 0.2 mg/l. Manganese concentration in the sample also breached UK PCV at 50 times the set value of 0.05 mg/l. This was also the only material sample that contained arsenic, uranium, and lead. Although not confirmed, it is suspected that this sample was exposed to a biocide solution, used as part of the cleaning process of SRs, prior to collection. This solution is used to strip biofilm and other accumulated materials off concrete and metal structures and in SRs is sprayed onto walls and internal structures [20]. It is likely that this particular sample was impacted by the removal of a large area of biofilms and the resulting effluent ending up in the collected sample, which could explain these large concentrations and the detectable presence of arsenic, uranium, and lead. These results suggest that some of the metals and elements that enter SRs may accumulate in small concentrations in biofilms and are not present in their loose, particulate forms in accumulated material, e.g., iron.

4 DISCUSSION

The aim of this research was to investigate how SRs affect drinking water quality as water travels through these essential assets. From the analysis of SR accumulated material, it was observed that iron was absent in the majority (87%) of collected material samples (Figures 3 and 4). This finding contradicts much of the current literature, which finds that iron is the predominant metal in the DWDS, both in bulk water and in accumulated material, but draws primarily on samples from pipes for these conclusions [6,7,8]. Results from the D4 SR clean showed that iron does not appear to remain within the SR in comparison to manganese, given that cleaning D4 SR made no difference to outlet iron levels but did for manganese (Figures 1 and 2). The lack of iron accumulation in D4 SR aligns with the results from the analysis of the collected material sample from this SR, which contained no iron in 89% of the subsamples while manganese was present in the majority (89%) (Figure 3).

Another example where SRs appear to be contributing manganese into the downstream network, but not iron, comes from higher manganese levels observed at SR outlets during the pandemic. In the UK, water utilities started using SR outlets as alternative sampling points to customer taps for regulatory monitoring for metals (44% samples taken at surrogate locations in England in 2020) [4]. This SR sampling resulted in fewer sample failures for iron, but more sample failures for manganese.

All of these results suggest that there are differences in how water quality parameters behave in SRs as compared to pipes. The work from this study suggests that iron potentially settles and remobilises more in pipes than SRs, and manganese settles and remobilises more in SRs than pipes. A similar observation highlighting different environments and consumption rates between pipes and SRs was made by Pick et al. (2021) [21], who found that the concentration of assimilable organic carbon decayed through pipes but increased slightly with travel through SRs. In pipes, both manganese and iron are known to accumulate in cohesive layers on pipe walls, assimilated into biofilms [22]. Potentially, the same thing is happening to iron in SRs, where small amounts are likely accumulating in biofilms on SR walls and internal structures over time, but not large amounts as loose deposits. The metals composition in the sample 'C7 SR - middle of compartment 2' that was likely exposed to biofilm-sloughing biocide, where not only was iron (and other metals) present, but concentrations also breached PCV (Figure 4) gives an indication of this iron accumulation at the SR wall. Internal corrosion of iron-based materials such as steel ladders within the SRs, was also observed to be taking place and may be contributing larger amounts of particulate iron to localised areas.

The findings in this study on iron and manganese behaviour in SRs are preliminary and highlight the need for further research into SR material accumulation and water quality changes at this stage of the network, for proactive management of associated risks.

Another valuable outcome of this research came from monitoring SR performance pre and post clean, which made it possible to, for the first time, measure and evaluate the effectiveness of SR maintenance on asset performance in terms of water quality and which demonstrated that the removal of material from SRs can reduce discolouration risk.

The above observations would not have been possible to make without the inclusion of inlet monitoring or monitoring parameters other than disinfectant residual and bacteriological indicators at SRs, or the collection and analysis of SR accumulated material. This highlights the importance and value of a more inclusive monitoring programme at this currently overlooked stage of the network.

5 CONCLUSIONS

- Inlet and outlet water quality monitoring at SRs indicates a difference in the behaviour of iron and manganese in that manganese appears to remain within the SR more than iron.
- Analysis of SR accumulated material shows high accumulation of manganese, but generally not iron in SRs.
- Conducting both inlet and outlet monitoring at SRs and for water quality parameters other than those stipulated by regulation enables the ability to assess SR performance and evidence the effectiveness of asset maintenance.

6 ACKNOWLEDGEMENTS

The authors extend their thanks to the participating water utilities on the project, namely Anglian Water, Dwr Cymru Welsh Water, Northumbrian Water, and Scottish Water, for their input and support, both financial and otherwise.

Gratitude is also extended to the staff at University of Colorado, Boulder, especially to Professor Fernando Rosario-Ortiz and Assistant Professor Julie Korak for their accommodation and assistance with the analysis of SR accumulated material samples.

Finally, a big thank you to Panton McLeod for making the collection of SR accumulated material samples possible.

7 FUNDING

This work was part funded by the Engineering and Physical Sciences Research Council (EPSRC) [National Productivity Fund Scholarship EP/R512175/1] and Wat Qual [European Commission Project 77813].

8 REFERENCES

[1] NRC. 2006. Drinking water distribution systems: assessing and reducing risks. Washington, DC: National Academy Press.

[2] Kirmeyer, G. J., Kirby, R., Murphy, B. M., Noran, P. F., Martel, K. D., Lund, T. W., Anderson, J. L., Medhurst, R. 1999. Maintaining Water Quality in Finished Water Storage Facilities. Denver, CO: AwwaRF and AWWA.

[3] Grayman, W. M., Rossman, L. A., Arnold, C., Deininger, R. A., Smith, C., Smith, J., Schnipke, R. 1999. Water quality modeling of distribution system storage facilities. Denver, CO: AwwaRF and AWWA.

[4] DWI. 2020. Summary of the Chief Inspector's report for drinking water in England. A report by the Chief Inspector of drinking water. London: Drinking Water Inspectorate.

[5] Brandt, M. J., Johnson, K. M., Elphinston, A. J., Ratnayaka, D. D. 2016. Twort's Water Supply (7th ed.). Oxford: Elsevier Science, Butterworth-Heinemann.

[6] Seth, A., Bachmann, R., Boxall, J., Saul, A., Edyvean, R. 2004. Characterisation of materials causing discolouration in potable water systems. Water Science and Technology 49(2):27-32.

[7] Gauthier, V., Rosin, C., Mathieu, L., Portal, J. M., Block, J. C., Chaix, P., Gatel, D., 1996. Characterization of the deposits in drinking water distribution systems. In: Proceedings of the WQTC of AWWA. Boston, Ma, USA.

[8] Lytle, D.A., Sorg, T., Wang, L., Chen, A. 2012. The accumulation of radioactive contaminants in drinking water distribution systems. Water Research 50:396–407.

[9] Grayman, W.M., Kirmeyer, G.J. 2000. Quality of water in storage. Water distribution systems handbook, 11-1. New York: McGraw-Hill.

[10] Vreeburg, J. H. G., Boxall, D. J. B. 2007. Discolouration in potable water distribution systems: A review. Water Research 41(3):519–529

[11] Husband, S. P., Boxall, J. B. 2016. Understanding and managing discolouration risk in trunk mains. Water Research. 107:127–140.

[12] DWI. 2012. Guidance on the implementation of the water supply (water quality) regulations 2000 (as amended) in England. Part 4 – Monitoring of Water Supplies, Regulation 8. London: Drinking Water Inspectorate. Last accessed 18/04/22. http://dwi.defra.gov.uk/stakeholders/guidance-and-codes-of-practice/WS(WQ)-regs-england2010.pdf

[13] WRAS. 2021. BS 6920:2014. Last accessed 18/04/22. https://www.waterregsuk.co.uk/guidance/legislation/glossary/bs6920/

[14] United Kingdom. 2018. The Water Supply (Water Quality) Regulations. 2018. No. 647 (W. 121). Stationery Office, London, UK.

[15] DWI. 2020. Guidance on the implementation of the water supply (water quality) regulations 2016 (as amended) in England and the water supply (water quality) regulations (Wales) 2018. https://www.dwi.gov.uk/water-companies/guidance-and-codes-of-practice/guidance-on-implementing-the-water-supply-water-quality-regulations/

[16] Lu, J., Struewing, I., Yelton, S., Ashbolt, N. 2015. Molecular survey of occurrence and quantity of Legionella spp., Mycobacterium spp., Pseudomonas aeruginosa and amoeba hosts in municipal drinking water storage tank sediments. Journal of Applied Microbiology 199:278–288.

[17] Qin, K., Struewing, I., Domingo, J.S., Lytle, D., Lu, J. 2017. Opportunistic Pathogens and Microbial Communities and Their Associations with Sediment Physical Parameters in Drinking Water Storage Tank Sediments. Pathogens 6(4):54.

[18] Peck, M., Gibson, R.W., Kortenkamp, A., Hill, E.M. 2004. Sediments are major sinks of steroidal estrogens in two United Kingdom rivers. Environmental Toxicology and Chemistry: An International Journal 23(4):945-952.

[19] Agilent. 2021. Agilent 7900 ICP-MS with Ultra High Matrix Introduction (UHMI). Last accessed 18/04/22 https://www.agilent.com/en/products/icp-ms/icp-ms-systems/7900-icp-ms

[20] Stainsby, F. M. 2014. Antibacterial testing to determine the efficacy of a Panton McLeod proprietary disinfectant. Report prepared for: Panton McLeod Limited. Scotland, Edinburgh: Edinburgh Napier University.

[21] Pick, F.C., Fish, K.E., Boxall, J.B. 2021. Assimilable organic carbon cycling within drinking water distribution systems. Water Research 198:117147.

[22] Husband, P. S., Boxall, J. B. 2010. Field studies of discoloration in water distribution systems: model verification and practical implications. Journal of Environmental Engineering 136(1)86–94.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14142

FLOW DYNAMICS IN A PIPE CONTAINING UNEVEN ROUGHNESS ELEMENTS

Katrin Kaur[1], Anatoli Vassiljev[2], Ivar Annus[3], Murel Truu[4] and Nils Kändler[5]

[1,2,3,4,5]Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn, Estonia

[1] *katrin.kaur@taltech.ee,* [2] *anatoli.vassiljev@taltech.ee,* [3] *ivar.annus@taltech.ee,*
[4] *murel.truu@taltech.ee,* [5] *nils.kandler@taltech.ee*

Abstract

The objective of this study is to improve the methodology for modelling aged pressurised urban networks, such as water distribution system (WDS) pipelines, with substantial and uneven wall roughness. Due to the complexity of the inner surface geometry of old rough pipes, WDS models are unable to predict the real velocities, which are requisite for estimating contamination propagation rates in the system. In order to analyse the impact of inner surface geometry a specifically designed experimental apparatus was built to investigate the flow dynamics in a pipe containing flow-obstructing elements. A transparent pipe was fitted with 3D-printed elements, in two different configurations, at the location of the measurement section and the velocity fields were captured using the particle image velocimetry technique at different flowrates. The experimental results revealed that in the case of one roughness element, the impact to the flow dynamics is local and can be expressed through minor losses, and in the case of multiple elements, the effective diameter of the pipe is reduced and a jet-type flow will develop.

Keywords
Particle image velocimetry, pipe wall build-up, roughness, water distribution system modelling.

1 INTRODUCTION

Degradation of conduits over time results in uneven internal boundaries and concurrently, deviation from design operating conditions. Examination of removed old pipe sections have revealed that the roughness height as well as the shape of the inner pipe surface may vary significantly, resulting in complex geometries. Furthermore, the broadly changing roughness height and distribution will change the pipe effective diameter and flow velocity. Therefore, the determination of real flow velocities in old rough pipes remains a challenge, as even with extensive data collection, it is implausible to determine the pipe roughness values for all links. Among other things, this problem is related to modelling of existing urban water drainage and distribution systems.

Pipes that were originally manufactured with smooth internal surfaces can over time develop irregular wall roughness elements that greatly complicate the flow dynamics [1]. The variation in the geometry of pipe interior due to wall build-up is not only time dependent but is affected by several factors, such as the pipe initial cross-section and material, the water quality, and flow velocity [2]–[5]. Various studies have analysed the growth of the pipe roughness over time, concurrence with flow velocities and quality parameters. [6] reported that the roughness growth depends primarily on the water pH and is linear, with rates ranging from 0.066 to 0.63 mm/yr. The range is in compliance with numerous investigations conducted in the US [7], and results reported by [8]; however, it is significantly lower than the growth rate reported by [9] for galvanized iron pipes (2.13 mm/yr.). The latter are concurrently less often used in WDSs compared to cast iron. A study carried out in Estonia, analysing 25-100 years old metal pipes (steel and cast iron) with diameters ranging from 75 to 200 mm revealed that the pipe cross-sections were reduced in average by 10% due to wall build-up [2].

Corrosion-induced hydraulic capacity failures due to the pipe wall build-up may prevent WDS from constantly ensuring customer water needs with satisfying quantity and quality [3]. The side effect of corrosion and low velocity can be discolouration of the water, which has been the main customer complaint about the water quality in the UK [4]. The formation and growth of particles in the WDS has diverse driving factors but the sedimentation of the particles is related to the hydraulic conditions of the network. It was shown that at low velocities, the sedimentation will take place in the lower half of the horizontal pipe while at higher velocities, the deposit will cover the entire pipe wall [4]. [5] suggested that the daily peak demands in combination with maximum flow velocity of at least 0.4 m/s can be used to prevent sediment accumulation and concurrent pipe wall build-up.

Data on nominal pipe diameters is generally used in the WDS model initialisation and the pipe wall build-up-induced change in the diameter is compensated by adjusting the pipe roughness [10] or the water consumption at the nodes [11], [12] in the calibration process. Adjusting the consumption is only applicable if it is not measured. Further, in the case of uneven changes in the internal boundary, the practice of adjusting roughness values can lead to estimated unrealistic values larger than pipe radius. Therefore, the latter approach is only suitable for surfaces with easily described geometry [13]. Contrarily, in real WDSs containing old pipes, the inner wall surface will vary significantly, producing complex geometries, and the common practice of using nominal diameters while adjusting the roughness, will result in relatively large errors in flow velocities. Information about real pipe diameters and flow velocities is of utmost importance in estimating the propagation rate of the contaminated zones in WDS in case of chemical or biological threats [14], [15].

Various studies have analysed the flow dynamics in pipes with regular roughness elements, with many of these focusing the main discussion on numerical investigations. [16], [17] investigated the turbulent flow in ribbed pipes and showed that the influence of vortical structures between the ribs on the core flow is dependent on the distance between the ribs. [18], [19] analysed the turbulent flow in corrugated pipes and concluded that the friction factor increases as both the Reynolds number and the groove length increase, and is not affected by the groove height. [20] conducted a study in corrugated pipes, and concluded that contrary to the known formulations of corrugated pipe friction factor, even if the groove width is greater than groove depth, a recirculation zone may occur.

Real old pipe surfaces generally have a range of roughness scales. [1] conducted a series of numerical investigations to model the flow through aged pipes at Reynolds numbers ranging from 6700 to 31 000. They found that the computational fluid dynamics (CFD) models underestimate the friction factors by 8-30% (dependent on Reynolds number and model used). Therefore, improved models, further validated for complex geometries are of high interest. [21] conducted a numerical and experimental investigation on flow in a pipe with an abrupt change in diameter. They concluded that the CFD model with standard k-epsilon turbulence allows to predict the flow dynamics in pipes with complex geometry if moderate changes to the turbulence closure coefficients are made and all geometrical elements that affect the flow are included in the model. [22] investigated the influence of irregular pipe wall roughness numerically and grouped the different types of roughness distributions common to old pipes. [23] investigated the different combinations of roughness elements further and proposed equations for velocity correction for the different wall build-up types based on the numerical study.

The ageing process of pipes should be described in the models not only by the increase of the absolute roughness but also by the reduction of their inner diameter in order to predict the actual flow behaviour more accurately. In addition, in water quality modelling the contamination propagation and dispersal are strongly coupled with the velocity and turbulence characteristics. Therefore, accurate models capable to simulate the pressurised flow in complex pipelines are crucial for understanding the limits, capacity and need for system's rehabilitation in a more general way. Currently, the modelling approach is also more and more frequently used for different real-time control and management systems and therefore the inaccurate presentation of flow dynamics will potentially hinder the safe operation of the water supply and drainage systems.

Therefore, applying evidence-based assumptions and advanced modelling to define the internal boundaries for the deteriorated pipeline sections is becoming more and more important as we are firstly, entering an era where the average life span of pipes in operation is running out [23] and secondly, more and more real-time control solutions are used to operate WDS. Many of the currently available studies relying on thorough analysis of the modelled velocity field, report validating the models by integral parameters, such as pressure at the in- and outflow of the section of interest. Therefore, more detailed experimental data is needed for the advanced numerical modelling to be practicable.

Herein, the effect of irregular pipe wall build-up to flow dynamics in old rough pipes is investigated experimentally. The aim of the study is to create a comprehensive database of different flow obstruction types and corresponding flow behaviours, and complemented by advanced numerical modelling, map those against different deviations from design operational conditions appearing in real water networks' datasets and give guidelines for the modelling of aged pipelines.

2 EXPERIMENTAL SET-UP

A pipeline apparatus was set up at the Laboratory of Fluid Mechanics, Tallinn University of Technology to investigate the flow dynamics in a pipe containing irregular roughness elements. The experimental investigations were conducted on a transparent pipeline containing different configurations of obstructing roughness elements in the flow domain. The velocity field in the vicinity of the elements was mapped using particle image velocimetry (PIV). The experiments were conducted at different flow rates (Table 1) to characterise the changes of effective flow section in the obstructed conduit.

Table 1. List of experiment. Uncertainty of flow rate measurement reading ±0.5%

Experiment number	Flow rate (Q, l/s)	Reynolds number of the full pipe cross-section ($D = 80$ mm)
1	3.0	47 900
2	3.8	59 800
3	4.4	69 800
4	5.0	79 300
5	5.6	89 300

The layout of the experimental apparatus is shown on Figure 1. It consists of a tank at the upstream end, a horizontal transparent polymethyl methacrylate pipe with a total length L = 18 m and an internal diameter D = 80 mm, and a PIV measurement section. The upstream tank is fed from, and the pipeline emptied into a 150 cubic meter underground reservoir. The pipeline is

CC BY NC SA

assembled of 2 m long pipes, hydraulically smooth and the outlet is open to atmosphere. Shut-off valves are situated at the up- and downstream end of the pipeline, adjacent to electromagnetic flow meters. The two flow meters are simultaneously used at the experiment initiation stage for the purpose of ensuring that all air is expulsed from the flow domain. The pipeline is pressure driven and the flow rate is controlled by the frequency of the upstream-end pump feeding the tank from the reservoir.

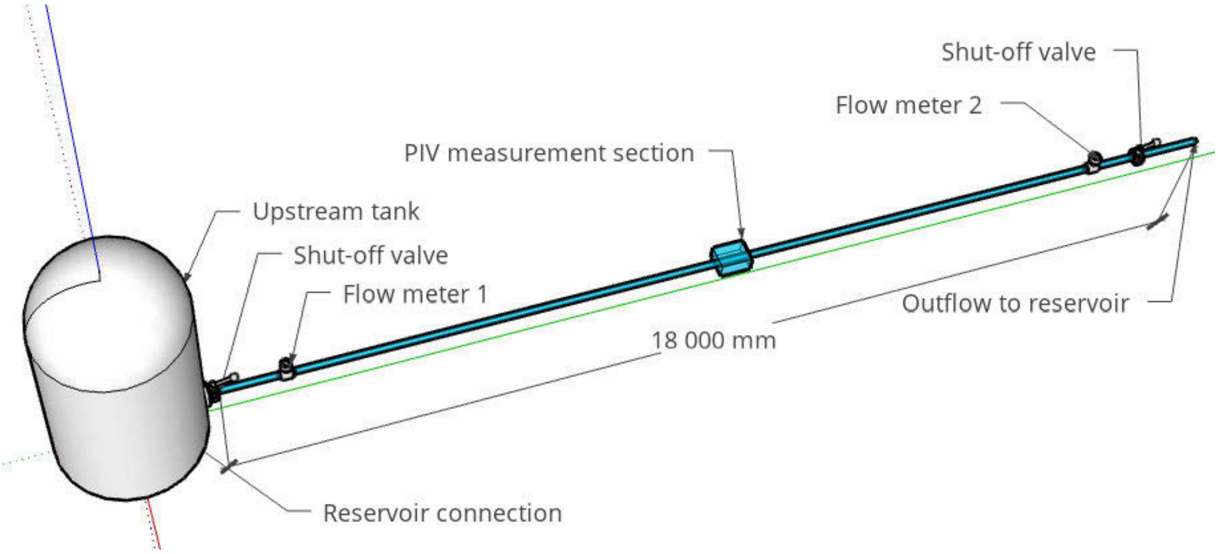


Figure 1. Schematic of the experimental apparatus

A random pattern for an irregular pipe wall build-up element (Figure 2) was created in a numerical study by [22] and obstructing the computational pipe flow domain with different numbers and combinations of these elements was further investigated by [24].

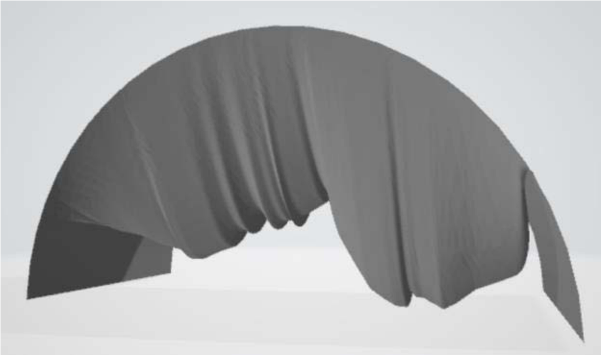


Figure 2. Irregular roughness element used in the experimental study

For this study, a set of the elements were 3D printed and attached inside the pipe of the experimental apparatus. Two configurations of the roughness elements were used for the velocity field measurements at five different flow rates. Firstly, one element was attached to the pipe obvert (Figure 3). Secondly, four elements were inserted into the pipe, two at the obvert and two at the invert, equidistantly (Figure 4).

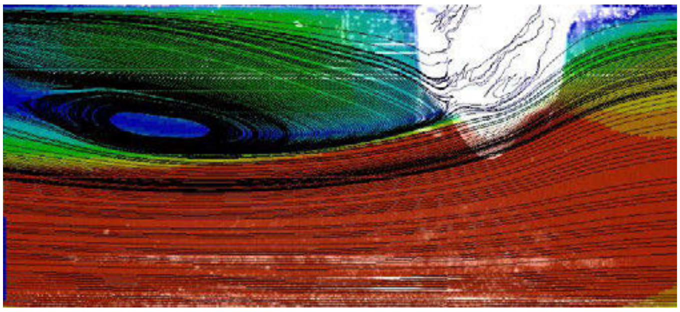


Figure 3. Central plane in the 80 mm pipe with one roughness element located at the pipe obvert. The visual comprises of the image mean, subtracted from the later analyses, streamlines in black, and a scalar map coloured by the axial velocity U. The flow direction is from right to left and a recirculation area occurs behind the element. Red corresponds to highest velocity, blue the lowest.

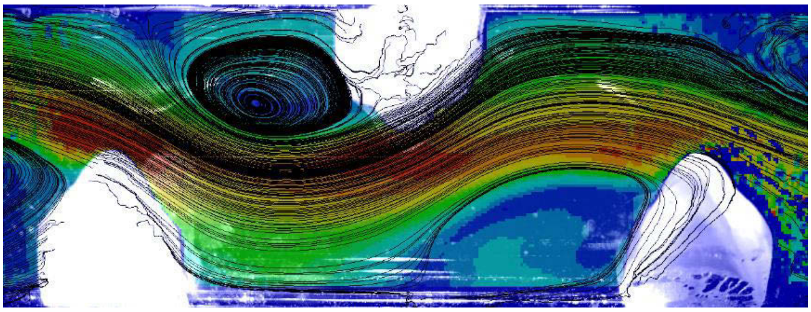


Figure 4. Central plane in the 80 mm pipe with four roughness elements in the investigated area. Three of the elements are in the view range and one is on the right-hand side affecting the inflow into the measurement section. The visual comprises of the image mean, subtracted from the later analyses, streamlines in black, and a scalar map coloured by the axial velocity U. The flow direction is from right to left and recirculation areas occur behind the elements. Red corresponds to highest velocity, blue the lowest.

Changes in the velocity field at the location of interest are measured in the streamwise central plane with two-dimensional PIV. The measurements were carried out using multi-frame single-exposure PIV with a high-speed camera for image capturing, and a continuous-wave laser for creating the light-sheet. The trigger frequency of the camera was adjusted 1000 Hz and the steady state flow was measured for 3 s, which corresponded to 3000 images per experimental run. The data was recorded and processed by Dynamic Studio 4.0 (Dantec Dynamics) software. The cross-correlation method was used to process the PIV data. As a result, velocity vector fields were calculated considering the displacement of the particles between the two frames. An interrogation window of 8 x 8 pixels with 50% overlap was used (1 pix=0.18 mm). The images were pre-processed by subtracting the image mean, getting rid of some minor reflections and the roughness element images in the vector maps. The calibration of the PIV was conducted by capturing a reference images series with rulers inserted into the measurement section. Any distortion in the capturing was avoided by encasing the pipe in a water-filled Plexiglas PIV box and mounting the camera in a horizontal position. Velocities, standard deviations, and variances were calculated using only valid vectors (i.e., gained from measured particle displacements not approximated by the software's algorithm). The average valid vector count for each data point was 1500.

3 RESULTS AND DISCUSSION

The velocity field mapping at different flow rates in the pipe section with one roughness element at the pipe obvert demonstrated how the local disturbance changes the flow dynamics. Namely, how a recirculation area forms and changes in scale with the changing flowrate, indicating a trend in the effective flow area development. The main purpose of these measurements is the creation of a comprehensive dataset for numerical model validation. Figure 5 and Figure 6 present the mapped velocity fields for two experiments.

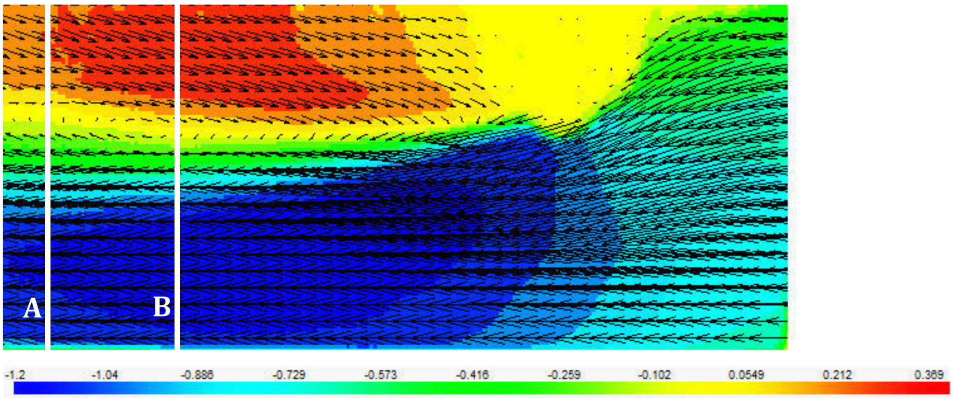


Figure 5. Vector and scalar map. Scalar map coloured by axial velocity U. The scale range -1.2 ... 0.4 m/s. Flow in negative x direction. Flowrate Q=3.0 l/s

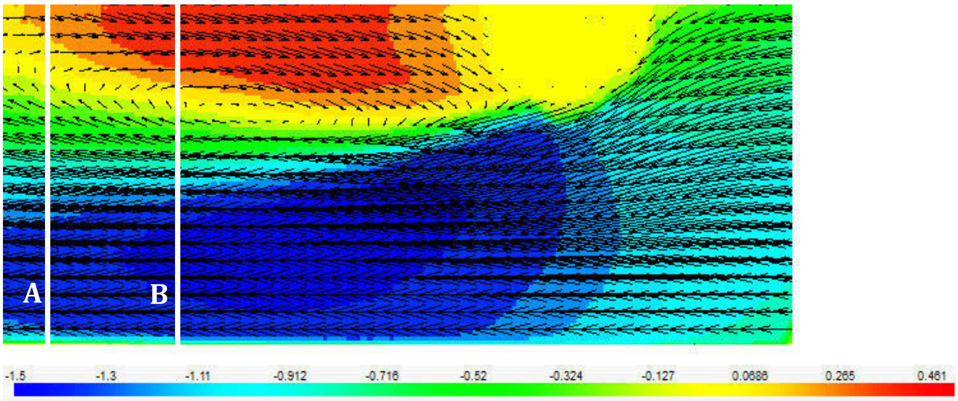


Figure 6. Vector and scalar map. Scalar map coloured by axial velocity U. The scale range -1.5 ... 0.5 m/s. Flow in negative x direction. Flowrate Q=3.8 l/s

Plotting the axial velocity at different sections *A* and *B* (Figure 7 and Figure 8) behind the obstruction reveals that at the nearer location, the higher flowrate yields a larger recirculation area. This agrees in general with the modelling results presented in [22], [24] (Figure 9), where roughness elements of the same type were investigated numerically. It must be noted that the numerical simulations were conducted with more elements in the pipe and at different pipe diameters. Therefore, it can be concluded that the trend is recurring.

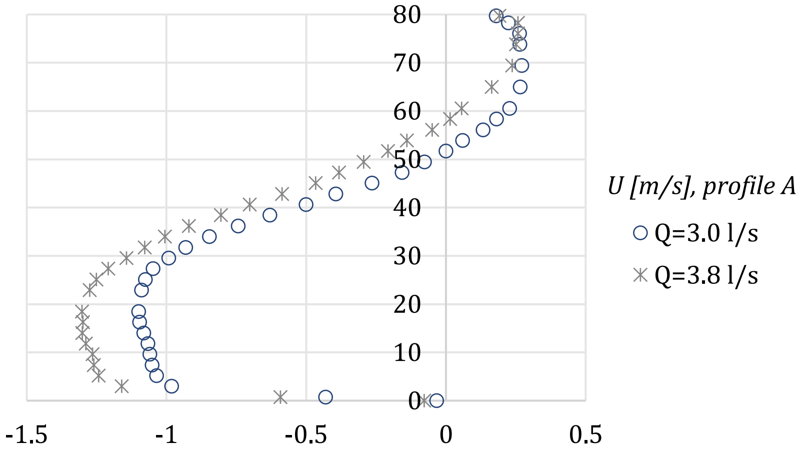


Figure 7. Axial velocity U [m/s] at two flowrates, at location A, in the 80 mm pipe (see Figure 5 and Figure 6)

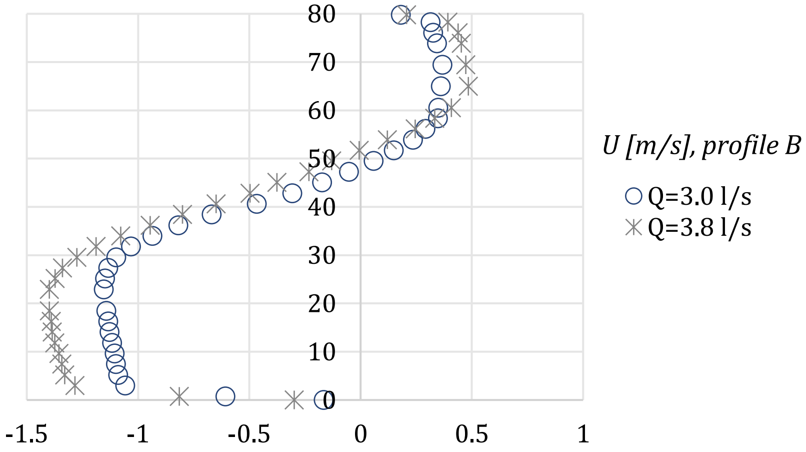


Figure 8. Axial velocity U [m/s] at two flowrates, at location B, in the 80 mm pipe (see Figure 5 and Figure 6)

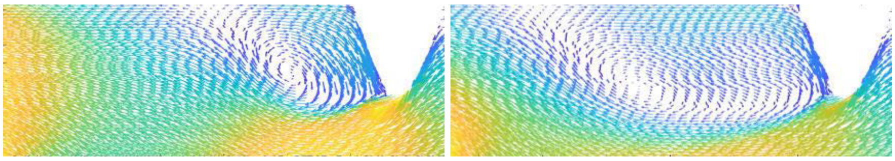


Figure 9. Extract of numerically obtained recirculation area data [22], [24] for qualitative comparison. Lower flowrate on the left hand side and higher on the right.

The velocity field mapping at different flow rates in the pipe section with four elements, two at the obvert and two at the invert, equidistantly, demonstrated clearly the formation of an effective flow area (Figure 10). While under conditions of rising flowrate the recirculation area on the central plane remains roughly the same size, the effective flow section changes. Dynamics at higher flowrates starts to demonstrate that the flow core is diverted from the central plane, forcing higher turbulence generation and forecasting conduit choking.

BY NC SA

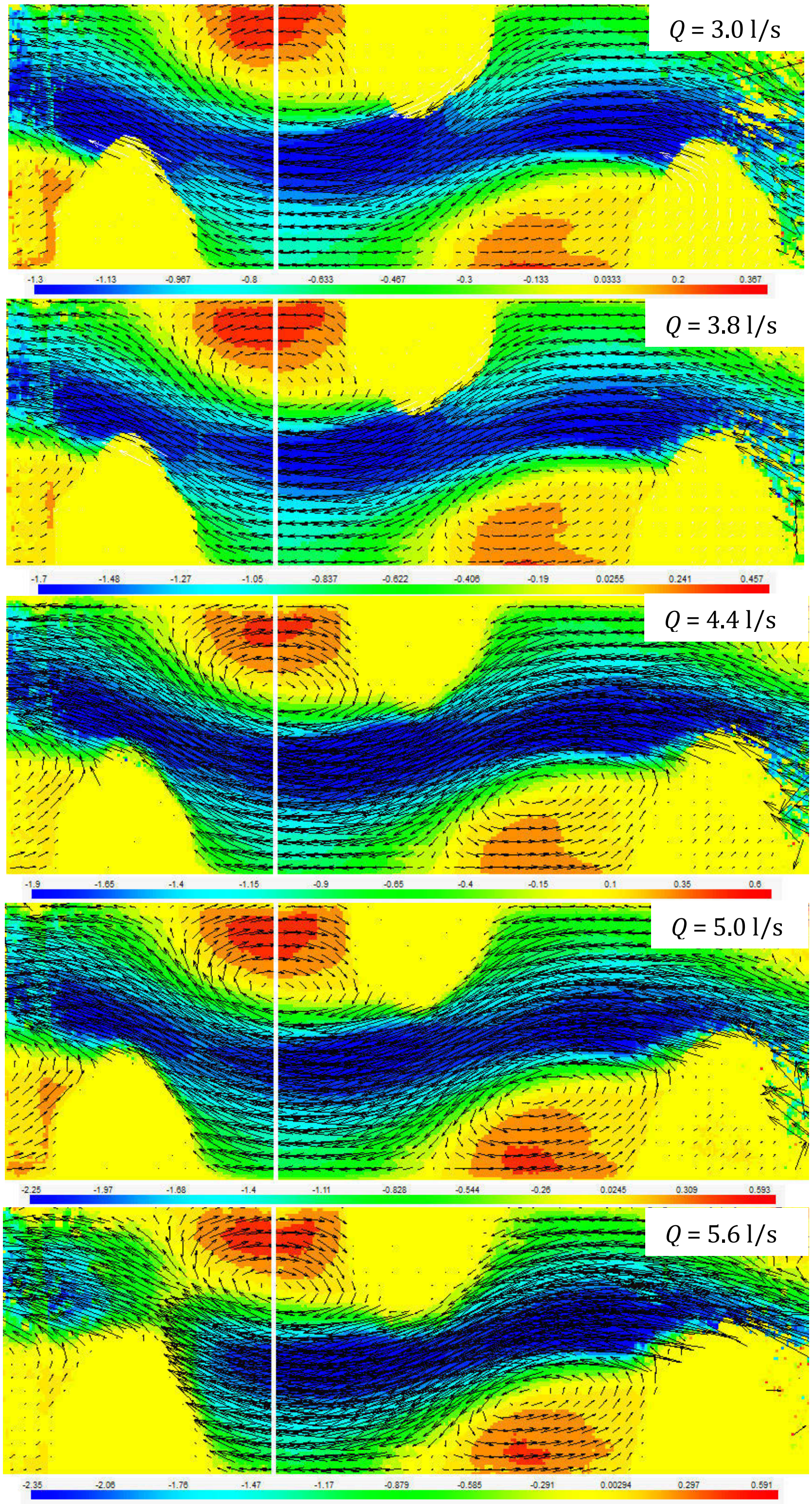


Figure 10. Vector and scalar map in the region of four roughness elements, three of which are in the camera field of view and one on the right hand side, obstructing inflow. Scalar map coloured by axial velocity U. For Q = 3.0 l/s, the scale range is -1.3 … 0.4 m/s; Q = 3.8 l/s, -1.7 … 0.5 m/s; Q = 4.4 l/s, -1.9 … 0.65 m/s; Q = 5.0 l/s, -2.25 … 0.6 m/s; Q = 5.6 l/s, -2.35 … 0.6 m/s. White vertical line – profile location for Figure 11.

BY NC SA

Plotting the axial velocity midway between two roughness elements, near the centre of the recirculation area further demonstrates that dead zone (0-velocity) does not shift with the changing of the flowrate (Figure 11). This agrees with the numerical modelling results obtained in [24]. However, it is noteworthy that the modelling failed to indicate the flow choking starting to occur in the experimental measurements.

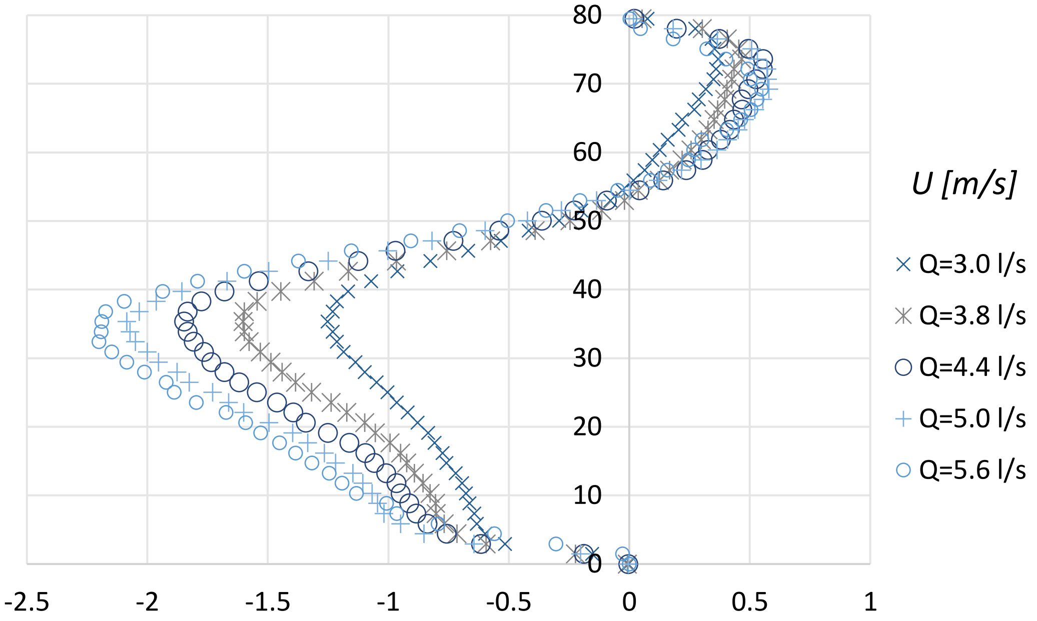


Figure 11. Axial velocity at the location indicated on (Figure 10) for the four roughness elements.

The experimental results reveal that in case of one roughness element the impact to the flow dynamics is local and can be expressed through minor losses. The velocity profiles converge to steady flow profiles downstream of the obstacle as expected. In case of multiple elements, the effective diameter of the pipe is reduced and a jet-type flow will develop at the center of the pipe. Therefore, the actual flow velocity can be times higher compared to the velocity in new pipes with nominal diameter. The changes in velocity increase the pressure drop in the system leading to unrealistic estimations of pipe roughness used in the hydraulic models of WDS.

4 CONCLUSIONS

Understating the flow dynamics in deteriorated pipeline sections is more relevant than ever as we are entering an era where the average life span of pipes in operation is running out, and concurrently, applying real-time control solutions is becoming more commonplace. While many authors recommend making use of advanced numerical modelling (viz. CFD), there are scarce experimental studies that would provide comprehensive calibration and validation opportunities for these complex models, where grid and time resolution, choice of turbulence models and correct treatment of boundaries, among other things, have the capacity to change the output not only qualitatively, but also quantitively.

Herein, the effect of irregular pipe wall build-up to flow dynamics in old rough pipes is investigated experimentally. The aim of the study is to create a comprehensive database of different flow obstruction types and corresponding flow behaviours. PIV was used to map the velocity fields in the pipe containing obstacles.

It was found that the experimental results preliminarily further support the previously published numerical studies of the authors. However, the experiments revealed some dynamics not

cc BY NC SA

witnessed in the computational results. The PIV measurements were conducted for two geometric configurations for five different flowrates. Plotting the axial velocity at different sections behind the one obstruction revealed that at the nearer location, the higher flowrate yields a larger recirculation area when comparing flowrates Q = 3.0 l/s and Q = 3.8 l/s. However, investigating the remaining flowrates of this experimental series revealed diverging results as compared to modelling and merit further investigation. Analysing the velocity field in the pipe with four roughness elements revealed that at higher flowrates the flow core is diverted from the central plane, forcing higher turbulence generation and forecasting conduit choking.

To create an extensive validation database and to develop a comprehensive methodology for determining the proper ratio of roughness and effective diameter in the WDS models based on pipe material and age, further experimental investigations are needed. The authors plan to conduct further experiments on the herein presented configurations of roughness elements and to test new configurations, too.

5 ACKNOWLEDGEMENTS

The research was supported by the Estonian Research Council grant PRG667.

6 REFERENCES

[1] R. T. Christensen, R. E. Spall, and S. L. Barfuss, “Application of Three RANS Turbulence Models to Aged Water Transmission Pipes,” Journal of Hydraulic Engineering, vol. 137, no. 1, pp. 135–139, 2011.

[2] N. Kändler, “Optimal algorithm for rehabilitation of a water distribution network,” Tallinn University of Technology, Tallinn, Estonia, 2002.

[3] V. K. Kanakoudis, “A troubleshooting manual for handling operational problems in water pipe networks,” Journal of Water Supply: Research and Technology-Aqua, vol. 53, no. 2, pp. 109–124, 2004.

[4] I. J. H. G. Vreeburg and Dr. J. B. Boxall, “Discolouration in potable water distribution systems: A review,” Water Research, vol. 41, no. 3, pp. 519–529, 2007.

[5] J. H. G. Vreeburg, E. J. M. Blokker, P. Horst, and J. C. van Dijk, “Velocity-based self-cleaning residential drinking water distribution systems,” Water Supply, vol. 9, no. 6, pp. 635–641, 2009.

[6] C. F. Colebrook and C. M. White, “Experiments with fluid friction in roughened pipes,” Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, vol. 161, no. 906, pp. 367–381, 1937.

[7] W. W. Sharp and T. M. Walski, “Predicting Internal Roughness in Water Mains,” Journal American Water Works Association, vol. 80, no. 11, pp. 34–40, 1988.

[8] G. S. Williams and A. Hazen, Hydraulic Tables. Chapman & Hall, 1920.

[9] G. Echávez, “Increase in Losses Coefficient with Age for Small Diameter Pipes,” Journal of Hydraulic Engineering, vol. 123, no. 2, pp. 157–159, 1997.

[10] K. E. Lansey, W. El-Shorbagy, I. Ahmed, J. Araujo, and C. T. Haan, “Calibration Assessment and Data Collection for Water Distribution Networks,” Journal of Hydraulic Engineering, vol. 127, no. 4, pp. 270–279, 2001.

[11] V. Kanakoudis and K. Gonelas, “Accurate water demand spatial allocation for water networks modelling using a new approach,” Urban Water Journal, vol. 12, no. 5, pp. 362–379, 2015.

[12] V. Kanakoudis and K. Gonelas, “Properly allocating the urban water meter readings to the nodes of a water pipe network simulation model,” Desalination and Water Treatment, vol. 54, no. 8, pp. 2190–2203, 2015.

[13] J. Jiménez, “Turbulent flows over rough walls,” Annual Review of Fluid Mechanics, vol. 36, no. 1, pp. 173–196, 2004.

[14] J. B. Boxall, A. J. Saul, and P. J. Skipworth, “Modeling for Hydraulic Capacity,” Journal - American Water Works Association, vol. 96, no. 4, pp. 161–169, 2004.

[15] I. Annus and A. Vassiljev, “Different Approaches for Calibration of an Operational Water Distribution System Containing Old Pipes,” Procedia Engineering, vol. 119, pp. 526–534, 2015.

[16] S. Vijiapurapu and J. Cui, “Simulation of Turbulent Flow in a Ribbed Pipe Using Large Eddy Simulation,” Numerical Heat Transfer, Part A: Applications, vol. 51, no. 12, pp. 1137–1165, 2007.

[17] S. Vijiapurapu and J. Cui, “Performance of turbulence models for flows through rough pipes,” Applied Mathematical Modelling, vol. 34, no. 6, pp. 1458–1466, 2010.

[18] H. Stel, R. E. M. Morales, A. T. Franco, S. L. M. Junqueira, R. H. Erthal, and M. A. L. Gonçalves, “Numerical and Experimental Analysis of Turbulent Flow in Corrugated Pipes,” Journal of Fluids Engineering, vol. 132, no. 7, p. 071203, 2010.

[19] H. Stel, A. T. Franco, S. L. M. Junqueira, R. H. Erthal, R. Mendes, M. A. L. Gonçalves, and R. E. M. Morales, “Turbulent Flow in D-Type Corrugated Pipes: Flow Pattern and Friction Factor,” Journal of Fluids Engineering, vol. 134, no. 12, 2012.

[20] F. Calomino, A. Tafarojnoruz, M. De Marchis, R. Gaudio, and E. Napoli, “Experimental and Numerical Study on the Flow Field and Friction Factor in a Pressurized Corrugated Pipe,” Journal of Hydraulic Engineering, vol. 141, no. 11, 2015.

[21] I. Annus, A. Kartushinsky, A. Vassiljev, and K. Kaur, “Numerical and Experimental Investigation on Flow Dynamics in a Pipe With an Abrupt Change in Diameter,” Journal of Fluids Engineering, vol. 141, no. 10, p. 101301, 2019.

[22] I. Annus, A. Vassiljev, N. Kändler, and K. Kaur, “Determination of the corresponding roughness height in a WDS model containing old rough pipes,” Journal of Water Supply: Research and Technology-Aqua, vol. 69, no. 3, pp. 201–209, 2020.

[23] AWWA, “Dawn of the Replacement Era: Reinvesting in Drinking Water Infrastructure,” United States of America Water Works Association, Denver, USA, 2001.

[24] K. Kaur, A. Vassiljev, I. Annus, N. Kändler, and J. Roosimägi, “Numerical investigation of the impact of irregular pipe wall build-up on velocity in the water distribution system,” Journal of Water Supply: Research and Technology-Aqua, vol. 69, no. 7, pp. 647–655, 2020.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14144

GIS-INTEGRATED PLUVIAL FLOOD RISK ASSESSMENT METHODOLOGY FOR URBAN AREAS

Murel Truu[1], Ivar Annus[2] and Nils Kändler[3]

[1,2,3] *Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn, Estonia*
[1] *murel.truu@taltech.ee,* [2] *ivar.annus@taltech.ee,* [3] *nils.kandler@taltech.ee*

Abstract

To different extents, most urban areas in Europe are exposed to pluvial flood hazards. The local communities, as well as city governments have to find measures either to cope with the consequences or look for solutions to prevent the possible damage of the floods. Moreover, in most cases cities cannot be transformed to be flood resilient with single isolated interventions, but need an adaptive approach for flood-conscious governance and management. For this, cities require up-to-date information on the flood risk, to make data-based decisions on how to avoid disastrous events, plan for flood resilient high-quality living environment, and where relevant, design and implement transformative interventions. The conventional definition of disaster risk combines the likelihood of potential hazard, exposure magnitude, and the level of vulnerability. However, when considering the multifaceted challenge of assessing the susceptibility and damage potential of urban pluvial flooding, this three-dimensional risk assessment methodology is not yet widely implemented. Additionally, a standardized risk management framework proposes an iterative risk assessment procedure, which could be well-suited for an adaptive governance approach. However, until now, the pluvial flood risk assessment has not been fitted to this framework. In the paper, we present a tiered pluvial flood risk assessment methodology, which can be applied to any urban area. The proposed solution couples the disaster risk function with the standardized iterative risk assessment procedure. This allows cities in various entry-level preparedness to improve their understanding of the city-wide pluvial flood susceptibility and identify the flood-prone urban watersheds in which more specific risk analyses are required. The methodology includes coupling a digital twin of an urban drainage system (UDS) and a geographical information system (GIS). By integrating the pluvial flood risk assessment procedure in the city GIS the cities can automatically determine the potential hazard and coping capacity of exposed areas, and analyse the concurrent vulnerabilities. The methodology has been tested in a small, but densely populated urban area in Estonia, Rakvere town.

Keywords
Pluvial flooding, risk assessment, flood modelling, GIS analysis, urban resilience.

1 INTRODUCTION

Cities need to consider the pluvial flood hazard in various governance decisions, both for routine everyday decision making as well as for far-sighted strategic planning. Examples of strategies in which municipalities need to have adequate up-to-date information on pluvial flood risk include city-level comprehensive spatial plans, sewer management plans, sustainable energy and climate plans, and various strategic investment plans. Moreover, also the routine permitting of various infrastructure and building projects as well as the drafting of related policies, should be based on informed decisions on floods. Therefore, risk-based management of the pluvial floods has become more relevant for the urban areas.

Pluvial flood-conscious cities assess urban flooding risk using a variety of methods, whereas many of them are based only on historical floods and community knowledge of such events [1], [2], [3]. With changing climate as well as transforming cityscapes, such methods fail in projecting future risks [4]. Only a limited number of studies (Table 1) responding to “pluvial flood” & “risk

BY NC SA

assessment" inquiry in the Web of Science platform apply the United Nations Disaster Risk Reduction Office (UNDRR) disaster risk framework. Whereas, the ones that do, fail in consistency in terms of interpreting the different risk parameters in the function. Moreover, the methods are developed and tested for cities of various sizes and complexities, some proposing solutions for only small towns [5],[6] while others are applied for large metropolitan regions [7],[8]. Data sources and methods, which are used to analyse the different risk parameters, vary significantly. However, most of the investigated solutions exploit the capacity of digital elevation/terrain models (DEM/DTM) and different public datasets available for the analysed area. None of the identified methods consider the hydraulic capacity of the urban drainage system (UDS) in case of extreme weather events. While more general approaches exist for assessing the effect of the coping capacity for the flood risk [9] and the adaptive iterative approach to flood management is supported in policy [10], then the adaptability of existing city-based risk assessment methods is limited.

Table 1 Studies presenting pluvial flood risk assessment methods that correspond with UNDRR risk function

Method	Case study area	Hazard	Exposure	Vulnerability
Othmer et al. 2020 [5]	53 km² area with a densely built urban centre covering ~17% of the territory (Olfen, Germany)	Flow path and sink analysis (DEM) combined with 2D surface runoff calculation	x	Potential damage to buildings weighed with the vulnerability of residents (age dependant)
Szewrański et al. 2018 [6]	5 km² village with a densely built urban centre covering ~16% of the territory (Dobrzykowice, Poland)	Precipitation forecast data combined with surface runoff estimation and sink evaluation (DEM)	Water level in lowpoint areas	Damage to buildings
Di Salvo et al. 2018 [7]	1285 km² metropolitan area of the city (Rome, Italy)	Flood susceptibility-combining observed floods with flood prediction based on DTM	x	Potential impact - damage to buildings, commercial activities, critical urban elements, potential pollution sources, heritage objects
Sperotto et al. 2016 [8]	416 km² densely populated urban area (Venice, Italy)	Intensity of future precipitation combined with maximum pluvial emergency thresholds	Exposed buildings and infrastructures	Vulnerability factors: slope, permeability, historically flooded areas

As seen from the selection of the available methods described above, concern about pluvial flood risk is universal to large metropolitan regions and small urban villages. While it is well understood that the pluvial flood damage to urban assets varies significantly depending on the size and complexity of the city, small rural towns can also face events with such disastrous impacts for which the risks have to be assessed and thereby addressed.

2 MATERIALS AND METHODS

2.1 Fitting the pluvial flood risk assessment into the standardized disaster risk assessment framework

In simple terms, the risk is understood as the potential for damage from unwanted events. As such, the term is broadly applied to quite a range of phenomena, covering both the causes and the effects of the unwanted event in focus. While risk perception can be subjective, risk governance should be standardized and, therefore, also framed by clear methods. The international standard ISO 31000 for Risk Management Principles [11] provides a comprehensive framework, principles, and a process description for risk management. The given framework places risk assessment into a holistic iterative cycle, where the risk assessment is only one step in a sequence of many (Figure 1).

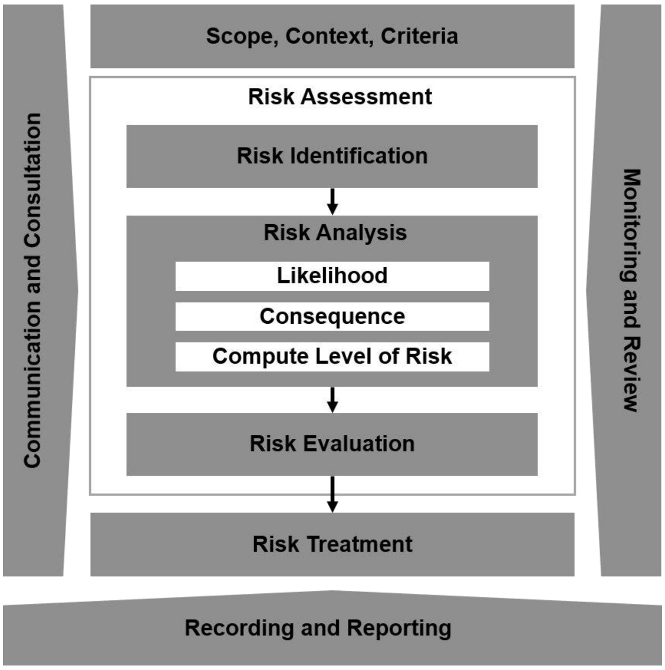

Scope, Context, Criteria
Risk Assessment
Risk Identification
Risk Analysis
Likelihood
Consequence
Compute Level of Risk
Risk Evaluation
Risk Treatment
Communication and Consultation
Monitoring and Review
Recording and Reporting

Figure 1 Risk management framework according to ISO 31000 [11]

The risk management principles listed in the standard foresee that the risk assessment is carried out using the state of the art methods and the best available data. Although for fluvial floods the EC Flood Directive [12] offers a widely applied and thus well-known methodology, the situation is much different in terms of the assessment of urban pluvial floods. In normal circumstances, the pluvial floods are considered minor to medium level inconveniences in the city. However, in case of extreme events, the damage from pluvial floods can be as disastrous. Therefore, when looking for methods for understanding the risks, clear interlinkages should be made with the risk definition provided by UNDRR. While in broad terms risk is understood as a two-parameter index combining the likelihood of non-anticipated events and their potential damages, then in the context of environmental disasters UNDRR describes risk [1] as a function of hazard, exposure, and vulnerability (1).

$$Risk \ = Hazard \ x \ \text{Exposure} \ x \ Vulnerability \qquad (1)$$

This function describes well the external risks of various disasters for which mitigative measures are not applicable. This, however does not stand for pluvial flood risks that are greatly affected by the urban development, in most cases planned long in advance. Furthermore, when placing the risk assessment in the iterative risk management framework, the risk reduction capacity is crucial for assessing the treatment alternatives. And when projecting the future risks it is necessary to analyse not only the positive coping capacity that would reduce the risks, but also the various negative development scenarios. The risk index function that considers coping capacity is well presented by Imamura, 2022 [9], developed for the generic country-scale flood risk assessment.

Previous studies that have applied the UNDRR function for the assessment of pluvial flood risk are inconsistent in interpreting the various risk parameters as presented in Table 1. However, the Intergovernmental Panel on Climate Change (IPCC) [14] and several review papers on the challenges of pluvial flood risk assessment [4], [15] have a common understanding. This allows to define risk parameters for any type of flooding as follows: 1) hazard is the likelihood of occurrence of a driving event; 2) exposure is the amount of people and assets that would be directly impacted (flooded zone), and; 3) vulnerability is understood as the severity of impacts.

2.2 Tiered pluvial flood risk assessment framework

The current paper presents a tiered pluvial flood risk assessment procedure (Figure 2), which combines the ISO 31000 risk management standard procedure with the UNDRR disaster risk function and integrates it into city GIS to institutionalize the further automatic iterations of the risk assessment. The iterative tiered pluvial flood risk assessment method developed in our research proposes a procedure on how to combine risk identification on a large metropolitan scale and application of detailed methods for small catchment-level risk analysis and unify the interpretation of the risk parameters. In the proposed procedure risk assessment is carried out in 3 tiers (Figure 3), whereas the various predefined scenarios allow us to carry out comparative iterations of the assessment or renew the assessment as baseline data changes.

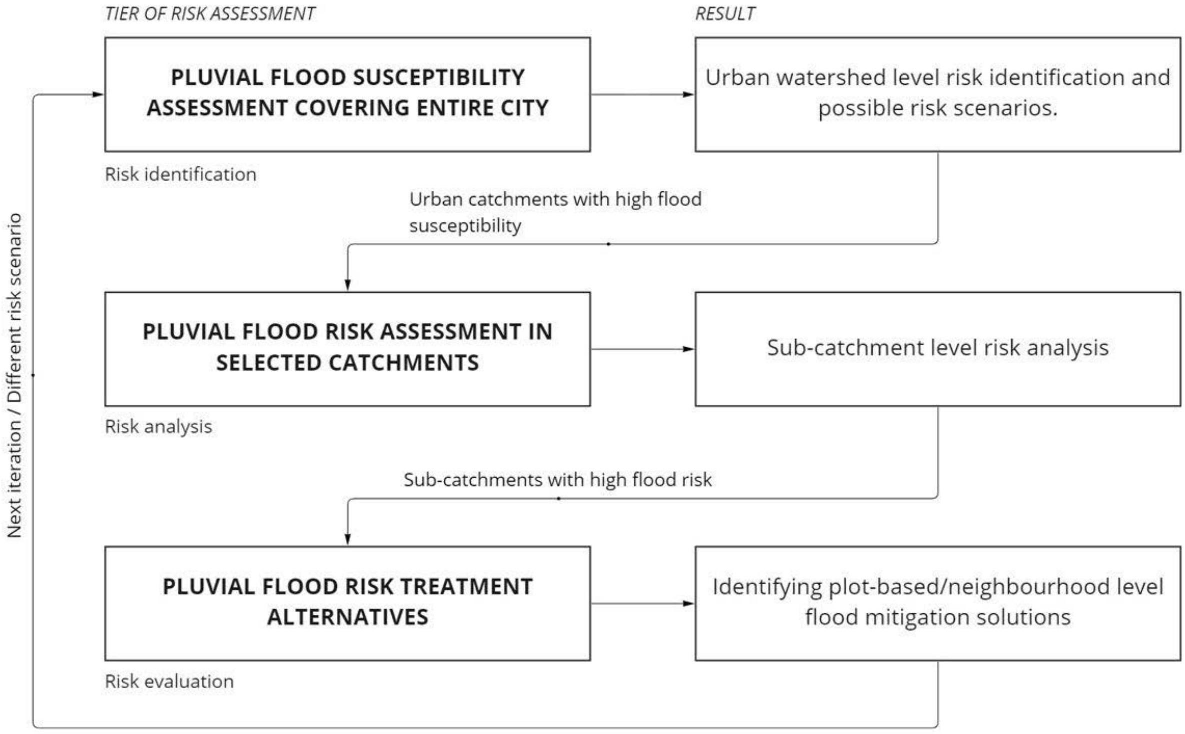

TIER OF RISK ASSESSMENT
RESULT
PLUVIAL FLOOD SUSCEPTIBILITY ASSESSMENT COVERING ENTIRE CITY
Urban watershed level risk identification and possible risk scenarios.
Risk identification
Urban catchments with high flood susceptibility
PLUVIAL FLOOD RISK ASSESSMENT IN SELECTED CATCHMENTS
Sub-catchment level risk analysis
Risk analysis
Sub-catchments with high flood risk
PLUVIAL FLOOD RISK TREATMENT ALTERNATIVES
Identifying plot-based/neighbourhood level flood mitigation solutions
Risk evaluation
Next iteration / Different risk scenario

Figure 2 Overview of the proposed tiered procedure to assess the risks of pluvial floods

Risk identification on tier 1 stands from the mapping pluvial flood susceptibility. Risk identification is carried out using widely accessible data and GIS methods. The overall objective of this step is to compare the flood susceptibility of different urban watersheds and identify areas where further analysis is needed. The identification procedure allows determining the specific factors that need to be considered in a detailed risk analysis carried out in Tier 2 and what methods to use for this. The main output of the risk identification stage of the assessment is a comparative flood susceptibility index of urban watersheds, which allows identifying of districts in the city where detailed risk analysis is required. The risk identification tier, if carried out with no further investigations, does not provide adequate answers about the potential consequences of the extreme events.

The risk analysis on tier 2 covers detailed risk assessment in catchments with higher pluvial flood susceptibility. This step of the analysis must consider the character of the catchment, due to which the resulting risk levels can be adequately compared only within the urban watershed being analysed, not on the city level. Different modelling methods can be applied for analysing the hydrodynamic features causing the urban flood risk in pipe-based systems or in peri-urban open-channel catchments. The main outputs of the risk analysis stage are the comparative risk index of sub-catchments and sub-catchment level estimates of potential damage.

Risk evaluation on tier 3 consists of weighing the alternatives of treating the pluvial flood risk. This step goes beyond the pluvial flood hazard, exposure, vulnerability, and predefined capacity to cope with consequences and investigates the level of acceptance of flood risk and thus weighs solutions to mitigate the flood risk in comparison with potential losses. This step also allows considering the residual risk and plan for measures for coping with this. In the evaluation stage, various thematic maps can be generated to visualize the impact on vulnerable urban features, as well as investigate possibilities to transform the city space with additional measures. If such solutions for risk treatment are found, new risk scenarios can be developed and assessed. The main outputs of the risk evaluation stage are decisive risk levels for high-risk sub-catchments/plots, determining whether risks are acceptable, need to be treated, or should be considered as residual risks that cannot be treated. The risk evaluation stage can be coupled with GIS-based decision support to localize the various flood mitigation solutions. Also, the risk evaluation stage provides suitable baseline data to monetize possible damage and assess alternative costs for not taking further action to treat the risks.

When handling the risk parameters, it is necessary to understand their connectivity (Figure 4) and to either consider the interdependencies or abandon the linkages decided based on the sensitivity of the analysis. Disassembling the risk factors also allows constructing of various risk scenarios based on the likelihood of the hazard or manifestation of concurrent hazards, with considerations of different development scenarios or governance decisions prioritizing critical urban vulnerabilities.

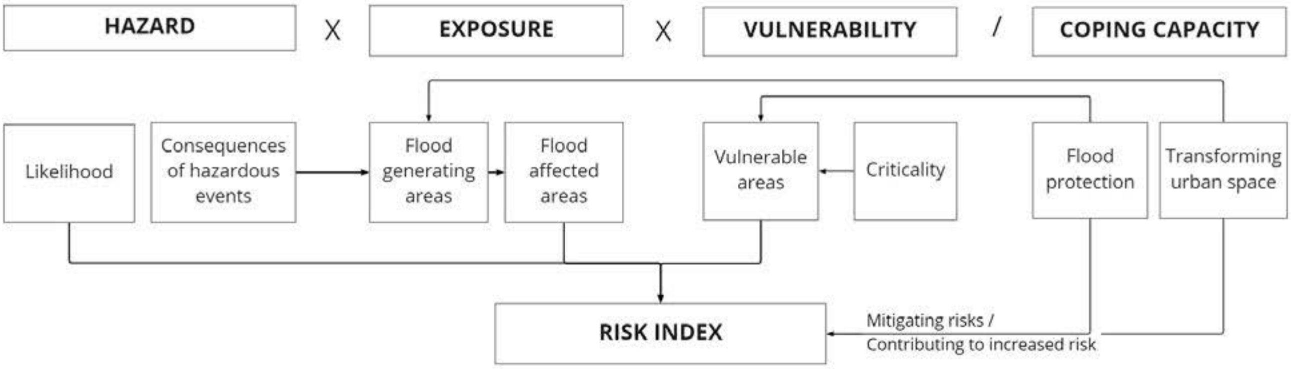


Figure 3 Simplified presentation of the interconnectivity of different risk factors

The risk scenarios to be assessed depending on hazard likelihood as well as preparedness to cope/mitigate the consequences, thus the risk scenarios are selected as combinations of manifested hazards (either independent extreme precipitation or concurrently manifesting hazards of fluvial floods, technical failures, etc.) and the planned/discussed interventions for dealing with the consequences.

In the current paper, the risk index is calculated using an expanded UNDRR function (2-3), which allows to separately weighing each sub-indicator to determine the risk parameters in the equation. For the baseline scenario to compare the risk indexes against, the coping capacity factor is abandoned. Weighing factors for the risk indicators are to be decided by an expert decision and can vary depending on the city. The indicators used in the function need to be normalized to a range of 0-1. Weighing factors of the risk sub-indicators - a_i, b_j, c_k, d_m. Risk sub-indicators for hazard (H), exposure (E) and vulnerability (V), coping capacity (CC) - H_i, E_j, V_k, CC_m. Modelled exposure indicator considering hazard magnitude and proposed coping measures as defined in analysed Scenario - $\frac{E_{H,j}}{CC_j}$;

$$Risk_{Tier\ 1} = \frac{\sum(a_i \cdot H_i)\ x \sum(b_j \cdot E_j)\ x\ \sum(c_k \cdot V_k)}{\sum(d_m \cdot CC_m)} \tag{2}$$

$$Risk_{Tier\ 2} = \sum\ (a_i \cdot H_i)\ x \sum\ (b_j \cdot \frac{E_{H,j}}{CC_j})\ x \sum\ (c_k \cdot V_k) \tag{3}$$

In tier 3 the decisions, whether the resulting risk level is to be considered as A - acceptable; T - treatable; or R – residual, are based on public risk perception and local governance decisions.

2.3 Materials and methods to understand the hydrodynamic nature of pluvial floods

The occurrence mechanism of the pluvial flood is a complex and dynamic problem. Not only are the pluvial floods dependant on hardly predictable extreme and very local downpours, but the floods are as much affected by the fine-scale character of the city-scape. As both the climate and the cities are subject to changes, historical evidence and present-day design standards are in many cases inadequate in solving the challenge. While calculating the risk levels, it is relevant to also consider future hazard scenarios as well as urban development projections to assess vulnerability, the key challenge in understanding the future risks lies in modelling of the pluvial flood exposure.

Model selection for the risk assessment depends on the risk assessment stage. Flood identification can be carried out using simplified methods such as topographic wetness index (TWI) [16] or rapid flood spreading models (RFS) [17]. Flood-prone district identification for city-level strategic planning can be less precise, as the simpler models are favoured also due to data availability as well as processing speed (run-time) while covering large territories. For a more comprehensive understanding of the complexity of the consequences as well as possible damages, more detailed models coupling overland flow and hydraulic capacity of the systems are required [18]. As the detailed models also demand precise baseline data, it would take decades to map, model, and calibrate all urban drainage systems in adequate quality to be used for comprehensive modelling for cities in need of flood inundation mapping. Identification of flood-prone areas should be carried out based on existing data in public registries and city GIS and only for the high-risk districts more specific studies are to be carried out.

Coupling pluvial flood modelling with geographic information systems (GIS) will add substantial value to risk assessment allowing fast and efficient adaptation of the risk calculation to the ever-changing urban environment and supporting risk communication to the stakeholders and citizens. The state-of-the-art GIS-based methods for pluvial flood risk assessment can interpret publicly available datasets such as digital elevation models, various state registries (environmental, heritage, building, etc.), and municipal datasets (population density, public transport). Large scale calculations can be carried out for translating vector data into raster layers for which multi-criteria ranking can be performed using user-defined weighs and site-specific concurring factors contributing to the risk. In such a manner the risk assessment has high automation potential.

3 CASE STUDY AREA

Rakvere is a small town (~11 km²) located in northern Estonia with a population of approximately 15 000 people. Two small streams, the Soolikaoja creek, and the main Tobia ditch flow through the city and the recreational forest covers approximately 15 % of the city's territory. The majority of the city is situated in the large watershed of Soolikaoja creek however, due to the urban drainage system and urban-space characteristics the town is better described by smaller catchments. In our work, these catchments have been generated in GIS based on 5x5m resolution DEM and modified according to the UDS (Figure 5). The waterbodies in the city are not prone to fluvial flooding. However, the flow rates in the streams affect significantly the capacity and performance of the UDS [19].

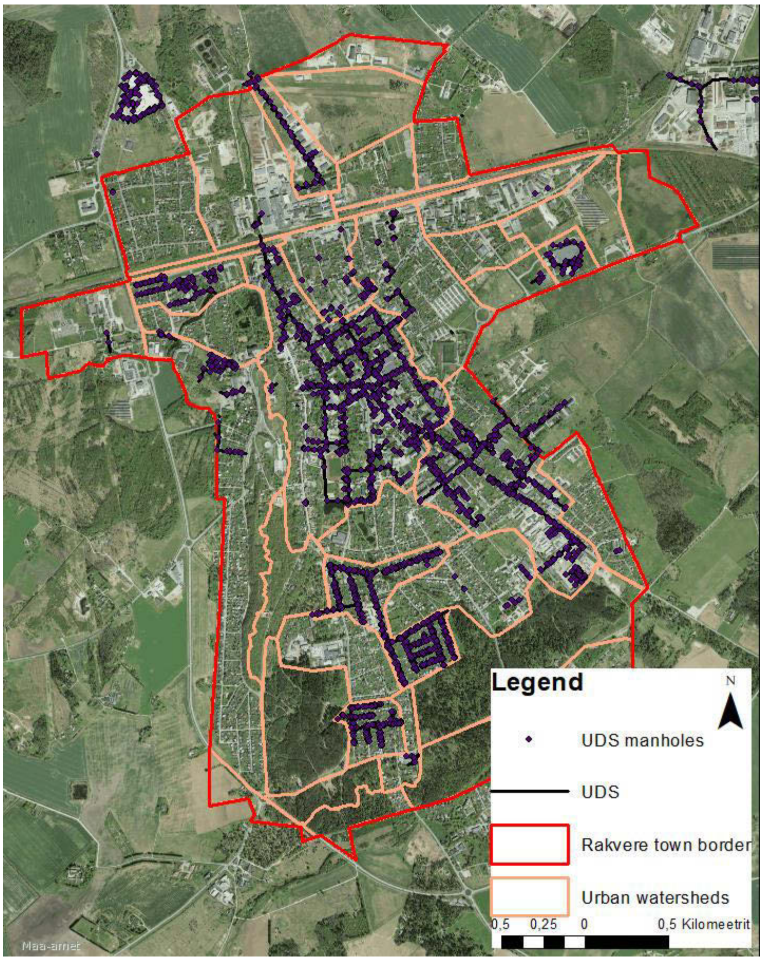


Figure 4 Rakvere city and the catchments generated with ArcMap Spatial Analyst tools based on 5x5m DEM and character of UDS.

Approximately half of the territory of the towns is covered with stormwater drainage system, while a large part of the town still deals with the stormwater runoff with combined sewage, open ditches and undirected infiltration areas (random and not specially designed permeable areas). The areas in the town with separate stormwater systems perform either as pipe-based systems or utilize open ditches, streams, and ponds.

4 RESULTS & DISCUSSION

The proposed methodology and procedure have been applied in the Rakvere case study area, using data available from public registries and the local municipal datasets (Table 2). In the current paper, we present the overall framework and demonstrate the potential of tier 1 and tier 2 assessments with a limited number of scenario iterations. Tier 3 evaluation is presented only as a concept and will be demonstrated in further papers.

4.1 Tier 1: Identifying pluvial flood susceptible districts/watersheds

Tier 1, identification of flood susceptible urban watersheds, can be carried out by urban planners or environmental consultants, with no comprehensive background in pluvial flood modelling. The procedure requires adequate DEM, land-cover data, pluvial flood design thresholds, and city-level understanding on urban flood vulnerabilities. Rapid analysis as described here can easily be set up and automated in GIS to be applied for various strategic planning documents, where such generic flood inundation mapping is required. For the iterative application of the flood identification stage, the different parameters can be updated as baseline data is updated or upgraded.

Table 2 Pluvial flood susceptibility parameters and the data used for the baseline scenario in Rakvere case study.

Risk factor	Concept
Hazard	Likelihood of occurrence for extreme precipitation/Variable intensity according to the cityscape. Baseline applied in case study: national design standard [20]
	Concurrent hazards: to be considered in case the pluvial flood is magnified by other natural or manmade hazards (e.g. fluvial floods, system failures)
Exposure	Topographic susceptibility of flood: DEM based surface flow modelling results (TWI, RFS or other) Baseline applied in case study: TWI based on 1x1m DEM [21]
	Infiltration capacity: Landcover based infiltration capacity estimate. Baseline applied in case study: national 1:10 000 base map landcover data [22] combined with national design standard surface runoff rate [20]
Vulnerability	City based estimate. According to the EC Flood Directive the flood risk needs to be assessed against economic, social, environmental and cultural vulnerability [12] Baseline applied in case study (weighing factor): density and value factor of built-up area (1), population density (1), UDS character (0.5), overlay of heritage monuments (0.5)
Coping capacity	Scenario based estimate. Indicators need to show the direction of the impact of the coping measures. Current paper demonstrates the impact of green factor policy to city level flood susceptibility.

The flood risk index is calculated using equation 3 and simple GIS raster calculation through various raster layers (figure 5) representing the indicators representing the risk parameters described in Table 2.

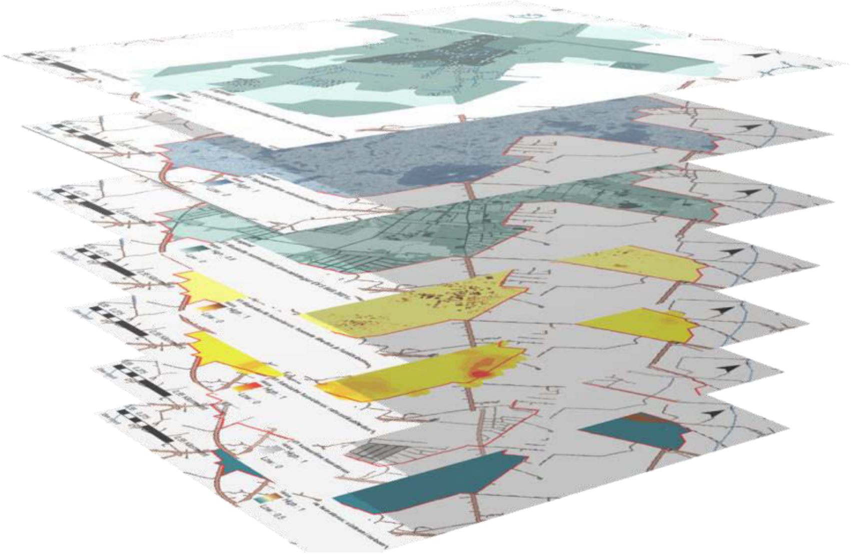


Figure 5 Risk index is calculated in GIS through various raster layers consisting of normalized flood risk indicators describing the specific parameters. If any of the baselines is updated, the risk level can be recalculated.

As a first step, the baseline scenario against what the further alternatives are to be analysed was defined. In Rakvere case study, the baseline scenario was set up as a current hazard in the current urban space, which meant that the indicator rasters were set up by national design standard, current land-use and current socio-economic and cultural-environmental characteristics of the town (Table 2). It must not be forgotten that the risk level in tier 1 is only applicable as comparative index. As the tier 1 is meant to identify the flood prone urban watersheds, then the risk levels have to be calculated for predefined watersheds or system units (Figure 6).

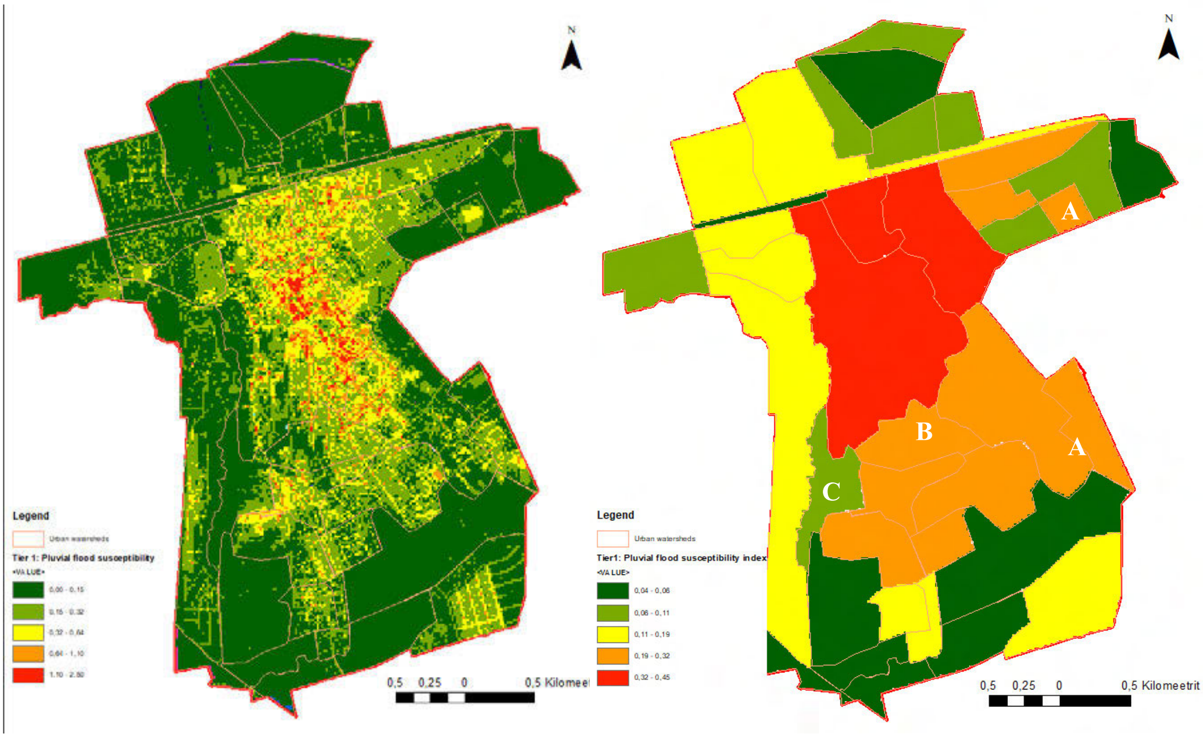

N
Legend
Urban watersheds
Tier 1: Pluvial flood susceptibility
<VA LUE>
0,00 - 0,15
0,15 - 0,32
0,32 - 0,64
0,64 - 1,10
1,10 - 2,50
0,5 0,25 0 0,5 Kilomeetrit
N
A
B
A
C
Legend
Urban watersheds
Tier1: Pluvial flood susceptibility index
<VA LUE>
0,04 - 0,06
0,06 - 0,11
0,11 - 0,19
0,19 - 0,32
0,32 - 0,45
0,5 0,25 0 0,5 Kilomeetrit

Figure 6 Tier 1 pluvial flood risk index raster (Left) and mean risk index of urban catchment (Right) for the baseline scenario. For avoiding misinterpretations only the urban watershed level risks should be used for decision-making.

The identified risk index allows to either contrast different sewer-sheds or compare different coping scenarios. When the mean risk index of the catchment is classified using natural breaks distribution then only 3 watersheds belong to the highest risk category (red). These are the most densely built and populated areas. 9 catchments belong to the no risk (dark green) category, whereas 8 catchments fall to the low risk level (light green). In most cases these are existing green areas in the city, recreational forests or yet undeveloped green spaces. Almost half of the territory of the city can be categorized as moderate (yellow) or significant (orange) risk levels. A large part of the moderate risk level areas is residential zones with private gardens. The reasons why watersheds fall in the significant risk levels however differ: in some cases the higher risk is caused by densely built-up area being situated in lowpoint areas (A, figure 6). In other cases the higher risk is a cause of higher vulnerability, which in the current case is the density of milieu-valuable wooden buildings and higher population density (B, figure 6). However, when comparing the latter another high vulnerability watershed (C, figure 6) with a relatively densely built up heritage protection area, the flood risk level there is lower as the topography does not favour floods.

The analysis of coping scenarios in Tier 1 can be carried out in a generic level. Various coping measures, which contribute to the reduction of any of the risk factors, can be assessed in such a manner. However, it must be acknowledged that in the risk function, the coping capacity indicators can only show the direction of how coping measures affect the risk level, not their absolute effect on flood reduction.

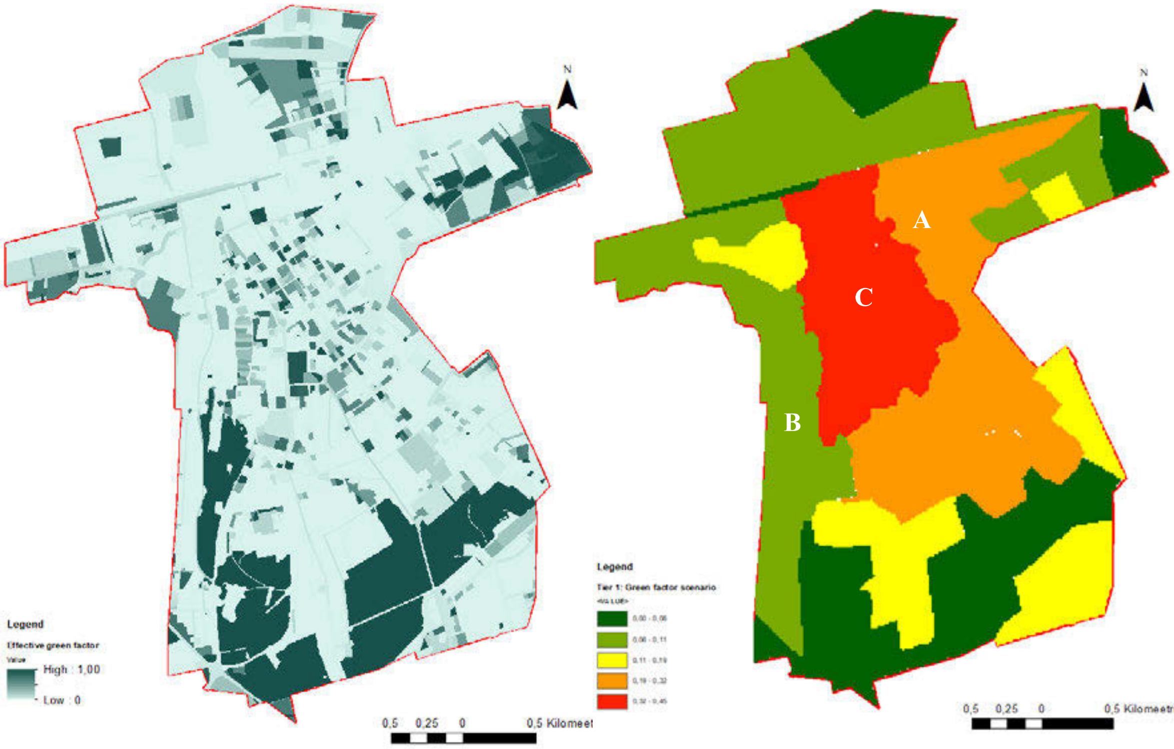


Figure 7 Effective green factor of property plots (Left). The risk level of catchments as policy would be enforced (Right).

This solution is useful for large territories in assessing city-wide policy effects or strategic planning decisions. To demonstrate the potential of the solution, a green factor scenario was prepared for the case study area. Several cities in the world have implemented a requirement of the green factor as a flood resilient measure for spatial planning of cities [23]. Until now, Rakvere town has not implemented a complete green factor requirement however, we analysed the current ratio of effective green areas in the city (recreational forests, greenfields, parks and gardens) and developed a raster layer representing the ratio of existing green area within a

CC BY NC SA

property plot (figure 7, left). Based on that an indicator raster layer was prepared, showing the gap between the current green factor and the potential policy requirement (30%). This allowed to analyse the coping capacity potential of the town if all plots in the city would fulfil the green factor requirement. The resulting map (figure 7, right), shows that such a policy would have flood risk mitigation potential in the eastern part of the central town (figure 7, A) as well as some residential zones (figure 7, B). However the increase of green areas reduces the absolute pluvial flood susceptibility risk in the central town, then the risk still remains high as in the baseline scenario (figure 7, C).

As described above, risk identification serves the role of identifying the hotspots that need more attention and in-depth modelling. Tier 1 helps to identify also the main causes why catchments fall into different risk classes allows to prepare for the next tier or of the assessment. In the current paper, we demonstrate the tier 2 analysis in the central part of the town, which fell into the highest risk class both in the baseline scenario as well as the green factor scenario. To advance with tier 2 analysis, a modelling method needs to be selected to refine the understanding of the causes and consequences of the pluvial floods. Many existing pluvial flood risk decision support tools do not consider the performance of the urban drainage system and expect it to fail in extreme weather events that exceed the design thresholds. While in many cases this is an adequate simplification, the malfunctions of urban drainage systems can significantly affect the consequences of extreme weather events [4].

4.2 Tier 2: Pluvial flood risk analysis in pipe-based urban catchments

The risk identification carried out in Tier 1 allowed several simplifications, for example, all the risk parameters were handled as separate indicator raster layers and potential feedback loops were abandoned. As the tier 2 analysis aims to establish an understanding of the damage potential of pluvial floods, not only the susceptibility of them, then more attention is given to the interconnectivity of the risk factors. As both the hazard magnitude and implemented coping measures define the exposed areas, then for every analysed risk scenario, a separate modelling simulation is required. In our study the modelling was carried out using EPA SWMM 5.1 modelling software [24], the sub-catchments for which the detailed analysis was carried out, were automatically generated using the GISto SWMM module [25]. A more detailed description of the used method is given by Truu et al. 2021 [18].

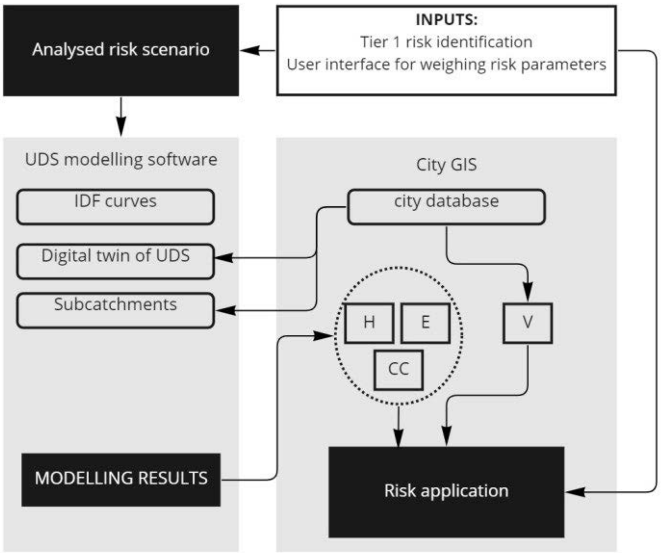


Figure 8 Concept for the risk analysis step for tier 2

BY NC SA

Truu et al. 2021 [18] presented the flood risk on three levels (low, moderate and high), however did not assess the different vulnerability of the cityscape for the various flooding events. In the current paper, the modelling is carried out to define the sub-catchment level indicators that show the combined effect of the magnitude of the hazard, applied countermeasures or implemented development scenarios to the pluvial flood exposure (figure 8). In Rakvere, flood volume and flood duration sub-indicators were calculated and converted to raster analysis. In such manner, these modelled features serve as combi-indicators that presents the scenario-based effect to the hazard-exposure and coping capacity. With an expanded analysis modelling can result additional sub-indicators as water-quality, ponding depth or other, relevant to different urban vulnerabilities.

The resulting sub-catchment level risk map is applicable for fine scale analysis that interlinks the hazard occurrence probability with its damage potential. Pluvial flood modelling results indicator against the resulting risk map. Modelling allows to determine the consequences of the extreme weather event and the coupled GIS analysis finalizes the analysis by determining the damage potential. In the example visualized in figure 9, it is shown that while flood volume in catchment A is classified as high, then as it occurs on a parking lot (figure 8, A), it is less relevant than the flood in catchment B, where similar flood volume affects several buildings (figure 8, B).

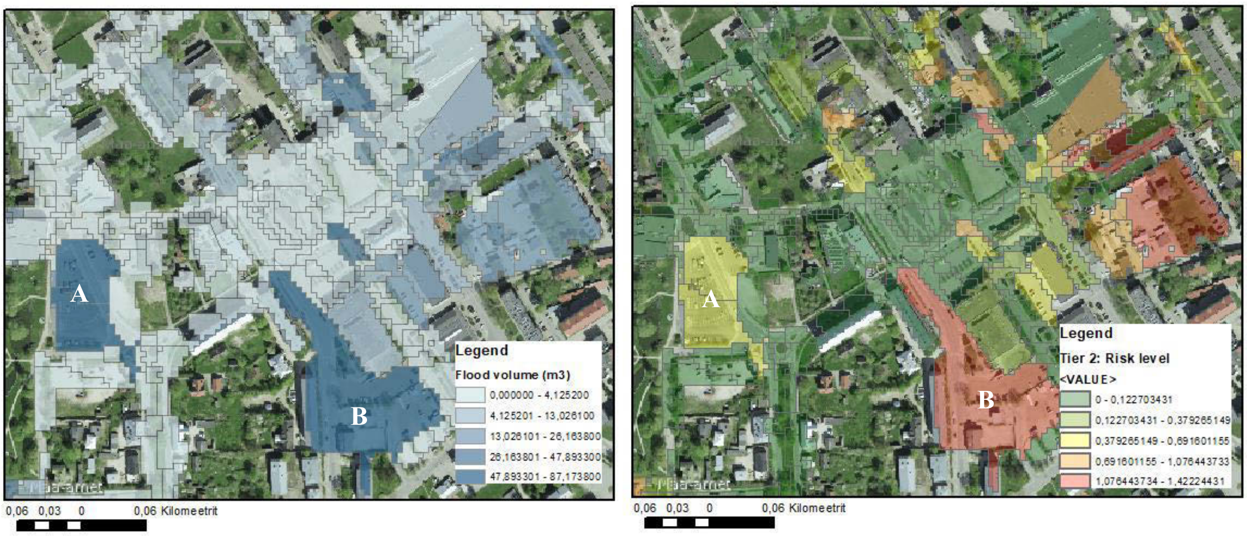


Figure 9 Fragment of pluvial flood combi-indicator raster showing the modelling results of flood volume in catchments in the baseline scenario (m^3) (left) and the risk level assessed based on vulnerability.

Additionally, the automation of the Tier 2 risk assessment tool in city GIS was piloted in Rakvere. The pilot tool combines drainage modelling software and a GIS database. All the input and output data is automatically exchanged between these two modules and necessary additional information like street objects and borders of the properties are acquired from the public databases.

EPA SWMM [24] modelling software engine with Python-based package PySWMM [26] was used for the analysis of the drainage system. ArcMap GIS software with built-in Python package ArcPy [27] was utilized to script data links between the modules. The stormwater catchments and drainage pipelines with manholes - compulsory data to model the runoff in SWMM - are kept in the GIS system which facilitates keeping this data up-to-date. Before each simulation, the information about the model elements and catchments can be automatically imported from GIS to the SWMM input file. User has to choose the pre-defined climate scenario (precipitation intensities) and after simulation the results, flood prone manholes are automatically imported back to the GIS with the data about the flood duration, depth, and volume. The results can be integrated into the risk function to allow also non-professional modellers to precept the risks in urban districts. The automated risk assessment module has also an interlinkage with public web

CC BY NC SA

map services (WMS) that allows the automatic update of spatial data (property borders, street names, etc.).

5 CONCLUSIONS

Cities need to consider the pluvial flood risk for different regular and routine governance decisions. While the development of the strategic documents requires a holistic understanding of the pluvial flood risk, then the routine land use decisions and various investments in the cityscape require a much more detailed understanding of the consequences of the floods. As the pluvial flood is a dynamic problem, then static risk assessments expire fast. The cityscape is not the only parameter subject to constant change, also other data (vulnerability parameters), climate scenarios, and design thresholds are regularly revised and if relevant also changed. All this requires an adaptive management approach, a concept which is mainstreamed also in the generic EU climate adaptation policy framework Climate-Adapt [10]. Moreover, the pluvial flood risk is a growing interest not only for Water Engineering field, but also for the disciplines of Social Sciences and Humanities. This means the baseline methods to determine the different risk parameters are constantly advancing [4], [15].

The GIS integrated risk assessment methodology proposed herein, fits well to the adaptive management framework. The proposed risk assessment solution couples ISO standard of risk management with UNDRR conceptual risk function. In a practical sense, it provides a set of risk parameters that are calculated based on information in public datasets already synchronized with city GIS and sets up a procedure how to calculate the risk index. The proposed procedure allows to deliver new risk assessments at request or as data updates.

The novelty of the described methodology lies in its iterative and dynamic nature allowing it to automatically calculate the summary risk for all land parcels in the analysed area. Any change in land use, UDS configuration, and vulnerabilities can be instantly referred to in a change in the risk levels. This allows municipal officers to deliver up-to-date risk assessment iterations with only minor effort and evaluate future risk scenarios with the same system to understand the effect of various urban development plans or also flood mitigation measures, i.e. increase of the permeable areas, planning detention facilities and improving the operation of UDS. Moreover, the iterative risk assessment method allows to upgrade the risk assessment by replacing any of the sub-indicators with new and improved understanding. Also, the results of risk assessment tiers can feed into plans, strategies and decisions of different level as not all governance decisions require supportive baseline data in same precision.

6 REFERENCES

[1] J. Chen, W. Chen, and G. Huang, "Assessing urban pluvial flood resilience based on a novel grid-based quantification method that considers human risk perceptions," Journal of Hydrology, vol. 601, p. 126601, Oct. 2021, doi: 10.1016/j.jhydrol.2021.126601.

[2] V. Norén, B. Hedelin, L. Nyberg, and K. Bishop, "Flood risk assessment – Practices in flood prone Swedish municipalities," International Journal of Disaster Risk Reduction, vol. 18, pp. 206–217, Sep. 2016, doi: 10.1016/j.ijdrr.2016.07.003.

[3] "Member States' examples of Flood Hazard and Flood Risk Maps," 2014. Accessed: Apr. 25, 2022. [Online]. Available: https://ec.europa.eu/environment/water/flood_risk/pdf/MS%20examples.pdf

[4] L. Cea and P. Costabile, "Flood Risk in Urban Areas: Modelling, Management and Adaptation to Climate Change. A Review," Hydrology, vol. 9, no. 3, Art. no. 3, Mar. 2022, doi: 10.3390/hydrology9030050.

[5] F. J. Othmer, D. Becker, L. M. Schulte, and S. Greiving, “A Methodological Approach to Municipal Pluvial Flood Risk Assessment Based on a Small City Case Study,” Sustainability, vol. 12, no. 24, Art. no. 24, Jan. 2020, doi: 10.3390/su122410487.

[6] S. Szewrański, J. Chruściński, J. Kazak, M. Świąder, K. Tokarczyk-Dorociak, and R. Żmuda, “Pluvial Flood Risk Assessment Tool (PFRA) for Rainwater Management and Adaptation to Climate Change in Newly Urbanised Areas,” Water, vol. 10, no. 4, Art. no. 4, Apr. 2018, doi: 10.3390/w10040386.

[7] C. Di Salvo, F. Pennica, G. Ciotoli, and G. P. Cavinato, “A GIS-based procedure for preliminary mapping of pluvial flood risk at metropolitan scale,” Environmental Modelling & Software, vol. 107, pp. 64–84, Sep. 2018, doi: 10.1016/j.envsoft.2018.05.020.

[8] A. Sperotto, S. Torresan, V. Gallina, E. Coppola, A. Critto, and A. Marcomini, “A multi-disciplinary approach to evaluate pluvial floods risk under changing climate: The case study of the municipality of Venice (Italy),” Science of The Total Environment, vol. 562, pp. 1031–1043, Aug. 2016, doi: 10.1016/j.scitotenv.2016.03.150.

[9] Y. Imamura, “Development of a Method for Assessing Country-Based Flood Risk at the Global Scale,” Int J Disaster Risk Sci, vol. 13, no. 1, pp. 87–99, Feb. 2022, doi: 10.1007/s13753-021-00388-w.

[10] “Home — Climate-ADAPT.” https://climate-adapt.eea.europa.eu/ (accessed Sep. 24, 2021).

[11] “ISO 31000:2018 - Risk management - A practical guide,” EVS. https://www.evs.ee/et/iso-31000-2018-risk-management-a-practical-guide (accessed Apr. 18, 2022).

[12] Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks (Text with EEA relevance), vol. OJ L. 2007. Accessed: Sep. 01, 2021. [Online]. Available: http://data.europa.eu/eli/dir/2007/60/oj/eng

[13] “UNDRR Understanding Risk,” UNDRR Understanding Risk. https://www.undrr.org/building-risk-knowledge/understanding-risk (accessed Mar. 12, 2022).

[14] “IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,” IPCC, 2022. Accessed: Apr. 12, 2022. [Online]. Available: https://report.ipcc.ch/ar6wg2/pdf/IPCC_AR6_WGII_FinalDraft_FullReport.pdf

[15] B. R. Rosenzweig et al., “Pluvial flood risk and opportunities for resilience - Rosenzweig - 2018 - WIREs Water - Wiley Online Library.” https://wires.onlinelibrary.wiley.com/doi/10.1002/wat2.1302 (accessed Mar. 12, 2022).

[16] C. Kelleher and L. McPhillips, “Exploring the application of topographic indices in urban areas as indicators of pluvial flooding locations,” Hydrological Processes, vol. 34, no. 3, pp. 780–794, 2020, doi: 10.1002/hyp.13628.

[17] “A rapid flood inundation model for hazard mapping based on least squares support vector machine regression”, doi: 10.1111/jfr3.12522.

[18] M. Truu, I. Annus, J. Roosimägi, N. Kändler, A. Vassiljev, and K. Kaur, “Integrated Decision Support System for Pluvial Flood-Resilient Spatial Planning in Urban Areas,” Water, vol. 13, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/w13233340.

[19] N. Kändler, I. Annus, and A. Vassiljev, “Controlling peak runoff from plots by coupling street storage with distributed real time control,” Urban Water Journal, vol. 19, no. 1, pp. 97–108, Jan. 2022, doi: 10.1080/1573062X.2021.1958235.

[20] “EVS 848:2021,” EVS. https://www.evs.ee/et/evs-848-2013 (accessed Sep. 01, 2021).

[21] Maa-amet, “Laadi kõrgusandmed alla.” https://geoportaal.maaamet.ee/est/Ruumiandmed/Korgusandmed/index.php?lang_id=1&page_id=614 (accessed May 05, 2022).

[22] Maa-amet, “Eesti põhikaart 1:10 000.” https://geoportaal.maaamet.ee/est/Ruumiandmed/Topokaardid-ja-aluskaardid/index.php?lang_id=1&page_id=30 (accessed May 05, 2022).

[23] S. Juhola, “Planning for a green city: The Green Factor tool,” Urban Forestry & Urban Greening, vol. 34, pp. 254–258, Aug. 2018, doi: 10.1016/j.ufug.2018.07.019.

[24] “SWMM-Docs: Open Water Analytics Stormwater Management Model.” http://wateranalytics.org/Stormwater-Management-Model/index.html (accessed Mar. 16, 2021).

[25] L. Warsta et al., “Development and application of an automated subcatchment generator for SWMM using open data,” Urban Water J., vol. 14, no. 9, pp. 954–963, 2017, doi: 10.1080/1573062X.2017.1325496.

[26] O. US EPA, “Storm Water Management Model (SWMM),” May 21, 2014. https://www.epa.gov/water-research/storm-water-management-model-swmm (accessed Sep. 01, 2021).

[26] PySWMM documentation, https://pyswmm.readthedocs.io/en/latest/overview.html, May 2022.

[27] ArcPy https://github.com/pmacMaps/ArcPy, May 2022.

7 ACKNOWLEDGEMENTS

The research was supported by the Estonian Research Council grant PRG667.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14145

TEMPORAL SCALE SIGMOID CURVE (TESIC): A TOOL TO CHARACTERIZE SHORT-TERM DEMAND VARIABILITY AT WATER SUPPLY SYSTEMS

Emilio Ruiz[1], **Sarai Díaz**[2] **and Javier González**[3]

[1]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain)
[2]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain)
[3]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain) &
Hidralab Ingeniería y Desarrollos, S.L., Spin-off UCLM, Hydraulics Laboratory, University of Castilla-La Mancha, Ciudad Real (Spain)

[1] *Emilio.Ruiz3@alu.uclm.es*, [2] *Sarai.Diaz@uclm.es*, [3] *Javier.Gonzalez@uclm.es*

Abstract

Water demand is the main random factor that conditions flow variability across water supply systems. Water demand measurements or pseudomeasurements (i.e. estimations based on historical data) are associated with a time interval (sampling rate), which affects the variability of water demands. Considering a long time interval implies losing information about water consumption within that temporal window. The variance (i.e. quantification of variability) computed from demand records is thus only "apparent", because the variability within the time interval is averaged and lost (i.e. "missed"). The relationship between missed and apparent variability can be assessed through the so-called TEmporal scale SIgmoid Curve (TESIC), which is here presented as a tool to characterize short-term demand variability. TESIC is used in this work to compute demand uncertainty for a given time resolution level and to estimate peak demands for different temporal resolutions in a realistic water supply case study. These applications show that TESIC provides a conceptual framework to explain and quantify the temporal resolution effect in hydraulic modelling applications.

Keywords
Water supply systems, smart meters, water consumption, uncertainty analysis, peak demand, high resolution.

1 INTRODUCTION

Water demands are stochastic in nature and constitute the main source of variability in water supply systems [1]. Consumption drives flow through the network, so it is paramount to have a good characterization of water demands to accurately simulate the hydraulic behaviour and associated water quality [2]. At a transport level, through the main arteries that supply water to District Metered Areas (DMAs), it may be enough to spatially and temporally average water demands [3]. This means that several users can be aggregated into a demand node (spatial averaging) and instantaneous variations in demand can be smoothed (temporal averaging). However, this approach is not suitable in the distribution mains that provide water to final users. Stochastic demand models were conceived to simulate the pulsating nature of water demand in these final pipelines (see [4] for references). Since these models particularly focus on what happens within a household (high spatial resolution), they are also associated with a very high temporal resolution, typically in the order of seconds (e.g. [5], [6]). This means that there is a large gap between the temporal resolutions typically adopted to model the macro network behaviour (15-60 minutes [7]) and the resolutions required to understand what happens on a micro household/fixture scale.

Temporal scale effects have been discussed in water supply systems for decades [8] and they are known for affecting the variability of water demands: considering longer time intervals (i.e. sampling rates) is associated with a loss of information from the consumption signal, leading to lower variance values [9]. Note that variance is here understood as a metric for demand variability. Several authors have worked on analysing the statistical properties of water consumption at different scales [1, 7, 10], but few have delved into how the sampling rate affects the ability to explain network behaviour and performance at different spatial aggregation levels. Note that the variance computed from demand/flow records is only the "apparent" variance, because the variability within the time interval is averaged and lost. This variability can therefore be accounted for as a "missed" variance. As presented by [11], the addition of the Apparent Variance (AV) and the Missed Variance (MV) provides the Total or instantaneous Variance (TV) for that time and spatial aggregation level, and both components are complementary. As a matter of fact, AV and MV describe a sigmoid curve when represented against the time interval. This S-shaped curved is called TESIC (TEmporal scale SIgmoid Curve) and can be computed either by progressively averaging measurement records with a very small time interval or by relying on a suitable stochastic demand model [12].

The aim of this work is to present the fundamentals of TESIC, as well as some related applications where this tool proves to be useful. More specifically, TESIC will be here used (1) to quantify the additional uncertainty derived from using larger time intervals than the monitoring time resolution [12], and (2) to compute peak demands considering different time intervals [13]. These application examples will show that TESIC provides a conceptual framework to explain short-term demand variability and combine information associated with different temporal scales. Therefore, it is relevant to bridge the gap between the micro and macro scales that coexist in water supply systems.

2 METHODOLOGY

2.1 TESIC conceptual framework and short-term variability scope

The relationship between AV, MV and TV was first explored in [11]. In this publication, an analytical approach to the end-use stochastic demand model SIMDEUM (SIMulation of water Demand, an End-Use Model) [6] was used to quantify not only the variance associated with a specific sampling rate (AV), but also the variability that is lost within the time interval (MV). This approach enables to compute AV, MV and TV at different times and for different sampling rates, proving that AV and MV are complementary. Figure 1 shows that AV is maximum for small time intervals (a high frequency sampling rate enables to capture most of the variability), keeping MV low. For long time intervals, there is a lot of variability not captured by the sampling rate (high MV), and AV is at its minimum. It is important to highlight that the curve presented in Figure 1 has two asymptotes at 0 and TV. This happens because TESIC focuses in "short-term" variability, which is unpredictable (i.e. fully random) and happens mostly within the 1-hour threshold.

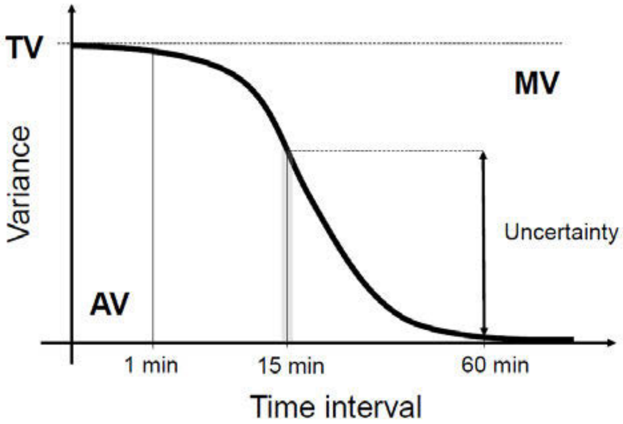


Figure 1. TESIC conceptualisation: Apparent (AV), Missed (MV) and Total (TV) Variances

Note that any water demand or flow time series can be decomposed into seasonality and noise. These components come determined by their period, so they can be associated with long and short-term variability. Short-term behaviour is here considered to take place with time periods below 1 hour, and periods above this threshold determine the mid and long-term behaviour [12]. Long-term variability is mainly associated with tendency, and it is more predictable because it comes mostly determined by the external factors that condition the population behaviour (e.g. working schedules, temperature, rainfall) [13]. Therefore, it is captured with usual demand sampling rates (above the hour) and/or it can be estimated with a predictive demand model (e.g. [14, 15, 16]). On the other hand, short-term variability is mostly random and associated with low correlation, i.e. it is poorly captured with usual sampling rates (above the hour) and more difficult to forecast. However, it can be estimated with a suitable stochastic demand model. Due to the limitations of the current models available to compute TESIC, these curves can only be applied to assess random short-term variability effects. TESIC definition is however universal, so it could be extended to cope with long and medium-term variability if micro demand models were combined with a suitable predictive model. This is a subject for further research.

TESIC has been presented up until now as a curve that can be derived from a stochastic demand model to characterize how short-term demand/flow variability changes with the temporal resolution. When computed like this, TESIC enables quantification of pseudomeasurement uncertainty [17], which may need to be additionally complemented with the medium/long-term component provided by a forecast model. However, TESIC can also be computed from available high-resolution measurement series. Infinitesimal time intervals are required to ensure that the original AV is representative of TV, but then the rest of the curve can be obtained by averaging AV over larger time intervals. Since flow metering devices with short time intervals are expensive and uncommon [2], TESIC estimation from a stochastic demand model is a good alternative to explore short-term variability. Note that if a model-based TESIC is to be compared with a measurement-based TESIC, the measurement series should be first treated to filter seasonality out [12], so that they both focus on short-term variability.

2.2 Demand time resolution uncertainty computation

TESIC provides the TV, AV and MV associated with a measured or pseudomeasured demand/flow at a specific time for different time intervals. Therefore, it can be useful to compute the uncertainty that would be induced in a hydraulic model by considering one sampling rate and not another. To explain this, it is important to highlight that, depending on the origin of the input, possible sources of uncertainty coexist [12]:

- Measurement uncertainty (for measured data): from the device metrological error

BY NC SA

- Pseudomeasurement uncertainty (for pseudomeasured data): from the inaccuracy of the considered estimation model
- Time resolution uncertainty (for either measured or pseudomeasured data): it appears when an input has a higher time interval than the time resolution required for modelling. This requires the input to be temporally disaggregated, introducing an additional uncertainty to the demand variable

Time resolution uncertainty can be quantified by making use of TESIC. Let us assume that a time resolution of 15 minutes is required for a specific hydraulic modelling application, but only demand measurements/pseudomeasurements with a 1-hour (i.e. 60 minutes) time interval are available. Time resolution uncertainty could be computed as the difference between the apparent variance associated with 15 minutes and the apparent variance for the 1-hour interval measurements/pseudomeasurements (see Figure 1). This uncertainty should be added to the measurement or pseudomeasurement uncertainty, and this total should be considered the input uncertainty for the modelling application. If measurements/pseudomeasurements with a time interval of 1 minute were available instead, averaging would be needed to adapt to the required time resolution, leading to a reduction in the perceived variability (AV).

It could be argued that in this case the best thing to do is to limit the temporal resolution of the model to 1 hour. However, this is simply not affordable for some applications. For example, this tool could be crucial to expand the scope of hydraulic state estimation from water transport networks to water distribution networks, where different sources of information with very different sampling rates coexist [12]. Since redundancy if of utmost importance for state estimation [18], TESIC is a necessary tool to maximize the number of measurements while coping with uncertainty correctly.

2.3 Peak demand computation

Peak demand has traditionally been defined as the maximum hourly consumption that takes place at a given spatial aggregation level (i.e. for a number of inhabitants N). Different empirical formulas are available to estimate the peak factor (ratio between the hourly maximum and the mean daily flow) based on the size of the population (see [19] for references), but such a deterministic approach is not in compliance with the now well-known stochastic nature of water demands. This explains why some authors have started to address peak demand analysis from a stochastic/probabilistic perspective (e.g. [20, 21]). The temporal and spatial resolution effect on peak factors is also a subject of ongoing research. Lower temporal scales are associated with greater peak demands, and 1-5 minute interval data have been found to be a good compromise solution to characterize maximum consumption [22, 23, 24].

Since the short-term variability of demand is characterized by TESIC, this curve can be used as a tool to explore the variability of peak demands for time intervals below 1 hour. Note that the demand/flow variance associated with a time resolution Δt and a number of inhabitants N (i.e. position of the network that provides water to N inhabitants) can be computed from the so called AV, i.e. $\sigma_Q(N,\Delta t) = AV$. This population size is also associated with a water demand mean value $\mu_Q(N)$, which remains constant regardless of the temporal framework adopted for the analysis and can be estimated according to [10] and [11]. Individual water consumption is not expected to follow a normal distribution, but this works assumes (and Section 3.3 shows) that when a sufficient number of inhabitants is considered, the aggregated water consumption can be assumed to follow a normal distribution with $\mu_Q(N)$ and $\sigma_Q(N,\Delta t)$. Therefore, the probability P of not exceeding a specific peak demand $Q_P(N,\Delta t)$ over a short time period where temporal homogeneity can be assumed [13] is:

$$P = \Pr[Q \leq Q_p(N,\Delta t)] = \Phi[Q_p(N,\Delta t), \mu_Q(N), \sigma_Q(N,\Delta t)] \tag{1}$$

Where Φ corresponds to the cumulative distribution function of the normal distribution. In order to analyse if peak demand values remain below $Q_P(N, \Delta t)$ not only for a time interval but over the whole temporal window (1 hour), and assuming that water demands behave independently, this probability should be modified as:

$$P = \Phi[Q_p(N, \Delta t), \mu_Q(N), \sigma_Q(N, \Delta t)]^{\frac{3600}{\Delta t}}; \forall N, \forall \Delta t \tag{2}$$

Note that $\mu_Q(N)$ and $\sigma_Q(N, \Delta t)$ refer to the statistical properties of the original normal distribution at peak hour. Both can be computed according to the microcomponent stochastic demand model presented in [10] and [11]. More specifically, the variance corresponds to the AV of the corresponding TESIC curve (position that supplies water to N inhabitants) at peak hour and for a Δt time interval.

The key interest of using TESIC rather than an empirical formula derived for a particular set of measurements is that this approach provides a conceptual framework, i.e. it is not site specific. Note that there are different ways of benefiting from TESIC depending on the available information [13]. If a consistent end-use model is available, $\mu_Q(N)$ and $\sigma_Q(N, \Delta t)$ can be computed at peak hour, and TESIC can be inferred. If there is no access to a microcomponent demand model, but a TESIC curve is available at peak hour (e.g. shape extrapolated from other similarly populated areas), the effect of short-term variability on peak demands can still be assessed. This is especially interesting when only low-resolution (i.e. per hour) measurements are available, because it enables to pre-evaluate the effect of adopting higher resolutions (e.g. installing new high resolution meters). If there is no access to a microcomponent demand model or a TESIC curve at all, TESIC (or a partial TESIC) could still be inferred from measurements. Extrapolation to other population sizes would be possible if the population is homogeneous (i.e. short-term effects predominate).

3 RESULTS

The archetypical network presented by [25] is here adopted (see Figure 2) to show the different TESIC curves (and so the order of magnitude of demand uncertainty) that would be obtained for different spatial aggregation levels in a realistic case study. A microcomponent demand model has already been adjusted at this network according to metered data and statistical information [12], so TESIC is computed directly from the demand model. The same case study is later on used to show the interest of the peak demand assessment strategy here presented.

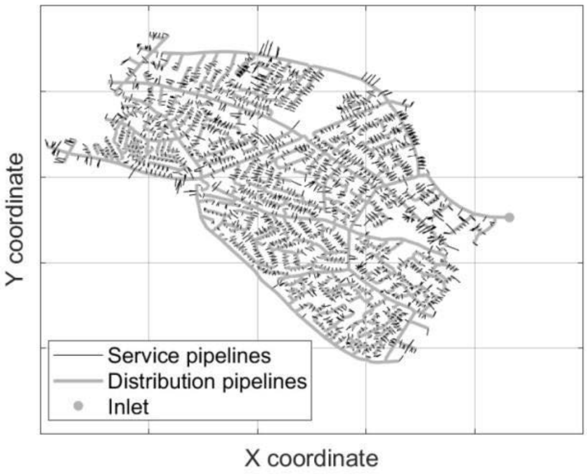


Figure 2. Case study layout

BY NC SA

3.1 TESIC

TESIC have been computed in this work at peak hour (10:00) for different spatial aggregation levels (N=100, 1000, 10000 and 79106 – total population). Figure 3a shows that the short-term variance lowers as the time interval increases, reaching the zero value for 1 hour (short-term threshold). Moreover, TV increases with the number of inhabitants, and so does AV for the different scales here considered. Figure 3b shows that the tendency is opposite in terms of Coefficient of Variation (CV): coefficients of variation are maximum for the minimum number of inhabitants here considered (N=100) and reduce as N increases. Results show that relative variability is bound to reduce when aggregating end-users. This highlights the importance of quantifying uncertainty when dealing with low spatial aggregation scales.

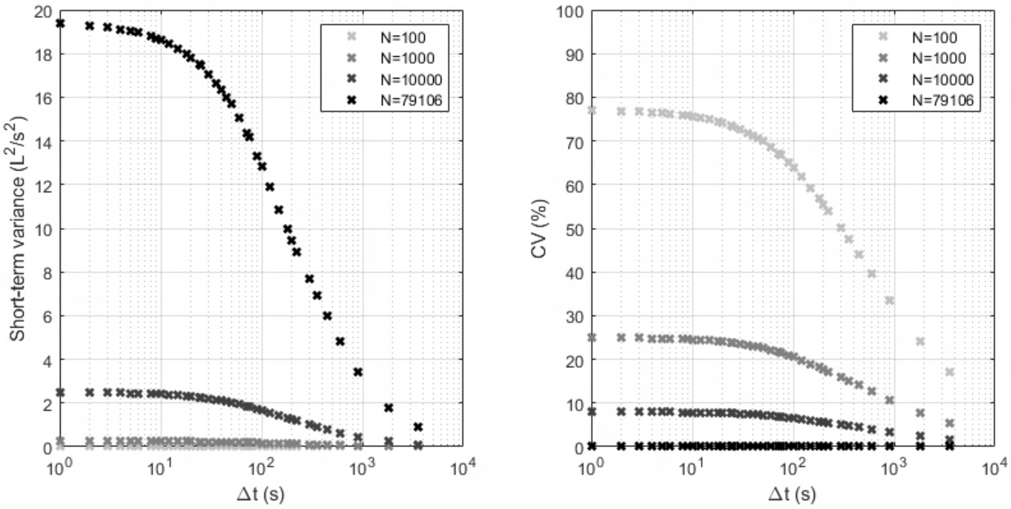


Figure 3. TESIC at peak hour for different spatial aggregation levels: a) original short-term variance curve and b) associated coefficient of variation

3.2 Demand time resolution uncertainty computation

In order to estimate how much uncertainty would be induced in a hydraulic model by changing the time resolution of aggregated water demands, Figure 4 presents a normalized version of the TESIC curve. Note that this curve remains approximately the same no matter the spatial aggregation level because population characteristics have been assumed homogeneous in the full network area in this example case study [10, 11]. This enables to estimate how much uncertainty should be added as a result of a change in temporal resolution. Note that if 1-hour (3600 s) time interval measurements were available and a 1-hour time resolution was required, the associated time resolution uncertainty would be null. However, if a 15-minute (900 s) time resolution was required, about 20% of the total variance for that spatial aggregation scale should be incorporated as uncertainty to the hydraulic model. If a 5-minute (300 s) time resolution was the target, this value would increase up until 36%, reaching 80% for 1 minute (60 s) and 100% if an instantaneous resolution was required. This figure shows the importance of having measurements with a sampling rate similar to the required time resolution (note that measurement or pseudomeasurement uncertainty should be added to compute the total uncertainty as explained in Section 2.2), and should serve a motivation to install convenient smart meters through the system. This conclusion is relatively intuitive and could have been anticipated, but the interest of TESIC lies in being able to quantify how much uncertainty should be included if there is no option to reduce the time interval of the available measurements/pseudomeasurements and/or if different sources of information with different sampling rates coexist.

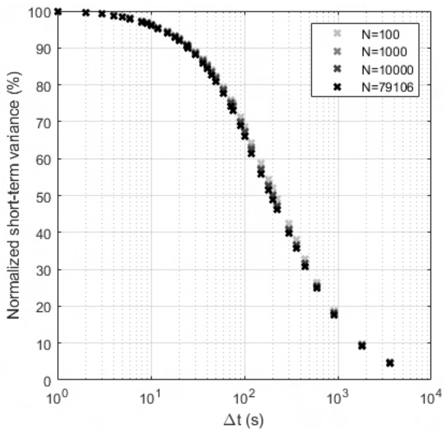


Figure 4. Normalized TESIC for different spatial aggregation levels

3.3 Peak demand computation

Probabilities of not exceedance have been computed for different threshold values $Q_P(N, \Delta t)$ in this case study: $\mu_Q(N)$; $\mu_Q(N) + \sigma_Q(N, \Delta t)$; $\mu_Q(N) + 2\sigma_Q(N, \Delta t)$ and $\mu_Q(N) + 3\sigma_Q(N, \Delta t)$. Figure 5 shows the analytically computed probabilities of not exceedance according to Eq (2) and 1000 Monte Carlo simulations. The probability of not exceedance has been computed in the numerical case by: 1) generating 1000 per-second demand series over the peak hour for each spatial aggregation level, 2) averaging these demand series over each Δt time interval and 3) counting the number of times that the maximum average value stays below each threshold value. Figure 5 shows similar results for the analytical and numerical simulations for time intervals above 1-2 minutes. Therefore, the analytical approach is adequate to quantify the temporal scale effect of peak demands in the range of time intervals that have been identified as a good compromise for maximum consumption assessment (1-5 minutes [20, 21, 22]). Note that the numerical and the analytical approach have a better correspondence as the number of inhabitants increases. Since this approach assumes that flows behave as a normal distribution and this is only true when a sufficient number of inhabitants is aggregated, the minimum number of inhabitants analysed in this work is 100 (equivalent to a few residential buildings).

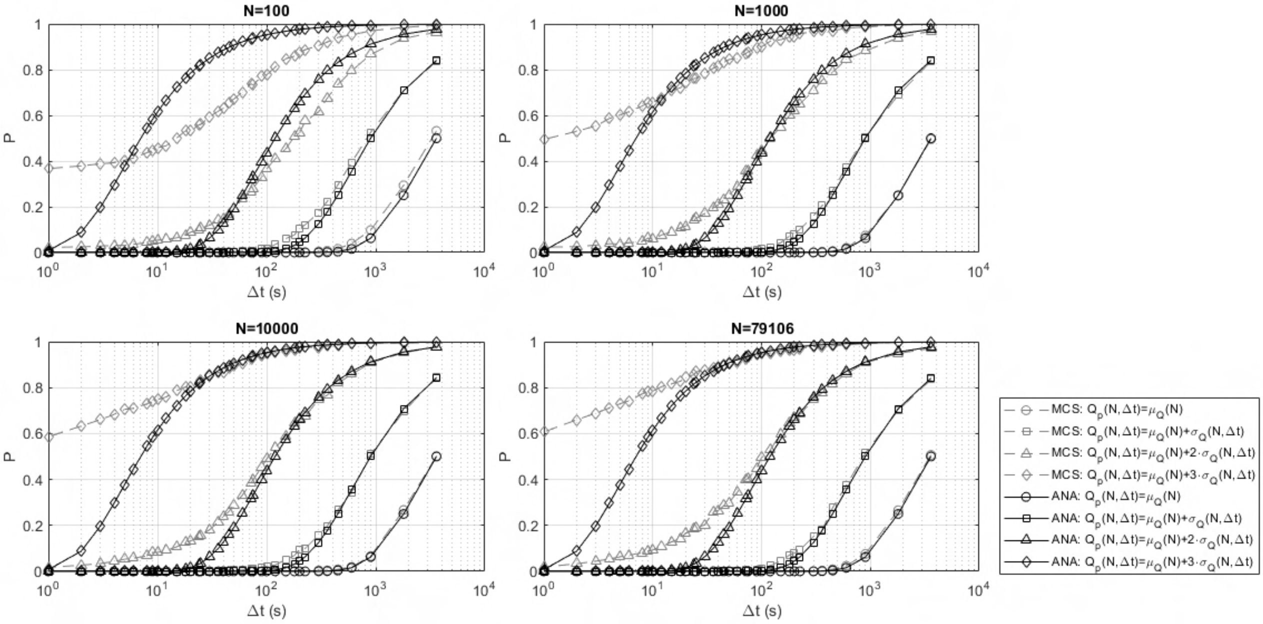


Figure 5. Probability of not exceedance for different peak demand values and spatial aggregation levels: analytical approach vs Monte Carlo simulation

4 CONCLUSIONS

TESIC has been here presented as a conceptual framework to explain and quantify the effect of adopting different time intervals/temporal resolutions to account for the variability of water demands. More specifically, it has been applied to quantify the uncertainty stemming from using measurements/pseudomeasurements with time intervals greater than the required temporal resolution in a hydraulic modelling application and to explore the temporal variability effect of peak demands. These applications show that now that smart meters are gaining importance, but technology is still limited to provide high spatial and temporal resolution measurements through entire systems continuously, TESIC could be a powerful tool to maximise the available information, because it allows to combine different sources of information by properly quantifying uncertainty.

5 ACKNOWLEDGEMENTS

The authors would like to thank the financial support provided by the Spanish Ministry of Science and Innovation - State Research Agency (Grant PID2019-111506RB-I00 funded by MCIN/AEI/10.13039/ 501100011033) and Junta de Comunidades de Castilla-La Mancha (Grant SBPLY/19/180501/000162 funded by Junta de Comunidades de Castilla-La Mancha and ERDF A way of making Europe).

6 REFERENCES

[1] R. Magini, I. Pallavicini and R. Guercio, “Spatial and Temporal Scaling Properties of Water Demand,” Journal of Water Resources Planning and Management, vol. 134, no. 3, 2008, pp. 276-284.

[2] E.J.M. Blokker, J.H.G. Vreeburg, S.G. Buchberger and J.C. van Dijk, “Importance of Demand Modelling in Network Water Quality Models: A Review,” Drinking Water Engineering and Science, vol. 1, 2008, pp. 27-38.

[3] S.G. Buchberger and L. Wu, “Model for instantaneous residential water demands,” Journal of Hydraulic Engineering, vol. 121, no. 3, 1995, pp. 232-246.

[4] E. Creaco, M. Blokker and S. Buchberger, “Models for generating household water demand pulses: Literature review and comparison,” Journal of Water Resources Planning and Management, vol. 143, no. 6, 2017, 04017013.

[5] S.G. Buchberger and G.J. Wells, “Intensity, Duration and Frequency of Residential Water Demands”, Journal of Water Resources Planning and Management, vol. 122, no. 1, 1996, pp. 11-19.

[6] E.J.M. Blokker, J.H.G. Vreeburg and J.C. van Dijk, “Simulating residential water demand with a stochastic end-use model”, Journal of Water Resources Planning and Management, vol. 136, no. 1, 2010, pp. 19-26.

[7] P. Kossieris, C. Makropoulos, “Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales”, Water, vol. 10, no. 1481, 2018.

[8] H. Tessendorff, “Problems of Peak Demands and Remedial Measures”, in Proc. 9th Congress of International Water Supply Association, Reston, Virginia, USA, 1972, pp. 10-14.

[9] S.G. Buchberger and G. Nadimpalli, “Leak Estimation in Water Distribution Systems by Statistical Analysis of Flow Readings”, Journal of Water Resources Planning and Management, vol. 130, no. 4, 2004, pp. 321-329.

[10] S. Díaz and J. González, “Analytical Stochastic Microcomponent Modelling Approach to Assess Network Spatial Scale Effects in Water Supply Systems”, Journal of Water Resources Planning and Management, vol. 146, no. 8, 2020, pp. 04020065.

[11] S. Díaz and J. González, “Temporal Scale Effect Analysis for Water Supply Systems Monitoring Based on a Microcomponent Stochastic Demand Model”, Journal of Water Resources Planning and Management, 2021, vol. 147, no. 5, 04021023.

[12] E. Ruiz, S. Díaz and J. González, “Potential Performance of Hydraulic State Estimation in Water Distribution Networks”, Water Resources Management, vol. 36, 2022, pp. 745-762.

[13] S. Díaz and J. González, "Short-Term Variability Effect on Peak Demand: Assessment Based on a Microcomponent Stochastic Demand Model", Water Resources Research, vol. 58, no.1, 2022, e2021WR030532.

[14] S.L. Zhou, T.A. McMahon, A., Walton and J. Lewis, "Forecasting Operational Demand for an Urban Water Supply Zone", Journal of Hydrology, vol. 259, 2002, pp. 189-202.

[15] I. Okeya, Z. Kapelan, C. Hutton and D. Naga, "Online Modelling of Water Distribution System Using Data Assimilation", Procedia Engineering, vol. 70, 2014, pp. 1261-1270.

[16] M. Xenochristou and Z. Kapelan, "An Ensemble Stacked Model with Bias Correction for Improved Water Demand Forecasting", Urban Water Journal, vol. 17, no. 3, 2020, pp. 212-223.

[17] A. Bargiela and G.D. Hainsworth, "Pressure and Flow Uncertainty in Water Systems", Journal of Water Resources Planning and Management, vol. 115, no. 2, 1989, pp. 212-229.

[18] E. Ruiz, S. Diaz and J. González, "Hydraulic State Estimation: Pilot Implementation in a Water Distribution System", in Proc. 2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry, Valencia, Spain, 2022.

[19] G. Balacco, A. Carbonara, A. Gioia, V. Iacobellis and A. Piccinni, "Evaluation of Peak Water Demand Factors in Puglia (southern Italy)", Water, vol. 9, no. 96, 2017.

[20] X. Zhang, S. Buchberger and J. van Zyl, "A Theoretical Explanation for Peaking Factors", in Proc. Of the ASCE EWRI Conference, Anchorage, Alaska, USA, 2005.

[21] E. Creaco, P. Signori, S. Papiri and C. Ciaponi, "Peak Demand Assessment and Hydraulic Analysis in WDN Design", Journal of Water Resources Planning and Management, vol 144, no. 6, 2018, 04018022.

[22] G. de Marinis, R. Gargano and A. Leopardi, "Un Laboratorio di Campo per il Monitoraggio di una Rete Idrica: Richieste di Portada – Primi Risultati" (in italian), in Proc. La Ricerca Delle Perdite a la Gestione Delle Reti di Acquedotto, Perugia, Italy, 2003.

[23] C. Tricarico, G. de Marinis, R. Gargano and A. Leopardi, "Peak Residential Water Demand", in Proc. Of the ICE – Water Management Journal, vol 160, 2007, pp 115-121.

[24] S. Gato-Trinidad and K. Gan, "Characterizing Maximum Residential Water Demand", WIT Transactions on the Built Environment, vol 122, 2012, pp. 15-24.

[25] E. Ruiz, Comprehensive Application of State Estimation Methodology to a District Metered Area, Master Thesis, University of Castilla-La Mancha (Spain), 2020.

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14152

MULTIDIMENSIONAL DATA GENERATION IN WATER DISTRIBUTION SYSTEMS USING THE CYCLE-GAN

Sehyeong Kim[1], Donghwi Jung[2]

[1]Master Student, School of Civil, Environmental and Architectural Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul, 02841, Republic of Korea

[2]Assistant Professor, School of Civil, Environmental and Architectural Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul, 02841, Republic of Korea

[1] kaiba0514@korea.ac.kr, [2] sunnyjung625@korea.ac.kr

Abstract

A water distribution system (WDS) consists of thousands of components interacting with each other. To analyze the status or to manage the operation of the WDSs, two main feature values have been mainly used: nodal pressure and pipe flow rate. However, insufficient data due to the malfunction of the sensors or economical limitations interrupts the collection of abundant information in many cases. Therefore, this study proposes a WDS data generation model based on the cycle-GAN (Generative Adversarial Networks). The proposed model learns the demand time series data and its corresponding time series data of pressure or flow. All training data is two-dimensionally constructed by considering the time series data for 24 hours of each component as a single row and arranging all row data vertically. After normalizing all the data to integers from 0 to 255, they become a greyscale image. Then, the cycle-GAN model consisting of two generators and two discriminators trains those image datasets, to translate the demand image data to the WDS feature image data (i.e., pressure or flow). Firstly, based on the random seeds, the first generator in the cycle-GAN model is trained to generate the demand image, and the second generator is trained to generate the WDS feature image. After the fundamental training for the generation of their own data, the second generator trained for the feature image data starts to use the synthesized demand image results from the first generator as its seeds, not using the random noise. This process makes the second generator have the ability to translate the demand images to the WDS feature images by using the convolutional and deconvolutional functions in the neural network layers. This model was demonstrated by applying to the Mays network, which is the benchmark network consisting of 13 nodes and 21 pipes with two reservoirs.

Keywords
Water distribution systems, Deep-learning-based hydraulic analysis, Multidimensional data generation, Image translation, Generative adversarial networks, Cycle-GAN.

1 INTRODUCTION

Improvements in monitoring technology have led it possible to set diverse sensors, including pressure and pipe flow rate, to measure the state of water distribution systems (WDSs). However, the cities have been also developed concentratively but extensively and the number of the critical points whose status data should be collected has been increased, making their WDSs more difficult to be completely analyzed. Even worse, the reliability of the sensors is not enough to observe the intact data if there are some malfunctions in the equipment or environmental fluctuations. Due to these limitations of the monitoring systems, the collectable data becomes more sparse and some studies have tried to supplement the missing values or empty space of the data that they cannot directly measure.

Some univariate and multivariate imputation methodologies using mean, median, and regression-based estimation were applied to the sparse WDS data and the sequential imputation that can

BY NC SA

consider the time variable was also utilized (Sankaranarayanan et al. 2019). Furthermore, machine learning techniques such as Kalman-filter-based nearest neighbor regressor or random forest regressor are also used to be assessed as statistical imputation models (Kabir et al. 2020, Rodriguez et al. 2021). These kinds of sparsity problems are not only in the WDS domain, and other fields like environmental engineering have widely used various techniques to construct the complete composition of the data-driven models (Zheng et al. 2016, Dumedah and Coulibaly 2011). For example, the subsurface environmental status data has been usually imputed using the techniques mentioned above, because it is not easy to monitor it in real-time. However, there is little effort to use the generative deep learning models for the imputation method although it has been appraised as one of the most important developments in the generation and the translation of the multidimensional data.

Therefore, this study suggests the multidimensional data translation model based on the cycle-GAN framework, especially transforming the demand time series data to the WDS feature data like nodal pressure or pipe flow rate. There are two generators in the cycle-GAN model, and the first generator is trained to synthesize the demand image data, while the second generator is trained to synthesize the WDS feature (i.e., pressure or flow) image data, based on the random seeds. After those two generators are fundamentally trained to synthesize their images in charge, the second generator starts to use the demand image outputs from the first generator as its seeds. Then, it can secure the ability to translate the demand data to the pressure or the flow data. As a result, this mechanism can provide the estimated WDS data if there is continuous missing data regardless of the cause.

2 METHODOLOGY

2.1 Generative Adversarial Networks (GAN)

Adversarial training is presented as a new way to train the generative models (Goodfellow et al. 2014). The generative adversarial networks (GAN) consist of two adversarial networks: generator and discriminator. The purpose of the generator G is to learn the existing data distribution to generate data as similar as possible, while the discriminator D is to distinguish whether the data given as input is synthesized by G or extracted from the real data (i.e., training data) distribution. The G constructs a mapping function that leads from a random noise distribution to a real data distribution in the form of a multi-layer perceptron. On the other hand, D receives the given input and derives the probability that the input came from the real data as a result. G and D are simultaneously learned. G learns to maximize the probability that the results generated from latent space (random noise) $\boldsymbol{z}$ used as an input is discriminated as real data by D. D learns to increase the discrimination performance between the real and the generated data itself. In other words, the result generated from the G is trained to minimize the probability that the result is determined as real data. It is in the form of a minimax game of two players, the generator and the discriminator, and it can be represented in equation (1):

$$\min_{G}\max_{D} V(D,G) = E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + E_{\boldsymbol{z}\sim p_{data}(\boldsymbol{z})}[\log\{1 - D(G(\boldsymbol{z}))\}] \tag{1}$$

2.2 Multidimensional Data Translation: Cycle-GAN

As a baseline model for data generation through image translation, cycle-GAN (Zhu et al. 2017) is adopted in this paper. The purpose of training cycle-GAN is to learn two generators and two discriminators that can travel between two different domains. Let the two domains be X and Y, respectively, and the generator translating X to Y, be G_X, while the generator translating Y to X, be G_Y. The discriminator D_X of the X domain aims to better discriminate between the data $\boldsymbol{x}$ in the X domain, and $G_Y(\boldsymbol{x})$ translated through the G_Y in the Y domain. Conversely, the discriminator D_Y

BY NC SA

in the Y domain aims to better discriminate between the data $\boldsymbol{y}$ in the Y domain and $G_1(\boldsymbol{x})$. These processes can be represented in equation (2) and (3):

$$\min_G \max_D V(D_X, G_Y) = E_{x\sim p_{data}(x)}[\log D_X(\boldsymbol{x})] + E_{y\sim p_{data}(y)}[\log\{1 - D(G_Y(\boldsymbol{y}))\}] \tag{2}$$

$$\min_G \max_D V(D_Y, G_X) = E_{y\sim p_{data}(y)}[\log D_Y(\boldsymbol{y})] + E_{x\sim p_{data}(x)}[\log\{1 - D(G_X(\boldsymbol{x}))\}] \tag{3}$$

In the proposed model, X domain will be the demand image data distribution and Y domain will be the feature image data distribution. Therefore, the model $G_{\boldsymbol{x}}$ will be trained to mimic and synthesize the demand image, and the feature data will be synthesized by the model G_y, while their corresponding discriminators try to classify the real dataset images and the synthesized images from $G_{\boldsymbol{x}}$ or G_y.

3 STUDY NETWORK

3.1 WDS Information

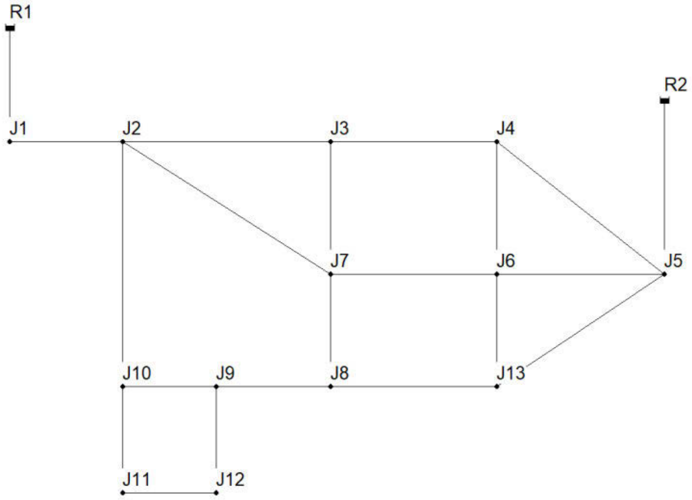


Figure 1. The configuration of Mays network

The proposed model is applied to the Mays network, a widely used benchmark network. Mays network consists of 13 nodes, 21 pipes, and two reservoirs. The configuration of the Mays network is represented in *Figure 1*, and the detailed information on the Mays network is the same as below.

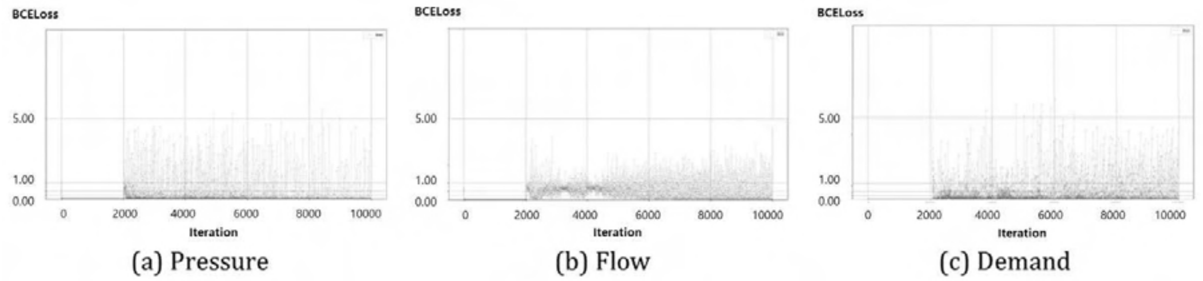
- The total levels of the two reservoirs are both 60.96m.
- Extended period simulation (EPS) is possible by setting a demand pattern for 24 hours.
- The demand pattern has a sub-pattern with one-hour intervals to make a distinctive characteristic with the stripe pattern in the two-dimensional time series image.
- The uncertainty (i.e., variance) of the demand is limited with around 1/100 of each data value, to avoid the unnecessary confusion of the image patterns.

3.2 WDS Image Construction

The training image data is constructed in a form of a two-dimensional matrix for each data. One row of the image represents the 24-hour time series data of one component (i.e., a node or a pipe) with 5-minute time steps, and one column of the image represents the status of all components at a single time step. Therefore, according to the feature of the Mays network, the pressure image and demand image has 13 rows and the flow image has 21 rows, the same as the number of the nodes and pipes, respectively. All data are normalized within the integer from 0 to 255, to transform them into greyscale images. Similar stripe patterns in the pressure and flow images can be observed due to the one-hour pattern of the demand. Total of 10,000 images for each data are generated to be trained.

4 APPLICATION RESULTS

When the cycle-GAN model receives the image as training data, they fundamentally train to generate them based on the random noise from the latent space. In the proposed model, the generative function in layers of the neural network was the 2-D deconvolutional function. Their training performance can be represented by the losses of the discriminators (*Figure 2*). It can be observed that the loss of each discriminator started to fluctuate after around a fifth of the total 10,000 iterations. It means that the discriminators' performance started to be deteriorated due to the synthesized results from the generators having become realistic successfully.

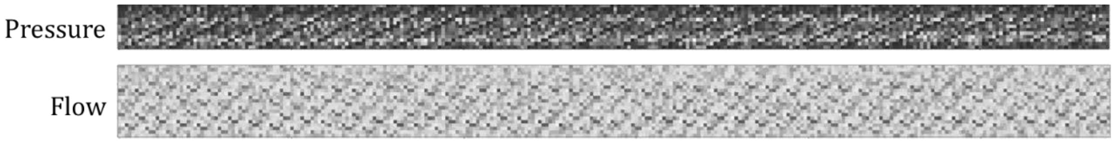


(a) Pressure (b) Flow (c) Demand

Figure 2. The loss graph of the discriminators based on the random noise in the latent space

Then, the data translation started after the fundamental training based on the random noise is completed. In principle, the cycle-GAN model can translate the data bidirectionally, but only demand-to-pressure and demand-to-flow cases were considered, taking into account the principle of the demand-based hydraulic analysis of the WDSs. After the training process for the image translation is finished, the model can generate the feature images based on the real demand images. The samples of the generated WDS feature values which were generated based on the real demand data are represented in *Figure 3*. It can be observed that the stripe patterns that can be found in the real image data used as the training datasets, although there was some unclearness.

Generated Demand Image (based on Random Noise)



Translated Quality Characteristics Images (based on Real Demand Image)

Figure 3. The generated demand image and the translated WDS feature images

To calculate the training performance of the translated results quantitatively, mean, median and mode with the class interval 10 were computed (Table 1). The differences between the original data and the generated data were not large, which means that the generators for each feature have good performance to synthesize the data. The mean values of the data had less than 3% of the normalized figures, and the median, and the mode values indicated the fine durability of the model representing stable distribution. Considering the results of this statistical analysis, it can be said that the suggested results showed that the proposed translation model can be used for data imputation in WDS, generating realistic data values with the original feature data.

Table 1. Statistically comparison of the normalized data values of the original and generated WDS features

Feature	**Mean**	**Median**	**Mode (Class interval: 10)**
Original Pressure	131.28	130	130-140
Generated Pressure	133.59	134	130-140
Original Flow	126.08	126	120-130
Generated Flow	124.89	123	120-130

5 CONCLUSIONS

This paper suggests the direction of utilizing multidimensional processing neural network technology as a translation tool for the hydraulic analysis of the WDSs, by using the cycle-GAN based on the demand and the WDS feature data. The proposed model learns the demand time series data and its corresponding time series data of pressure or flow. All training data is two-dimensionally constructed by considering the time series data for 24 hours of each component as a single row and arranging all row data vertically. After normalizing all the data to integers from 0 to 255, they become to a greyscale image. Then, the cycle-GAN model consisting of two generators and two discriminators trains those image datasets, to translate the demand image data to the feature image data (i.e., pressure or flow). The process of the suggested model makes the generator have the ability to translate the demand images to the feature images by using the convolutional and deconvolutional functions in the neural network layers. This model was demonstrated by applying to the small benchmark WDS, the Mays network. Finally, in the image data translation, the results indicated that the demand-to-flow translation is more successfully worked than the demand-to-pressure translation, showing the lesser RMSE errors than the demand-flow translation. The suggested results showed that the proposed translation model can be used for data imputation in WDS.

Besides, some possibilities for future studies were suggested based on the results of this study. Firstly, the bidirectional translation can be considered such as pressure(flow)-to-demand or the pressure-to-flow. As mentioned in the paper, the generators in the cycle-GAN model can travel between the two different distribution domains, but the proposed model only demonstrated one-side direction considering the analyzing approach. If this bidirectional translation is applied, data generation for the other feature estimations will be enabled. Additionally, the pressure-driven analysis (PDA) approach for the hydraulic data generation using the EPANET program can be considered. Due to the effectiveness of practical and realistic analysis of the WDSs, PDA has been intensively studied to be used more generally, especially in the simulation of abnormal circumstances. If this approach is applied to the proposed model, the various multidirectional translation and their corresponding data generation in the WDS data can be more valuable.

6 ACKNOWLEDGEMENT

This work was supported by the Korean Ministry of the Environment as a Subsurface Environment Management (SEM) project (No. 2021002170003)

7 REFERENCES

[1] Dumedah, G., & Coulibaly, P. (2011). Evaluation of statistical methods for infilling missing values in high-resolution soil moisture data. Journal of Hydrology, 400(1-2), 95-102.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[3] Hu, P., Tong, J., Wang, J., Yang, Y., & de Oliveira Turci, L. (2019, June). A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. In 2019 IEEE Congress on evolutionary computation (CEC) (pp. 1088-1094). IEEE.

[4] Jung, D., & Lansey, K. (2015). Water distribution system burst detection using a nonlinear Kalman filter. Journal of Water Resources Planning and Management, 141(5), 04014070.

[5] Jung, D., Yoo, D. G., Kang, D., & Kim, J. H. (2016). Linear model for estimating water distribution system reliability. Journal of Water Resources Planning and Management, 142(8), 04016022.

[6] Kabir, G., Tesfamariam, S., Hemsing, J., & Sadiq, R. (2020). Handling incomplete and missing data in water network database using imputation methods. Sustainable and Resilient Infrastructure, 5(6), 365-377.

[7] Kang, D., & Lansey, K. (2010). Optimal meter placement for water distribution system state estimation. Journal of Water Resources Planning and Management, 136(3), 337-347.

[8] Kumar, S. M., Narasimhan, S., & Bhallamudi, S. M. (2008). State estimation in water distribution networks using graph-theoretic reduction strategy. Journal of Water Resources Planning and Management, 134(5), 395-403.

[9] Misiunas, D., Lambert, M., Simpson, A., & Olsson, G. (2005). Burst detection and location in water distribution networks. Water Science and Technology: Water Supply, 5(3-4), 71-80.

[10] Perelman, L., & Ostfeld, A. (2007). An adaptive heuristic cross-entropy algorithm for optimal design of water distribution systems. Engineering Optimization, 39(4), 413-428.

[11] Rodríguez, R., Pastorini, M., Etcheverry, L., Chreties, C., Fossati, M., Castro, A., & Gorgoglione, A. (2021). Water-quality data imputation with a high percentage of missing values: A machine learning approach. Sustainability, 13(11), 6318.

[12] Romano, M., Kapelan, Z., & Savić, D. A. (2014). Automated detection of pipe bursts and other events in water distribution systems. Journal of Water Resources Planning and Management, 140(4), 457-467.

[13] Sankaranarayanan, S., Swaminathan, G., Radhakrishnan, T. K., & Sivakumaran, N. (2019). Missing data estimation and IoT-based flyby monitoring of a water distribution system: Conceptual and experimental validation. International Journal of Communication Systems, e4135.

[14] Shuang, Q., Liu, Y., Tang, Y., Liu, J., & Shuang, K. (2017). System reliability evaluation in water distribution networks with the impact of valves experiencing cascading failures. Water, 9(6), 413.

[15] Suribabu, C. R. (2010). Differential evolution algorithm for optimal design of water distribution networks. Journal of Hydroinformatics, 12(1), 66-82.

[16] Tabesh, M., Jamasb, M., & Moeini, R. (2011). Calibration of water distribution hydraulic models: A comparison between pressure dependent and demand driven analyses. Urban Water Journal, 8(2), 93-102.

[17] Tabesh, M., Shirzad, A., Arefkhani, V., & Mani, A. (2014). A comparative study between the modified and available demand driven based models for head driven analysis of water distribution networks. Urban Water Journal, 11(3), 221-230.

[18] Xing, L., & Sela, L. (2022). Graph Neural Networks for State Estimation in Water Distribution Systems: Application of Supervised and Semisupervised Learning. Journal of Water Resources Planning and Management, 148(5), 04022018.

[19] Ye, G., & Fenner, R. A. (2011). Kalman filtering of hydraulic measurements for burst detection in water distribution systems. Journal of pipeline systems engineering and practice, 2(1), 14-22.

[20] Yi, X., Zheng, Y., Zhang, J., & Li, T. (2016, June). ST-MVL: filling missing values in geo-sensory time series data. In Proceedings of the 25th International Joint Conference on Artificial Intelligence.

[21] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14153

AN EXPERIMENTAL STUDY ON EARLY LEAK LOCALIZATION IN DRINKING WATER NETWORKS USING PRESSURE MEASUREMENTS

Yannick Deleuze[1], Arley Nova-Rincon[2], Yves-Marie Batany[3], Teodulo Abril[4], Damien Chenu[5] and Nicolas Roux[6]

[1]Veolia - Scientific & Technological Expertise Department, Maisons-Laffitte, France,
[2]Veolia - Scientific & Technological Expertise Department, Limay France,
[3]Veolia - Scientific & Technological Expertise Department, Limay, France,
[4]Veolia - Technical and Performance Department Latam, Bogotá, D.C., Colombia,
[5]Veolia - Scientific & Technological Expertise Department, Limay, France
[6]Veolia - Scientific & Technological Expertise Department, Aubervilliers, France

[1] *yannick.deleuze@veolia.com, [2]arley.nova-rincon@veolia.com, [3]yves-marie.batany@veolia.com,*
[4] *teodulo.abril@veolia.com, [5] damien.chenu@veolia.com, [6]nicolas.roux@veolia.com*

Abstract

Leaks represent a major issue impacting the management and efficiency of Drinking Water Networks (DWN) in cities worldwide. According to the Development Bank of Latin America, by 2018 the losses in DWN range from 40 to 60% in the region. In Europe, the OECD reports a wider range with few losses in cities like Amsterdam (4%) up to 37% in Naples. With this context, some regional policies have emerged like the 2020 european drinking directive “Right2Water”, that aims to encourage major suppliers (more than 50000 users), to develop tools to measure and reduce leakages by 2025. Considering this situation, we introduce here a systematic approach for leak management that combines field data, hydraulic models (HM) and machine learning.

Model-based and data driven methods have been of great interest for leak location methodologies in DWN. This research will design energy-efficient and cost-efficient leak localization hotspots in the DWN. The approach is intended for sectorized DWN, equipped with a SCADA system and where a calibrated hydraulic model (e.g. EPANET) is available. This latter serves to evaluate the sensitivity of the system to leaks and identify potential points for pressure measurements in order to optimise the number of installed sensors.

Given a detected leak in the network, a multiclass classifier using pressure data is developed to reduce the inspected pipe length for the leak location. The leakage localization method is implemented combining multiple individual classifiers using ensemble learning methods and a reduced number of decision variables.

The methodology is tested on a real case study from a Colombian site. The method faces challenges in (a) collecting correctly labelled real leak data, and (b) modelling and calibrating hydraulic models. Those challenges are being addressed. The outcome shows that the length of pipes inspected can be reduced by one third with high performance in accuracy with few sensors required (low capital expenditures) and low computational effort (low energy and low operational expenditures).

Keywords
Water distribution systems, Leak zone location, Hydraulic modelling, Mixed model-based/data driven methods.

1 INTRODUCTION

The constant increase of urban populations requires a more efficient management of the water resources. One of the aspects that affects the efficient use of water resources are leaks in DWNs [1], representing the main source of water physical losses in these systems [2]. They can cause significant economic losses in fluid transportation leading to increased reparations and therefore, a potential extra cost for the final user [3]. Since it is a major issue in the performance of DWN, leak detection and management have been a subject of studies with different approaches and important evolutions as detailed in different literature reviews [2, 4, 5].

Leak localization is frequently performed using acoustic methods using hardware such as leak correlators and leak noise loggers and other non-acoustic methods [2]. These methods are accurate, however, it can take extensive labour effort to find a leak, even in small District Metered Areas (DMAs). Leak localization processes can benefit from software-based methods using mathematical modelling in addition to pressure and flow field measurements. Usually, the leak is localised in a zone (a group of network nodes). Hardware-oriented technologies can then be used to identify the exact location of the leak at pipeline level. Hu *et. al*. [6] provide a detailed literature review on model based and data-driven approaches for leak detection and localization. Some leak localization approaches can reduce the area of interest and facilitate traditional DMA techniques to locate leaks. In addition to the possibility of saving large volumes of water, software-based methods also represent financial savings when using only a few pressure sensors.

Mixed model-based/data driven approaches have been developed aiming to predict leak locations. These methods consist in the (i) construction of a hydraulic model, (ii) calibration of the hydraulic model, (iii) leak detection, and (iv) leak localization. After a leak has been detected, it is necessary to localise it within the network. The step (iv) characterises the mixed model-based/data-driven approach using Machine Learning algorithms. Compared to a leak-free situation, a leak results in larger flow in a pipe, with larger head loss, and differences in pressure within the DWN. In a monitored DWN, signatures on pressure data can be used to find leak location [7]. A calibrated hydraulic model can be used to generate synthetic data based on a multitude of demand scenarios. Subsequently the synthetic data can be used to train a classifier. Researchers have used a variety of classification methods. Rojek and Jan Studzinski [8] train Artificial Neural Networks (ANNs) such as Multi-Layer Perceptron (MLP) and Self-Organising Map (SOM). Their goal is to find the exact node where a leak is occurring. De Silva *et. al*. [9] use Support Vector Machines (SVMs) and claim better results compared to Neural Networks. Gupta [10] trains SVMs, ANNs, Random Forest classifiers and Ensemble methods using all the previous methods to improve leak localization predictions. Ares-Milian *et. al.* [1,11] train a SVM classifier to find the leak zone. The authors then apply a model/optimization-based approach to refine and reduce the leakage detection zone. Soldevila and al. [3] present a slightly different approach, where the modelling of the DWN is made and compared to the measured pressures in real time. The comparison is made using residuals. A residual represents the difference between the pressure measured and a reference leak-free scenario pressure from the model. Bayesian classifiers and k-Nearest Neighbour classifiers are trained for the localization task. In this work, the authors overcome the difficulty related with the implicit difference of synthetic data from noisy measured data. The method requires measurements to control all the boundary conditions on the hydraulic model. These constraints can be hard to obtain on most hydraulic networks. Other improvements include the selection of better locations for pressure sensors [12]. From the aforementioned contributions, we can also point out the difficulty to assess the classifiers' performances on real scenarios as collection of leak and leak-free data is hard. To date, model-based leak detection methodologies have not reached the maturity expected by the water industry.

Considering this, this article presents an improved model-based approach for early leak localization in DWNs from pressure measurements. Field data includes days of normal operation without leaks and a set of nine controlled leaks (leak campaigns). On the other hand, a synthetic dataset gathers the result of hundreds of thousands of scenarios generated from a calibrated hydraulic model (EPANET [14]), including demand variations on each node, and leaks of different sizes on every pipe of the network. Classifiers for the leak localization tasks are trained on residuals computed against leak-free scenarios from simulations and known days of normal operation. Then, we propose a domain adaptation technique to correct the differences between synthetic and measured data distributions by searching for a new representation of the data in which both domains (measured and synthetic) are similar. The performance of the trained classifier is evaluated on measured data only. The performance on synthetic data is not of interest. The methodology is tested in a real district metered Area (DMA).

2 METHODS

2.1 Early leak localization methodology overview

The methodology proposed for early leak localization is presented in Figure 1. The proposed method localises leaks in a DWN once a leak has been detected with any procedure available (examples in the literature [5]).

The requirements for the proposed methodology in a DWN application are:

- A hydraulic model (EPANET) of a DWN or a District Metered Area (DMA);
- A set of pressure sensors (installed on the network using sensitivity matrices-based methods[13])
- Inlet and outlet (if applicable) flowmeters to capture the hydraulic information;
- An IT system for data acquisition and processing;
- A calibration procedure for the hydraulic model.

The classifying task for early leak localization is computed in two main stages. First the network is partitioned into a set of candidate leak zones using sensitivity matrices [13] considering the topological relationships between the network nodes. Leak zones correspond to groups of pipes with similar hydraulic response to leaks (see section 2.3). Then a trained multiclass classifier is used to estimate a potential leak zone (see sections 2.4).

2.2 Training and testing sets

Machine learning algorithms require an important number of training examples. Nevertheless, obtaining real-world DWN leak data is near-to-impossible, since it would imply collecting data for leaks in all pipes multiple times. The lack of historical data is a great challenge. We will have at our disposal only a limited set of samples. In addition, these records are, most of the time, not properly labelled. For example, start and end days of the leak and/or location might not be properly logged.

The training database will be constructed from simulated scenarios of variable demands. Simulations are run using EPANET Programmer's Toolkit. On the other hand, the test database will be created with measured pressure data. Leaks are created by opening purge valves or fire hydrants. The advantage of synthetic data is that it is properly labelled and balanced for machine learning applications. Although DWN models remain only an approximation (even if calibrated), the use of synthetic data is commonly accepted in the field of machine learning [15].

CC BY NC SA

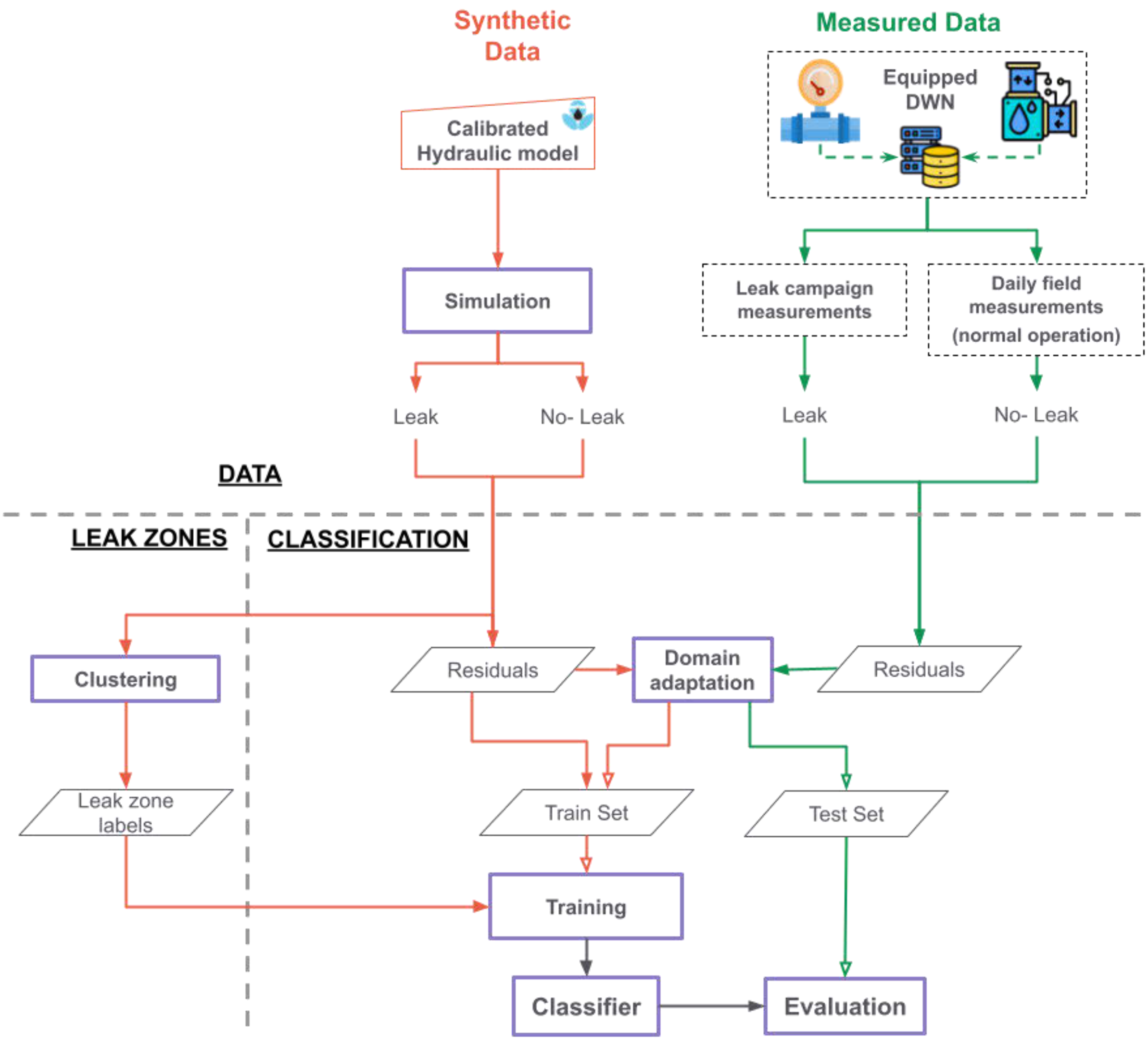

Synthetic Data
Measured Data
Calibrated Hydraulic model
Equipped DWN
Simulation
Leak campaign measurements
Daily field measurements (normal operation)
Leak
No- Leak
Leak
No- Leak
DATA
LEAK ZONES
CLASSIFICATION
Clustering
Residuals
Domain adaptation
Residuals
Leak zone labels
Train Set
Test Set
Training
Classifier
Evaluation

Figure 1. Methodology overview for early leak localization

2.3 Leak Zone labels

Using sensitivity matrices methodologies [13], multiple groups of pipes are defined from the position of sensors and their sensitivity to leaks. A sensitivity leak analysis is done by simulating leaks with different flow rates at all possible pipes within the DWN and calculating the corresponding sensitivity in flow and pressure at each point by comparing with the base-case scenario (i.e. no leaks in the network). The two main limitations to directly use these groups are that they are often groups of one pipe and groups of multiple pipes can include pipes not located in the same part of the DMA.

The spectral clustering method [16] has proven itself to be effective in many different situations. The network of pipes is considered as a mathematical graph with a set of edges and a set of vertices. The mapping from the hydraulic model to the graph model is as follows: pipes (and valves) correspond to edges and junctions (such as pipe intersections, water sources) correspond to nodes. Spectral clustering is applied considering only the distance between nodes.

The search zones can be constructed from groups of pipes defined by hydraulic sensitivity to leaks and groups of pipes defined from the spectral clustering with an ensemble method [17]. Finally, labels for the multiclass classifier are defined by matching every sample with the zone that contains the actual leak node.

CC BY NC SA

2.4 Classifier and input variables

In this study, we propose the use of the Gradient Boosting Classifier [18], that is a set of machine learning algorithms which include several weaker models that are combined together to get higher predictive output. Gradient Boosting algorithms are popular due to their ability to classify datasets effectively. Finding the leak zone where a leak occurred can be framed as a multiclass classification task. The Scikit-learn [20] project provides a set of machine learning tools that can be used for the multiclass classification task.

In order to remove long-term seasonality, the classification tasks proposed will not be carried out directly on the pressure and flow raw variables. We use residuals to remove a short-term mean behaviour of the network from the measured signal. To do so, we take the average of M days to classify and subtract the mean over N past days labelled as having normal behaviour (no leaks, no outliers). A process to identify days with normal conditions is thus essential. The residuals are defined by Equation 1:

$$X_{res}(t) = \frac{\bar{X}_M(t) - \bar{X}_N(t)}{\sigma(X_N(t))} \tag{1}$$

where:

- $\bar{X}_M(t) = \frac{1}{M}\sum_{m=1}^{M} X_m(t)$ is the mean value of a sensor at time t over the M consecutive days to classify;
- $\bar{X}_N(t) = \frac{1}{N}\sum_{n=1}^{N} X_n(t)$ is the mean value of a sensor at time t over N previous days that exhibit a normal behaviour; and
- $\sigma(X_N(t))$ is the standard deviation of values of a sensor at time t over N previous days that exhibit a normal behaviour.

Parameters N and M are chosen depending on the network hydraulics and precisions of the sensors. Repetition helps isolate signals from noise. Note that for synthetic data, N and M do not represent a number of consecutive days. Simulations are unordered as each simulation with the EPANET software is independent of others. Days can be selected randomly for the simulation scenarios. The times t can be either chosen during the day, a window during night time, on a precise time of the day such as the time at which the minimal night flow is observed.

One issue we face when using synthetic data is that our training and testing data are different. Training data is drawn from simulation from the EPANET model and thus has a bias (from modelling) and variability different from the testing data drawn from measurements. The purpose of transfer learning and domain adaptation methods [19] is to handle this common issue encountered in machine learning. In the domain adaptation setting, one considers, on one hand a source domain from which a large sample of labelled data is available. And on the other hand, a target domain from which no labelled data are available. One refers to unsupervised domain adaptation. The goal is to build an estimator on the target domain by leveraging information from the source domain.

CORAL (CORrelation ALignment) [21] is a feature based domain adaptation method which minimises domain shift by aligning the second-order statistics of source and target distributions. Feature-based methods consist of searching for common features which have similar behaviour with respect to the classification task on source and target domain. A new feature representation (often called encoded feature space) is built with a projecting application which aims to correct the difference between source and target distributions. The CORAL method transforms source features to minimise the Frobenius norm between the correlation matrix of the input target data and the one of the transformed input source data. CORAL can use only labelled source and unlabeled target data. The ADAPT Python package [22] provides tools that can be used for domain adaptation. The classification task is then trained in this encoded feature space instead of the initial input space.

2.5 Performance metrics

2.5.1 Accuracy score

The accuracy score is defined in Equation 2 as the percentage of leak samples that were correctly located by the classifier for N samples:

$$accuracy = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{L_i}(\hat{L}_i) \tag{2}$$

where L_i is the actual leak zone for sample i, $\hat{L}_i$ is the predicted leak zone for sample i, and $\mathbb{I}$ the indicator function.

2.5.2 Top-k accuracy score

For the multiclass classification problem, another metric of interest is the top-k accuracy score. The leak zone prediction is considered correct as long as the true leak zone is associated with one of the k highest leak zones predicted by the classifier. For N samples, the top-k accuracy is defined in Equation 3 as:

$$topk\ accuracy = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} \mathbb{I}_{L_i}(\hat{L}_{i,j}) \tag{3}$$

where L_i is the actual leak zone for sample i, $\hat{L}_{i,j}$ is the j-th largest score of the predicted leak zones for the sample i, and $\mathbb{I}$ the indicator function. This metric is interesting for leak localization as it can measure the leak localization performance along a path to search the DWN.

2.5.3 Matthews correlation coefficient

The Matthews correlation coefficient (MCC) is used in machine learning as a measure of the quality of classification. The MCC ranges between 1 and -1. A coefficient of 1 represents a perfect prediction, 0 an average random prediction, and a negative score is linked to an inverse prediction. The MCC is defined for N samples by Equation 4:

$$Matthews\ correlation\ coefficient = \frac{c \times N - \sum_{k=1}^{N_L} p_k \times \gamma_k}{\sqrt{(N^2 - \sum_{k=1}^{N_L} {p_k}^2) \times (N^2 - \sum_{k=1}^{N_L} {\gamma_k}^2)}} \tag{4}$$

where γ_k is the number of times a leak truly occurred in leak zone k, p_k is the number of times a leak was predicted in leak zone k, c the count of leak zones correctly predicted, and N_L the number of leak zones. Having an unbalanced and few number of samples of each leak zone, this measure provides a metric on leak zones not being predicted by randomness.

2.5.4 Leak Search Zone Size

On top of machine learning metrics, we propose to look at the expected gain in terms of reduced leak search zone. The expected leak search zone size is defined in Equation 5 as the percentage of the total DWN linear length to explore to truly find a leak:

$$Leak\ Search\ Zone\ Size = \frac{D - \sum_{k=1}^{N_Z} (\frac{c_k}{N} \times \sum_{j \neq k}^{N_Z} d_j)}{D} \times 100 \tag{5}$$

where c_k is the number of times a leak is correctly predicted in leak zone k, d_j is the pipe length of leak zone j in km, D the total pipe length in the DWN , N the number of samples, and N_Z the number of leak zones. A correct prediction by the classifier would avoid the need to inspect the rest of the network.

3 CASE STUDY AND RESULTS

3.1 Case study

The methodology was developed and implemented on a Colombian DWN operated by Veolia Colombia. The DWN is located in a residential sector, and the INP model provided by the operator accounts for 288 nodes (users with individual demands), 303 pipes for a total length of 13.5 km approximately, a reservoir representing the unique source of water for the whole sector, and 3 pressure reduction valves (PRV). In 2019 the sector reported a total water loss of 18% (leaks +apparent losses). Figure 2 shows topology of the sector of analysis.

The set of data used in this work, contains the measured values of 12 pressure sensors and 1 flowmeter, whose locations are also detailed in Figure 2. Data was collected for a year (September 2020 to August 2021), including mostly normal operation measurements, as well as data of 2 leak campaigns, over 5 different points (1 to 5 in Figure 2). The field data frequency is 1 measurement every 15 minutes.

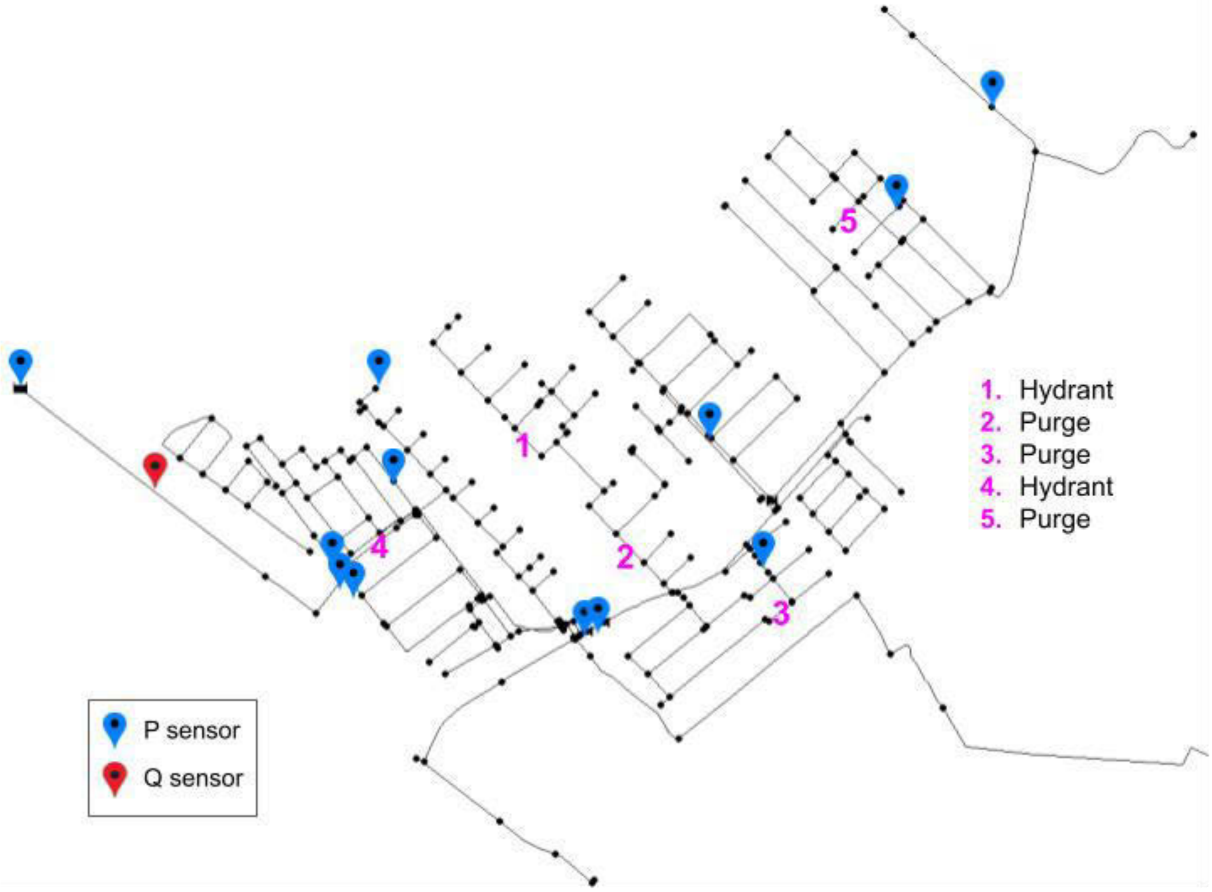


Figure 2. Location of the sensors and controlled leaks in the case study sector

The leak campaigns consisted of a total of 9 leak configurations (Table 1), each one with a duration of 26 hours. The first leak campaign was launched in February 2021 (configurations A and B) and the second one in June 2021. Three of the configurations considered simultaneous leaks in two of the five points. For the sake of consistency, at least two days of normal operation were considered between each consecutive configuration to avoid any possible effect of the precedent controlled leak.

Table 1. Leak campaign configurations

Config	Location	Max flow rate [m^3/h]	Config	Location	Max flow rate (m^3/h)
A	2	1	G	3	5
B	5	3		5	2
C	2	5	H	2	5
D	3	3		4	2
E	1	2	I	5	3
F	2	5		1	2

3.2 Experimental results

When implementing the methodology in the case study, we tested the division of the sector into different numbers of clusters ranging from 2 to 8. Figure 3 details the resulting zones for 2, 4, 6 and 8 leak zones. Residuals have been computed with M=1, N=5, at time t when the minimum night flow is observed. The training dataset contains 15000 simulations. Performance of the proposed method is evaluated on the leak campaign test dataset.

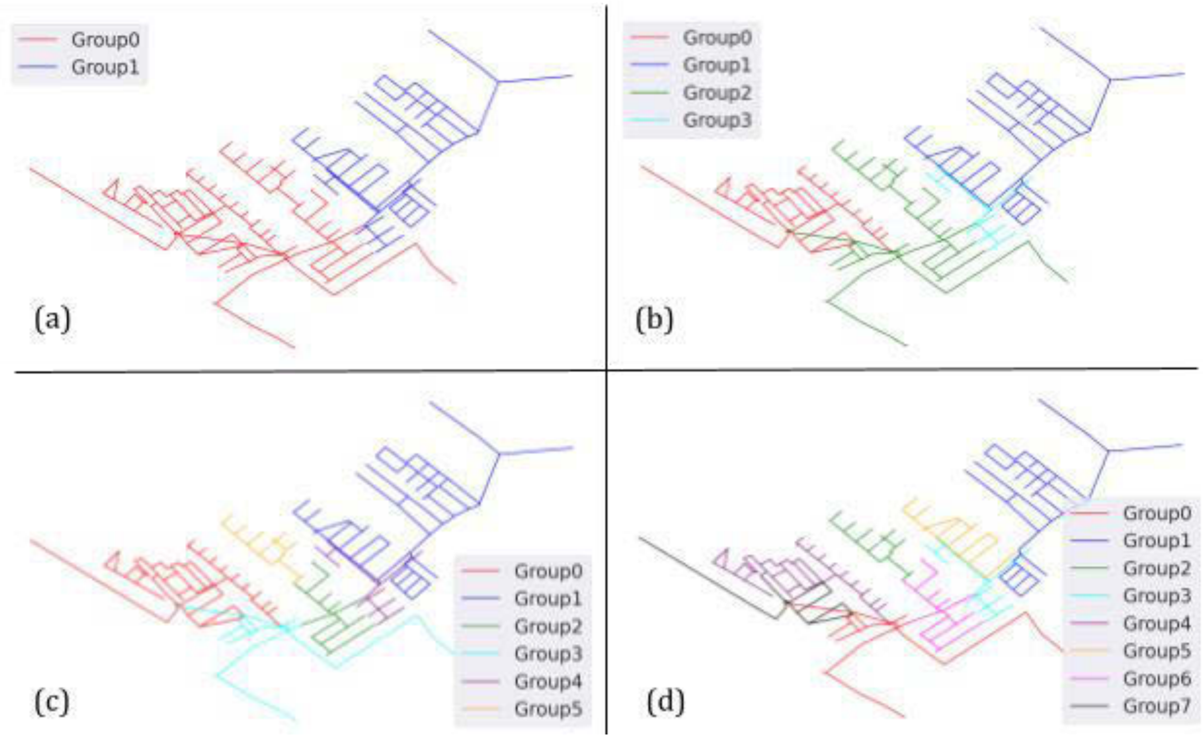

Group0
Group1
(a)
Group0
Group1
Group2
Group3
(b)
(c)
Group0
Group1
Group2
Group3
Group4
Group5
(d)
Group0
Group1
Group2
Group3
Group4
Group5
Group6
Group7

Figure 3. Leak zones for different numbers of clusters [(a) 2, (b) 4, (c) 6, (d) 8]

Figure 4 and Figure 5 present respectively the accuracy score and the top-2 accuracy score for the different number of leak zones to search within. As expected, the lesser number of zones, the higher the accuracy of the classifier. Using domain adaptation to compute a feature encoded space before training the classifier seems to improve the accuracy of the classifier. Without domain adaptation, the classifier is trained directly on the pressure residuals. With domain adaptation, the 2-zone classifier reports an accuracy of 78%. The top-2 accuracy score remains over 80% up to 4 zones.

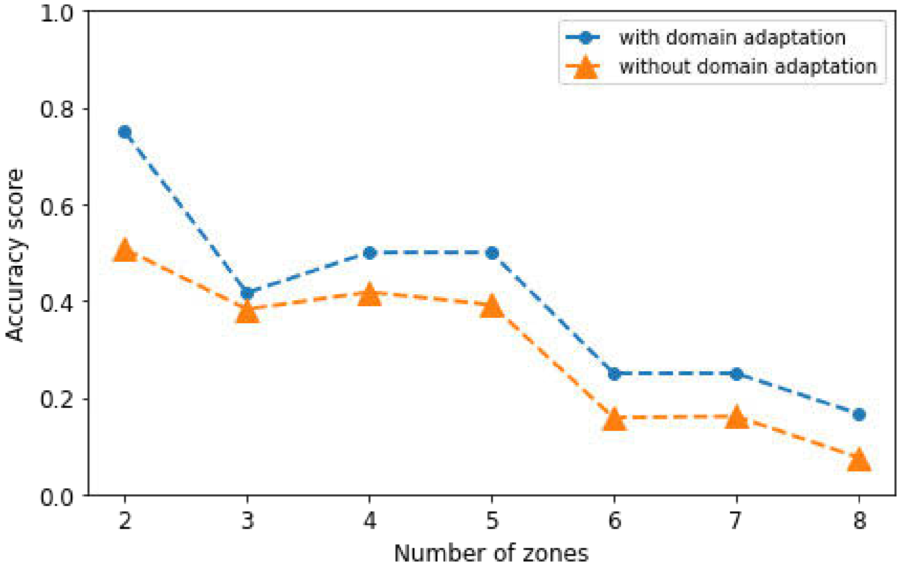


Figure 4. Accuracy score by number of zones

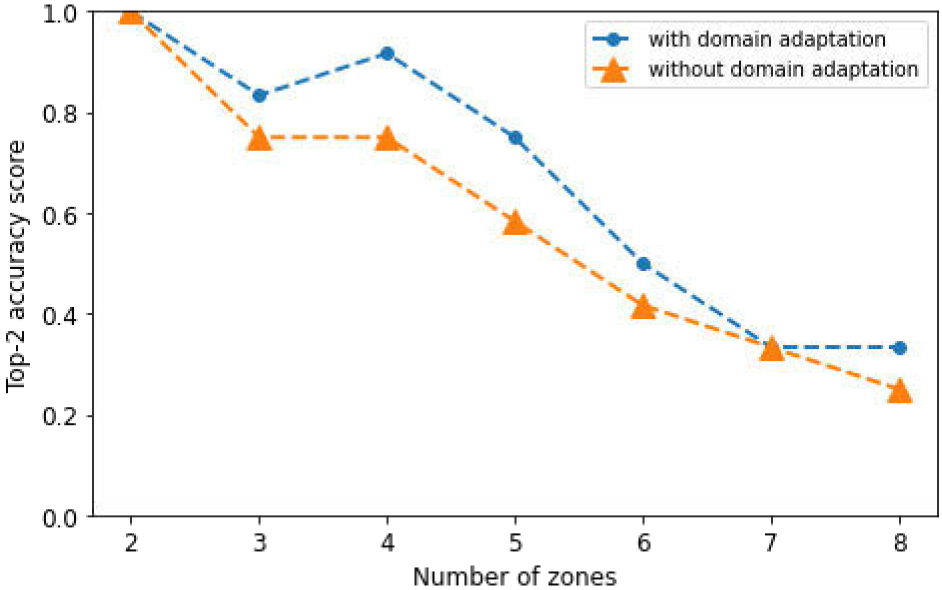


Figure 5. Top-2 Accuracy score by number of zones

The Matthews correlation coefficient score is depicted in Figure 6. The MCC score can help to identify the ineffectiveness of the classifier. For 6 and more zones, the score is below 0.2. In this case, the classifier is very close to a random guess classifier. For 2 to 5 zones, the classifier can be considered better than a random guess classifier.

On a field perspective of the method's performance, the percentage of the expected leak zone search is presented in Figure 7. The 2-zones classifier yields a leak search zone of just above 60% of the total DWN pipe length. With the proposed method, the length of pipe inspected is expected to be reduced by one third. The 4-zones classifier also yields promising gains with a leak search zone just below 70% of the total DWN.

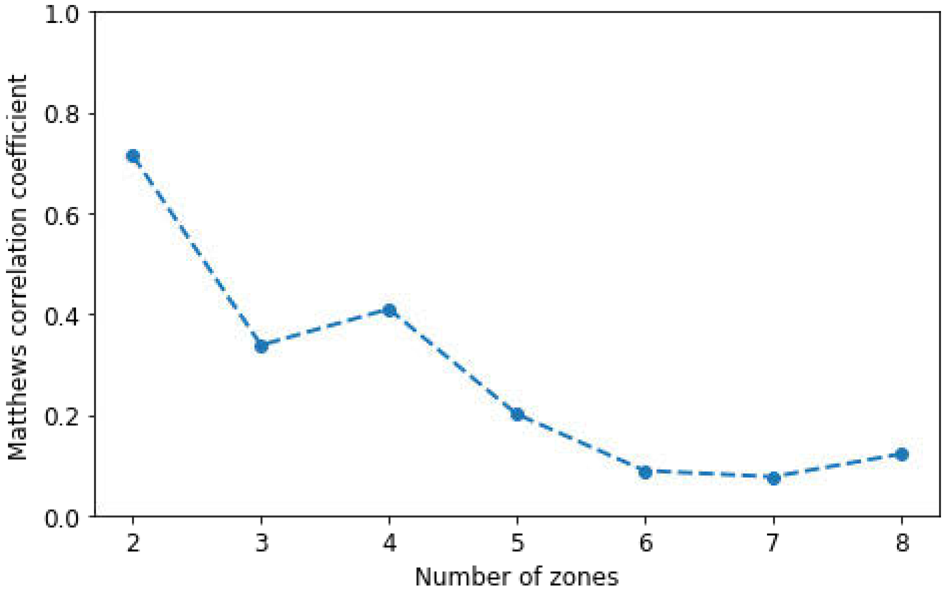


Figure 6. Matthews correlation coefficient by number of zones

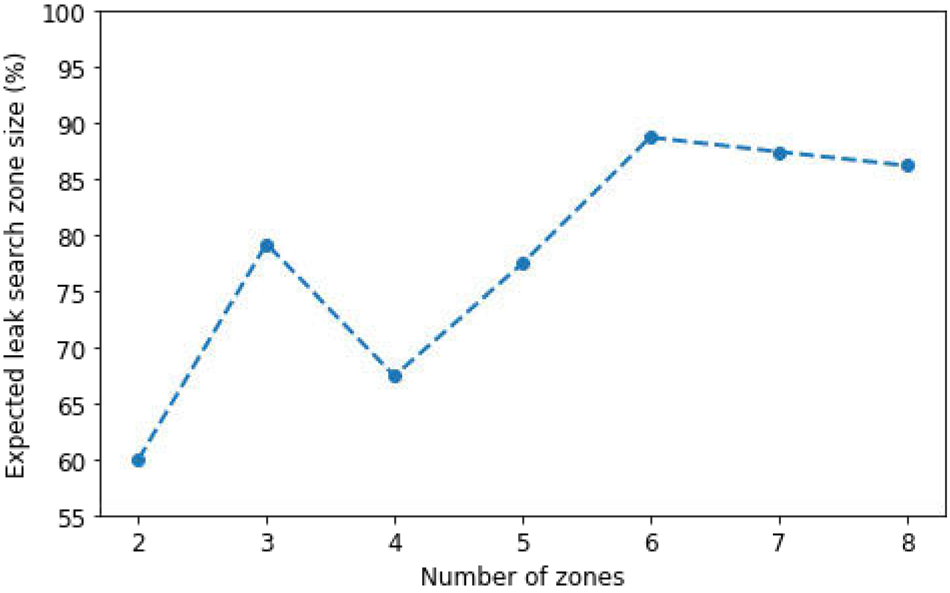


Figure 7. Expected leak search zone size in percentage of the total DWN pipe length by number of zones

4 CONCLUSIONS

Reducing costs and time searching leakages in DWNs is known to be a difficult task for the water utility industry. The presented work demonstrates the benefits of using a hybrid model-based and data-driven approach for early leak localization. Dividing the DWN into two or four zones doesn't seem like the target, but shows success in contributing to early leak localization by reducing the leak search space. This low cost solution requires only pressure sensors on top of common inlet and outlet flowmeters, and it relies on more frugal (less data, parameters, and computational resources required) algorithms than neural networks. The field experiment considers 12 sensors, but with few zones, the number of required sensors could be lower.

Proper data collection of network hydraulic behaviour for data-driven approaches is a challenging task. Historical data are either non-existent or not properly labelled. To overcome this issue, our experimental approach focuses on evaluating the performance only on measured data from a designed leak campaign. This strategy allows us to go further than using synthetic data or toy DWN models. To reduce the differences between synthetic data samples and the measured data samples, a domain adaptation approach was implemented. Moreover, the proposed residuals don't require such high fidelity simulation of the DWN. The current limitation lies in the identification of previous normal operation days to be able to compute a mean pressure profile for each of the sensors.

In conclusion, the proposed methodology allows to improve early leak detection performance, while meeting operational requirements. The field experiments show that the total pipe length required for inspection in the DMA can be reduced by one third with a significant gain of performance in accuracy, while requiring a limited number of pressure sensors installed on the network.

5 REFERENCES

[1] M. Quiñones-Grueiro, M. Ares Milián, M. Sánchez Rivero, A. J. Silva Neto, and O. Llanes-Santiago, "Robust leak localization in water distribution networks using computational intelligence," Neurocomputing, vol. 438, pp. 195–208, May 2021, doi: 10.1016/j.neucom.2020.04.159.

[2] A. Gupta and K. D. Kulat, "A Selective Literature Review on Leak Management Techniques for Water Distribution System," Water Resour Manage, vol. 32, no. 10, pp. 3247–3269, 2018, doi: 10.1007/s11269-018-1985-6.

[3] A. Soldevila, J. Blesa, S. Tornil-Sin, E. Duviella, R. M. Fernandez-Canti, and V. Puig, "Leak localization in water distribution networks using a mixed model-based/data-driven approach," Control Engineering Practice, vol. 55, pp. 162–173, Oct. 2016, doi: 10.1016/j.conengprac.2016.07.006.

[4] T. K. Chan, C. S. Chin, and X. Zhong, "Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection," IEEE Access, vol. 6, pp. 78846–78867, 2018, doi: 10.1109/ACCESS.2018.2885444.

[5] S. El-Zahab and T. Zayed, "Leak detection in water distribution networks: an introductory overview," Smart Water, vol. 4, no. 1, p. 5, Jun. 2019, doi: 10.1186/s40713-019-0017-x.

[6] Z. Hu, B. Chen, W. Chen, D. Tan, and D. Shen, "Review of model-based and data-driven approaches for leak detection and location in water distribution systems," Water Supply, vol. 21, no. 7, pp. 3282–3306, Apr. 2021, doi: 10.2166/ws.2021.101.

[7] S. Sophocleous, D. Savić, and Z. Kapelan, "Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction," Journal of Water Resources Planning and Management, vol. 145, no. 7, p. 04019024, Jul. 2019, doi: 10.1061/(ASCE)WR.1943-5452.0001079.

[8] I. Rojek and J. Studzinski, "Detection and Localization of Water Leaks in Water Nets Supported by an ICT System with Artificial Intelligence Methods as a Way Forward for Smart Cities," Sustainability, vol. 11, no. 2, p. 518, Jan. 2019, doi: 10.3390/su11020518.

[9] D. De Silva, J. Mashford, and S. Burn, "Computer aided leak location and sizing in pipe networks (second edition)," Urban Water Security Research Alliance, Technical Report EP112167, Apr. 2011.[Online]. Available: http://hdl.handle.net/102.100.100/104745?index=1

[10] G. Gupta, "Monitoring Water Distribution Network using Machine Learning," KTH Royal Institute of Technology, Stockholm, Sweden, 2017. Accessed: Apr. 12, 2022. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-221832.

[11] M. J. Ares-Milián, M. Quiñones-Grueiro, C. Verde, and O. Llanes-Santiago, "A Leak Zone Location Approach in Water Distribution Networks Combining Data-Driven and Model-Based Methods," Water, vol. 13, no. 20, p. 2924, Jan. 2021, doi: 10.3390/w13202924.

[12] A. Soldevila, J. Blesa, R. M. Fernandez-Canti, S. Tornil-Sin, and V. Puig, "Data-Driven Approach for Leak Localization in Water Distribution Networks Using Pressure Sensors and Spatial Interpolation," Water, vol. 11, no. 7, p. 1500, Jul. 2019, doi: 10.3390/w11071500.

[13] R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, and A. Peralta, “Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks,” Control Engineering Practice, vol. 19, no. 10, pp. 1157–1167, Oct. 2011, doi: 10.1016/j.conengprac.2011.06.004.

[14] L. A. Rossman, “The EPANET Programmer’s Toolkit for Analysis of Water Distribution Systems,” pp. 1–10, Apr. 2012, doi: 10.1061/40430(1999)39.

[15] M. Hittmeir, A. Ekelhart, and R. Mayer, “On the Utility of Synthetic Data: An Empirical Evaluation on Machine Learning Tasks,” in Proceedings of the 14th International Conference on Availability, Reliability and Security, New York, NY, USA, août 2019, pp. 1–6. doi: 10.1145/3339252.3339281.

[16] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888–905, Aug. 2000, doi: 10.1109/34.868688.

[17] L. Rokach, “Ensemble-based classifiers,” Artif Intell Rev, vol. 33, no. 1, pp. 1–39, Feb. 2010, doi: 10.1007/s10462-009-9124-7.

[18] J. Friedman, “Greedy function approximation: A gradient boosting machine.,” 2001, doi: 10.1214/AOS/1013203451.

[19] F. Zhuang et al., “A Comprehensive Survey on Transfer Learning,” Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[20] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, no. 85, pp. 2825–2830, 2011, Accessed: Apr. 15, 2022. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[21] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona, février 2016, pp. 2058–2065.

[22] A. de Mathelin, F. Deheeger, G. Richard, M. Mougeot, and N. Vayatis, “ADAPT : Awesome Domain Adaptation Python Toolbox,” arXiv:2107.03049 [cs], Jul. 2021, Accessed: Apr. 15, 2022. [Online]. Available: http://arxiv.org/abs/2107.03049

BY NC SA

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14154

WATER DISTRIBUTION NETWORK OPERATION OPTIMIZATION: AN INDUSTRIAL PERSPECTIVE

Marco Lauricella[1], Felix Lenders[2] and Matthias Biskoping[3]

[1,2,3]ABB AG, Corporate Research Germany
Wallstadter Str. 59, 68526 Ladenburg, Germany

[1]*marco.lauricella@de.abb.com,*[2] *felix.lenders@de.abb.com,* [3]*matthias.biskoping@de.abb.com*

Abstract

A significant portion of the operating costs associated with drinking water distribution networks is related to energy usage, which is mostly employed to drive pumps. One strategy to improve energy efficiency and to reduce energy cost is to operate water pumps in an optimal manner that allows a reduction in energy consumption. This produces also environmental benefits, since decreasing energy consumption contributes to the reduction of the associated greenhouse gas emissions, helping utility providers to reduce their carbon footprint and to reach sustainability goals. However, numerical optimization of water distribution network operation is a difficult problem to solve, given the combination of non-linear hydraulic dynamics and the presence of discrete decision variables, corresponding to pumps and valves having on/off or open/closed characteristics. In this work, we address such problem from an industrial perspective, reformulating the mathematical program that is at the core of such operation optimization solutions using complementarity constraints to transform the resulting mixed integer nonlinear program into a nonlinear program having only continuous variables. This allow us to obtain a tractable optimization problem, that could be solved in a short amount of time even for large-scale water networks, making it compatible with industrial implementation and real-time optimization.

Keywords
Water distribution networks, water network operation optimization, mixed integer nonlinear programming, nonlinear programming.

1 INTRODUCTION

Given the increasing awareness of environmental problems, resulting in the call for a substantial reduction of greenhouse gas emission, and considering the increasing energy cost, there is a significant benefit coming from a more efficient use of energy. This is particularly true for drinking water distribution networks, where a significant portion of the operating costs is related to energy usage, and up to 70% of that energy is employed to drive pumps [1], [2], [3]. The adoption of optimization algorithms to schedule pump operation in water distribution networks brings several benefits: it helps reducing energy consumption, thus lowering energy cost and the carbon footprint associated with energy use; it contributes to a better control of water pressure across the distribution network, lowering peak pressure and consequent leakage, and mitigating the risk of pipe burst. Energy and cost optimization of water distribution networks represents an important practical problem [4], and optimal pump operation enables utility providers to reduce non-revenue water, to decrease operating costs, to reduce their carbon footprints, and to reach sustainability goals. Moreover, operation optimization problems can easily be formulated to take into account additional cost components like electricity price, energy production from renewable sources, CO2 production of the current energy mix, allowing utility providers to develop custom control strategies to target their economic and environmental goals.

It is generally known that the numerical optimization of water distribution networks operation is a difficult problem to solve, since it combines the non-linear and non-convex hydraulic dynamics

BY NC SA

of water networks with the presence of discrete decision variables, corresponding to pumps and valves having an on/off or open/closed characteristic [4], [5], [6], [7]. Moreover, the optimization of a medium-sized real-world water distribution network over a one-day horizon, already leads to large-scale, mixed integer nonlinear programs (MINLP) that need to be solved in a relatively short amount of time. The practical challenges range from the computational effort required to numerically solve such a class of optimization problem, to the presence of multiple local minima. An additional difficulty comes from the presence of uncertainties associated with water demand, renewable power generation and volatility in energy prices.

For the described reasons, both in the literature and in industrial applications based on mathematical programming approaches, it is common to rely on heuristics to simplify the problem structure, or to improve convergence time of the optimization. Heuristics-based solutions often require a significant number of ad hoc decisions, which negatively affects the feasibility of the resulting solutions in face of uncertainties and its generality, that limits the possibility to easily extend the solution to different water networks without an extensive engineering effort. In addition to deterministic methods based on mathematical programming, since the 1990s, metaheuristic algorithms have been applied to the problem of water distribution network optimal operation, among which we can cite genetic algorithms [8], [9], ant colony optimization [10], and simulated annealing approaches, see e.g. the review paper [4] and the references therein. However, recent years showed an increasing interest for mathematical programming approaches, which are more suitable for real-time control and can now be applied more easily on industrial products thanks to the growth of the available computational power.

In this work, we present the results of the application of nonlinear programming techniques to the problem of pump operation scheduling for medium-sized water distribution networks. We propose an approach developed for industrial case studies, but we demonstrate the potential showing only results pertaining to open-source network models for confidentiality reasons. We discuss the reasoning that led to the current problem formulation, and we highlight the practical challenges. The proposed approach relies on standardized EPANET [11], [12] hydraulic modelling, without creating alternative surrogate models for optimization. Despite the use of standard EPANET description of water networks, our approach allows to automatically generate analytical gradient information relative to the hydraulic model, and it allows to simultaneously obtain dynamics simulation and optimization results for the operation of drinking water distribution network.

2 MOTIVATIONS AND METHODOLOGY

The problem addressed here is how to optimally operate pumps in water distribution network, managing flow, pressure and storages in the network to minimize operational costs, while complying with operational requirements, i.e. satisfaction of customer demands at all network nodes, and capability to maintain nodes pressure above a minimal threshold under all demand conditions. Here, as a starting point, we assume to have at our disposal a network model defined by its topology and hydraulics parameters according to EPANET standard description.

Operational costs could be mathematically described by combining the different elements contributing to the total costs, among which one could consider pumping energy costs, costs related to water losses and water treatments, costs related to planned or unplanned maintenance and reparation works after failures or damages. The operation optimization is then performed on a day-ahead basis, where we want to schedule pumps operation over a 24 hours time horizon. We assume to have at our disposal reliable demand patterns for all network nodes, thus the problem of demand forecast will not be addressed here. Moreover, in our analysis we are not considering water treatment processes since our target is water distribution. As stated in the introduction, we resort to mathematical programming to tackle the problem of pump operation optimization, where the main control variables are discrete decision variables associated with pumps, which define if a pump is on or off at every given time instant.

In this work, we consider the full operation optimization problem, which includes binary and continuous decision variables in discrete time, resulting in a mixed integer nonlinear program. We then reformulate such optimization problem as a nonlinear program (NLP) using a relaxation-based approach that resorts on complementarity constraints to transform the original MINLP into a more easily tractable NLP. To do so, the binary decision variables associated with pumps operation are substituted by continuous variables, and appropriate slack variables and constraints are added to the problem formulation. This allows us to obtain an optimization problem that can be solved in a limited amount of time even for medium and large-scale network, and that is suitable for the implementation on industrial applications. More details regarding the mathematical programming formulation and the problem relaxation using complementarity constraints are described in Section 2.2.

2.1 Network model

Here, we present the modelling strategy that we adopt to describe the hydraulics of water networks, which sits at the core of the operation optimization algorithm. Water distribution networks can be described by a directed graph $G = (N, L)$ consisting of vertices, or nodes (N), and arcs, or links (L). Here, following the nomenclature commonly adopted in the literature [7], [13] and by EPANET [12], we classify as nodes all tanks (N_t), reservoirs (N_r), junctions and end points/demand points (N_j), while pumps (L_{pu}) and pipes (L_{pi}) are considered links, thus $N = N_t \cup N_r \cup N_j$ and $L = L_{\text{pu}} \cup L_{\text{pi}}$.

Table 1. Basic notation

Symbol	Explanation	Value	Unit
q_{lt}	Flowrate at link *l* at time *t*		$\mathrm{m^3/s}$
q_{pt}	Flowrate at pump *p* at time *t*		$\mathrm{m^3/s}$
d_{nt}	Demand flowrate at node *n* at time *t*		$\mathrm{m^3/s}$
h_{nt}	Head at node *n* at time *t*		m
$\bar{H}_n$	Constant head at reservoir *n*		m
Δh_{lt}	Headloss at pipe *l* at time *t*		m
Δh_{pt}	Head increase at pump *p* at time *t*		m
ρ	Water density	1000	$\mathrm{kg/m^3}$
g	Gravity acceleration	9.81	$\mathrm{m/s^2}$

To obtain a tractable model that can be used for optimization, we adopt a discrete time model, where the considered time horizon is divided into T equidistant intervals indexed by $t \in \{1, \dots, T\}$. Here, we consider an optimization horizon of one day, divided into 24 periods of one hour each, which gives a discretization interval $\Delta t = 1$ hour. This usually corresponds to the discretization with which demand forecast and electricity prices are provided [6], [7].

For each link $l = (i, j), l \in L$, we denote its flow variable with q_{lt}, which is positive if the flow is directed from i to j, and negative otherwise, with $i, j \in N$ representing network nodes. The flow is always non-negative for pumps and for certain pipes allowing only unidirectional flow via check valves. It is customary for water distribution network problems to measure pressure as the sum of geodetic height and elevation difference $\Delta h = \frac{p}{\rho g}$ due to hydraulic pressure. This goes under the name of nodes head, and here it is denoted by h_{nt}, with $n \in N$. The head increase for pumps is defined by Δh_{pt}, with $p \in L_{\mathrm{pu}}$.

The water distribution network is then described by a quasi-stationary, discrete-time, hydraulic model [13], and here we present the equations adopted to model the various network elements. More details regarding hydraulic modelling can be found e.g. in [5], [13].

Reservoirs are considered as unlimited sources of water, where the head is always equal to a known constant value $\bar{H}_n$ with $n \in N_r$, resulting in:

$$h_{nt} - \bar{H}_n = 0, \forall n \in N_r. \tag{1}$$

Tanks are modelled via a discrete-time flow balance equation:

$$h_{nt+1} - h_{nt} - \frac{\Delta t}{A_n} \Big(\sum_{l \in L^{in}(n)} q_{lt} - \sum_{l \in L^{out}(n)} q_{lt} \Big) = 0, \forall n \in N_t, \tag{2}$$

where A_n represents the cross-section area of the tank, and $L^{\mathrm{in}}(n)$ and $L^{\mathrm{out}}(n)$ are respectively the set of incoming and outgoing links for node n.

The flow balance in node $n \in N_j$, having a given demand profile d_{nt}, is described by:

$$\sum_{l \in L^{\mathrm{in}}(n)} q_{lt} - \sum_{l \in L^{\mathrm{out}}(n)} q_{lt} - d_{nt} = 0, \forall n \in N_j. \tag{3}$$

Here, we assume that the demand d_{nt} is positive if the water flow is leaving the network at node n, and it is fixed to zero for nodes having no demands.

Friction losses in pipes can be generically described using the formula expressing the head loss as a function of the flow and of the pipe resistance coefficient a_l:

$$\Delta h_{lt} = a_l {q_{lt}}^B, \forall l \in L_{pi} \tag{4}$$

where B is the generic flow exponent. Different headloss expressions are available, and Hazen-Williams and Darcy-Weissbach formulas are among the most widely adopted both in industry and literature [5], [7]. In the Hazen-Williams formula, (4) becomes:

$$\Delta h_{lt}(q_{lt}) = \frac{10.67 L_l}{{C_l}^{1.852} {d_l}^{4.871}} \mathrm{sign}(q_{lt}) |q_{lt}|^{1.852}, \forall l \in L_{\mathrm{pi}}, \tag{5}$$

where L_l and d_l are respectively the pipe length and diameter, and C_l is the Hazen-Williams pipe roughness coefficient. In case of the Darcy-Weissbach formula, (4) becomes:

$$\Delta h_{lt}(q_{lt}) = \frac{8 L_l \lambda_l}{g\, \pi^2 {d_l}^5} \mathrm{sign}(q_{lt}) {q_{lt}}^2, \forall l \in L_{\mathrm{pi}}, \tag{6}$$

where $\lambda_l = \lambda_l(q_{lt})$ represents the pipe friction coefficient and depends on the Reynolds number, which nonlinearly depends on the flow, see [5], [6], [14], [15], [16] for more details. Additional minor losses in pipes, that could be caused by turbulence induced by the network layout (e.g. bends, fittings, etc.), are given by:

$$\Delta h_{m_{lt}}(q_{lt}) = \frac{\text{sign}(q_{lt}){q_{lt}}^2}{2g(\pi(d_l/2)^2)^2}, \forall l \in L_{\text{pi}}, \tag{7}$$

and can be added to the main headloss term described by (5) or (6) to obtain the total headloss. Thus, for every pipe, it is possible to write the following head balance:

$$h_{it} - h_{jt} - \Delta h_{lt}(q_{lt}) - \Delta h_{m_{lt}}(q_{lt}) = 0, \forall l = (i,j) \in L_{\text{pi}}. \tag{8}$$

The pumps head increase Δh_{pt} for fixed-speed pumps is obtained by the head-flow characteristic diagrams of the pump, given the flow defined by the network current operating point. Here, we assume that the pump characteristic curves are defined by a number of operating points provided by the pump constructor of measured by dedicated "experiments". We then fit a continuous function of the form:

$$\Delta h_{pt}(q_{pt}) = A_p - B_p {q_{pt}}^{C_p}, \tag{9}$$

obtaining the constants A_p, B_p and C_p describing the head-flow curve for each fixed-speed pump $p \in L_{pu}$. Then, since in the graph $G = (N, L)$ representing the water network pumps are modelled as links, we obtain the following equation:

$$h_{jt} - h_{it} - \Delta h_{pt}(q_{pt}) = 0, \text{with } \forall p = (i,j) \in L_{\text{pu}}. \tag{10}$$

At present time, we are not considering variable speed pumps in the optimization problem, since we use EPANET hydraulic simulation engine to validate the optimization results, and EPANET does not directly support variable speed pumps.

To conclude, pump efficiency can be provided either as a constant value $\eta_p(q_{pt}) = \eta_p$, or as an efficiency curve describing the relationship between efficiency and flowrate, which we then represent using a continuous linear or quadratic function, resulting respectively in $\eta_p(q_{pt}) = B_\eta q_{pt} + C_\eta$ or $\eta_p(q_{pt}) = A_\eta q_{pt}^2 + B_\eta q_{pt} + C_\eta$. Finally, the electric power [W] consumed by the pump is obtained as:

$$P_{pt} = \frac{\rho g q_{pt} \Delta h_{pt}(q_{pt})}{\eta_p(q_{pt})}. \tag{11}$$

2.2 Optimization

The goal of this work is to schedule daily pump operations to minimize the associated operating costs, and to guarantee the fulfilment of demand requirements. To do so, we start by formulating a mixed integer nonlinear optimization problem, having a linear objective function and nonlinear constraints, with both binary and continuous variables, generically described by:

$$\begin{aligned} &\min_x f(x) \\ \text{s.t.}\ & x_{\text{lb}} \leq x \leq x_{\text{ub}} \\ & g_{\text{lb}} \leq g(x) \leq g_{\text{ub}} \end{aligned} \tag{12}$$

where x represent the vector of optimization variables, x_{lb} and x_{ub} its lower and upper bounds respectively. To reduce the computational complexity of (12), we opt for a linear objective function $f(x) = c^T x$, moving all nonlinearities into the constraints defined by $g(x)$. In (12) we use a generic constraints formulation, which contains both equality and inequality constraints. The cost vector c can be easily tuned to account for different forms of operative costs, from pump energy cost to CO2 and non-revenue water cost terms. Here, for the sake of simplicity, we focus on pumping energy as main term of the objective function.

In addition to the constraints derived from the model equations (1)-(10), further constraints can be included in (12) to take into account minimum pressure requirements at nodes and tank level limits, to impose initial and final conditions (e.g. tank initial and final level):

$$h_{nt} - \underline{H}_{nt_{\min}} \geq 0, \forall n \in N_j, \tag{13}$$

$$h_{n0} - H_{n_{\text{initial}}} = 0, \forall n \in N_t, \tag{14}$$

$$h_{nT} - H_{n_{\text{final}}} = 0, \forall n \in N_t. \tag{15}$$

Finally, (12) can then be rewritten as:

$$\begin{gathered} \min_x c^T x \\ \text{s.t.} \begin{gathered} x_{\text{lb}} \leq x \leq \mathrm{x}_{\text{ub}} \\ (1)-(3),(5),(7)-(10),(13)-(15), \forall t \in \{1,\ldots,T\} \end{gathered} \end{gathered} \tag{16}$$

Considering the presence of discrete variables, which scale with the number of pumps and the number of considered time samples, and the nonlinear, non-convex nature of the hydraulics constraints, the resulting MINLP is notoriously an NP-hard problem [7], [17], which can easily become intractable for medium and large-scale water networks. Considering that, when a pump is turned off, no head increase constraints should be imposed for said pump, those constraints should be removed at every time instant when the pump is not operating. However, from a mathematical perspective, relaxing a constraint is better than removing it, so an approach often adopted in MINLP is the so-called big-M formulation, where the head constraints is relaxed up to a large value M when the corresponding pump is off. This means that in (16) the set of constraints of the type of (10) is replaced by the following constraints:

$$\begin{gathered} -M(1-\omega_{pt}) \leq h_{jt} - h_{it} - \Delta h_{pt}(q_{pt}) \leq M(1-\omega_{pt}), \forall p = (i,j) \in L_{\text{pu}} \\ 0 \leq q_{pt} \leq q_{\max}\,\omega_{pt} \end{gathered} \tag{17}$$

where ω_{pt} is the pump status binary indicator, and M is a constant whose value express how much the head constraint can be relaxed when the pump is not operating. In the literature there are several approaches that could be used to relax the mixed integer problem into a continuous one. Here we adopt a complementarity formulation see e.g. [18], [19], [20], [21], [22]. Using suitable slack variables to construct the complementarity constraints, (17) becomes:

$$\begin{gathered} -s_{pt}^{-} \leq h_{jt} - h_{it} - \Delta h_{pt}(q_{pt}) \leq s_{pt}^{+} \\ s_{pt}^{+} q_{pt} = s_{pt}^{-} q_{pt} = 0 \\ s_{pt}^{+} \geq 0, s_{pt}^{-} \geq 0, q_{pt} \geq 0, p \in L_{\text{pu}}, t \in \{1,\ldots,T\} \end{gathered} \tag{18}$$

This formulation allows us to drop the pump binary indicator, transforming the mixed integer nonlinear (MINLP) optimization problem into a nonlinear (NLP) one, adding a penalty function $p(x)$ to the objective function:

$$\begin{gathered} \min_x c^T x + p(x) \\ \text{s.t.} \begin{gathered} x_{\text{lb}} \leq x \leq x_{\text{ub}} \\ (1)-(3),(5),(7)-(9),(13)-(15),(18), \forall t \in \{1,\ldots,T\} \end{gathered} \end{gathered} \tag{19}$$

The mathematical program with complementarity constraints in itself is difficult to solve, since standard regularity assumptions are violated and the resulting feasible region is connected only by one point, i.e. the origin. Several approaches have been proposed in the cited literature to solve such issues, including the use of smoothing functions, of penalization terms, or the use of approaches based on enumeration of branches. Here, we implement a complementarity formulation based on penalisation, see [22]. This formulation provides practical advantages, but comes with a potential drawback, since a local minimum of the complementarity optimization program with penalty reformulation might not be a local minimum of the original problem, see [18]. However, the penalty formulation allows us to solve the operation optimization problem for medium and large-scale water networks (up to 13000 nodes) in a short amount of time, making it feasible for practical implementation and for real-time optimization. In the next section, we present some results to showcase the performance of such formulation on three different water network models.

3 NETWORKS DESCRIPTION AND RESULTS

The complementarity formulation of the pump operation optimization problem was tested on several water network models. Here, we present results related to open-source academic network models, which constitute a good benchmark to showcase the benefits and the practical applicability of the proposed formulation. We selected three network models which, we believe, are representatives of a small-scale (BWSN), a medium-scale (C-Town), and a large-scale (DWES) water distribution network. These network models have been used extensively in the literature as benchmark for various engineering problems, including water distribution network design and operation optimization, see e.g. [23], [24].

3.1 Water networks description

BWSN is a small-scale network with 126 junctions, 2 tanks, 1 reservoir, 168 pipes and 2 pumps. It includes different demand patterns, and pump pressure-flowrate characteristic curves. We added quadratic efficiency curves, obtained by appropriately rescaling the efficiency-flowrate curves of a different network.

C-Town is a medium-scale network with 388 junctions, 7 tanks, 1 reservoir, 429 pipes and 11 pumps, including demand patterns at nodes, and pump pressure-flowrate and efficiency-flowrate characteristic curves.

DWES is representative of a large-scale network, it has 12523 junctions, 2 tanks, 2 reservoirs, 14822 pipes and 5 pumps, with head-flowrate characteristic curves. Also in this case, we included efficiency curves taken from a different network model, which are rescaled to fit DWES's pumps operating regions.

The networks called BWSN and DWES are respectively Network 1 and Network 2 proposed as benchmark in [25]. These network models were created for the Battle of the Water Sensor Networks initiative, an engineering design challenge aimed at addressing the problem of optimal placement of sensors in water distribution systems. The network model called C-Town is based on a real-world medium-sized water network, and was proposed as benchmark for the Battle of Water Calibration Network initiative [26], a challenge regarding the problem of water network model calibration, i.e. the process of comparing model results with measurements, making the appropriate adjustments so that model results and data provide a correct fit.

3.2 Results

The NLP resulting from the use of the complementarity formulation discussed in Section 2.2 is implemented with CasADi, an open-source tool for nonlinear optimization including a symbolic framework for algorithmic differentiation, used to construct gradients, Jacobians and Hessians, which allows rapid and efficient implementation of NLP [27]. IpOpt [28], [29] is then used as solver, as it can be directly called from within the CasADi environment. IpOpt is an open-source solver that implements an interior point line search filter method to tackle large-scale nonlinear optimization problems.

Here, we report an extract of the results of the pump operation optimization. In Table 2, a comparison of the required computational effort for the three network models is presented, measured in terms of number of iterations needed to reach convergence. For every model, Table 2 also lists its dimension in terms of number of nodes and links, of number of variables and constraints of the related optimization problem, and of number of complementarity constraints.

Table 2. Results of pump operation optimization based on the complementarity formulation.

Name	N. pumps	N. nodes	N. links	N. variables	N. constr.	N. compl. constr.	N. iter.
DWES	4	12527	14831	970855	970704	184	872
C-Town	11	396	444	30965	30873	506	969
BWSN	2	129	178	11433	11249	92	927

Figure 1 shows the topology of C-Town network, while Figures 2, 3 and 4 depict a comparison between the optimized and the non-optimized pump operation of C-Town, showing the flowrate values of pumps over the considered 24 hours.

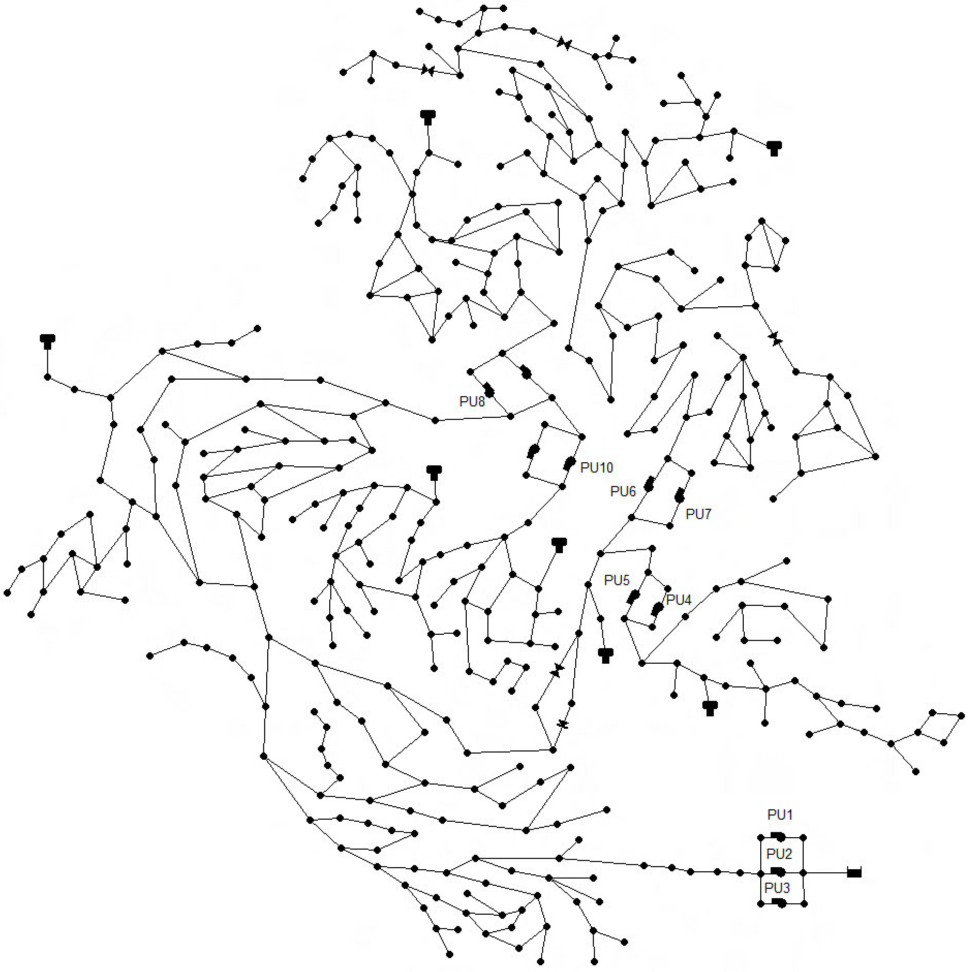


Figure 1. C-Town – Network topology

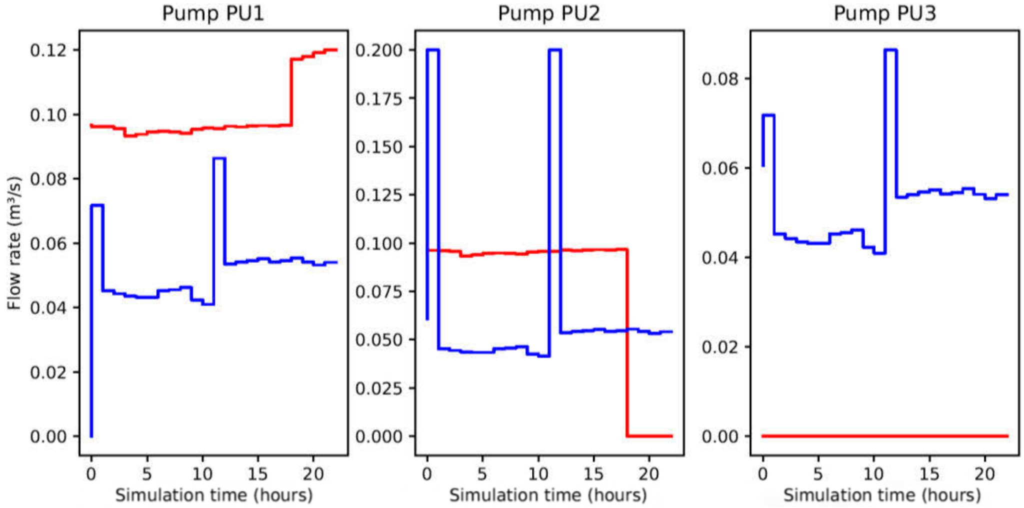


Figure 2. C-Town – Daily schedule for pumps PU1, PU2, PU3. Comparison between pump flowrate with optimized scheduling (blue) and non-optimized operation (red).

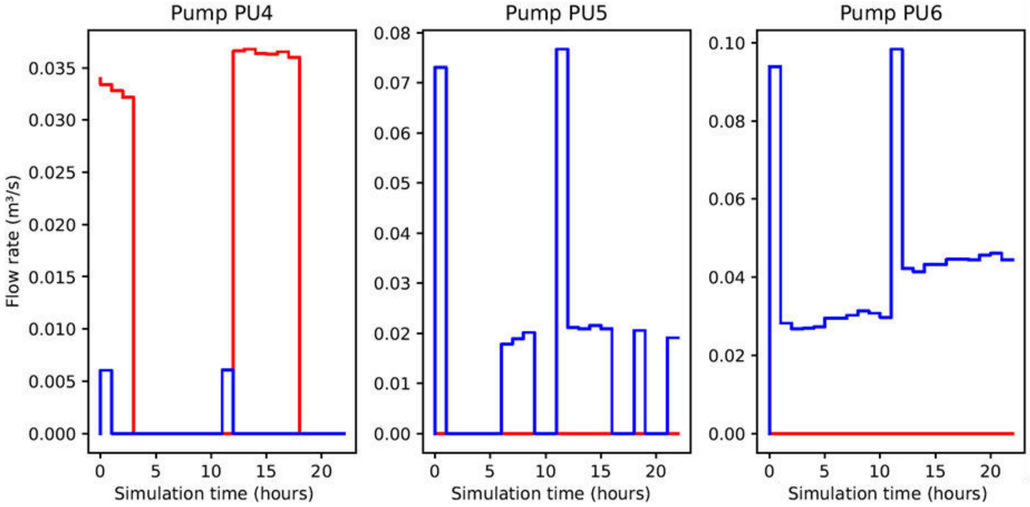


Figure 3. C-Town – Daily schedule for pumps PU4, PU5, PU6. Comparison between pump flowrate with optimized scheduling (blue) and non-optimized operation (red).

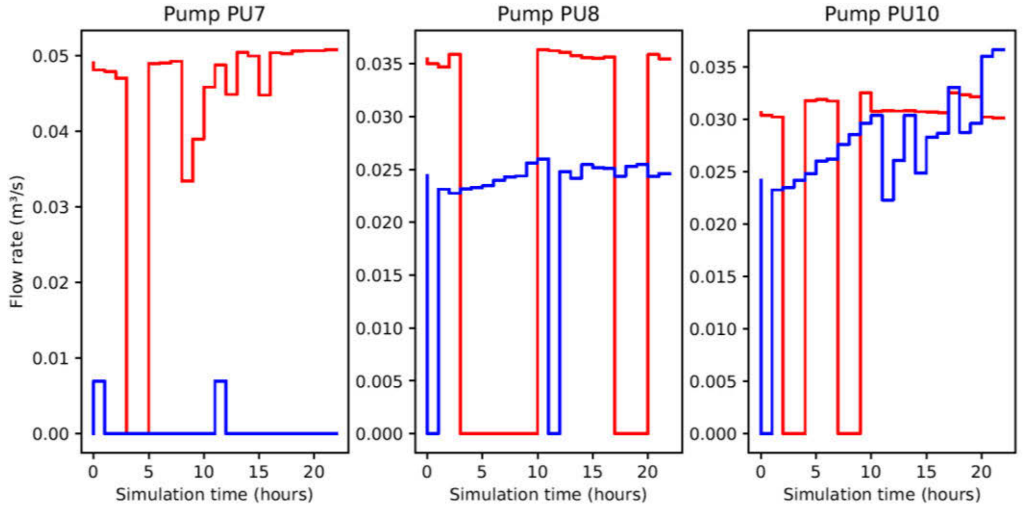


Figure 4. C-Town – Daily schedule for pumps PU7, PU8, PU10. Comparison between pump flowrate with optimized scheduling (blue) and non-optimized operation (red).

4 DISCUSSION

The results presented in Table 2 and in Figures 2-4 provide interesting insights on the benefits and performance of the proposed water distribution network (WDN) operation optimization approach based on a complementarity formulation with penalty terms for the pump constraints, which results in a nonlinear program. This formulation combines the advantages of deterministic optimization approaches based on mathematical programming, with a reduced computational effort that allows its industrial implementation even for large-scale water networks.

From Table 2 it is possible to appreciate that the number of iterations required for the solver IpOpt to converge to a solution scales nicely with the network dimension and with the number of pumps, and the optimization remains tractable even for DWES, the largest water network used here as benchmark. This suggests that the presented approach could be suitable to tackle the operation optimization problem even for large-scale WDN.

The results obtained by applying the complementarity-based optimization to several real-world and open-source water models demonstrated the potential to achieve a reduction of energy consumption and subsequent pumping energy costs that is in line to what described in the relevant literature, e.g. [4] and [7]. The pump flowrate comparison shown in Figures 2-4 exhibits a clear and significant reduction of pump usage when the optimization approach is used, with respect to the non-optimized scenario which is a representative of the current network operation standard. In all the addressed case studies, the use of the presented optimization approach provided energy savings, while guaranteeing satisfaction of water demand and operational requirements. Additional benefits of a reduced pump usage are the increased pump lifetime and less maintenance costs.

5 CONCLUSION

In this work we presented an industrial perspective on the problem of water distribution network operation optimization. A deterministic optimization approach based on mathematical programming with complementarity constraints was proposed, and its application to case studies constructed using open-source WDN models demonstrated the benefits and the performance of the approach. In particular, the presented optimization program achieves a significant reduction of pump usage, lowering energy consumption and the associated costs, while guaranteeing satisfaction of operational constraints. Moreover, the proposed complementarity formulation allowed to move from a mixed integer nonlinear optimization program to a nonlinear one having only continuous variables. This resulted in a tractable problem requiring reduced computational effort, even for large-scale networks. These results demonstrate the potential of the industrial implementation of such optimization approach to medium and large-scale real-world water distribution networks, thus providing a new tool for water utilities, to reduce their energy costs and their carbon footprint.

6 REFERENCES

[1] P. Boulos et al., "Optimal pump operation of water distribution systems using genetic algorithms," in Distribution system symposium, 2001.

[2] D. Torregrossa, Dario and F. Capitanescu, "Optimization models to save energy and enlarge the operational life of water pumping systems," Journal of cleaner production, vol. 213, 2019, pp. 89-98.

[3] T. Luna et al, "Improving energy efficiency in water supply systems with pump scheduling optimization," Journal of cleaner production, vol. 213, 2019, pp. 342-356.

[4] H. Mala-Jetmarova, N. Sultanova, and D. Savic, "Lost in optimisation of water distribution systems? A literature review of system operation," Environmental modelling & software, vol. 93, 2017, pp. 209-254.

[5] A. Martin et al., Mathematical optimization of water networks, Birkhäuser-Science, Basel, 2012.

[6] J. Burgschweiger, B. Gnädig, and M. Steinbach, "Optimization models for operative planning in drinking water networks," Optimization and Engineering, vol. 10, n. 1, 2009, pp. 43-73.

[7] C. D'Ambrosio et al., "Mathematical programming techniques in water network optimization," European Journal of Operational Research, vol. 243, n. 3, 2015, pp. 774-788.

[8] S. Abkenar, et al., "Evaluation of genetic algorithms using discrete and continuous methods for pump optimization of water distribution systems," Sustainable Computing: Informatics and Systems, vol. 8, 2015, pp. 18-23.

[9] J. Van Zyl et al., "Operational optimization of water distribution systems using a hybrid genetic algorithm," Journal of water resources planning and management, vol. 130, n. 2, 2004, pp. 160-170.

[10] M. López-Ibáñez, T. Prasad and B. Paechter, "Ant colony optimization for optimal control of pumps in water distribution networks," Journal of water resources planning and management, vol. 134, n. 4, 2008, pp. 337-346.

[11] EPANET software webpage, https://www.epa.gov/water-research/epanet.

[12] L. Rossman et al., "EPANET 2.2 User Manual," US Environmental Protection Agency, Washington, DC, EPA/600/R-20/133, 2020.

[13] A. Pirsing and A. Morsi, Decision Support Systems for Water Supply Systems, European Mathematical Society, Berlin, 2020.

[14] T. Walski et al., "Advanced water distribution modeling and management," Haestad press, 2003.

[15] C. P. Liou, "Limitations and proper use of the Hazen-Williams equation," Journal of Hydraulic Engineering, vol. 124, n. 9, 1998, pp. 951-954.

[16] G. Brown, "The history of the Darcy-Weisbach equation for pipe flow resistance," Environmental and water resources history, 2003, pp. 34-43.

[17] J. Burgschweiger, B. Gnädig and M. Steinbach, "Nonlinear programming techniques for operative planning in large drinking water networks," Open Applied Mathematics Journal, vol. 3, 2009, pp. 14-28.

[18] H. Scheel, and S. Scholtes, "Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity," Mathematics of Operations Research, vol. 25, no. 1, 2000, pp. 1-22.

[19] D. Ralph and S. Wright, "Some properties of regularization and penalization schemes for MPECs," Optimization Methods and Software, vol. 19, no. 5, 2004, pp. 527-556.

[20] R. Fletcher and S. Leyffer, "Solving mathematical programs with complementarity constraints as nonlinear programs," Optimization Methods and Software, vol. 19, no. 1, 2004, pp. 15-40.

[21] A. Raghunathan and L. Biegler, "An interior point method for mathematical programs with complementarity constraints (MPCCs)," SIAM Journal on Optimization, vol. 15, no. 3, 2005, pp. 720-750.

[22] M. Anitescu, "Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints," SIAM Journal on Optimization, vol. 16, no. 1, 2005, pp. 120-145.

[23] J. Sousa et al., "Optimal management of water distribution networks with simulated annealing: the C-town problem," Journal of Water Resources Planning and Management, vol. 142, n. 5, 2016.

[24] E. Creaco, S. Alvisi and M. Franchini, "Multistep approach for optimizing design and operation of the C-town pipe network model," Journal of Water Resources Planning and Management, vol. 142, n. 5, 2016.

[25] A. Ostfeld et al., "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," Journal of water resources planning and management, vol. 134, n. 6, 2008, pp. 556-568.

[26] A. Ostfeld et al., “Battle of the water calibration networks,” Journal of Water Resources Planning and Management, vol. 138, n. 5, 2012, pp. 523-532.

[27] J. Andersson et al., “CasADi: a software framework for nonlinear optimization and optimal control,” Mathematical Programming Computation, vol. 11, n. 1, 2019, pp. 1-36.

[28] IpOpt webpage, https://coin-or.github.io/Ipopt/.

[29] A. Wächter and L. T. Biegler, “On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming”, Mathematical Programming, vol. 106, n. 1, pp. 25-57, 2006.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14158

BIOKINETIC AND ARTIFICIAL INTELLIGENCE MODELS FOR THE SIMULATION OF NITROUS OXIDE EMISSIONS FROM WASTEWATER TREATMENT PLANTS

Siddharth Seshan[1], Johann Poinapen[2], Jules B. van Lier[3] and Zoran Kapelan[4]

[1,2]KWR Water Research Institute, Nieuwegein, The Netherlands
[1,3,4]Department of Water Management, Delft University of Technology, Delft, The Netherlands

[1] *siddharth.seshan@kwrwater.nl*, [2] *johann.poinapen@kwrwater.nl*,[3] *j.b.vanlier@tudelft.nl*,
[4] *z.kapelan@tudelft.nl*

Abstract

Nitrous oxide (N2O) is considered a potent and very harmful greenhouse gas (GHG), and wastewater treatment plants (WWTPs) are considered a potent source of it. Predicting N2O emissions is a first step in reducing these. One way of doing this is by using a process-based biokinetic model, based on Activated Sludge Models (ASMs) that have been extended to include the N2O production pathways. Alternatively, data-driven Artificial Intelligence (AI) models can be used to predict N2O emissions. In this paper, a biokinetic model has been built and calibrated for the Amsterdam West WWTP (1.1 Million PE; 168 MLD), using the EnviroSim software, BioWin®. A comprehensive monitoring campaign was conducted to characterise the common quality parameters (COD, TKN, TP, TSS, etc.) into their fractions, which were then used as BioWin model inputs. The calibration was conducted in two stages to predict effluent quality followed by model calibration to predict N2O emissions. Additionally, an Artificial Neural Network (ANN) based model was developed using pertinent process parameters, such as the influent flowrate, and NH4 in the aerobic tank as inputs to predict the N2O concentration in the gas phase. Preliminary results demonstrate that the ANN model outperforms the BioWin model in terms of prediction accuracy. Still further work is required to better understand the pros and cons of the two modelling approaches.

Keywords
Nitrous oxide, wastewater treatment, biokinetic modelling, artificial intelligence, artificial neural network.

1 INTRODUCTION

1.1 Background

With the increasing effects of climate change and global warming becoming more apparent, great efforts are being made in reducing the carbon footprint of industries and society in general. Nitrous oxide (N_2O) is considered a potent and very harmful greenhouse gas (GHG); in addition, N_2O has been considered to contribute to the depletion of the ozone layer in the stratosphere [1]. While the global anthropogenic GHG emission contribution from N_2O can be considered minor (6.2% in terms of CO_2eq), the global warming potential of N_2O is very high, 298 times greater than that of CO_2 on a 100-year time scale [2]. In the past decades, wastewater treatment plants (WWTPs) are increasingly considered to be one of the potent sources of N_2O. Therefore, there is a global call for action to invest and investigate in advance wastewater treatment technologies and operational strategies to reduce the generation of the harmful gas [3]. As a result, the quantification of N_2O emissions from full-scale plants have been considered greatly, and monitoring campaigns of varying durations have been conducted. Additionally, biokinetic models, primarily using the widely known activated sludge models (ASMs) have been extended to include the N_2O production pathways. In parallel, like many other domains, more and more data-driven based analytics are being utilised and adopted in the wastewater field, to solve complex process challenges. Artificial Intelligence (AI) models also have been used in the prediction of key

CC BY NC SA

wastewater parameters including N_2O in data-rich systems. In this study, a biokinetic model was calibrated using long-term (1 year) N_2O emissions data from a full-scale WWTP and its predictive capabilities were assessed. To acquire specific data on the raw influent wastewater that is necessary as input into the model, a comprehensive sampling campaign was also conducted. Finally, initial investigations of training AI models to predict the N_2O emissions were conducted. Preliminary comparison of the performance with the biokinetic model predictions have been discussed.

1.2 Biokinetic Modelling of N_2O Emissions

Activated sludge models (ASMs) are widely and successfully used for process modelling, subsequentially supporting in finding solutions to process design and operational problems [4]. The production of N_2O emissions in WWTPs has been associated with the process of biological nitrogen removal, where three production pathways are prominent, of which two are attributed to the ammonia oxidising bacteria (AOB), i.e., hydroxylamine oxidation and nitrifier denitrification; and one is attributed to incomplete heterotrophic denitrification. Over the past decade, such biokinetic models have been extended to also include the production pathways of N_2O, for its prediction and to test control and mitigation strategies. While the current N_2O biokinetic models have been able to predict the general trend of the observed N_2O emissions, the prediction accuracies are still unsatisfactory [5, 6]. Furthermore, when a calibrated model is confronted with unseen data for validation, the prediction accuracy has been reported to be much lower, thereby questioning the capabilities of the models to future data [4, 5, 6, 7, 8].

The modelling investigations were mostly carried out using datasets obtained from a controlled environment as a lab-scale or pilot-scale setup, barring a handful of cases that used data from full-scale systems [5, 6, 7, 9, 10]. Therefore, the applicability of such biokinetic N_2O models in a full-scale WWTP can be questioned given the controlled operating and process conditions that are administered in lab/pilot scale setups. Furthermore, in the cases of full-scale based investigations, the duration of data used for the calibration and validation purposes can be considered short-term (< 1 month) or medium-term (< 1 year). In the prediction of N_2O emissions using biokinetic models, there is a clear requirement to analyse the performance of the models when calibrated on long-term data containing seasonal variations and under full-scale operating conditions.

1.3 Data-driven Modelling of N_2O Emissions

The use of data-driven based analytics or AI models to predict N_2O emissions in WWTPs is still sparse. Even though advanced information extraction methods and dimensionality reduction techniques have been used on WWTP data, a handful of investigations have used these methods to analyse data from N_2O monitoring campaigns [11]. For example in the investigation in [12], a Random Forest (RF) analysis, a machine learning method, was used to identify the primary effecters of N_2O emissions from a full-scale BNR system. In [13], Support Vector Machine (SVM) classifiers were trained to predict, with high accuracy (95% - 99%), whether the dissolved N_2O will be consumed during the anoxic and anaerobic phases and subsequently, used the information from the classifiers to predict the average dissolved N_2O concentration in the anaerobic and aerobic phases by training a Support Vector Regression (SVR) model for each. The SVR models were reported to predict with good accuracy with R^2 values ranging from 0.85 – 0.94 on the training dataset and 0.75 – 0.82 in the test dataset. The use of Deep Learning (DL) models for N_2O predictions is still a novelty, with only a handful of studies being conducted [14, 15]. A Deep Neural Network (DNN) model was developed to predict the N_2O in the liquid phase using over a year of operational data from a WWTP, while using the influent flowrate, DO, NH_4, NO_3, air flowrate and temperature as inputs. While using a trailing moving average to smoothen the data, a significantly high R^2 value of 0.9 was achieved. Furthermore, the DNN model was compared with a Long Short Term Memory (LSTM) based Recurrent Neural Network (RNN) model, that utilised historical N_2O data to predict future values. The two models were assessed for their forecasting capabilities over a 1-day prediction horizon. It was seen that the DNN model's performance was limited ($R^2 = 0.76$), while the LSTM model achieved better results ($R^2 = 0.94$). Accordingly, the application of data-driven models could present a suitable alternative to biokinetic modelling, especially to overcome the latter's limitations.

2 METHODOLOGY

2.1 Brief Description of WWTP

The case study for this investigation is the Amsterdam West WWTP (**¡Error! No se encuentra el origen de la referencia.**) that is operated by the water company Waternet. Amsterdam West WWTP has a capacity of 1.1 million population equivalent (168 MLD) and serves the Amsterdam city and its neighbouring regions. The raw influent wastewater, post the grit chambers and primarily settlers, is distributed to 7 treatment lanes to conduct the activated sludge (AS) process for biological nitrogen and phosphorus removal. The process configuration applied for the AS is the modified University of Cape Town (mUCT) process.

Figure 1: Amsterdam West WWTP

For this study, data from legacy online sensors measuring key process parameters of the influent flow, wastewater quality parameters in the bioreactor, recirculation flowrates, sludge flowrates, and effluent flowrates, were used. Most importantly, time series online sensor data of the N_2O emissions measured in one treatment lane is also available. A distinctive condition in Amsterdam West is that the bioreactor units are covered, which allows for the accurate capture of the off-gas emissions from the AS processes, providing direct measurements of the N_2O in the gaseous phase. All signals have a time resolution of 1 minute. Furthermore, historical laboratory measurements are also available for TKN, TN, Total P, BOD, COD for all stages in the wastewater treatment process.

2.2 Biokinetic Modelling to Predict N_2O Emissions

A comprehensive sampling campaign was conducted to characterise the influent as well as to obtain additional effluent data to calibrate the biokinetic model. Flow proportional daily composite samples and diurnal sampling were taken for the raw influent and effluent wastewater, where the following parameters were monitored: COD_{total}, $COD_{filtered}$, BOD_{total}, $BOD_{filtered}$, TKN, NH_4, NO_3, Total P, Ortho-PO_4, ISS and TSS. Numerous samples were also taken at various locations including the anaerobic, anoxic, and aerobic zones of the bioreactor, sludge treatment lines and return streams.

For this investigation, a biokinetic model using the ASDM model coupled with a N_2O model was calibrated for 1 treatment lane of the Amsterdam West WWTP. The simulation software by EnviroSim, BioWin® was used to conduct the research. Initially, the process configuration (Figure 2) of the WWTP was set up in BioWin using the plant design and operating parameters.

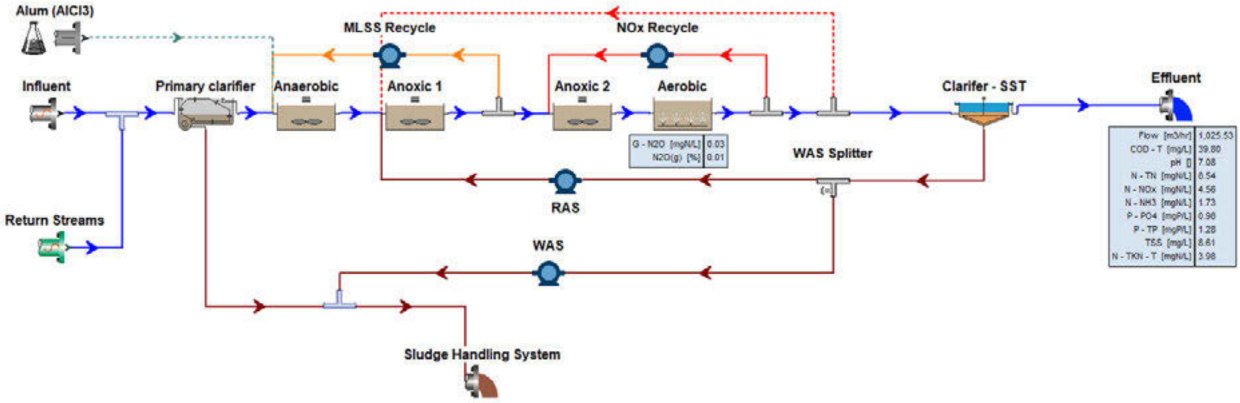


Figure 2: Process Configuration of 1 treatment lane in Amsterdam West WWTP, as setup in the simulation software BioWin.

Subsequently, the calibration procedure was conducted in two stages. In the first stage, the model was calibrated using the wastewater characteristics and fractions obtained from the sampling campaign as well as historical datasets. Subsequently, steady-state and dynamic simulations were performed using the N_2O model default values in BioWin. In the second stage, further calibration was conducted with a goal to match the observed plant N_2O emission data. The N_2O model provided in BioWin includes the description of all three N_2O production pathways. The kinetic parameters specific to N_2O production from the nitrification and denitrification processes were then fine-tuned to be able to match the observed N_2O emissions from the gaseous phase. During the calibration of the N_2O specific kinetic parameters, it was ensured that the effluent quality is matched at all times with the observed data which served as a strict boundary to adhere to.

The calibration of the model for dynamic simulations was conducted using a variety of datasets as acquired from the sampling campaign, historical laboratory measurements and from online sensors. Simulations were performed by inputting the flowrate and quality parameters for the influent wastewater and return streams from laboratory measurements. Daily values of duration 1 year from 11/2020 – 10/2021, covering all 4 seasons, were used.

2.3 Preliminary AI Modelling Investigations

As an alternative to biokinetic modelling of N_2O emissions, preliminary investigations were also made to assess the predictive capabilities of AI models. An Artificial Neural Network (ANN) model was developed. As input to the model, 3 parameters were used, namely the raw influent flowrate to 1 treatment lane, NH_4 concentration levels in the aerobic tank of the bioreactor and the N_2O concentration levels (in ppmv). The target variable is the N_2O concentration levels. The data was resampled to a resolution of 15 minutes. The model structure included 1 hidden layer containing 64 units. The model was trained to take a certain amount of historical input of the variables to make a one-step ahead prediction of the target. As a result, the data was prepared into sequences where data amounting to 2 days (192 values) of all input variables was used as historical input into the model. The data of the target variable were prepared accordingly to also facilitate the training of the ANN model to perform a one-step-ahead prediction. The model was trained on 1 year of data (11/2020 – 10/2021), and 3 months of data was used for testing.

3 RESULTS AND DISCUSSIONS

The results of the N_2O emission predictions from the biokinetic model (black line) compared with the observed data (blue dots) are depicted in Figure 3, where a seasonal N_2O peak is distinctive. To simulate this peak in the biokinetic model (in BioWin), season-specific fine-tuning of the N_2O related parameters was necessary. Specifically, the parameters related to the nitrifier

cc BY NC SA

denitrification production pathway were adjusted along with the free nitrous acid inhibition parameter to allow for more production of N_2O due to incomplete denitrification. Furthermore, the Arrhenius value related to the mass transfer (Kl) of N_2O was also changed (reduced by half) to allow for more stripping of N_2O from the liquid to gaseous phase. While these changes resulted in successfully and accurately predicting the N_2O seasonal peak, as shown in Figure 3, there are currently no proper scientific justification for these changes in the model. For such a justification, extended research and investigation on these kinetic and temperature dependent parameters are required, which is beyond the scope of this study. Moreover, the occurrence of the N_2O seasonal peak could point in the direction that there are still a lot of uncertainties with regard to the biokinetic processes related to the N_2O production pathways currently in the model. In the seasons where low N_2O emissions were observed, the default parameters available in BioWin were sufficient to accurately predict the N_2O emissions, except for the Arrhenius value for the NOB maximum specific growth rate. Such a result would be expected as the default values available in BioWin were calibrated on datasets obtained from bench-scale studies during summer-like conditions [16, 17]. Adjusting some of the N_2O biokinetic model parameters without scientific justification in order to satisfactorily predict the seasonal peak implies the limitation of the biokinetic model.

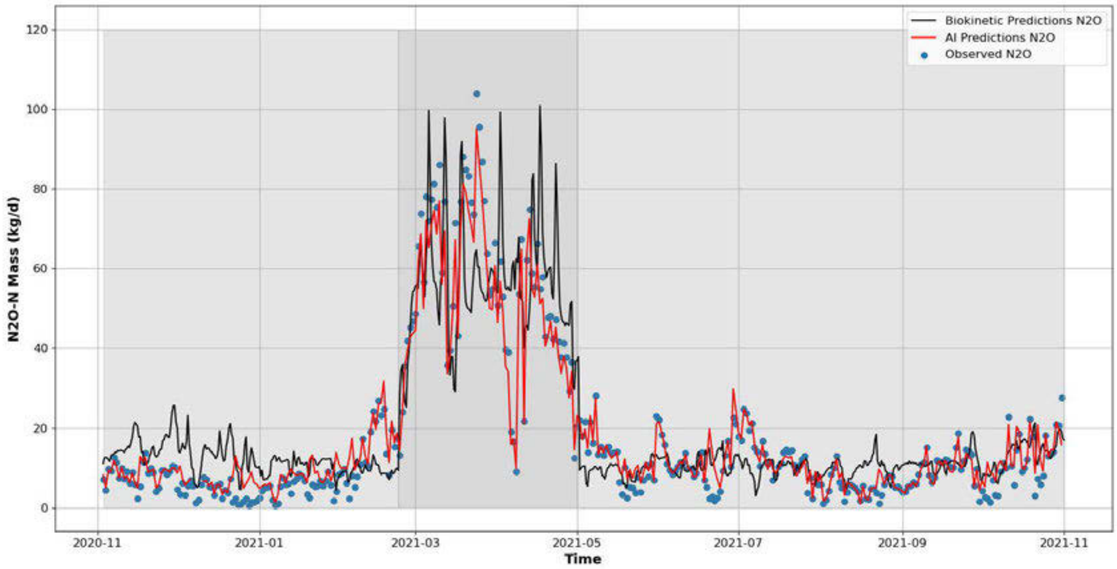


Figure 3: Comparing the Biokinetic Model N_2O Emissions Predictions (black line) and AI Model Predictions (red line) with Observed N_2O Emissions (blue dots) for a Period of 1 Year. The dark grey area signifies the seasonal N_2O emissions peak that was observed.

A similar simulation exercise was conducted using the developed AI model in order to compare with the biokinetic model predictions. From the initial results illustrated in Figure 3, the predictions from the AI model (red line) resulted in a good fit to the observed data, suggesting that the AI model outperforms the biokinetic model. This positive and encouraging outcome from the AI model presents a suitable alternative to overcome the limitations of the biokinetic model in satisfactorily predicting N_2O emissions for all seasonal conditions.

4 CONCLUSIONS

In this study, a biokinetic model was developed first using the BioWin software. This model was calibrated on 1 year of data from a real-life WWTP. Additional data were collected during a comprehensive sampling campaign conducted to characterise the raw influent into its fractions. During the calibration of the biokinetic model, it was established that the default values of the physical and kinetic parameters were unable to adequately predict the seasonal variations observed in the N_2O emissions. As a result, an AI based model in the form of an Artificial Neural Network was developed and trained on the same data set as an alternative. Preliminary results show that the AI model is outperforming the biokinetic model in terms of prediction accuracy. Therefore, further research on developing data-driven and possibly hybrid models should be pursued. An improved N_2O prediction model can then be used for the development of (near) real-time control strategies to mitigate the production of N_2O emissions from WWTPs.

5 ACKNOWLEDGMENTS

The presented research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 821036.

6 REFERENCES

[1] Intergovernmental Panel on Climate Change, R.K Pachauri, and A. Reisinger, "Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change". IPCC, Geneva, Switzerland, 2007, pp 104.

[2] Intergovernmental Panel on Climate Change, T.F., Stocker, D. Qin, G.K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013, pp. 1535

[3] Intergovernmental Panel on Climate Change, O. Edenhofer, R. Pichs-Madruga, Y. Sokona, J. Minx, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schloemer, C. von Stechow, and T. Zwickel, "Climate Change 2014 Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, 2014.

[4] G. Sin and R. Al, "Activated sludge models at the crossroad of artificial intelligence—A perspective on advancing process modelling," Npj Clean Water, vol. 4, no. 1, 2021, pp. 1–7

[5] B.J. Ni, L. Ye, Y. Law, C. Byers and Z. Yuan, "Mathematical modeling of nitrous oxide (N2O) emissions from full-scale wastewater treatment plants," Environmental Science and Technology, vol. 47, no. 14, 2013, pp. 7795–7803.

[6] M. Spérandio, M. Pocquet, L. Guo, B.J. Ni, P. A. Vanrolleghem, and Z. Yuan, "Evaluation of different nitrous oxide production models with four continuous long-term wastewater treatment process data series," Bioprocess and Biosystems Engineering, vol. 39, 2016.

[7] K.E. Mampaey, M. Spérandio, M. C. M. van Loosdrecht and E. I. P. Volcke, "Dynamic simulation of N2O emissions from a full-scale partial nitritation reactor," Biochemical Engineering Journal, vol. 152, 2019.

[8] B. J. Ni, L. Peng, Y. Law, J. Guo and Z. Yuan, "Modeling of Nitrous Oxide Production by Autotrophic Ammonia-Oxidizing Bacteria with Multiple Production Pathways," Environmental Science and Technology, vol. 48, no. 7, 2014, pp. 3916-3924.

[9] K. Blomberg, P. Kosse, A. Mikola, A. Kuokkanen, T. Fred, M. Heinonen, M. Mulas, M. Lübken, M. Wichern and R. Vahala, "Development of an Extended ASM3 Model for Predicting the Nitrous Oxide Emissions in a Full-Scale Wastewater Treatment Plant," Environmental Science and Technology, vol. 52, no. 7, 2018, pp. 5803–5811.

[10] B.J Ni, Y. Pan, B. van den Akker, L. Ye and Z. Yuan, "Full-Scale Modeling Explaining Large Spatial Variations of Nitrous Oxide Fluxes in a Step-Feed Plug-Flow Wastewater Treatment Reactor," Environmental Science and Technology, vol. 49, no. 42, 2015, pp. 9176-9184.

[11] V. Vasilaki, T.M. Massara, P. Stanchev, F. Fatone and E. Katsou, "A decade of nitrous oxide (N2O) monitoring in full-scale wastewater treatment processes: A critical review," Water Research, vol. 161, 2019, pp. 392–412.

[12] M. J. Song, S. Choi, W. B. Bae, J. Lee, H. Han, D. D. Kim, M. Kwon, J. Myung, Y. M. Kim and S. Yoon, "Identification of primary effecters of N2O emissions from full-scale biological nitrogen removal systems using random forest approach," Water Research, vol. 184, 2020.

[13] V. Vasilaki, V. Conca, N. Frison, A. L. Eusebi, F. Fatone and E. Katsou, "A knowledge discovery framework to predict the N2O emissions in the wastewater sector," Water Research, vol. 178, 2020.

[14] S. Hwangbo, R. Al and G. Sin, "An integrated framework for plant data-driven process modeling using deep-learning with Monte-Carlo simulations," Computers & Chemical Engineering, vol. 143, 2020.

[15] S. Hwangbo, R. Al, X. Chen and G. Sin, "Integrated Model for Understanding N 2 O Emissions from Wastewater Treatment Plants: A Deep Learning Approach," Environmental Science and Technology, vol. 55, 2021, pp. 2143-2151.

[16] D. Houweling, P. Wunderlin, P. Dold, C. Bye, A. Joss and H. Siegrist, "N 2 O Emissions: Modeling the Effect of Process Configuration and Diurnal Loading Patterns," Water Environment Research, vol. 83, no. 12, 2011, pp. 2131–2139.

[17] D. M. Shiskowski, "Investigation of Nitrous Oxide As a Nitrification Monitoring and Aeration System Control Parameter in Sequencing Batch Reactor Wastewater Treatment Systems", Ph.D. Thesis, University of British Columbia, Vancouver British Columbia, Canada, 2004.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14161

FLOW RATIO AND PRESSURE CONTROL IN A LOW LOSSES WDS – ANALYSIS OF DYNAMICS

Ulanicki, B.[1], Beaujean, Ph.[2] Plancke, J.[3], Leruth, B.[4]

[1] De Montfort University, Water Software Systems, Leicester, LE2 4NW, UK
[2,4] Société wallonne des eaux, Rue de la Concorde,41 4800 Verviers, Belgium
[3] SOFTEAU srl, www.softeau.be, Belgium

[1] bul@dmu.ac.uk , [2] philippe.beaujean@swde.be, [3] juliaan.plancke@skynet.be, [4] bastien.leruth@swde.be

Abstract

Société wallonne des eaux (SWDE) is the most important water production and distribution company in Wallonia, one of the regions in Belgium. Its distribution network stretches over 40,000 km. It covers nearly 200 municipalities and has more than one million connections. SWDE supplies drinking water to nearly 2.4 million people, constituting over 70% of the Walloon population. Recently, SWDE has launched a major multimillion-Euro project to improve security of the water supply in Wallonia. The project involves a 500/800 mm extension of a 1000 mm wide, over 100 km long pipeline as a backbone of the distribution network, starting from the Eupen and Gileppe reservoirs. The project created a number of engineering challenges and one of them is considered in this paper. It is concerned with the design of a feedback control scheme to achieve satisfactory operation in terms of water mixing from two sources and pressure stability in the new water distribution system (WDS).

A feedback control strategy to maintain this constant ratio have been investigated. The objective was to check stability of the control scheme. The concern was about potential interactions between the PRV operation and the pump speed control loop. Matlab/Simulink which allows to model generic dynamic systems as well as control components has been selected as a modelling environment. A dynamic model of the WDS has been implemented following the general methodology developed by the authors and presented earlier in the literature. Pipes have been represented by the rigid column model, the PRV by behavioral model described by a first order differential equation and a pump by a first order inertial equation. In a low losses system, small changes of the pressure at the PRV outlet may cause significant changes in the flow, making the flow ratio control difficult. An important innovation was to introduce an additional TCV downstream to the PRV to increase head losses and reduce the interactions between the flow ratio control loop and the PRV. The paper presents results of an extensive simulation study of the control system behavior for different control parameters and time varying demands.

Keywords
Flow control, Pressure control, Rigid Water Column model, Dynamic, Stability.

cc BY NC SA

1 INTRODUCTION – CASE STUDY DESCRIPTION

A schematic of the system fragment where the mixing takes place is presented in Figure 1.

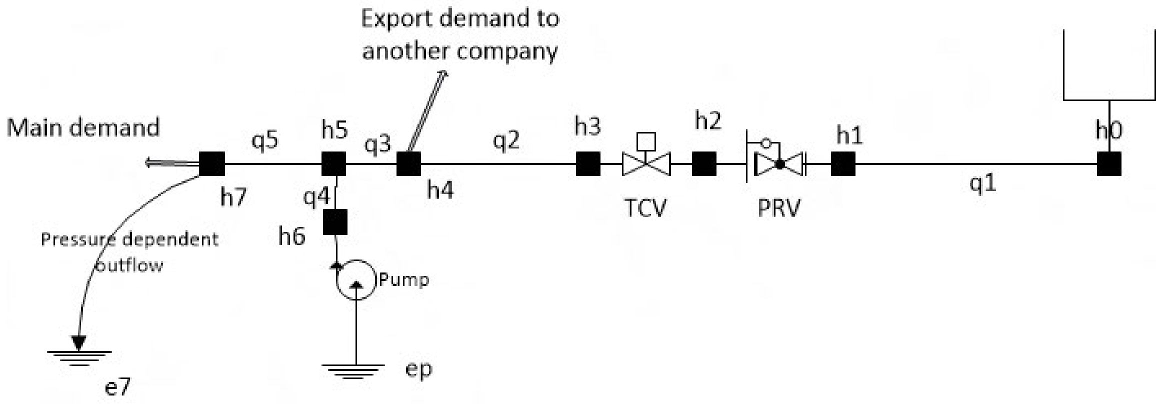


Figure 1. Schematic of the water supply system

The source flow comes from the Eupen reservoir with capacity of 50,000 m^3. In order to maintain the required pressure in the downstream part of the water system a PRV with the outlet head of 254m (pressure 28m) is planned to be installed. There is a significant water demand at node h4 where the water is exported to another water company. At node h5 the Eupen water is mixed with the water imported from another water company through a pumping station. It is necessary to maintain a constant ratio of 0.6 between the q4 and q5 flows due to water quality requirements, and at same time to maintain feasible positive pressure at node h7 in order to provide enough pressure at a critical point downstream of the system not shown in Figure 1. The pipe data are presented in Table 1.

Table 1. Pipe data]

Pipe id	Length [m]	Diameter [mm]	D-W friction factor
q1	46212	800	0.0175
q2	9749	800	0.015
q3	10	800	0.015
q4	10	800	0.015
q5	21179	800	0.015

Pipes q1, q2 and q5 are long pipes connecting important sites along the main pipeline and pipes q3 and q4 are short pipes around the pump shaft.

The node data are included in Table 2.

Table 2. Node data]

Node id	Elevation [m]	Total demand [m3/d]	Comment
h0	322.5	-	Constant Head Source
h1, h2, h3	226	-	Valve connection nodes
h4	129.5	5000	Demand pattern in Figure 2
h5	129.5	-	Connection node
h6	120	-	Destination node of the pump station
h7	180	7218	Main demand+ outflow area $=10^{-7}$ m^2

The node h0 represents a fixed head source and the node h7 is a boundary node representing the whole downstream part of the system with the main demand equal to $7218\,{m^3}/{d}$ and an additional pressure dependent outflow with the equivalent area of $10^{-7}m^2$.

In this study, Cla-Val DN250 was selected as a PRV and the Monovar DN250 valve was selected as a TCV, the respective valve capacity characteristics are displayed in Figures 2a and 2b .

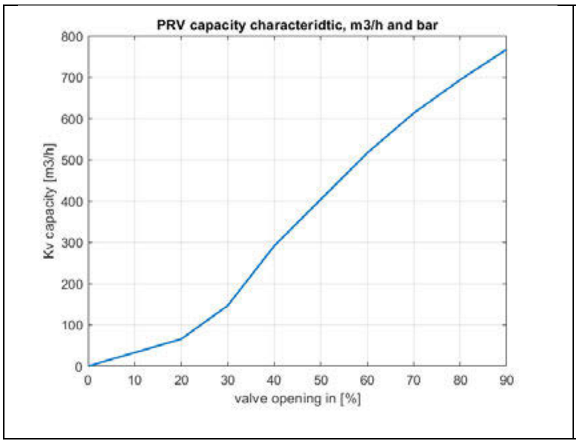


Figure 2a. PRV capacity characteristic

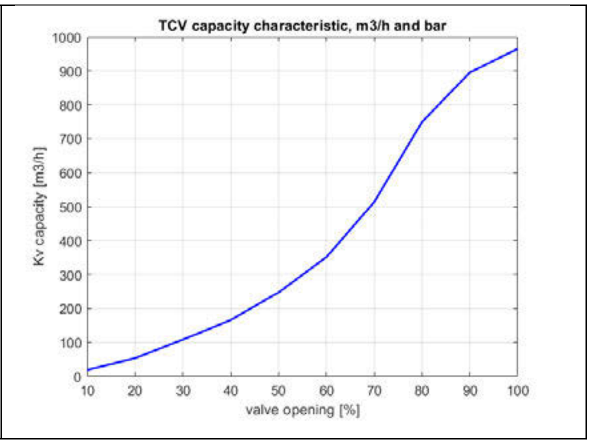


Figure 2b. TCV capacity characteristic

The characteristics were approximated by piece-linear functions in the Simulink model to preserve accuracy of the original data from the catalogue.

The pump model was derived from the manufacturer data for the KSB pump, type W12-200/T14 and is displayed in Figure 3.

cc BY NC SA

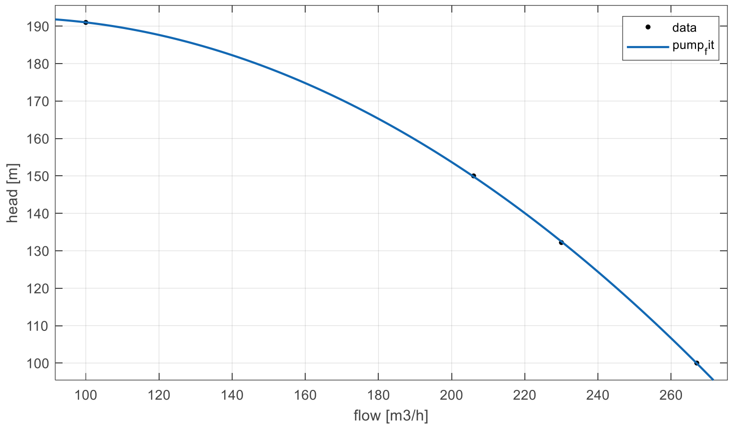


Figure 3. KSB pump characteristic

Analytically the model was represented by a second order polynomial given in Equation 1,

$$\Delta h = \; -0.0026 * q^2 + \; 0.3974 * q * s + 177s^2 \tag{1}$$

where q is the pump flow in $[{}^{m^3}/_{h}]$, Δh is the head increase across the pump in $[m]$ and finally s is the relative pump speed, where value 1 corresponds to the nominal speed.

The overall objective was to check feasibility of achieving stable and robust operation of the system to maintain the flow ratio ${}^{q4}/_{q5} = 0.6$. The specific objectives were identified as follow:

- Design an algorithm to control pump speed
- Investigate the need for an additional head loss (TCV)
- Investigate operation of the system over time varying flow conditions.

The reminder of the paper is organised as follows. The Methodology section describes an approach adopted in this study, simulation results to answer the research questions are presented in the Results section, the paper finishes with conclusions and final recommendations.

2 METHODOLOGY

The methodology is based on the theory presented in [1]. A mathematical model is formulated in the form of differential-algebraic equations (DAEs) and the model is simulated using the Matlab/Simulink environment.

The rigid water column (RWC) model given in Equation 2 is used for pipes,

$$\dot{q} = -Rq|q| + M\Delta h \tag{2}$$

where $q =$ pipe flow; $\Delta h =$head loss along the pipe; $R = f_D({}^{1}/_{2DA})$; $M = {}^{gA}/_{L}$; $f_D =$Darcy friction factor; $D =$pipe hydraulic diameter; $A =$pipe cross section area; $L =$pipe length; and $g =$ gravitational acceleration. The RWC can be used for analysis of slow transients, [2] or for analysis of control problems in WDSs, [1].

The PRV is represented by a behavioural model from [3] and is given in Equations 3 and 4 (differential part) and in Equation 5 (algebraic part),

BY NC SA

$$\dot{x}_m = \alpha_{open}(h_{set} - h_d) \tag{3}$$

$$\dot{x}_m = \alpha_{close}(h_{set} - h_d) \tag{4}$$

where x_m =valve opening in percent; α_{open} and α_{close}=rates of the valve opening and closing, respectively; h_{set} =PRV setpoint and h_d = PRV outlet head. The algebraic part is given by a standard valve equation, Equation 5

$$q = K_v(x_m)\sqrt{\Delta h} \tag{5}$$

where $K_v(x_m)$=valve capacity, which depends on the valve opening x_m, the $K(x_m)$ characteristic provided by the manufacturer is shown in Figur2a.

The TCV is a static element represented by an algebraic equation of the same form as Equation 5 and with the capacity characteristic displayed in Figure 2b.

The pump is represented by the two equations. The first equation, Equation 6

$$\dot{s} = -\frac{1}{T}s + \frac{1}{T}v \tag{6}$$

is a differential equation that describes the pump inertia, where s =pump speed, v = speed set-point, and T = time constant. The second is an algebraic equation that describes the head increase along the pump scaled by the pump speed s, as discussed in[4],

$$\Delta hh_o = Aq^2 + Bq + Cs^2 \tag{7}$$

where the specific values for coefficients A, B, C are given in Equation 1.

According to the theory presented in [1] it is necessary to select pipe independent flows, in this case they are *q1* and *q4* while the reminder flows can be evaluated from mass balance at the connection nodes as follow:

$$\begin{gathered} q_2 = q_1 \\ q_3 = q_2 - d_{export} \\ q_5 = q_3 + q_4 \\ q_{out} = q_5 - d_{main} \end{gathered} \tag{8}$$

where d_{export} is the water export to another water company at node h4 in Figure 1, $q_{out} = c_{out}(h7 - e7)^{0.5}$ is the pressure dependent outflow in the downstream system and d_{main} is the main flow of the downstream system, c_{out} is the outflow coefficient and $e7$ is the elevation of the h7 node.

The independent pipe flows *q1* and *q4* are explicitly represented by the differential Equation 2. To make sure that the dependent flows in other pipes, i.e., $q2, q3, q5$ satisfy the RWC model it is necessary, according to [1], to introduce the 'differential mass balance equations for the connection nodes with pipes as follow:

$$\dot{q}_2 = \dot{q}_1$$

$$\dot{q}_3 = \dot{q}_2 - \dot{d}_{export}$$

$$\dot{q}_5 = \dot{q}_3 + \dot{q}_4$$

which are subsequently is transformed into Equations 9 with the help of the relationship from

Equation 2.

$$\begin{gathered} -R_2q_2|q_2| + M_2\Delta h_2 = -R_1q_1|q_1| + M_1\Delta h_1 \\ -R_3q_3|q_3| + M_3\Delta h_3 = -R_2q_2|q_2| + M_2\Delta h_2 - \dot{d}_{export} \\ -R_5q_5|q_5| + M_2\Delta h_2 = -R_3q_3|q_3| + M_3\Delta h_3 - -R_4q_4|q_4| + M_4\Delta h_4 \end{gathered} \tag{9}$$

The mass balance equations need to be complemented by the energy balance equations for the two pseudo-loops, the first between the source h_0 to the ground level of the pump branch and the second between the source and the ground level of the outflow branch $e7$.

Starting from the pump branch and following the clockwise direction yields,

$$e_p + \Delta h_{pump} - \Delta h_4 + \Delta h_3 + \Delta h_2 + \Delta h_{TCV} + \Delta h_{PRV} + \Delta h_1 - h_0 = 0 \tag{10}$$

and then starting from the outflow branch provides,

$$e_7 + \Delta h_{leak} + \Delta h_5 + \Delta h_3 + \Delta h_2 + \Delta h_{TCV} + \Delta h_{PRV} + \Delta h_1 - h_0 = 0 \tag{11}$$

The differential part of the DAE model is represented by Equation 1 for the independent flows q_1 and q_4, Equations 2 for the valve opening, x_m and Equation 6 for the pump speed s.

Consequently, the state vector of the model $\boldsymbol{x}$ is made of the variables listed in Equation (12).

$$\boldsymbol{x} = [q_1 \quad q_4 \quad x_m \quad s]^T \tag{12}$$

The algebraic part of the model is represented by Equations 5, 7, 8, 9, 10 and 11 and the vector of the algebraic variables $\boldsymbol{y}$ is composed of the dependent flows and the head losses/gains on the all components depictured in Figure 1, i.e.,

$$\boldsymbol{y} = [q_2 \quad q_3 \quad q_5 \quad \Delta h_{PRV} \quad \Delta h_{TCV} \quad \Delta h_{pump} \quad \Delta h_{out} \quad \Delta h_1 \quad \Delta h_2 \quad \Delta h_3 \quad \Delta h_4 \quad \Delta h_5]^T. \tag{13}$$

The considered DAE system has index 1,as discussed in [1], which means that for the given value of the vector $\boldsymbol{x}$ the vector $\mathbf{y}$ can be evaluated in a unique way by solving equations of the algebraic part of the model In this case the solution of the algebraic part can progress in the following steps: calculate the dependent flows from Equations 8; calculate the head losses/gains on control elements from Equations 5 and 7; calculate head losses on the pipes $\Delta h_1, \Delta h_2, \Delta h_3, \Delta h_4,$, Δh_5 from the system of linear simultaneous equations composed of Equations 9, 10 and 11.

The Simulink model is depicted In Figure 4.

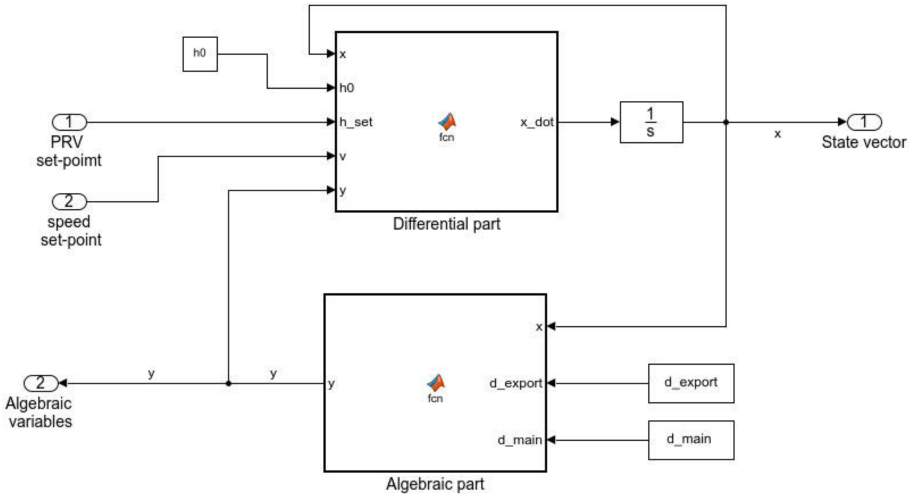


Figure 4. Simulink model of the physical WDS from Figure 1.

The Simulink model comprises the two major Matlab functions, 'Differential part' and 'Algebraic part'. Differential part calculates derivatives of the state variables, i.e., right sides of Equations 2, 3, 4 and 6. The input variables and the parameters are: the state vector $\boldsymbol{x}$**,** the head of the source h_0, the PRV set-point h_{set}, the pump speed set-point v, and the vector of the algebraic variables $\mathbf{y}$. These signals are used to calculate the mentioned derivatives, the commands inside the blocks are written in the Matlab language. The block with the symbol $\frac{1}{s}$ contains integrators which calculate the state vector elements from its derivatives. Algebraic part calculates the current value of the

CC BY NC SA

algebraic vector **y** for a given value of the state vector $\boldsymbol{x}$ and the demand data, including the export demand, d_{export} and the downstream demands, d_{main}. Note, that in order to calculate the head losses along the pipes it is necessary to solve a the system of five linear algebraic equations, Equations 9, 10 and 11.

The feedback loop to control the pump speed is shown in Figure 5.

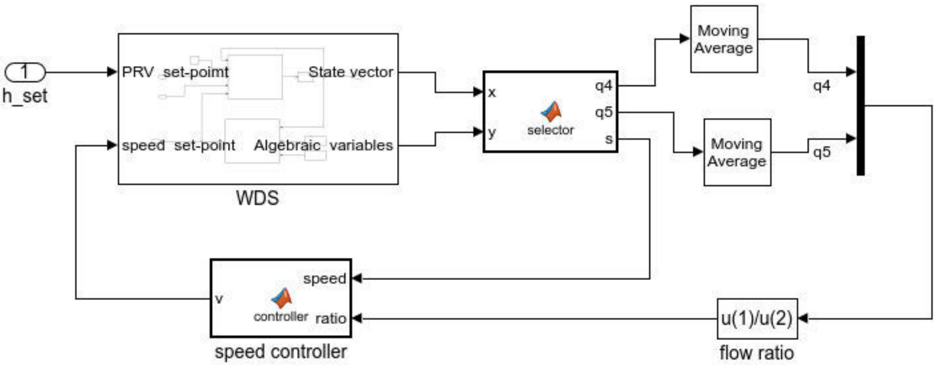


Figure 5. Simulink model of the WDS with the speed controller.

A subsystem named WDS is the masked model of the WDS depicted in Figure 4. The Selector block selects the required elements from the state vector and from the algebraic vector which are used by the speed controller. The averaged flows q_4 and q_5 are used to calculate the flow ratio which needs to be equal to 0.6, the pump speed is s another input to the controller. This part of the model mimics measurements which are to be implemented in a physical system The program which is inside the controller is shown in Table 3 in the next section.

3 RESULTS

The first task was to propose an algorithm for the speed controller depicted in Figure 5.

3.1. Selection of the pump speed controller

The two candidates were considered, a PID controller readily available as a Simulink block and a heuristic algorithm shown in Table 3. Simulation experiments showed that the PID controller required very high sampling rate which was not acceptable for practical reason and it was decided to use the heuristic controller.

Table 3. Pump speed control algorithm

```
function v = speed_rule(ratio,s)
if ratio<0.58
v=s+eps;
elseif ratio>0.62
v=s-eps;
else
v=s;
end
```

The algorithm calculates the set-point for the pump speed where value 1 corresponds to the nominal speed. If the flow ratio is smaller than 0.58 then the setpoint v is set eps above the current speed, if the ratio is greater than 0.62 then the set-point v is set eps below the current speed otherwise the set point assumes the same value as the current speed. The value of eps can be interpreted as a gain of the controller, by the trial-and-error simulations the value was selected equal to 0.004 (0.4%), greater eps values made the control system unstable whilst the smaller eps values made the control system sluggish. The next important parameter to be decided was the sampling rate i.e., how often the pump speed should be adjusted. The simulation experiments suggested that 4 minutes is a good compromise, it is acceptable for a practical reason and still provided stable and robust behaviour.

3.2. Requirement for an additional head loss (TCV)

The WDS in Figure 1 has a very low head loss between the PRV outlet h_2 and the flow mixing point h_5 ,which means that small variations in the PRV output head causes significant variations in the main flow which in turn can cause problems with maintaining the constant flow mixing ratio $\frac{q_4}{q_5}$. An additional TCV was inserted between h_2 and h_3 as illustrated in Figure 1 and simulation experiments were carried out for different had losses across the TCV. The required head at h_3 is 254 m due to the downstream system conditions, which corresponds to pressure of 28 m as the elevation at this point is 226 m. The simulation results are summarised in Table 4.

Table 4. Additional head loss requirements

Case	TCV head loss [m]	PRV setpoint head [m]	Flow ratio signal
a	<2.52	264	unstable
b	2.52	264	borderline (Figure 7a)
c	10	264	Stable and robust (Figure 7b)

The two demands in the WDS have time varying pattern, low demand at night and high demand during the day. In order to maintain approximately a constant head loss on the TCV it was necessary periodically, each 15 minutes, to adjust the valve position following the changing flow. The flow ratio signal for the case b) is shown in Figure 6a and for the case c) in Figure 6b.

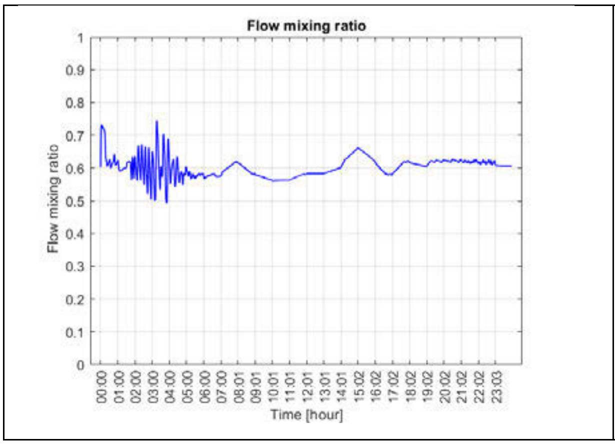


Figure 6a. Ratio signal for the low TCV head loss

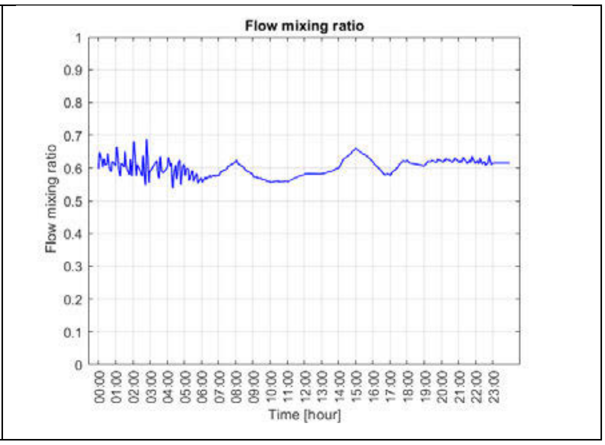


Figure 6b. Ratio signal for the high TCV head loss

The ratio signal in Figure 6a for the TCV head loss of 2.52 m shows significant oscillations at night with minimum value of 0.5 and the maximum values of 0.75. The corresponding values in Figure 6b for the TCV head loss of 10 m are 0.55 and 0.65. The ratio signals in Figure 6a and 6b look similar for the higher flows during the day. This can be explained by the fact that PRVs are less stable for small flows, i.e., small valve openings as demonstrated in [5]. When the output head from the PRV slightly oscillates the main flow oscillates as well and subsequently causes difficulty for the pump speed control system to maintain the constant flow mixing ratio.

3.3. Simulation results for the normal demand conditions

In order to understand the system behaviour, signals from different parts of the model are presented and discussed. The following conditions were assumed in the simulations:

- Main demand and Export demand follow the same pattern shown in Figure 7 with total values over the day of 7218 $^{m^3}/_{day}$ and 5000 $^{m^3}/_{day}$ respectively.
- The pump speed was controlled by the algorithm displayed in Table 3 with the sampling rate of 4 $minutes$ and $eps = 0.4\%$
- The hydraulically controlled PRV, Cla-Val DN250 operated in a continuous manner
- The TCV, Monovar DN250 valve was adjusted at 15 minutes intervals to maintain the head loss approximately of 10 m

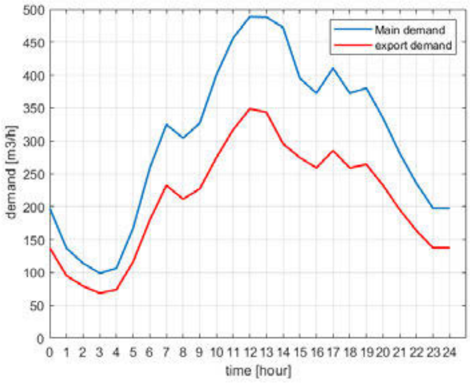


Figure 7. Main demand and Export demand

The flow mixing ratio for these conditions is presented in Figure 8. The desired average value of 0.6 is maintained well over the whole 24 hour horizon. The more erratic behaviour at the night when the flow is low can be explain by the variability of the PRV output and its impact on the main flow and in turn on the flow ratio control system.

CC BY NC SA

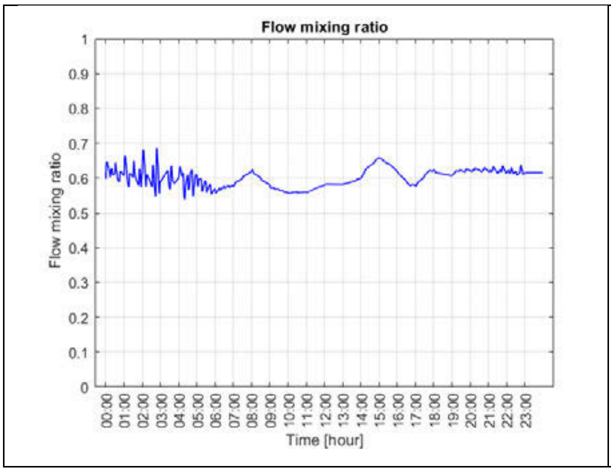


Figure 8. Flow ratio signal

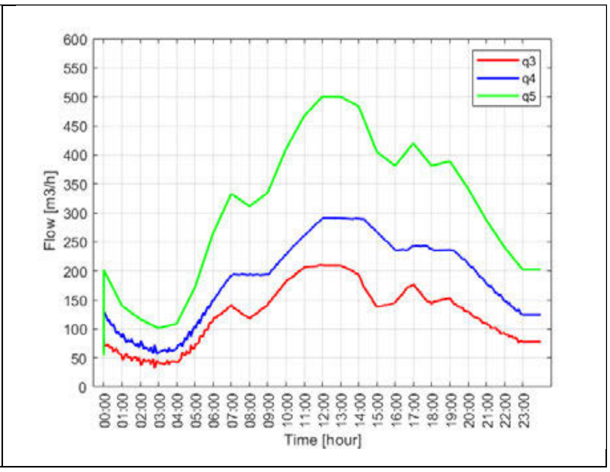


Figure 9. Flows at the mixing node

The flows at the flow mixing node h5 are featured in Figure 9, incoming flow $q3$ is in red, the pump flow $q4$ is in blue and the downstream flow $q5$ which is a sum of $q3$ and $q4$ is in green. The minimum flow through the pump $q4$ is circa 55 m^3/h which is minimum allowed for this pump, see also Figure 15.

Figure 10 illustrates the three important heads, the PRV inlet, the PRV outlet and the TCV outlet and the related Figure 11 depicts the corresponding flow through the PRV and the TCV.

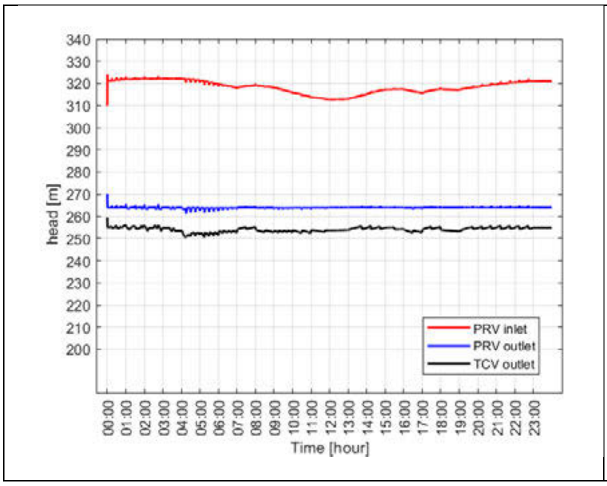


Figure 10. Heads at h1, h2 and h3 nodes

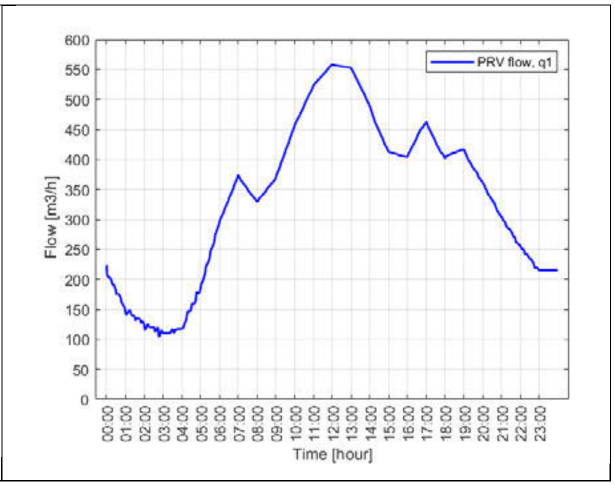


Figure 11. Flow through the PRV

The PRV outlet maintain the set-point 0f 264 m with frequent blips of $\mp 2\ m$. The TCV output is kept around 254 m , the value required to sustain the feasible pressure in the downstream system. The almost constant head loss of 10 m is sustain by adjusting the TCV position each 15 $minutes$. The PRV inlet head shown in red is equal to 322 m for the low flows and drops to 313 m for the peak flow due to the higher head loss on the long supplying pipe. The PRV flow in Figure 11 varies widely from 110 m^3/h up to 520 m^3/h caused by the variable demands. Small oscillations can be observed in the flow signal for the low flows whilst for the higher flows the signal is quite smooth.

The PRV opening and the TCV opening signals are displayed in Fig.12 and the corresponding PRV valve sensitivity coefficient is displayed in Figure 13.

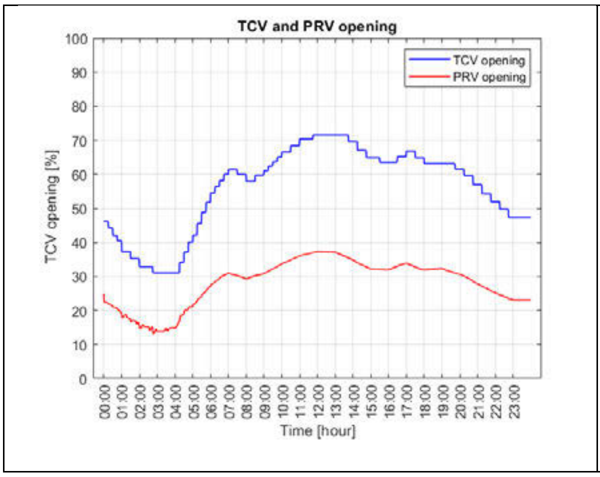


Figure 12. PRV and TCV opening signals

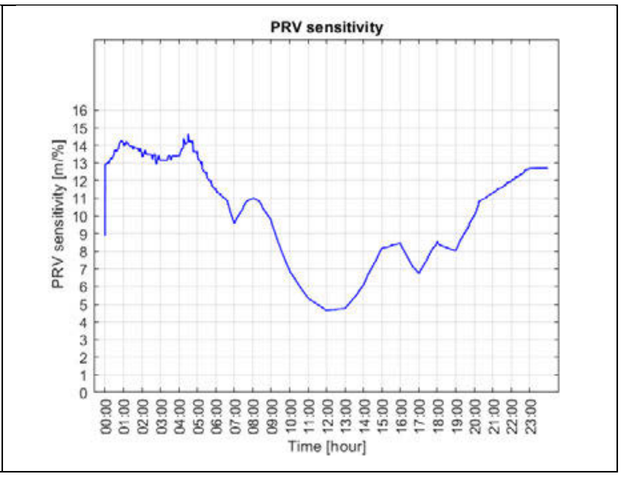


Figure 13. PRV sensitivity signal

The minimum PRV opening is around 15% which is acceptable for this type of valve. The TCV opening varies widely between 30% and 70%. It would be difficult to keep the TCV opening constant and not to adjust its position because in in such a situation the outlet head would drop well below the required value for the high flows. The concept of the PRV sensitivity was introduced in [5], it indicates the change in the PRV output head when the valve position changes by 1%. It can be interpreted as a measure of robustness of the system. The signal in Figure 13 is an upper bound of the sensitivity coefficient after ignoring influence of the rest of the WDS. Presence of the pressure dependent outflow has a dumping effect on the sensitivity, the higher the outflow the less sensitive is the entire system [5]. It can be observed that the sensitivity is low for high flows and relatively high for the low flows. This phenomenon impacts all signals in the system including the ratio signal in Figure 8, the flows in Figures 9 and 11 and also the valve outlet heads in Figure 10, some oscillations are present for the low flows period.

The pump characteristics and the operating points are shown in Fig.14 below.

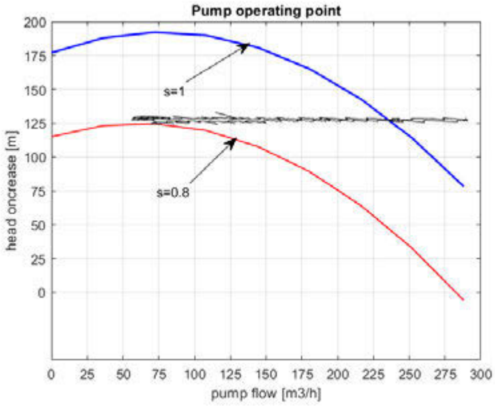


Figure 14. Operating points of the pump

The pump characteristic for the nominal speed $s = 1$ is depicted in blue and the pump characteristic for the minimum speed of $s = 0.8$ is in red. The black line represents a collection of operating points over the entire 24 hour period. The operating points align along the system curve which in this case is a straight line with a small slope characteristic for the low losses system. The pump flow which displayed in Figure 9 (blue line) changes from 55 $m^3/_h$ to 290 $m^3/_h$.

It can be seen that it is not possible to deliver flows below 55 $m^3/_h$ and maintain the desired delivered head, the operating points for the higher flows require speed above 1 which would necessitate to switch the second pump on

4 CONCLUSIONS

It was possible to design a pump speed controller which interacts in a stable manner with the hydraulically controlled PRV. The controller adjusts the pump speed at 4 minutes intervals in steps of 0.4% of the nominal speed. It was necessary to introduce an additional head loss on the main pipe between the PRV and the flow mixing node in order to provide a stable operation. The head loss of 2.52 m is a borderline between the stable and unstable behaviour. In the presented study the 10 m head loss was implemented via a TCV adjusted at 15 minutes intervals when the flow changes significantly. It was observed that the mixing ratio signal exhibited some oscillations during the low flow period. It was the result of higher PRV sensitivity at that time and subsequent interactions between the PRV and the flow ratio control system.

The study considered normal level of demands. In reality the instantaneous value of the PRV flow can vary from 40 $m^3/_h$0 up to 1500 $m^3/_h$. It was proposed to replace the 250 mm valve with two 200 mm valves working in parallel. The analysis of the two hydraulically controlled PRVs connected in parallel is presented in the associated paper, [6] in this conference.

5 REFERENCES

[1] Ulanicki, B. and Beaujean, Ph., “Modeling Dynamic Behavior of Water Distribution Systems for Control Purposes”, J. Water Resour. Plann. Manage., 2021, 147(8): 04021043, DOI: 10.1061/(ASCE)WR.1943-5452.0001403.

[2] Nault, J. D., and B. W. Karney. 2016a. “Adaptive hybrid transient formulation for simulating incompressible pipe network hydraulics.” J. Hydraul. Eng. 142 (11): 04016050. https://doi.org/10.1061/(ASCE)HY .1943-7900.0001195.

[3] Prescott, S.L. and Ulanicki, B. (2003). “Dynamic modeling of pressure reducing valves”, ASCE Journal of Hydraulic Engineering, 129(10), 804-812.

[4] Ulanicki, B., J. Kahler, and B. Coulbeck. 2008. “Modeling the efficiency and power characteristics of a pump group.” J. Water Resour. Plann. Manage. 134 (1): 88–93. https://doi.org/10.1061/(ASCE)0733-9496 (2008)134:1(88).

[5] Janus, K. T. and Ulanicki, B.2018. ‘Improving Stability of Electronically Controlled Pressure-Reducing Valves through Gain Compensation’, J. Hydraul. Eng., 2018, 144(8): 04018053, DOI: 10.1061/(ASCE)HY.1943-7900.0001498.

[6] Ulanicki, B., Beaujean, Ph. Plancke, J. and Leruth, B. 2022.’PRVs Working in Parallel Configuration – Analysis of Dynamics’, 2nd International CCWI / WDSA Joint Conference, Valencia (Spain), 18-22 July 2022 doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.*****

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14164

PRVS WORKING IN PARALLEL CONFIGURATIONS – ANALYSIS OF DYNAMICS

Ulanicki, B.[1], Beaujean, Ph.[2] Plancke, J.[3] and Leruth, B.[4]

[1] De Montfort University, Water Software Systems, Leicester, LE2 4NW, UK
[2,4] Société wallonne des eaux, Rue de la Concorde,41 4800 Verviers, Belgium
[3] SOFTEAU srl, www.softeau.be, Belgium

[1] *bul@dmu.ac.uk* , [2] *philippe.beaujean@swde.be,* [3] *juliaan.plancke@skynet.be,*
[4] *bastien.leruth@swde.be*

Abstract

Société wallonne des eaux (SWDE) is the most important water production and distribution company in Wallonia, one of the regions in Belgium. Its distribution network stretches over 40,000 km. It covers nearly 200 municipalities and has more than one million connections. SWDE supplies drinking water to nearly 2.4 million people, constituting over 70 % of the Walloon population. Recently, SWDE has launched a major multimillion-Euro project to improve security of the water supply in Wallonia. The project involves a 500/800 mm extension of a 1000-mm wide, over 100 km long pipeline as a backbone of the distribution network, starting from the Eupen and Gileppe reservoirs. The project created a number of engineering challenges and one of them is considered in this paper. The work represents a design of a pressure control station which provides efficient operation over a wide range of flows from 40 m3/h to 1500 m3/h. Additionally, the pressure control scheme needs to work in a stable manner with the downstream flow control system. It is proposed to use two identical hydraulically controlled PRVs connected in parallel without any external PLC control loop. To make sure that such a configuration can be installed in practice it was necessary to carry out extensive simulations to understand the system dynamic behavior for different scenarios.

Keywords
Pressure control, PRV parallel connection, Rigid Water Column model, Dynamics, Stability.

1 INTRODUCTION – CASESTUDY DESCRIPTION

A schematic of the system fragment with the PRV chamber with two PRVs connected in parallel is pictured in Figure 1.

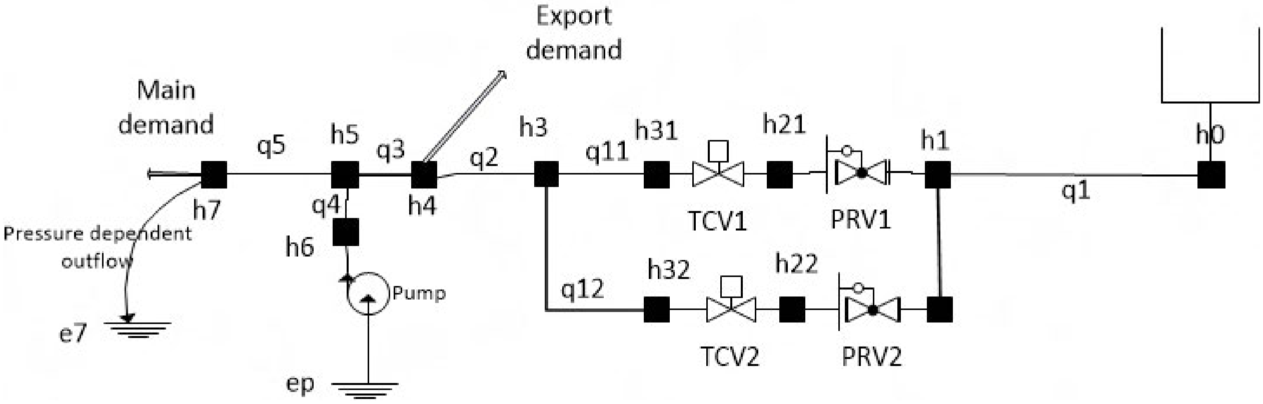


Figure 1. Schematic of the water distribution system

The source flow comes from the Eupen reservoir with capacity of 50,000 m^3. In order to reduce the required pressure in the downstream part of the water system a PRV station with the two valves connected in parallel with the outlet head of 254 m (pressure 28 m) is planned to be installed. There is a significant water demand at node h4 where the water is exported to another water company. At node h5 the Eupen water is mixed with the water imported from another water company through a pumping station. It is necessary to maintain a constant ratio of 0.6 between the $q_4$and q_5 flows due to water quality requirements, and at same time to maintain feasible positive pressure at node h7 in order to maintain the required pressure at a critical point downstream of the system not shown in Figure 1. The design of the flow ratio control system was considered in an associated paper [2] presented also at this conference. An important conclusion was that it was necessary to introduce additional head losses in order to prevent strong interactions between the flow control system and the PRVs. These additional head losses are represented by TCV1 and TCV2 in Figure 1. The focus of this paper is on dynamics of the two hydraulically controlled PRVs connected in parallel for different setting scenarios summarized in Table 1. If there is a need to connect two PRVs in parallel some manufacturers, for instance Bermad, [1] recommends setting the main valve 1m-2m above the second valve. In this way the main valve operates all the time and the second valve with the lower settings becomes active only for peak flows when the main valve is fully open (saturated). However, in this study because of the presence on the TCVs the situation is more involving and all four scenarios given in Table 1 for different PRV and TCV settings are investigated.

Table 1. Valve setting scenarios

Id	PRV1 head setpoint [m]	PRV2 head setpoint [m]	TCV1 head loss [m]	TCV2 head loss2 [m]
a	264	264	10	10
b	264	262	10	8
c	264	260	10	6
d	264	260	10	10

The description of the scenarios is provided below:

a) The two PRVs has the same set-point of 264 m and the two TCVs are adjusted to have the same head loss of 10m, the situation is symmetrical for the two PRV branches.

b) PRV2 has a lower set-point of 262 m and TCV2 is adjusted to have 8 m head loss to make sure that the target head value at the connection node h3 is consistent for the two branches and equal to 254 m.

c) PRV2 has a lower set-point of 260 m and TCV2 is adjusted to have 6 m head loss to make sure that the target head at the connection node h3 is consistent for the two branches and equal to 254 m.

d) PRV2 has a lower set-point of 260 m and TCV2 is adjusted to have 10 m head loss in order to create a conflict between the two PRV branches. The target head at h3 is still 254 m for Branch 1 but it is 250 m for Branch 2.

The TCV settings are corrected at 15 minutes intervals when the flow and subsequently head loss changed significantly. The PRVs operate continuously controlled by hydraulic pilot loops.

BY NC SA

Pipes q1, q2, q3 and q5 in Figure 1 are transport pipes connecting important sites along the main pipeline and their data are displayed in Table 2. Remaining pipes, q11, q12 and q4 are short connection pipes in the PRV chamber and the pump chamber respectively.

Table 2. Pipe data of main pipes

Pipe id	Length [m]	Diameter [mm]	D-W friction factor
q1	46212	800	0.0175
q2	9749	800	0.015
q3	10	800	0.015
q5	21179	800	0.015

The node data are included in Table 3.

Table 3. Node data]

Node id	Elevation [m]	Demand [m^3/d]	Comment
h0	322.5	-	Fixed Head Source
h1, h21, h31, h31, h32, h3	226	-	Valve connection nodes
h4	129.5	5000	Demand pattern is in Figure 2
h5	129.5	-	Connection node
h6	120	-	Destination node of the pump station
h7	180	7218	Main demand+ outflow area =10^{-7} m^2

The node h0 represents a fixed head source and the node h7 is a boundary node representing the whole downstream part of the system with Main demand equal to $7218\, {m^3}/{d}$ and an additional pressure dependent outflow with the equivalent area of $10^{-7} m^2$.

The selected PRVs for this study are 200mm Cla-Val valves with the « Low Flow system » option. The PRV capacity characteristic in original units, i.e.. flow expressed in l/s and pressure in bar is displayed in Fig.2.

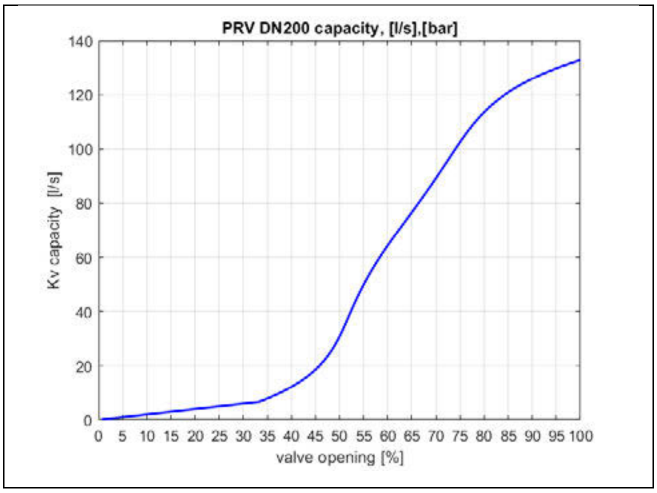


Figure 2. PRV DN200 capacity characteristic in original units

It was suggested to try the same 200mm Cla-Val valves but without control pilot loops as the TCVs. The characteristic was approximated by a piece-linear function in the Simulink model to preserve accuracy of the original data from the catalogue.

The pump model was derived from the manufacturer data for the KSB pump, type W12-200/T14 and is displayed in Figure 3.

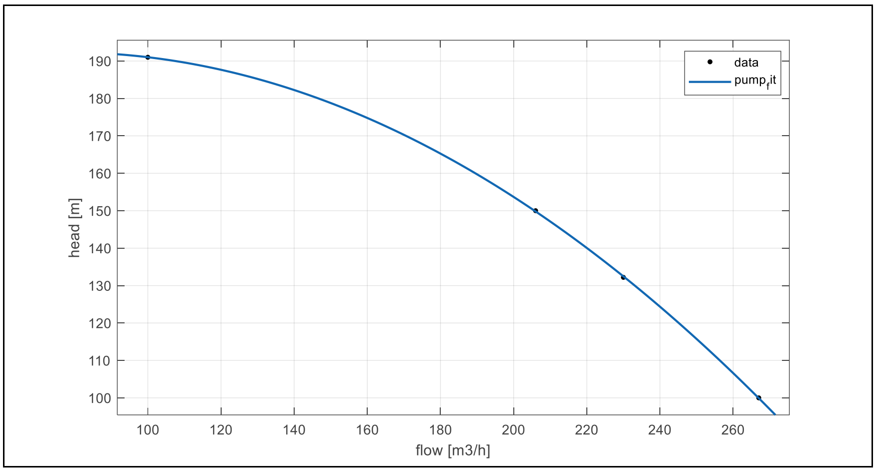


Figure 3. KSB pump characteristic

Analytically the model was represented by a second order polynomial given in Equation 1,

$$\Delta h = -0.0026 * q^2 + 0.3974 * q * s + 177s^2 \tag{1}$$

where q is the pump flow in $[{m^3}/{h}]$, Δh is the head increase across the pump in $[m]$ and finally s the relative pump speed, where value 1 corresponds to the nominal speed.

The overall objective was to investigate dynamics of the PRVs connected in parallel for different setting scenarios given in Table 1 over the range of flows from 300 ${m^3}/{h}$ up to 1450 ${m^3}/{h}$., i.e., which represents the high flow situation

The reminder of the paper is organised as follows. The Methodology section describes an approach adopted in this study. The simulation results for the high flow conditions which represents the most interesting case are described in the Results section. The paper finishes with conclusions from the modelling study.

2 METHODOLOGY

The methodology is based on the theory provided in [3]. A mathematical model is formulated in the form of differential-algebraic equations (DAEs) and the model is simulated using the Matlab/Simulink environment.

The rigid water column (RWC) model given in Equation 2 is used for pipes,

$$\dot{q} = -Rq|q| + M\Delta h \tag{2}$$

where $q =$ pipe flow; $\Delta h =$head loss along the pipe; $R = f_D(1/_{2DA})$; $M = {}^{gA}/_{L}$; $f_D =$Darcy friction factor; $D =$pipe hydraulic diameter; $A =$pipe cross section area; $L =$pipe length; and $g =$ gravitational acceleration.

The RWC model can be used for analysis of slow transients, [4] or for analysis of control problems in a WDS, [3].

The PRVs are represented by a behavioural model from [5] and is given in Equations 3 and 4 (differential part) and in Equation 5 (algebraic part),

$$\dot{x}_m = \alpha_{open}(h_{set} - h_d) \tag{3}$$

$$\dot{x}_m = \alpha_{close}(h_{set} - h_d) \tag{4}$$

where x_m =valve opening in percent; and α_{open} and α_{close}=rates of the valve opening and closing, respectively; h_{set} =PRV setpoint and h_d = PRV outlet head. The algebraic part is given by a standard valve equation, Equation 5,

$$q = K_v(x_m)\sqrt{\Delta h} \tag{5}$$

where $K_v(x_m)$=valve capacity, which depends on the valve opening x_m, the $K(x_m)$ is a characteristic provided by the manufacturer and is shown in Figur2.

The TCV is a static element represented only by an algebraic equation of the same form as Equation 5 and with the capacity characteristic displayed in Figure 2.

The pump is represented by the two equations. The first equation, Equation 6

$$\dot{s} = -\frac{1}{T}s + \frac{1}{T}v \tag{6}$$

is a differential equation that describes the pump inertia, where s =pump speed, v = speed set-point, and T = time constant. The second is an algebraic equation that describes the head increase along the pump scaled by the pump speed s, [6],

$$\Delta hh_o = Aq^2 + Bqs + Cs^2 \tag{7}$$

where the specific values for the coefficients A, B, C are given in Equation 1.

According to theory presented in [3] it is necessary to select pipe independent flows, in this case they are q_1, q_{11} and q_4 while the dependent flows can be evaluated from the mass balance at the connection nodes as follow:

$$\begin{gathered} q_{11}+q_{12} = q_1 \\ q_2 = q_{11}+q_{12} \\ q_3 = q_2 - d_{export} \\ q_5 = q_3 + q_4 \end{gathered} \tag{8}$$

$$q_{out} = q_5 - d_{main}$$

where d_{export} is the water export to another water company at node h4 in Figure 1, $q_{out} = c_{out}(h_7 - e_7)^{0.5}$ represents pressure dependent outflow in the downstream system and d_{main} is the main flow of the downstream system, c_{out} is the outflow coefficient and e_7 is the elevation of the h_7 node.

The independent pipe flows, q_1, q_{11} and q_4 are explicitly represented by the differential Equation 2. To make sure that the dependent flows in other pipes, i.e., $q_{12}, q_2, q_3,\ q_5$ obey the RWC model it is necessary, according to [3], to introduce the 'differential mass balance' equations for the connection nodes with pipes as follow,

$$\begin{aligned} \dot{q}_{11} + \dot{q}_{12} &= \dot{q}_1 \\ \dot{q}_2 &= \dot{q}_{11} + \dot{q}_{12} \\ \dot{q}_3 &= \dot{q}_2 - \dot{d}_{export} \\ \dot{q}_5 &= \dot{q}_3 + \dot{q}_4 \end{aligned}$$

which are subsequently are transformed into Equations 9 with the help of the relationship from Equation 2.

$$\begin{gathered} -R_{11}q_{11}|q_{11}| + M_{11}\Delta h_{11} - R_{12}q_{12}|q_{12}| + M_{12}\Delta h_{12} = -R_1q_1|q_1| + M_1\Delta h_1 \\ -R_2q_2|q_2| + M_2\Delta h_2 = -R_{11}q_{11}|q_{11}| + M_{11}\Delta h_{11} - R_{12}q_{12}|q_{12}| + M_{12}\Delta h_{12} \\ -R_3q_3|q_3| + M_3\Delta h_3 = -R_2q_2|q_2| + M_2\Delta h_2 - \dot{d}_{export} \\ -R_5q_5|q_5| + M_5\Delta h_5 = -R_3q_3|q_3| + M_3\Delta h_3 - -R_4q_4|q_4| + M_4\Delta h_4 \end{gathered} \tag{9}$$

The mass balance equations need to be complemented by the energy balance equations for one proper loop and the two pseudo-loops. The proper loop involves all components in the PRV chamber as depicted in Equation 10.

$$\Delta h_{11} + \Delta h_{TCV1} + \Delta h_{PRV1} - \Delta h_{PRV2} - \Delta h_{TCV1} - \Delta h_{12} = 0 \tag{10}$$

The first pseudo-loop is between the source h_0 to the ground level of the pump branch e_p and the second between the source and the ground level of the outflow branch e_7. Starting from the pump branch and following the clockwise direction yields,

$$e_p + \Delta h_{pump} - \Delta h_4 + \Delta h_3 + \Delta h_2 + \Delta h_{11} + \Delta h_{TCV1} + \Delta h_{PRV1} - h_0 = 0 \tag{11}$$

and then starting from the outflow branch provides the third energy balance equation,

$$e_7 + \Delta h_{leak} + \Delta h_5 + \Delta h_3 + \Delta h_2 + \Delta h_{11} + \Delta h_{TCV1} + \Delta h_{PRV1} - h_0 = 0 \tag{12}$$

The differential part of the DAE model is represented by Equation 2 for the independent flows q_1, q_{11} and q_4, Equations 3 and 4 for the valve openings, x_{m1} and x_{m2} and Equation 6 for the pump speed, s; altogether six differential equations.

Consequently, the state vector of the model $\boldsymbol{x}$ is made of the six variables listed in Equation (13).

$$\boldsymbol{x} = [q_1 \quad q_{11} \quad q_4 \quad x_{m1} \quad x_{m2} \quad s]^T \tag{13}$$

The algebraic part of the model is represented by Equations 5, 7, 8, 9, 10, 11 and 12 and the vector of the algebraic variables $\boldsymbol{y}$ is composed of the dependent flows and head losses/gains on all components. Altogether seventeen variables displayed in Equation 14.

$$\begin{gathered} \boldsymbol{y} = [q_{12} \quad q_2 \quad q_3 \quad q_5 \quad \Delta h_{PRV1} \quad \Delta h_{TCV1} \quad \Delta h_{PRV2} \quad \Delta h_{TCV2} \quad \Delta h_{pump} \quad \Delta h_{leak} \quad \Delta h_1 \quad \dots \\ \Delta h_{11} \quad \Delta h_{12} \quad \Delta h_2 \, \Delta h_3 \quad \Delta h_5 \quad \Delta h_6]^T. \end{gathered} \tag{14}$$

The considered DAE system has index 1as indicated in [3], which means that for a given value of the vector $\boldsymbol{x}$, the vector $\mathbf{y}$ can be evaluated in a unique way by solving equations of the algebraic part of the model. In this case, the solution of the algebraic part can progress in the following steps: 1) calculate the dependent flows from Equations 8; 2) calculate the head losses/gains on the control elements from Equations 5 and 7; 3) calculate the head losses on the pipes $\Delta h_1, \Delta h_{11}$, Δh_{12}, $\Delta h_2, \Delta h_3, \Delta h_4, \Delta h_5$ from the system of linear simultaneous equations composed of Equations 9, 10, 11 and 12.

The Simulink model is depicted in Figure 4.

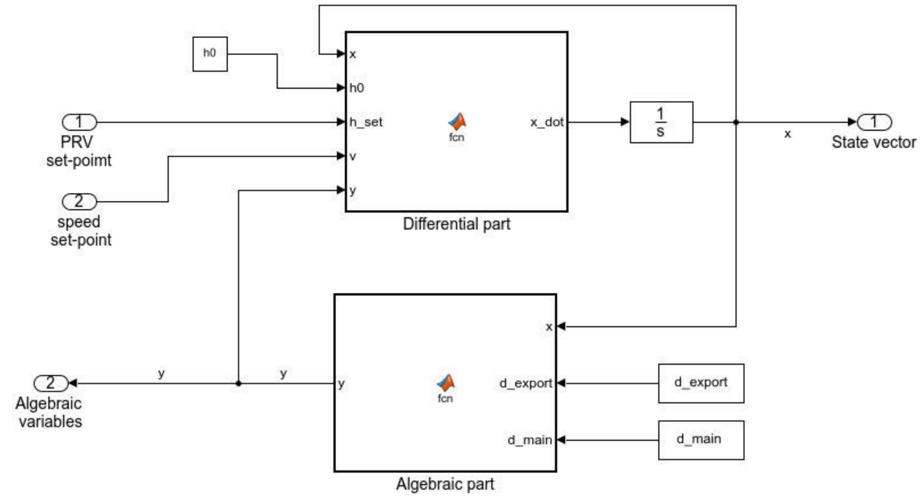


Figure 4. Simulink model of the physical WDS from Figure 1.

The Simulink model comprises the two major Matlab functions, 'Differential part' and 'Algebraic part'. Differential part calculates derivatives of the state variables, i.e., right sides of Equations 2,3, 4and 6. The input variables and the input parameters to this block are: the state vector $\boldsymbol{x}$**;** the head of the source h_0;the vector of the PRV set-points $\boldsymbol{hset}$; the pump speed set-point v ; and the vector of algebraic variables $\mathbf{y}$. These signals are used to calculate the mentioned derivatives, the commands inside the block are written in the Matlab language. The block with symbol $\frac{1}{s}$ contains integrators which calculate the state vector elements from its derivatives. Algebraic part calculates the current value of algebraic vector $\mathbf{y}$ for given values of the state vector $\boldsymbol{x}$ and the demand data, including the Export demand d_{export} and the downstream demands d_{main}.Note, that in order to calculate the head losses along the pipes it is necessary to sole the system of seven linear algebraic equations, Equations 9, 10, 11 and 12.

3 RESULTS

The simulation results are presented for the high flow conditions. The high demand conditions are more interesting than normal or low flow conditions as they show how the two valves interact during the whole day when the flow vary widely from 300 $m^3/_h$ up to its maximum value of 1450 $m^3/_h$. In the low flow conditions, for instance it is sufficient to use one PRV branch over the entire day. There are limits imposed on the PRV opening, the maximum opening is 89% (recommended by the manufacturer) and the minimum opening is 1%, if these limits are violated during simulations the PRVs act like fixed valves with the constant opening of 89% or 1%, respectively. The lower limit of 1% was assumed for the modelling purposes, in the physical system the valve would be completely closed.

The simulations were carried out for all four scenarios defined in Table 1. Scenario a) is symmetrical for the two PRV branches with identical settings for PRV1 and PRV2 and also for TCV1 and TCV2. Scenarios b) and c) are not symmetrical with respect to the PRV set-points and the TCV head losses but are consistent with respect to the target head at the node h_3 of $254m$

which is the same for the both branches. A PRV with a higher set-point, in this case PRV1 will be called a leader and the PRV with a lower set-point, i.e., PRV2, a follower. Scenario d) is not symmetrical and is not consistent with respect to the target head at h_3, Branch 1 tries to impose the head of $254m$ whilst Branch 2, $250m$., in such a conflict situation the two PRVs cannot be active at the same time. The behaviour of the two branches is investigated in details in this section.

The high flow conditions were mimicked by multiplying the normal profile of Export demand by factor of 3.45. Main demand and the modified Export demand are depicted in Figure 5.

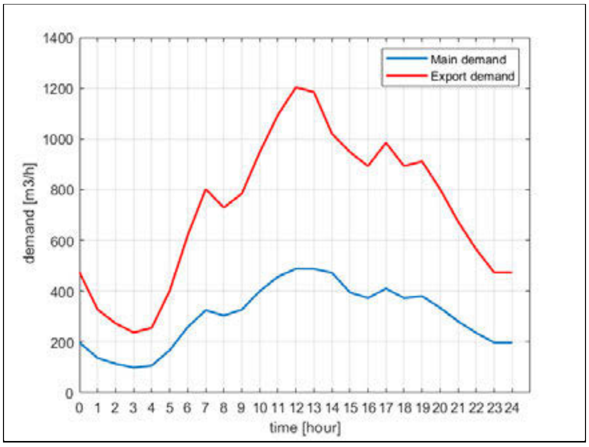
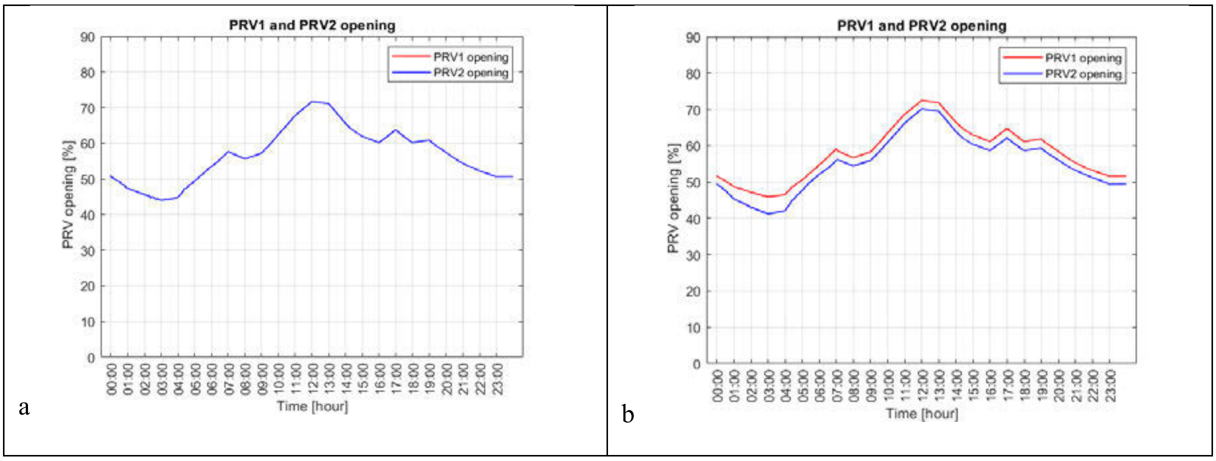


Figure 5. Main demand and Export demand

The value of the demand factor has been adjusted by trial and error to achieve the maximum q_1 flow of 1450 $^{m^3}/_{h}$ see Figure 8. The valve opening signals for different scenarios are displayed in Figure 6.



BY NC SA

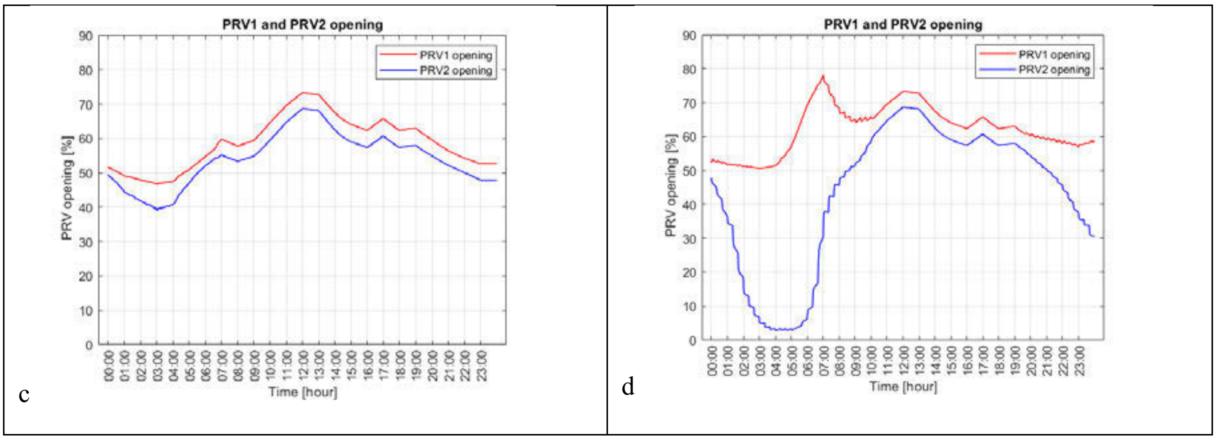


Figure 6a, 6b, 6c, 6d. PRV1 and PRV2 opening signals for different scenarios

The PRV opening signals are consistent with engineering intuition, the behaviour of the valves for each scenario are summarised in the bullet points below:

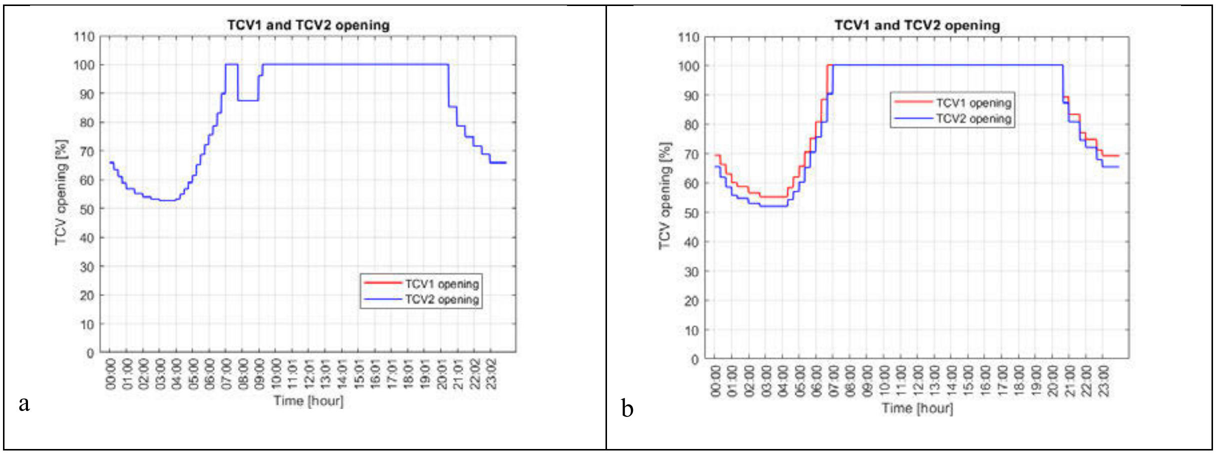
a) The opening signals from the two PRVs are overlapping due to the symmetry and identical initial conditions assumed in the both valves.

b) The symmetry has been broken with PRV2 (follower) having lower set-point of $hset(2) = 262m$. The PRV2 opening is slightly smaller than for PRV1 (leader) however both branches aim the same head of 254 m at the connection node h_3.

c) Ditto but effects are even stronger than in point b).

d) PRV2 closes at night in the enforced conflict situation between the PRV branches, PRV1 has enough capacity at that time to convey the entire flow. However, during the day when the flow increases and TCV1 fully opens, PRV2 opens gradually and then during the peak hours the two PRVs convey almost equal flow.

The corresponding TCV opening signals are presented in Figure 7.



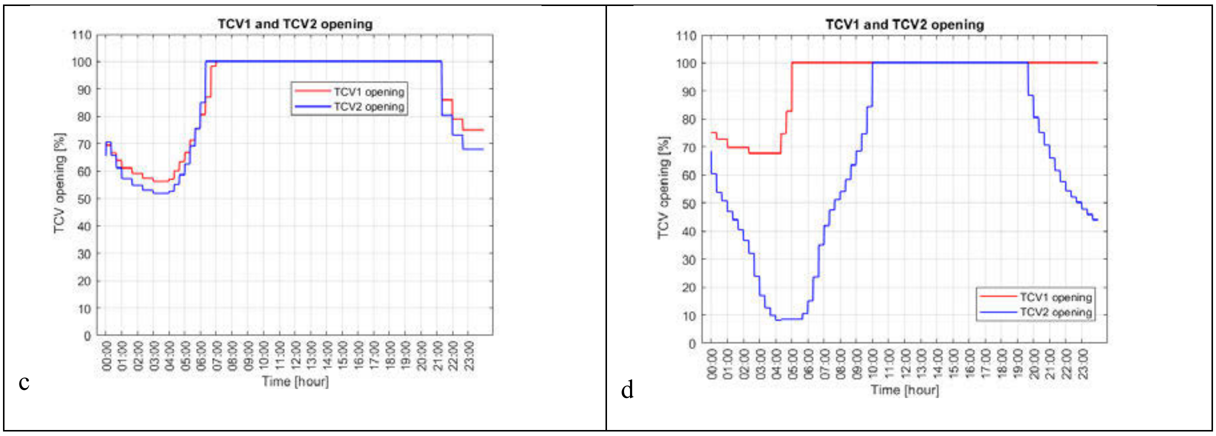


Figure 7a, 7b, 7c, 7d. TCV1 and TCV2 opening signals for different scenarios

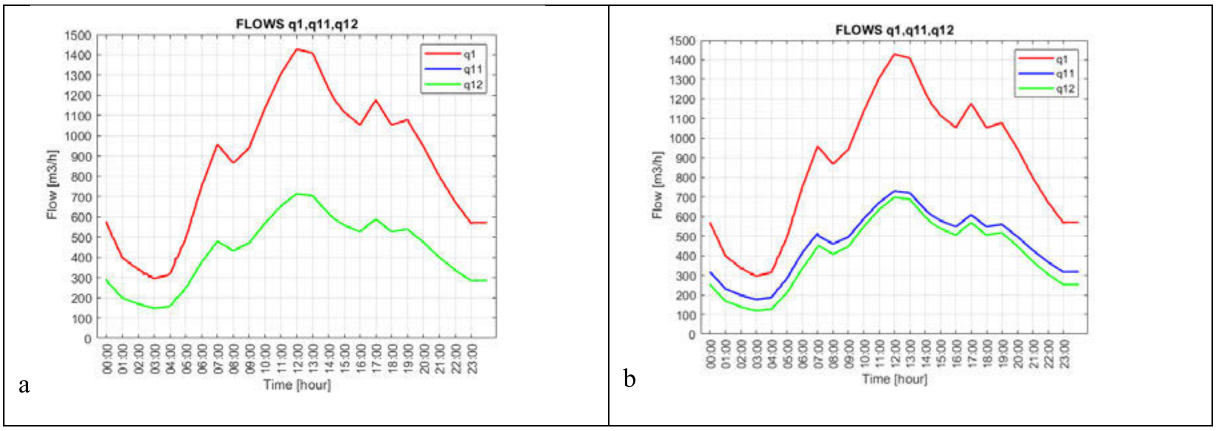
The behaviour of the valves for each scenario are summarised in the bullet points below.

a) b) c) The TCV s are fully open (100%) over the high flow period but even so the head loss on the valves is higher than the desired 10m and their outputs drop to 240m as it is shown in Figures 9a, 9b and 9c.

d) In the conflict situation TCV2 follows the same pattern as PRV2. It closes at night and then opens gradually during the day. TCV1 fully opens at 06.00am and stays fully open (100%) over the rest of the day.

Clearly, the TCVs are undersized and even fully opened generate a head loss higher than 10 m for peak flows. These valves are adapted for low flows and here they should be replaced by valves with different capacity characteristic and possibly bigger diameter.

The main Eupen flow and the two PRV branch flows are shown in Figure 8.



BY NC SA

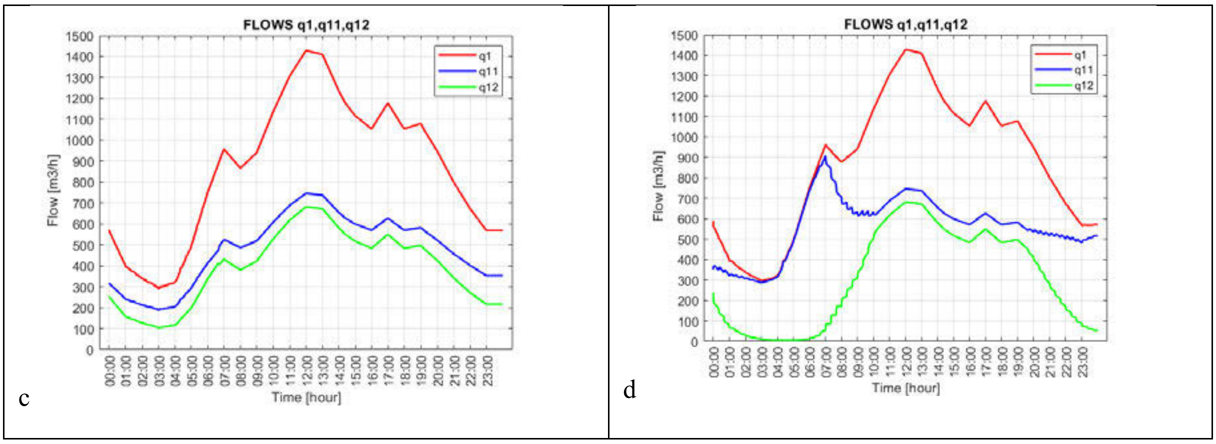


Figure 8a,8b,8c,8d. Flows q1, q11 and q12 for different scenarios

The Eupen (q_1) flow is in the red line, PRV1 (q_{11}) flow is in the blue line and the PRV2 (q_{12}) flow is in the green line. The behaviour of the flows for each scenario are described below

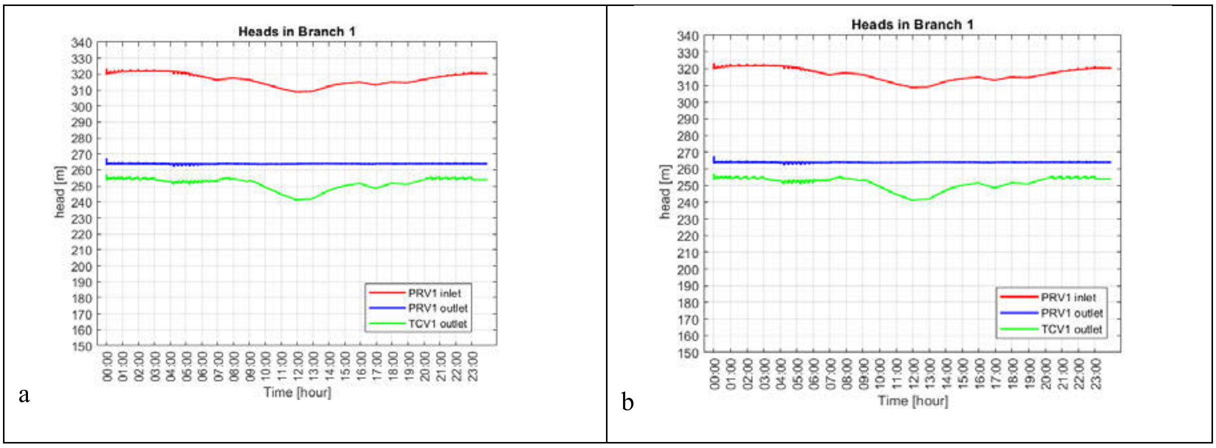
a) The flows through both valves are identical due to symmetry of the arrangement.

b) The flow through PRV1 (leader) has slightly increased compared to a) and the flow through PRV2 (follower) has slightly decreased. This difference can be explained by the flow selecting a path with a lower resistance.

c) The phenomenon is more visible in Figure 8c. for $hset(2) = 260m$, the flow in PRV1 increased even more and subsequently the PRV2 flow has decreased.

d) For lower flows at night PRV1 is open and PRV2 is closed, PRV1 is sufficient to convey the required flow. However, when the flow increases to a significant value during the peak hours PRV2 opens (see Figure 6) and the main flow is distributed evenly between the both valves. This happens when TCV1 saturates at 100% opening (see Figure 7) and cannot be adjusted any more. Branch 1 cannot enforce the required head of 254m at the TCV1 output.

Figure 9. depicts the three important heads in Branch 1, the PRV1 input, the PRV1 output and the TCV1 output



cc BY NC SA

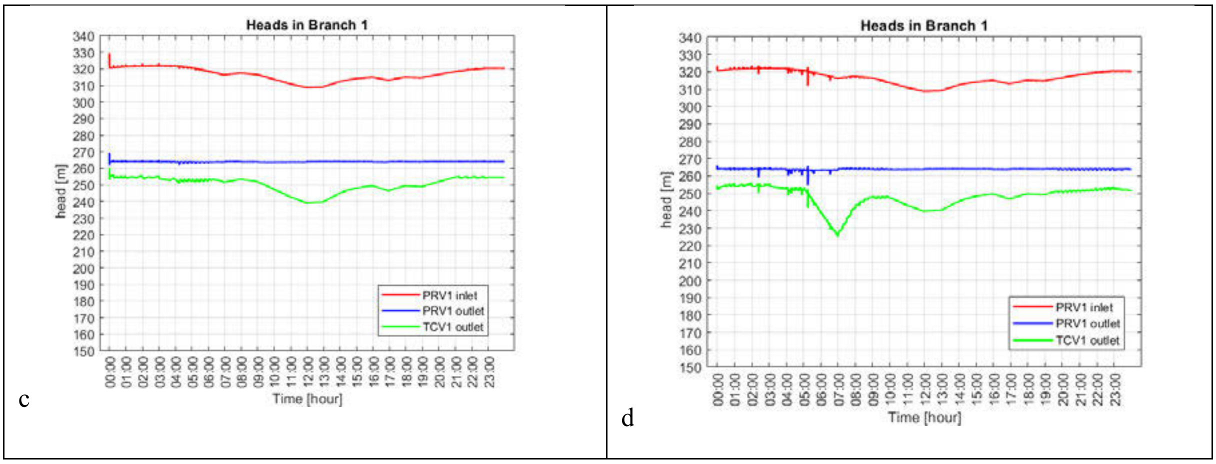


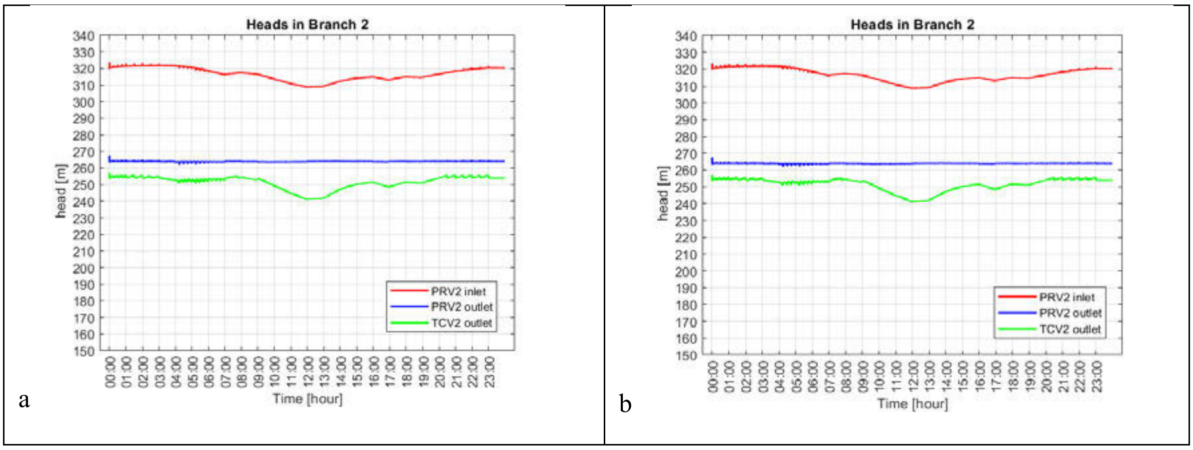
Figure 9a, 9b, 9c, 9d. Head signals in Branch 1 for different scenarios

The PRV1 inlet head is represented by the red line, the PRV1 outlet head by the blue line and the TCV1 output by the green line. The following can be observed for each scenario.

a) b) c) Independently of the PRV2 settings PRV1 maintains constant head of 264m following its set-point. The TCV1 output drops to 240m due to very high flow which gives the head loss higher than the desired 10m even for the fully open TCV valve.

d) PRV1 maintains the 264m set-point however the output from TCV1 drops to 225m at 06.00am during the transition period when PRV2 is slowly opening. PRV1 conveys the substantial flow of 950 m^3/h which causes big head loss on TCV1 despite the valve being fully open. TCV1 acts as a fixed valve and consequently the conflict situation ceases to exist.

The head signals from Branch 2 are given in Figure 10.



cc BY NC SA

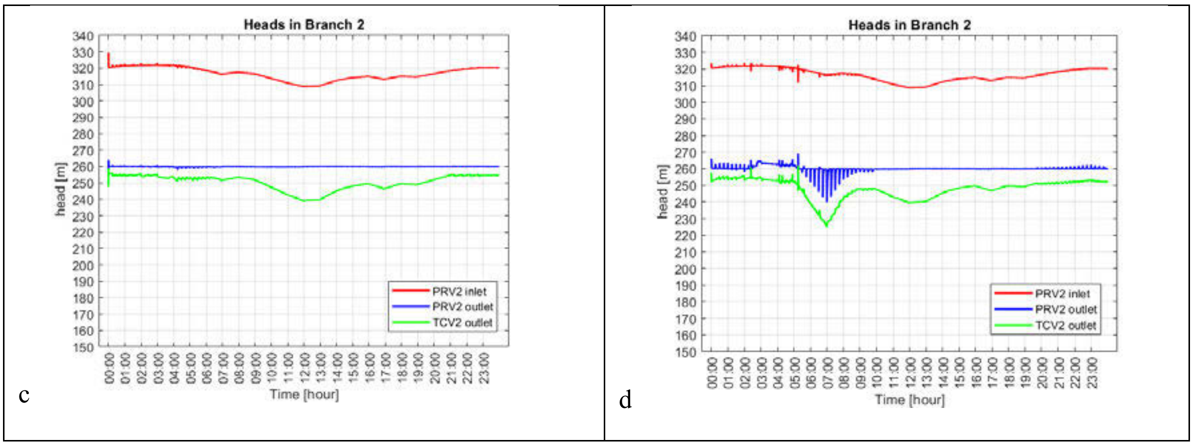


Figure 10a, 10b, 10c, 10d. Head signals in Branch 2 for different scenarios

The figure presents the head signals in the PRV2 branch, PRV2 input (red line), PRV2 output head (blue line) and the TCV2 output (green line). The observations for each scenario are as follow:

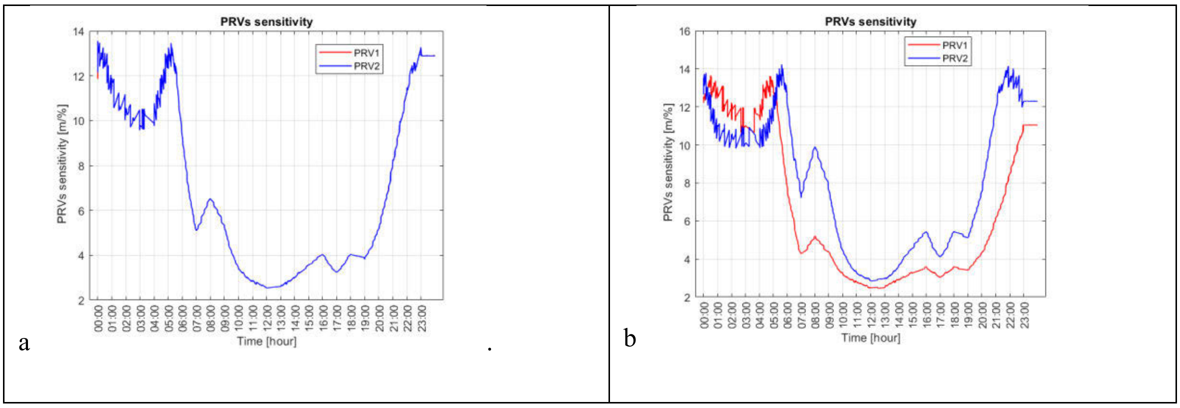
a) PRV2 maintains its 264 m set-point, the TCV2 output drops to 240 m despite the valve being fully open for peak hours

b) PRV2 maintains its 262 m set-point, the TCV2 output drops to 240 m despite the valve being fully open for peak hours

c) PRV2 maintains its 260 m set-point, the TCV2 output drops to 240 m despite the valve being fully open for peak hours

d) PRV2 is closed at night and does not control its output. It starts opening at 06.00am and during this transition period correcting the TCV2 settings result in significant jumps in the PRV2 output. The TCV2 output is the same a TCV1 output and drops to 225m at the peak hours. After 10 am when the both TCVs are fully open the TCVs output is similar to scenarios a), b) and c).

Finally, the sensitivity signals for the PRVs for different scenarios are displayed in Figure 11.



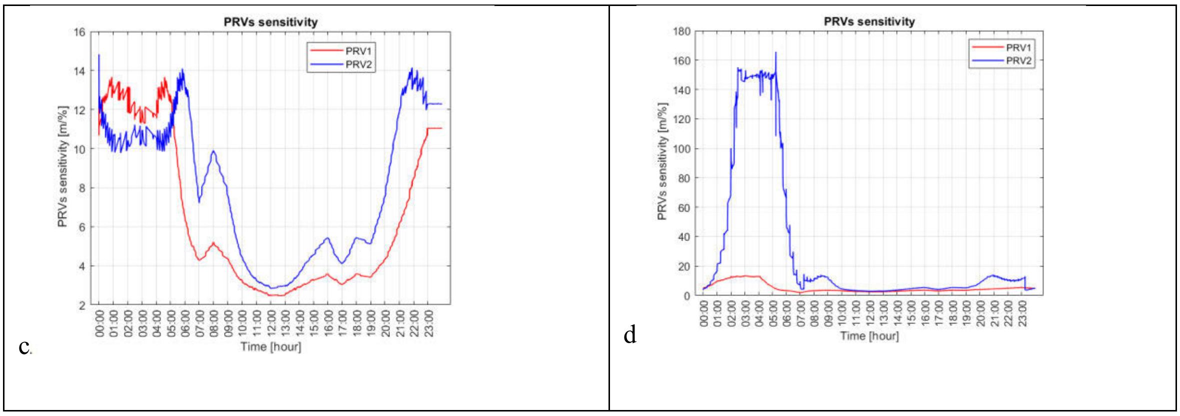


Figure 11a, 11b, 11c, 11d. PRV1 and PRV2 sensitivity signals for different scenarios

The sensitivity signal is an indicator of how stable is operation of the PRVs. The following observations can be made for each scenario.

a) b) c) The sensitivity is higher for the low flows, up to 14 ${}^{m}/_{\%}$ and lower for the peak flows, down to 2 ${}^{m}/_{\%}$. These are upper bond estimates and the real values are lower due to damping effects of the pressure dependent outflow in the WDS as it is explained in [7].

d) The PRV1 sensitivity is in normal range. The PRV2 sensitivity jumps to 120 ${}^{m}/_{\%}$ but it is practically closed at night and the notion of the sensitivity does not then applies. During the day the both PRVs have acceptable level of sensitivity.

The sensitivity values for lower the flows are not only higher but also noisier compared to the period with higher flows.

4 CONCLUSIONS

A need for parallel connection of two PRVs arises when the range of flows is very wide and one PRV cannot cope with the range. Normally it is recommended setting the main valve to higher set-point which results in the main valve being active most of the time and the second valve opens only for peak flow when the main valve is already saturated. In this case-study an additional feature was the presence of the TCVs. The PRVs operated continuously controlled by hydraulic pilot loops. The TCVs were adjusted periodically at 15 minutes intervals to maintain constant head loss following the changing flows. The TCVs were introduced to reduce the interactions between the PRVs and the downstream flow ratio control system. In such a situation behaviour of the entire system depends on both the PRV settings and the TCV settings. The detailed observations for scenarios a), b) c) and separately for scenario d) are summarised in the two paragraphs below.

Observations for scenarios a), b) and c) PRV2 settings affects flow distributions among PRVs. If the two set-points are equal $hset(1) = hset(2) = 264m$ there is a symmetrical situation and the valve openings and the valve flows are identical. For $hset(2) = 262m$ and $hset(2) = 260m$ the PRV1 opening is bigger than the PRV2 opening and also flow through PRV1 is higher than through PRV2. The both PRVs maintain their respective set-points well, $hset(1) = 264m$ and $hset(2) = 262m, 262m \; and \; 260m$ respectively. The both TCVs fully open for the high flows but even so the head loss on the valves is much higher than expected 10 m. At node h3 where the both PRV branches merge the head varies from 240 m to 255 m. The outlet head from TCVs of 240 m is well below the required value of 254 m. The two PRVs for all PRV2 settings work in a stable manner because the target head at h3 for the two branches is the same and equal to 254 m. The downstream part of the system after the pump connection is affected by the low head of 240 m at node h3 nevertheless, the flow mixing ratio is relatively stable.

BY NC SA

Observations for Scenario d). This scenario corresponds to the enforced conflict situation where Branch1 tries to enforce head 254 m at the node h3 whilst Branch 2 tries to enforce the head equal to 250 m. In this situation Branch 2 with the lower PRV2 settings of 260 m closes and water is conveyed only by PRV1 at night when the main flow is low. However, during the day around 06.00am when the main flow increases and saturates TCV1, PRV2 regains control and starts to open. The main flow is now split between the two valves. The flow in PRV2 gradually increases until the both flows are almost identical and both TCVs are fully open. The lowest head at node h_3 is at 07.00am and is equal to 225 m. This impacts the pump operating point down to 100m head increase nevertheless, the flow mixing ratio is maintained at the required level in relatively stable fashion.

In scenarios a), b), c) in which the target head at node h3 is the same, both PRV branches are active all the time. In scenario a) the two PRV flows were identical while in scenarios b) and c) the PRV with the higher setting conveyed slightly higher flow. If the intention is to use mainly one PRV branch and open another only for the peak flow then scenario d) is recommended. For normal and low flow situations not presented in this case-study the valve with lower setting closes immediately and remained closed over the entire 24 hour period.

5 REFERENCES

[1] Bermad, 2022, ‘How to design a pressure reducing station with valves in parallel’, https://www.bermad.com.au/training-resources/how-to/design-pressure-reducing-station-valves-parallel/ accessed on 06/05/2022/

[2] Ulanicki, B., Beaujean, Ph. Plancke, J. and Leruth, B. 2022.’Flow Ratio and Pressure Control in a Low Losses WDS – Analysis of Dynamics, 2nd International CCWI / WDSA Joint Conference, Valencia (Spain), 18-22 July 2022 doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.*****

[3] Ulanicki, B. and Beaujean, Ph., “Modeling Dynamic Behavior of Water Distribution Systems for Control Purposes”, J. Water Resour. Plann. Manage., 2021, 147(8): 04021043, DOI: 10.1061/(ASCE)WR.1943-5452.0001403.

[4] Nault, J. D., and B. W. Karney. 2016a. “Adaptive hybrid transient formulation for simulating incompressible pipe network hydraulics.” J. Hydraul. Eng. 142 (11): 04016050. https://doi.org/10.1061/(ASCE)HY .1943-7900.0001195.

[5] Prescott, S.L. and Ulanicki, B. (2003). “Dynamic modeling of pressure reducing valves”, ASCE Journal of Hydraulic Engineering, 129(10), 804-812.

[6] Ulanicki, B., J. Kahler, and B. Coulbeck. 2008. “Modeling the efficiency and power characteristics of a pump group.” J. Water Resour. Plann. Manage. 134 (1): 88–93. https://doi.org/10.1061/(ASCE)0733-9496 (2008)134:1(88).

[7] Janus, K. T. and Ulanicki, B.2018. ‘Improving Stability of Electronically Controlled Pressure-Reducing Valves through Gain Compensation’, J. Hydraul. Eng., 2018, 144(8): 04018053, DOI: 10.1061/(ASCE)HY.1943-7900.0001498.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14168

HYDRAULIC STATE ESTIMATION: PILOT IMPLEMENTATION IN A WATER DISTRIBUTION SYSTEM

Emilio Ruiz[1], Sarai Díaz[2] and Javier González[3]

[1]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain)
[2]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain)
[3]Department Of Civil Engineering, University of Castilla-La Mancha, Ciudad Real (Spain) &
Hidralab Ingeniería y Desarrollos, S.L., Spin-off UCLM, Hydraulics Laboratory, University of Castilla-La Mancha, Ciudad Real (Spain)

[1] *Emilio.Ruiz3@alu.uclm.es*, [2] *Sarai.Diaz@uclm.es*, [3] *Javier.Gonzalez@uclm.es*

Abstract

Hydraulic State Estimation (HSE) is an effective tool for water supply systems monitoring. This technique provides the most likely hydraulic state of the network by making use of the available measurements through the system and the associated hydraulic model, enabling to track uncertainty through the process. HSE has been applied to water transport networks before, but extending its application to distribution systems currently constitutes a challenge. The deployment of smart meters offers a unique opportunity to gain distributed information within the distribution level, but it also poses questions regarding the temporal resolution that is available/achievable and how to combine different sources of information. This work presents a real HSE implementation in a Spanish water distribution pilot area (approximately 1,000 inhabitants). Implementation is possible thanks to the systematic measurement of water consumption at each dwelling (with 1 minute and 1 L resolution) and flow and pressure at the inlet to the pilot area. This application is conceived to provide a better understanding of the challenges and opportunities for HSE implementation in water distribution systems.

Keywords
Water supply systems, distribution systems, smart meters, water consumption, household, high resolution.

1 INTRODUCTION

Water supply systems are currently being modernized thanks to the installation of metering devices and telemetry systems. These configurations are providing water utilities with large amounts of data, which must be processed to convert isolated readings into real information about how the system works. Hydraulic State Estimation (HSE) is an effective tool to process this data. It is posed as an optimization problem that provides the most likely hydraulic state by minimizing the difference between uncertain measurements (i.e. pressure measurements, flow measurements, demand measurements or estimations) and network variables (i.e. pressures and flows) while considering the hydraulic model equations as constraints [1]. This approach is more flexible than traditional hydraulic modelling and better suited for monitoring because it enables to consider any combination of measurements and their associated uncertainty to determine the flow regime [2].

HSE implementation requires that the network is observable, i.e. the number of measurements must be equal or greater than the number variables to be determined (i.e. unknowns) [3, 4]. However, this condition is not enough to ensure a successful state estimation. Since measurements are uncertain, uncertainty assessment is essential to quantify the quality of results [5, 6]. The interest and usefulness of HSE increases with the number of measurements, because it is when there are redundant measurements (i.e. more measurements than unknowns state variables) that state estimation really helps to identify the most likely state of the system based on all the available noisy inputs.

Measurement availability is therefore crucial to make the most of state estimation implementation. Since water systems have traditionally lacked metering devices, HSE deployment in this field has been slow. State estimation has been discussed in water supply systems on a scientific level for decades (see [7] for references), but its operational implementation is relatively scarce [8, 9, 10, 11]. The few applications that exist in the water domain focus in water transport networks [12, 13, 14]. Transport networks constitute the main arteries that convey water to District Metered Areas (DMA). DMA inflows are usually metered, which guarantees that a minimum number of measurements is available to run state estimation. The situation within DMAs or unstructured water distribution networks is different, because instrumentation is not so frequently available in distribution pipelines. The smart revolution is promoting the installation of meters at water service connections (i.e. at the entrance to each household, public building, industry, etc), offering a unique opportunity to gain distributed information throughout the system. However, the resolution of the new volume meters that may be located at water connections (which may provide 1 hour resolution as opposed to the traditional monthly volume read [15]) is usually poorer than the resolution of the flow meters located at the entrance to the DMA or other points within the transport/distribution network (1-15 minutes, e.g. [16]). These resolutions may also differ from the time resolutions usually adopted for monitoring or modelling purposes (15-60 minutes [17]). The impact of temporal resolution [18] and the need to level different temporal resolutions to implement HSE [19] has already been explored from a theoretical point of view in some archetypical examples, but according to the authors' knowledge it has never been explored in a real system.

The aim of this work is to present how a real HSE implementation is being put together in a pilot area (approximately 1,000 inhabitants) within a Spanish water distribution system. Apart from presenting the formulation required to implement HSE (and computing HSE uncertainty), the main characteristics of the measurement scheme (every minute 1 L resolution) are here presented. It is important to highlight that previous works have reached higher temporal resolutions in the past for specific demand characterization/disaggregation research purposes (up until the order of seconds, see [20] for references), but the emphasis of this work is to address the multi-scale problem rather than to focus on consumption particularities within the household. Therefore, this implementation will contribute to bridge the gap between the macro (network level) and the micro (household level) scales that coexist in water distribution systems.

2 METHODOLOGY

2.1 Hydraulic State Estimation

HSE minimizes the difference between available measurements or pseudomeasurements (i.e. estimations based on historical data [21]) and the equivalent network hydraulic variables, which are connected to each other through the hydraulic model equations (energy and continuity) and the so called state variables (set of independent variables that describe the network hydraulic state). The vector of differences between measured variables and corresponding measurements $\boldsymbol{\varepsilon} \epsilon \mathbb{R}^m$ to be minimized can be generally defined as:

$$\boldsymbol{\varepsilon} = \boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{z} \epsilon \mathbb{R}^m$ is the measurement vector, $\boldsymbol{x} \epsilon \mathbb{R}^n$ is the state variable vector and $\boldsymbol{h}(\boldsymbol{x}): \mathbb{R}^n \to \mathbb{R}^m$ represents the function of non-linear relationships between the state and measured variables. Different criteria might be adopted to minimize the error as defined in Eq (1). According to the Weighted Least Squares (WLS) criterion, which is one of the most popular criteria for HSE [7], the problem can be posed as:

$$\min_{\boldsymbol{x}} \; G(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{\varepsilon}^T \boldsymbol{C}_{\boldsymbol{z}}^{-\mathbf{1}}\boldsymbol{\varepsilon} \tag{2}$$

$$\text{Subject to } \boldsymbol{f}(\boldsymbol{x}) = \mathbf{0} \tag{3}$$

Where $J(\boldsymbol{x})$ is the objective function that is to be minimized, $\boldsymbol{C}_{\boldsymbol{z}} \in \mathbb{R}^{m\times m}$ is the variance-covariance matrix of measurements and $\boldsymbol{f}(\boldsymbol{x})\colon \mathbb{R}^n \to \mathbb{R}^c$ is the function of equality constraints (e.g. transit nodes with null demand). Note that Problem (2)-(3) minimizes the errors between measurements and estimations, providing the most likely state of the system. This implies that, thanks to redundancy, the algorithm can probabilistically estimate the state of the system considering as more reliable those measurements that are consistent among each other and with the hydraulic model (see Figure 1). Inconsistent measurements may be detected as outliers, which are potentially useful to identify abnormal behaviours [14].

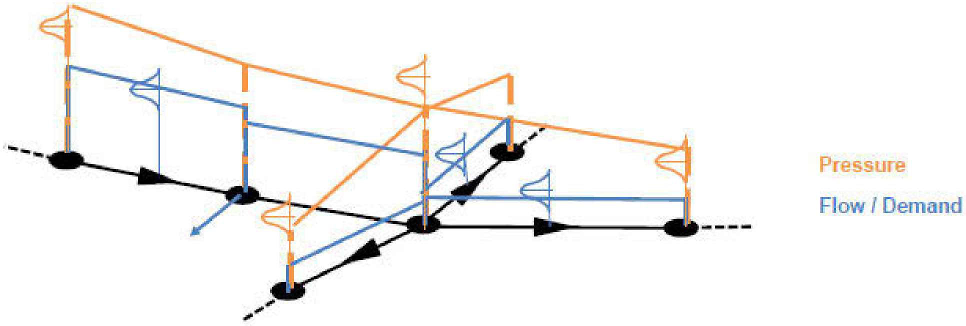


Figure 1. State estimation conceptualization in water supply systems

Problem (2)-(3) can be solved as a Newton iterative process:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \Delta\boldsymbol{x}_{i+1} \tag{4}$$

$$\begin{bmatrix} \boldsymbol{H}_i^T \boldsymbol{C}_{\boldsymbol{z}}^{-\mathbf{1}} \boldsymbol{H}_i & \boldsymbol{F}_i^T \\ \boldsymbol{F}_i & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x}_{i+1} \\ \boldsymbol{\lambda}_{i+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_i^T \boldsymbol{C}_{\boldsymbol{z}}^{-\mathbf{1}}[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}_i)] \\ -\boldsymbol{f}(\boldsymbol{x}_i) \end{bmatrix} \tag{5}$$

Where i is an iteration counter, $\Delta\boldsymbol{x} \in \mathbb{R}^n$ is the vector of increments for the state variables, $\boldsymbol{H} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{F} \in \mathbb{R}^{c\times n}$ are the Jacobian matrices of $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x})$ respectively, $\boldsymbol{\lambda} \in \mathbb{R}^c$ represents the vector of dual variables and $\hat{\boldsymbol{x}}$ is the optimal solution to the problem. These equations have been particularised for different sets of state variables at water supply systems in [2].

2.2 Uncertainty quantification

HSE results are associated with uncertainty because the process nourishes from uncertain measurements/pseudomeasurements (matrix $\boldsymbol{C}_{\boldsymbol{z}}$). The variance-covariance matrix for the state variables $\boldsymbol{C}_{\boldsymbol{x}} \in \mathbb{R}^{n\times n}$ can be computed once $\hat{\boldsymbol{x}}$ is known by applying the First-Order Second-Moment (FOSM) method [6]:

$$\boldsymbol{C}_{\boldsymbol{x}} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \boldsymbol{C}_{\boldsymbol{z}} \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}\right)^T \tag{6}$$

Where the sensitivity of the state and dual variables can be obtained as:

$$\begin{bmatrix} \dfrac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \\ \dfrac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{z}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}^T \boldsymbol{C}_{\boldsymbol{z}}^{-\mathbf{1}} \boldsymbol{H} & \boldsymbol{F}^T \\ \boldsymbol{F} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{H}^T \boldsymbol{C}_{\boldsymbol{z}}^{-\mathbf{1}} \\ \mathbf{0} \end{bmatrix} \tag{6}$$

The variance-covariance matrix of any other set of variables can be computed by applying the FOSM again [6].

3 RESULTS

The pilot area is located within a water distribution supply system that provides water to a town located in Castilla-La Mancha (Spain). The hydraulic model of the distribution system where the pilot area is located is currently being built. Apart from the model definition, this involves a socioeconomic analysis that will support the water consumption records that the water utility collects by default for billing purposes (every 3/6 months).

At the same time, sensor and mesh communication deployment are currently underway. Volume meters with a 1 litre resolution were already available in the case study area and/or being installed. These conventional volume meters have been upgraded with data loggers that record water consumption with a 1-minute time resolution, although data is sent with a smaller frequency (60 minutes) to extend battery life. Pressure and flow are also measured at the entrance to the pilot area with a 1-minute temporal resolution. The HidraIoT solution developed by Hidralab Ingeniería y Desarrollos, S.L. is being adopted for the communication technology and mesh deployment. Different temporal resolutions for the HSE output will be tested to explore the temporal resolution effect as in [19].

4 CONCLUSIONS

Hydraulic State Estimation is a powerful tool to monitor water systems, but it has only been applied to water transport networks until now in real practice. This work presents the particularities of HSE implementation in a pilot area within a distribution system in Spain. This application will enable to explore the required temporal and spatial resolution levels needed to achieve different uncertainty thresholds within the distribution level. Therefore, it will enable to fully understand flow and water quality dynamics within the pilot area, but also to identify challenges and opportunities for HSE systematic implementation in other real (and not so well-instrumented) case studies.

5 ACKNOWLEDGEMENTS

The authors would like to thank the financial support provided by the Spanish Ministry of Science and Innovation - State Research Agency (Grant PID2019-111506RB-I00 funded by MCIN/AEI/10.13039/ 501100011033) and Junta de Comunidades de Castilla-La Mancha (Grant SBPLY/19/180501/000162 funded by Junta de Comunidades de Castilla-La Mancha and ERDF A way of making Europe).

6 REFERENCES

[1] S. Díaz, R. Mínguez and J. González, "Topological State Estimation in Water Distribution Systems: Mixed-integer Quadratic Programming Approach," Journal of Water Resources Planning and Management, vol. 144, no. 7, 2018, 04018026.

[2] E. Ruiz, S. Díaz and J. González, "State Variable Implications on Hydraulic State Estimation", Water Resources Management, https://doi.org/10.1007/s11269-022-03152-x

[3] A. Nagar and R. Powell, "Observability Analysis of Water Distribution Systems Under Parametric and Measurement Uncertainty", Building Partnership, 2000, pp.1-10.

[4] S. Díaz, J. González and R. Mínguez, "Observability Anaysis in Water Transport Networks: Algebraic Approach," Journal of Water Resources Planning and Management, vol. 142, no. 4, 2016, 04015071.

[5] S.G. Vrachimis, S. Timotheou, D.G. Eliades and M.M. Polycarpou, "Iterative Hydraulic Interval State Estimation for Water Distribution Networks", Journal of Water Resources Planning and Management, vol. 145, no. 1, 2019, 04018087.

[6] S. Díaz, J. González and R. Mínguez, "Uncertainty Evaluation for Constrained State Estimation in Water Distribution Systems", Journal of Water Resources Planning and Management, vol. 142, no. 12, 2016, 06016004.

[7] K.S. Tshehla, Y. Hamam and A.M. Abu-Mahfouz, "State Estimation in Water Distribution Network: A Review", in Proc. IEEE 15th International Confeerence on Industrial Informatics (INDIN), Embeden, Germany, IEEE, 2017, pp. 1247-1252.

[8] A. Bargiela, On-line Monitoring of Water Distribution Networks, PhD Thesis, University of Durham, UK, 1984.

[9] R.S. Powell, On-line Monitoring for Operational Control of Water Distribution Networks, PhD Thesis, University of Durham, UK, 1992.

[10] P. Carpentier and G. Cohen, "State Estimation and Leak Detection in Water Distribution Networks", Civil Engineering Systems, vol. 8, no. 4, 1991, pp. 247-257.

[11] S. Díaz, Comprehensive Approach for On-line Monitoring Water Distribution Systems Via State Estimation Related Techniques, PhD Thesis, University of Castilla-La Mancha, Spain, 2017.

[12] J. González, R. Mínguez and S. Díaz, Observability Study for Hydraulic State Estimation of the Sectorised Water Supply Network, Booklet Research & Development & Innovation 23, Madrid: Canal de Isabel II, 2016.

[13] S.G. Vrachimis, D.G. Eliades and M.M. Polycarpou, "Real-time Hydraulic Interval State Estimation for Water Transport Newtorks: A Case Study", Drinking Water Engineering and Science, vol. 11, no. 1, 2018, pp. 19-24.

[14] J. González, S. Díaz and J. Campos, "Pilot Implementation of a Hydraulic State Estimator for the Strategic Supply Network of Madrid (Spain)", in Proc. 17th International Computing & Control for the Water Industry Conference, Exeter, UK, 2019.

[15] L. Pastor-Jabaloyes, F.J. Arregui and R. Cobacho, "Water End Use Disaggregation Based on Soft Computing Techniques", Water, vol. 10, no. 46, 2018.

[16] R. Gargano, C. Tricarico, F. Granata, S. Santopietro and G. de Marinis, "Probabilistic Models for the Peak Residential Water Demand", Water, vol. 9, no. 417, 2017.

[17] P. Kossieris, C. Makropoulos, "Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales", Water, vol. 10, no. 1481, 2018.

[18] S. Díaz and J. González, "Temporal Scale Effect Analysis for Water Supply Systems Monitoring Based on a Microcomponent Stochastic Demand Model", Journal of Water Resources Planning and Management, 2021, vol. 147, no. 5, 04021023.

[19] E. Ruiz, S. Díaz and J. González, "Potential Performance of Hydraulic State Estimation in Water Distribution Networks", Water Resources Management, vol. 36, 2022, pp. 745-762.

[20] A. Di Mauro, A. Cominola, A. Castelletti and A. Di Nardo, "Urban Water Consumption at Multiple Spatial and Temporal Scales. A Review of Existing Datasets", Water, vol. 13, no. 36, 2021.

[21] A. Bargiela and G.D. Hainsworth, "Pressure and Flow Uncertainty in Water Systems", Journal of Water Resources Planning and Management, vol. 115, no. 2, 1989, pp. 212-229.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14169

MODELLING A COMPREHENSIVE RELATION BETWEEN WATER QUALITY AND CEMENT DEGRADATION IN THE DRINKING WATER DISTRIBUTION NETWORK

Karel van Laarhoven[1], Alex Hockin[2] and Martin Korevaar[3]

[1, 2, 3]KWR Water Research Institute, Nieuwegein, Utrecht, the Netherlands

[1] *karel.van.laarhoven@kwrwater.nl,* [2] *alex.hockin@kwrwater.nl,* [3] *martin.korevaar@kwrwater.nl*

Abstract

Over a quarter of the Dutch drinking water distribution network consists of cementitious pipe materials (e.g. asbestos cement, concrete and cement-coated cast iron). Leaching of cement components into the drinking water may lead to changes in water quality and to a reduction of a pipe's structural integrity. Many utilities worldwide condition their drinking water to prevent leaching. Several indices can be derived from a given water composition to estimate how well the water composition will protect against leaching. While water utilities have so far achieved good results using for instance the Langlier Saturation Index (LSI) and the Calcium Carbonate Precipitation Potential (CCPP) as an indicators, scientific authors have already posed in the past that the protection against leaching depends on a complex interplay between many chemical components in both the drinking water and the cement, and that no single index is sufficient to guarantee protection. In the presented work, a comprehensive model of the chemical and physical interactions between a cement pipe and drinking water is constructed. The model takes any given drinking water composition for input and returns a description of the chemical and microstructural changes that develop in the pipe wall over time as leaching progresses. The model shows reasonable agreement with leaching experiments described in literature. Using such a model, one can evaluate the protective capabilities of a given water quality based on all its aspects rather than just the aspects captured by derived indices. The model can be a powerful research tool for studying the interactions between drinking water and pipe materials, but this approach should also already prove to be a valuable aid for process engineers to optimize their conditioning strategies in practice, and for asset managers to translate historical water quality data to an estimation of the structural integrity of their cementitious pipes.

Keywords
Leaching, Water Quality, Asbestos Cement, PHREEQC, Transport in Porous Materials.

1 INTRODUCTION

Over a quarter of the Dutch drinking water distribution network consists of cementitious pipe materials (e.g. asbestos cement (AC), concrete and cement-coated cast iron). Given time, however, cement dissolves into wate. This process, leaching, is known as one of the most common degradation mechanisms of such pipe materials. Leaching is a problem because, first of all, it leads to components of the pipe wall entering the drinking water, changing the water quality. Additionally, leaching removes load-bearing material from the pipe wall, reducing the structural integrity of the distribution network. Leaching, therefore, must be prevented.

Many utilities worldwide condition their drinking water to prevent leaching. A key strategy is to aim at a water composition that is conducive to the precipitation of a thin layer of calcite that obstructs the contact between water and cement, thereby preventing dissolution 0. Several indices can be derived from the water composition to estimate the likelihood of such a layer forming, such as the Langlier Saturation Index (LSI) and the Calcium Carbonate Precipitation Potential (CCPP) 0, 0. While the Dutch utilities have so far achieved good results using the LSI as an indicator, scientific authors have already posed in the past that the formation of protective layers depends on a complex interplay between many chemical components in both the drinking water and the cement, and that no single index is sufficient to guarantee protection 0, 0, 0.

The goal of the presented research is to construct and validate a numerical model of the chemical and physical interactions between a cement pipe and drinking water of any given composition. Using such a model, water utilities can evaluate the protective capabilities of a given water quality based on all its aspects, rather than just the aspects captured by derived indices such as the LSI and the CCPP. Ultimately, this should allow utilities to prevent water quality changes during distribution more effectively, and to better gauge the condition and remaining lifetime of their network.

2 MODEL

2.1 General approach

Cement is a complex composite material. It consists of several types of mineral crystals, held together by an amorphous, chemically complex gel of calcium silicate hydrates: the C-S-H gel. The crystal clusters provide a large part of the cement's structural strength. Aggregates may be added to cement to provide additional strength, such as pebbles in the case of concrete and asbestos fibres in the case of asbestos cement. Pores of different length scales are presents in the material as well, forming a microscopic network of interconnected voids that allow for some measure of fluid transport through the material.

When cement comes into contact with water, mineral crystals and C-S-H may dissolve. This makes the cement more porous and reduces the cohesion of the cement components and aggregates. This in turn reduces the overall strength of the material. Calcium hydroxide (portlandite, or CH) is one of the first minerals to dissolve and already leads to a dramatic reduction of material quality 0.

The general sequence of mechanisms that will be modelled is illustrated in figure 1. The cementitious material of the pipe is in contact with the drinking water transported by it. Due to porous transport, it is assumed that the initial pores present in the material are filled with water (1). The water in the pores equilibrates with the cement around it, so that cement components dissolve in the pore water (2). Due to concentration differences between the pore water and the drinking water outside the pipe wall, the dissolved cement components diffuse through the pore network into the drinking water (3). The flowing drinking water removes the dissolved minerals from the modelled system (4). More and more pores are formed as more and more material dissolves, making degraded parts of the material more permeable to the diffusive transport of dissolved components. This creates a positive feedback loop that has partially degraded material degrade faster and faster.

Four aspects of this general process require detailed modelling in particular: the initial cement composition; the thermodynamic stability of the different cement components, which governs how they dissolve; the diffusive transport of dissolved components through the system; and the model of the cement microstructure that governs how diffusive transport changes as the leaching progresses. These aspects are discussed in the following sections. After that, a more detailed description of the technical implementation of the model is given as well.

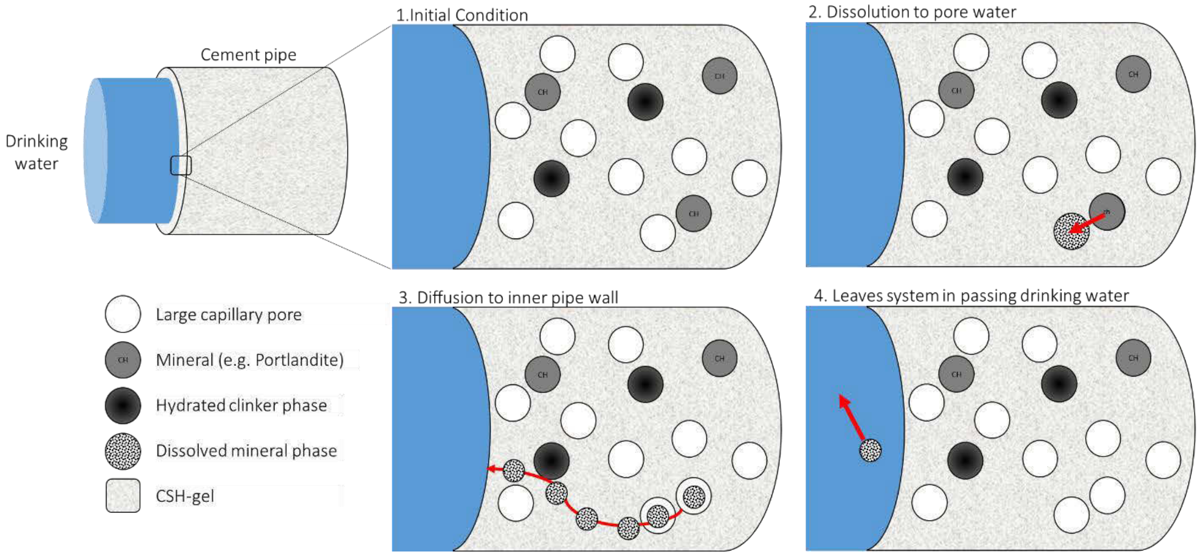

Drinking water
Cement pipe
1.Initial Condition
2. Dissolution to pore water
3. Diffusion to inner pipe wall
4. Leaves system in passing drinking water
CH
Large capillary pore
Mineral (e.g. Portlandite)
Hydrated clinker phase
Dissolved mineral phase
CSH-gel

Figure 1. Schematisation of the process of cement leaching for a cement drinking water pipe. Initially, (1) cement is in contact with drinking water at the left boundary. Cement minerals and hydrated phases are present in the cement. (2) Cement components dissolve into the water present in the pores, leaving additional pores behind. (3) The dissolved mineral diffuses from the inner cement through the pipe wall towards the drinking water. (4) The dissolved mineral is removed from the system by the passing drinking water.

2.2 The thermodynamic stability of cement minerals

The dissolution or precipitation of minerals in cement is determined by the mineral thermodynamics. Minerals precipitate or dissolve to achieve chemical equilibrium in the system. When a system is in equilibrium, the Gibbs free energy of the system is at a minimum and all chemical reactions are in equilibrium. Mass action equations relate changes in the Gibbs free energy of reaction the mass action constants 0. In chemical equilibrium modelling, the mass action equations and conservation of mass are used to calculate the equilibrium state of a system 0. In order to calculate the chemical equilibrium of a system, chemical modelling software is used 0. This work uses one example of such software: PHREEQC.

PHREEQC requires a database that contains data (aqueous speciation and activity coefficients) of the elements in the modelled system, to evaluate their equilibrium. A specialized cement database (CEMDATA07), compiled by Lothenbach et al., 0, 0 and converted for use in PHREEQC by Jaques, 0, is used for the purposes of this model. This database compiles mass action constants for cement minerals from literature and practice. Importantly, the database also implements a model to represent the complex C-S-H gel as a system of a reduced number of specific hydrates: tobermorite-II ($Ca_5Si_6O_{16}(OH)_2\cdot 4H_2O$), jenite ($Ca_9Si_6O_{18}(OH)_6\cdot 8H_2O$) and amorphous silicagel (SiO_2). Tabel 1 summarizes the different phases that are selected as relevant components of cement for PHREEQC to evaluate in the model.

Table 1. Mineral phases and elements in the model.

Mineral	**Description**
CALCIUM-SILICATE-HYDRATES (CSH)	The C-S-H gel is modelled as an ideal solid solution of Jennite ($Ca_9Si_6O_{18}(OH)_6\cdot 8H_2O$), tobermorite-II ($Ca_5Si_6O_{16}(OH)_2\cdot 4H_2O$) and a pure phase of amorpheous silciumdioxide (SiO_2).
AFT-PHASES	An ideal solid solution of ettringite-tricarboaluminate is present as the most important Aft-phase.

OTHER PHASES	Other cement minerals that are included are: hydrotalcite, calcite, portlandite, monocarboaluminate, stratlingite, hydrotaliceOH. Included elements are: Al, C, Ca, K, Mg, Na, S, Si. Iron was not included in the model as Jacques et al., 0, found it led to unnecessarily complicated calculations and its inclusion had no effect on the evolution of pH, element concentrations and solid phase composition.

2.3 The chemical composition of the cement

Ordinary Portland Cement (OPC) was chosen as the default cement composition for the model, as this is the type of cement used in the majority of Dutch asbestos cement pipes 0. OPC consists of four major clinker phases: alite (C3S), belite(C2S), aluminate (C3A) and ferrite (C4AF) 0. The exact composition of elements used as a default is CEM I 42.5, as reported in Table 2, but can be modified to match other types of cement if so desired. When a defined amount of water is added, these clinker phases slowly hydrate to form the various cement hydrates.

The initial composition of the hydrated cement is obtained during the initialization of the model by entering the OPC composition from table 2 in PHREEQC as a solution of the components in a certain amount of water. The amount of water is chosen according to a desired 'water to cement mass ratio' (w/c). PHREEQC then converts this solution to the corresponding composition of hydrated mineral crystals and C-S-H gel present in table 1, assuming complete hydration (α) of the clinker products. In fully hydrated cement (α=1), all cement clinker has reacted and no unreacted clinker is left. Under regular mixing conditions, the degree of hydration is less than 1 and can be calculated from empirical formulas from literature (e.g. 0). The main producer of Dutch AC pipes, however, reports that their material is produced under circumstances (high pressure and humidity) that are conducive to full hydration 0. Note that a w/c of at least more than approximately 0.27 is required to provide sufficient water molecules for complete hydration (higher w/c may influence the specific composition of hydrated components and the microstructure of the cement, as discussed below).

Table 2. CEM I compositions of Ordinary Portland Cement according to 0.

	CEM I 42.5 N 0	**CEM I 52.2 N 0**
Mineral	**g/100 g OPC**	**g/100 g OPC**
CaO	62.4	67.4
SiO_2	18.9	23.4
Al_2O_3	4.4	30.5
Fe_2O_3	2.5	2.15
CaO (free)	0.6	-
MgO	1.4	0.7
K_2O	0.95	0.15
Na_2O	0.10	0.1

CO_2	2.1	-
SO_3	3.0	2.1

2.4 Microstructure model of degrading cement

The porosity of the cement plays a key role in the diffusive transport of solutes (as will be discussed below). The porous structure of the cement must therefore be addressed in the model, especially since the leaching of cement will result in progressive changes in the porous structure. The well-established cement models of Powers et al., 0, as reviewed by Brouwers, 0, are used to estimate the initial volume fractions of the cement and its pores. As discussed, the cement consists of minerals (e.g. portlandite, calcite etc.) and a CSH-gel. The porous structure of the cement contains roughly two types of pores with different length scales: 'capillary pores' (0.1-1 µm), which are voids between the crystals and gel; and the smaller 'gel pores' (5-50 nm), which are voids between the molecules that constitute the C-S-H gel.

The capillary porosity depends on the circumstances under which cement is produced: it is a function of the water to cement ratio and the degree of hydration 0. With increasing w/c and decreasing α, the initial capillary porosity of cement increases. This is because water consumed by hydration becomes part of the cement material, but the water in excess of the amount consumed by hydration remains distributed across the curing cement, resulting in the formation of capillary pores 0. Powers et al., 0, provide empirical formulas for the individual volume fractions of the capillary pores, θ_{cp}, and the gel pores θ_g, that arise in cement under regular mixing conditions:

$$\theta_{cp} = \frac{w/c - 0.36\alpha}{w/c + 0.32} \tag{1}$$

$$\theta_g = \frac{wc - 0.17\alpha}{wc + 0.32} - \theta_{cp} \tag{2}$$

As mentioned above, the cement in AC pipes was not produced under regular mixing conditions, but under circumstances (high pressure and humidity) that are conducive to full hydration at very low w/c, so that very few capillary pores are formed 0. Because of this, (1) and (2) are likely not suitable to use directly for AC. To determine the initial microstructure of AC specifically, first regular cement was simulated with PREEQC (assuming a w/c of 0.4-0.5 and an α of 1). Evaluating (1) and (2) for this cement, it follows that e gel pore volume is approximately equal to 70%-95% of the volume of the C-S-H minerals. Subsequently, it is assumed that the AC production process negates the creation of capillary pores but leaves the voids in the C-S-H gel unchanged. The total volume and porosity of the AC is then recalculated from: the volume of the hydrated products, the volume of the gel pores, and the volume of an added amount of asbestos aggregate.

In cement-aggregate systems, such as concrete or asbestos cement, the porosity of the cement close to the aggregate may be modified. This leads to a interfacial transition zone (ITZ), a zone of heightened porosity and permeability, between the bulk cement paste and aggregate particles like pebbles or asbestos fibers 0. This zone is important in composite systems, because it may substantially facilitate diffusive transport. Focussing on asbestos cement, however, the importance of the ITZ is negligible because the asbestos fibers are not thick enough to influence the porosity of the surrounding cement 0. The ITZ is therefore not considered in this study, but may need to be included when studying for instance concrete pipes.

To evaluate how the microstructure changes during leaching, it is assumed that the dissolution of a certain volume of cement components results in the creation of an equal volume of capillary

BY NC SA

pores. When C-S-H components dissolve, however, also a proportional volume of gel pores is converted to capillary pore volume. As capillary pores are far larger than gel pores, they would form the main transport path for diffusion. If the production circumstances lead to a low capillary porosity, however, capillary pores may not form a connected network, whereas gel pores are ubiquitous as they are linked to the C-S-H gel. Diffusion through gel pores therefore also contributes and even dominates transport at very low capillary porosity 0. A critical capillary porosity (θ_{ccp}) may be identified below which there is no transport through the capillary pores (Figure 2).

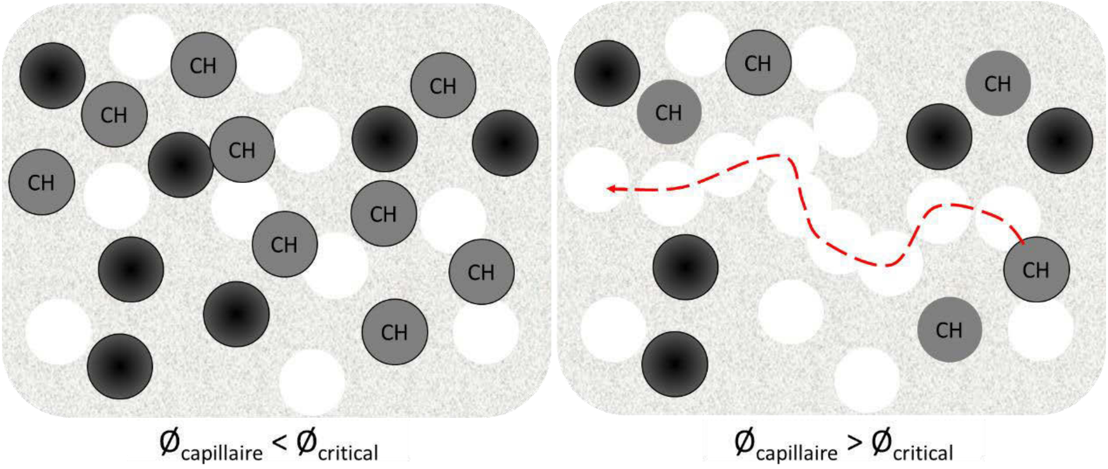

CH
$\emptyset_{capillaire} < \emptyset_{critical}$
$\emptyset_{capillaire} > \emptyset_{critical}$

Figure 2. Visualization of the connectivity of the capillary pores (open circles). Grey and black cicles represent the cement minerals and the light grey background represents the porous CSH-gel. Left: below the critical capillary porosity, capillary pores are disconnected and diffusion (red line) through capillary pores does not occur. Right: above the critical capillary porosity, capillary pores are connected and diffusion through capillary pores occurs.

2.5 Diffusive transport of dissolved minerals through the degrading porous material

Diffusion is the resulting movement of particles from regions of higher to lower concentration, as described by Fick's Law 0:

$$F = -D\frac{\partial c}{\partial x} \tag{3}$$

Where F is the flux [mol/s/m^2], D is the diffusion coefficient [m^2/s] and c is the concentration [mol/m^3]. The diffusion coefficient depends on, among other things, the ion size, the temperature and interaction with other diffusing species due to species charge. Oppositely charge species can accelerate while similarly charged species can retard diffusion 0. Detailed data on the diffusion coefficients for elements is not always available, therefore for multi-component diffusion an average diffusion coefficient in free water is currently used 0. Diffusion of elements in porous systems must travel longer distances than in bulk water. The TRANSPORT block of PHREEQC accounts for this with a hardcoded relation for an effective diffusivity, D_e:

$$D_e = [\theta]^n D_0 \tag{4}$$

Where θ is the total (water-filled) porosity [-] and n is the Archie's constant [-]. As described in the previous section, however, not only the overall porosity of the cement changes during leaching, but also the connectivity between (capillary) pores changes. To account for this, a more suitable effective diffusion coefficient is instead taken to be D_{GEM}, obtained with the general effective medium (GEM) equation reported by Oh et al., 0:

$$D_{GEM} = [\theta_{GEM}\,]^n D \tag{5}$$

$$\theta_{GEM} = \mathrm{m}_\theta + \sqrt{m_\theta^2 + \frac{\theta_{ccp}}{1-\theta_{ccp}}\left(\frac{D_s}{D_0}\right)^{\frac{1}{n}}} \tag{6}$$

$$\mathrm{m}_\theta = \frac{1}{2}\left[\left(\frac{D_s}{D_0}\right)^{\frac{1}{n}} + \frac{\theta_{ccp}}{1-\theta_c}\left(1-\left(\frac{D_s}{D_0}\right)^{\frac{1}{n}}\right) - \frac{\theta_{ccp}}{1-\theta_{ccp}}\right] \tag{7}$$

Where θ_c is the capillary porosity, D_s/D_0 is the normalized diffusivity through solid C-S-H (when the capillary porosity is zero), D_0 and D_s are the diffusivities [m^2/s] in, respectively, bulk water (i.e. the maximum effective diffusivity possible), and in the solid C-S-H (i.e. the minimum effective diffusivity possible). For Portland cement pastes, $n \approx 2.7$ and $D_s/D_0 \approx 2.0*10^{-4}$ 0. The critical capillary porosity is approximately 0.18 0. Inserting θ_{GEM} into (3) provides PHREEQC's transport module with a suitable, leaching-dependent diffusivity for cement.

2.6 Technical implementation of the model

The concepts in the sections above were combined into a numerical model using the geochemical code of PHREEQC 3.3.7, 0, through the python package PhreeqPython 0. CEMDATA07 was used as the chemical database for PHREEQC. The model is set up and initialized as follows:

1. A representative volume of (asbestos) cement is initialized by creating a solution of an amount of clinker (e.g. with the composition in table 2) in an amount of water according to a chosen w/c and evaluating it with PHREEQC (resulting in a composition of the components listed in table 1). The initial composition of the pore water is taken to be the composition that is in equilibrium with the initialized cement.
2. The diffusive transport is implemented as 1D diffusive transport using the TRANSPORT keyword in PHREEQC. Electro-diffusive phenomena are neglected. Consequently each aqueous species has the same diffusion coefficient 0. Using the multi-component diffusion option in PHREEQC, overall charge balance is maintained 0.
3. For the purposes of transport, the volume of cement is discretized in a number of cells of thickness *l* with a combined thickness L in the direction of transport.
4. One boundary condition is set as closed on one side (Neumann type boundary condition), representing the external wall surface of the cement pipe (which is not in contact with the drinking water). The other boundary condition (inner pipe wall) is set as a constant set of concentrations, representing exposure to any (continuously refreshed) drinking water composition (Dirichlet type boundary condition).
5. The initial capillary porosity and gel porosity are determined from the initialized cement as described in paragraph 2.4, with equations (1) and (2) and specific assumptions in the case AC is being modelled.

Simulation of the leaching in this system is then achieved by the following cycle for each numerical timestep:

1. For each cell, the effective porosity θ_{GEM} is determined with equation (6), based on each cell's current θ_{cp}.
2. The θ_{GEM} is implemented in the transport block using the built-in BASIC command program of PHREEQC. To save computational time, the GEM porosity may be updated less

frequently. No difference was observed with performing the calculation after each time step versus after every 100th time step.

3. Diffusive transport (mediated by the θ_{GEM}) between the pore water solutions of neighbouring cells (and the drinking water boundary condition) is evaluated with PHREEQC.
4. Equilibration between the cement in each cell and its new pore water composition is evaluated with PHREEQC.
5. Based on possible precipitation or dissolution of components in a cell during step 4, the θ_{cp} of each cell is updated.

3 PRELIMINARY OUTCOMES AND DISCUSSION

3.1 Example results and grid sensitivity

The raw output of the model consists of the chemical composition of the cement and pore water in each cell, and the derived porosities, for each timestep. Figure 3 display examples of data from a series of simulations performed for the purposes of a grid sensitivity study of the model. The initial cement composition from 0 in table 2 was used (and was not modified to resemble AC as discussed in paragraph 2.4). Other model parameter values, as mentioned earlier, were: w/c = 0.5; D_s/D_0 = 2.0*10^{-4}; α = 1, θ_{ccp} = 0.18, n = 2.7. Pure water was used for the water composition at the Dirichlet boundary condition. A system length L of 5 mm was chosen. A simulation of 114 days of leaching in this system was repeated for different discretization length scales (l = 1, 0.5, 0.25, 0.125, 0.0625).

Figure 3 shows two profiles of aspects across the system after the simulation: the number of moles of portlandite remaining in each cell and the pH of the pore water in each cell. With respect to the different l, it can be seen that succeeding solutions become more similar with decreasing l, and that further reduction of l below 0.125 do not lead to substantial changes in the profiles. Moreover, it can be seen that – as is expected to occur during leaching – a sharp leaching front in terms of portlandite depletion forms during leaching, but that in the degraded part of the cement does have variation in its chemical composition, as can be gleaned from the different pHs in the degraded cells.

3.2 Verification from literature

The model was verified by using it to emulate a number of laboratory leaching experiments and comparing the outcomes. Moranville et al., 0, reported leaching experiments of 114 days with cement of different w/c (0.25, 0.4, 0.5), exposed to different water qualities (pure water and mineral water with the composition reported in table 3). The composition of their cement differs slightly from the CEM I previously discussed and is included in table 2. The model was used to reproduce these conditions, further using a grid size of $l = 0.1$ mm. Moranville reported the overall porosities of their samples, and the initialized cement was modified by adjusting the θcp of the initialized cement to match this. Moranville defined the results of the experiments in terms of the, visually observed, depth of portlandite leaching after the 114 days.

Table 3. mineral water composition (pH = 7) used in the experiments reported in 0.

Element	**Ca**	**Si**	**Mg**	**Na**	**K**	**Cl**	**SO_3**	**HCO_3**
Concentration [mg/L]	11.5	31.7	8	11.6	6.2	13.5	8.1	71

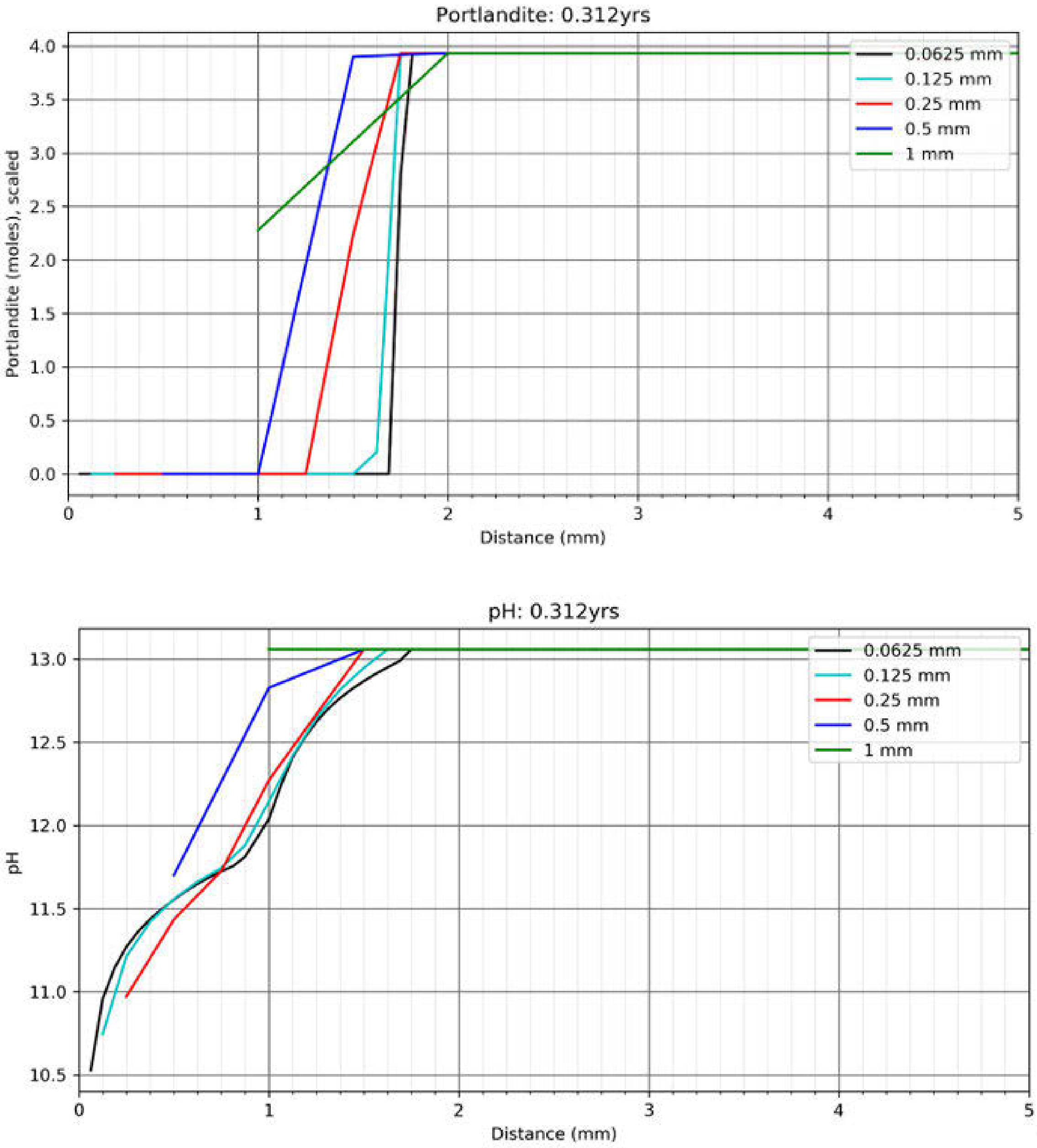


Figure 3. Outcomes of grid independence simulations of 114 days of 1D leaching of Portland cement exposed to pure water at distance 0. Shown are profiles across the system of portlandite content (upper) and pore water pH (lower). With decreasing grid size, the depletion front for portlandite become sharper and the change in pH between cells smoother.

Figure 4 compares the results of Moranville with the model outcomes. The progression of the portlandite front over time for the model simulations is shown by the green curves, and the snapshot measurements reported by Moranville are shown by the black markers. The model results follow the same trend as the lab results. The lower *w/c* leads to a reduced initial capillary porosity, which in turn leads to slower leaching. On top of that, exposure to mineral water results in even slower leaching; presumably due to slower dissolution and diffusion of cement components in the pores and due to deposition of calcite scaling obstructing diffusion even further.

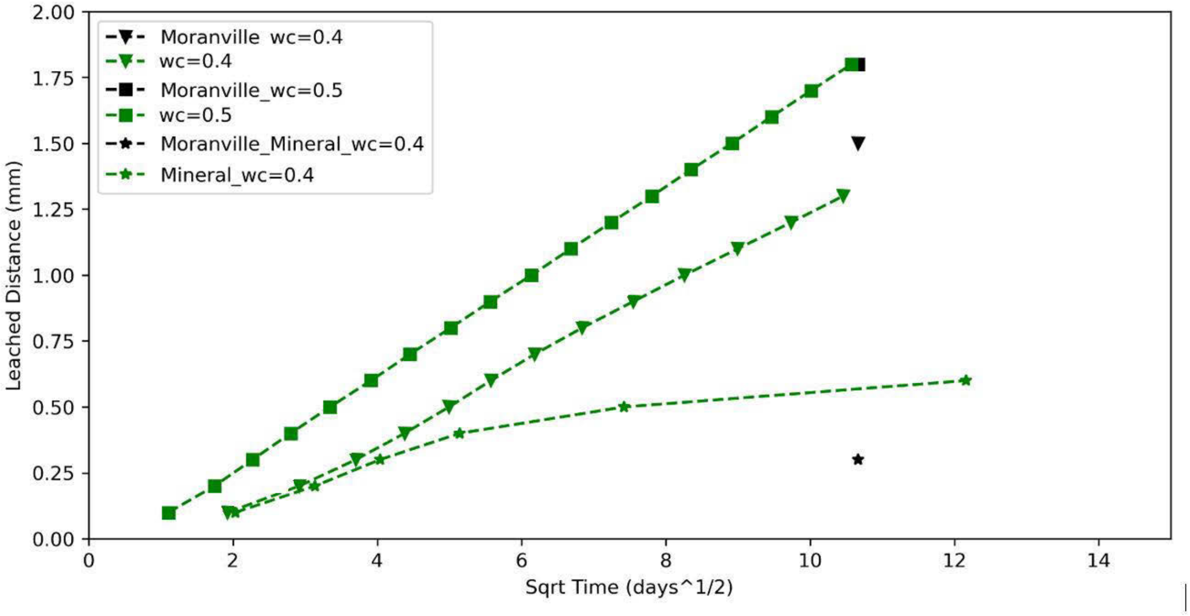


Figure 4. Comparison of the model outcomes to the experiments reported by 0. Green curves indicate the modelled progression over time of the leached dept. Black markers indicate the reported leached depths at the end of the experiments. The different symbols indicate which scenario was modelled: w/c = 0.5, exposed to pure water (squares); w/c = 0.4, exposed to pure water (triangles); and w/c = 0.4 exposed to the mineral solution summarized in table 3 (stars).

Carde et al., 0, reported a different set of leaching experiments, in which cement was exposed to demineralized water of pH = 4.5 and the leaching depth was measured after 90 and after 210 days. Carde used the CEM I 42.5 cement reported in table 2, with a *w/c* of 0.5, but with no measured porosities reported. The porosities of the cement were therefore estimated with equations (1) and (2). Because preliminary tests with the model showed a substantial sensitivity of results to the initial porosity, and additional simulation with *w/c* = 0.55 was carried out. The initial capillary porosities for *w/c* 0.5 and 0.55 are, respectively, 17 and 21.8 %.

Figure 5 shows the leached depth reported in Carde (black) and in the simulations with w/c =0.5 (blue) and w/c=0.55 (red). The results shows reasonable agreement with both simulations, with closer results for the simulations with the high w/c ratio (w/c = 0.55). Indeed, the model is sensitive to the initial capillary porosity, therefore without reported porosities to validate the initial porosity it is more difficult to get a very good fit, and the estimations of porosities with the empirical equations of Powers may not be completely suitable. However, even the simulations with the w/c of 0.5 (reported value in 0) the simulations are within 0.2 and 0.3 mm of the leaching after 90 and 210 days respectively. This is improved to 0.1 mm in the simulation with w/c = 0.55 for both the leached depth after 90 and 210 days.

4 CONCLUSIONS AND OUTLOOK

A numeric model for simulating the leaching and scaling processes in cementitious pipe materials exposed to a specific drinking water quality was presented. This model can be a powerful research tool for studying the interactions between drinking water and pipe materials, but this approach should already also prove to be a valuable aid for process engineers to optimize their conditioning strategies in practice, and for asset managers to translate historical water quality data to an estimation of the remaining structural integrity of their cementitious pipes.

At the moment of writing, steps are being taken to further refine different aspects of the model, such as bringing it in line with the most recent version of the CEMDATA database (CEMDATA18);

implementing individual diffusion rates for separate chemical species; the influence of the context of the larger hydraulic network on local leaching; and further refining the microstructure model of the cement. Apart from validation from experiments reported in the literature, the next step for validation is to model the leaching of real pipes that have been in use for several decades, based on several historic and recent condition measurements with phenolphthalein. Given the substantial sensitivity of the system to the initial porosity of cement, lab experiments probing the microstructure of actual pipe material would be preferable to provide the best model input assumptions.

After these refinement, the next research step is to use the model to carry out an extensive sensitivity analysis to start searching for the water compositions that optimally protect against leaching, and to find out how well the effectiveness of compositions can be gauged from the various existing index parameters.

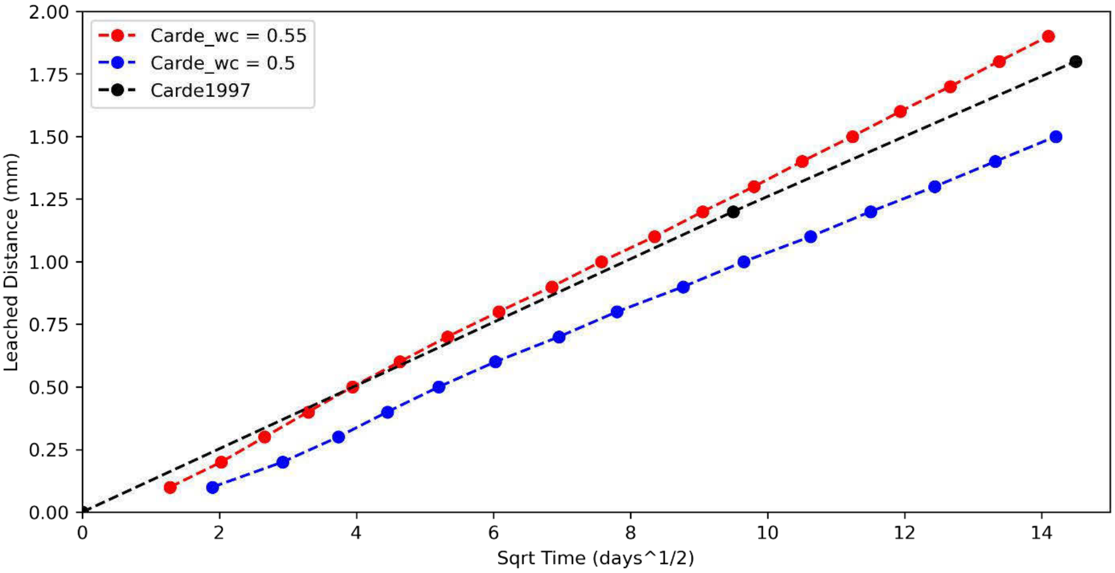


Figure 5. Comparison of the model outcomes to the experiments reported by 0. The red and blue curves indicate the modelled progression over time of the leached dept. The black markers indicate the reported leached depths during the experiments.

5 ACKNOWLEDGEMENTS

This research was funded by the BTO joined research programme of the Dutch drinking water utilities under project number 402045-211. The authors thank Frank Schoonenberg, Mario Castro Gama, Menno van Leenen, Herman Smit and Jantinus Bruins for many discussions on the drinking water practice behind the subject.

6 REFERENCES

[1] P. Leroy, M.R. Schock, I. Wagner, and H. Holtschulte, Cement-based materials in Internal Corrosion of Water Distribution Systems, 1st ed., Denver: AWWA Research Foundation, 1996.

[2] J.R. Rossum and D.T. Merrill, An evaluation of the calcium carbonate saturation indexes, AWWA Journal, vol. 75, no. 2, February 1983, pp 95-100.

[3] P.J. de Moel, A.W.C. van der Helm, M. van Rijn, J.C. van Dijk, and W.G.J. van der Meer, Assessment of calculation methods for calcium carbonate saturation in drinking water for DIN 38404-10 compliance, Drinking Water Engineering and Science, vol. 6, November 2013, pp. 115-124.

[4] WHO, Guidelines for Drinking-water Quality, 4th ed., Geneva: WHO-Press, 2011.

[5] P. Torres-Lozada, K.A. Bueno-Zabala, L.G. Delgado-Cabrera, L.E. Barba-Ho, and C.H. Cruz-Velez, Corrosion control using hydroxide and bicarbonate alkalising agents in water drinking processes, Drinking Water Enginering and Science, vol. 8, March 2015, pp. 53-76.

[6] C. Carde, G. Escadeillas, and R. Francois, Use of ammonium nitrate solution to simulate and accelerate the leaching of cement pastes due to deionized water, Magazine of Concrete Research, vol. 49, no. 181, December 1997, pp.295-301.

[7] Jacques, D., Benchmarking of the Cement Model and Detrimental Chemical Reactions Including Temperature Dependent Parameters : Project Near Surface Disposal of Category A Waste at Dessel, report NIROND-TR 2009–30 E, Brussels: ONDRAF/NIRAS.

[8] C.A.J. Appelo, and D. Postma, D., Geochemistry, groundwater and pollution, 2nd ed., Leiden: Balkema Publishers, 2005.

[9] D.L. Parkhurst, and C.A.J. Appelo, Description of Input and Examples for PHREEQC Version 3 — A Computer Program for Speciation , Batch-Reaction , One-Dimensional Transport , and Inverse Geochemical Calculations. U.S., in Geological Survey Techniques and Methods, book 6, Florida: U.S. Geological Survey, 2013.

[10] B. Lothenbach, T. Matschei, G. Möschner, and F.P. Glasser, Thermodynamic modelling of the effect of temperature on the hydration and porosity of Portland cement. Cement and Concrete Research, vol. 38, no. 1, 2008, pp. 1–18.

[11] B. Lothenbach, and F. Winnefeld, Thermodynamic modelling of the hydration of Portland cement. Cement and Concrete Research, vol. 36, no. 2, 2006, pp. 209–226.

[12] S. Bejaoui, and B. Bary, B., Modeling of the link between microstructure and effective diffusivity of cement pastes using a simplified composite model. Cement and Concrete Research, vol. 37, no. 3, 2007, pp. 469–480.

[13] Eternit, Asbestcement Tabellen, product catalogus, Goor: Eternit, 1980.

[14] T. Powers, and T.L. Brownyard, Studies of the Physical Properties of Hardened Portland Cement Paste. ACI Journal Proceedings, vol. 43, no. 9, 1946.

[15] H.J.H. Brouwers, The work of Powers and Brownyard revisited: Part 1, Cement and Concrete Research, vol. 34, no. 9, 2004, pp. 1697–1716.

[16] R.A. Patel, J. Perko, D. Jacques, G. De Schutter, G. Ye, and K. Van Bruegel, Effective diffusivity of cement pastes from virtual microstructures: Role of gel porosity and capillary pore percolation, Construction and Building Materials, vol. 165, 2004, pp. 833–845.

[17] B.H. Oh, and S.Y. Jang, Prediction of diffusivity of concrete based on simple analytic equations. Cement and Concrete Research, vol. 34, 2004, pp. 463–480.

[18] A. Bentur, S. Mindess, and N. Banthia, N., The Interfacial Transition Zone in Fiber Reinforced Cement and Concrete State-of-the-Art Report of RILEM, In Engineering and Transport Properties of the Interfacial Transition Zone in Cementitious Composites.Paris: RILEM, 2004.

[19] E.J. Garboczi, and D.P. Bentz, Computer simulation of the diffusivity of cement-based materials. Journal of Materials Science, vol. 27, no. 8, 2004, pp. 2083–2092.

[20] Vitens, Vitens/phreeqpython: Object-oriented python wrapper for the VIPhreeqc module, https://github.com/Vitens/phreeqpython, May 2021.

[21] D. Jacques, J. Šimůnek, D. Mallants, J. Perko, and S.C. Seetharam, Evaluating changes of transport properties of chemically degrading concrete using a coupled reactive transport model, In 1st International Symposium on Cement-Based Materials for Nuclear Wastes, Avignon, France: Oct 2011.

[22] M. Moranville, S. Kamali, and E. Guillon, Physicochemical equilibria of cement-based materials in aggressive environments - Experiment and modelling, Cement and Concrete Research, vol. 34, no. 9, 2004, pp. 1569–1578.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14170

ON THE USE OF SINDY FOR WDN

Mario Castro-Gama[1]

[1] Specialist Infrastructure for Water Expertise Center (WEC) at Vitens N.V., Zwolle (The Netherlands)

[1] *Mario.Castrogama@Vitens.nl*

Abstract

With the growing interest of water utilities on digitalization, running multiple scenarios can become cumbersome with limited budget and short data collections. The total number of hydraulic simulations required (usually using commercial software), becomes a burden for near real-time operation. In order to circumvent the computational burden (limitation), since a couple of decades, several Machine Learning techniques have been used to create a meta-model or surrogates of a Water Distribution Networks (WDN) based on a subset of data available through SCADA. Among the many possible surrogates a Sparse Identification of Non-linear Dynamics (*SINDy*) method is presented. The method is applied to two datasets: i) to obtain a surrogate of a benchmark network and ii) real data of water consumption of different District Metered Areas (DMA) of a real water utility. The method is: i) computationally inexpensive, ii) less data demanding for calibration than other modern methods, iii) parsimonious, and iv) could be used to infer physical relations among data.

Keywords
Water demand, WDN, Surrogates, SINDy, DMD.

1 INTRODUCTION

Since a couple of decades ago there has been an increase on the number assets which are registered by utilities and used as elements of hydraulic models. Such models are then used to for real-time applications. With the development of Digital Twins [1] by utilities the number of scenarios and decisions which can be pursued by operators increases exponentially the requirement of hydraulic simulations. This creates a trade-off between model size and number of simulations which can be carried out to answer a specific question (i.e. leakage detection, anomaly detection, long-term and short-term planning, condition assessment). A way of circumventing this trade-off is the use of additional computer processing, parallelization of model runs, skeletonization or model simplification or the use surrogates or meta-models of the Water Distribution Network (WDN) model. Although pushing additional computer processing and capacity in data warehouses it's a possibility for some utilities, it is not sustainable in the long run. Within the meta-models category several applications for meta-modelling of WDN are available. Methods ranging from neural networks [2] [3], from simple types such as generalized and perceptron multilayers [4] [5], to more recent developments such as Deep learning networks (DNN) [6] are available. Most of these meta-models encapsulate a large amount of data (i.e. pipe flows, pressures or heads and demands) as black-box and their physical interpretation gets lost in the inner workings of these non-linear regressions. Since a few years ago some methods have been (re)discovered for the identification of principal modes from complex systems such as turbulent flows in the form of Koopman operators.

There is a large amount of such methods such as Principal Component Analysis (PCA), ERA, PDO, ICA, KIC, Dynamic Mode Decomposition (DMD) [7] [8] and Sparse Identification Non-Linear Dynamics (SINDy) [9] [10]. Here the latter and its possible applications for WDN are presented.

2 SPARSE IDENTIFICATION NON-LINEAR DYNAMICS - SINDY

SINDy states that given a set of measurement data $\{x(t)\}_{i \in I}$, it is possible to accurately learn a function $f(x(t))$ so that $\frac{dx}{dt} = f(x(t))$ is identified. Two assumptions are required, i) the full state measurements, and ii) that f only has a few active terms, (i.e. f is sparse) in the space of all possible functions of $x(t)$.

$$X = \begin{bmatrix} x^T(t_1) \\ x^T(t_2) \\ \vdots \\ x^T(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \cdots & x_n(t_m) \end{bmatrix} \tag{1}$$

The rate of change of X can be estimated using finite differences or total variation derivatives. Or simply by assuming an Euler update by taking the next time step as the outcome when dt is small.

$$\dot{X} = \begin{bmatrix} \dot{x}^T(t_1) \\ \dot{x}^T(t_2) \\ \vdots \\ \dot{x}^T(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \cdots & \dot{x}_n(t_m) \end{bmatrix} \tag{2}$$

In order to solve this the first step is to construct library Θ(X) of candidate nonlinear functions of X :

$$\Theta(\mathrm{X}) = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X & X^2 & X^3 & \cdots & X^p \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}$$

$$\text{i.e. for } p = 2,\ X^2 = \begin{bmatrix} x_1^2(t_1) & x_1(t_1)x_2(t_1) & \cdots & x_2^2(t_1) & \cdots & x_n^2(t_1) \\ x_1^2(t_2) & x_1(t_2)x_2(t_2) & \cdots & x_2^2(t_2) & \cdots & x_n^2(t_2) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_1^2(t_m) & x_1(t_m)x_2(t_m) & \cdots & x_2^2(t_m) & \cdots & x_n^2(t_m) \end{bmatrix} \tag{3}$$

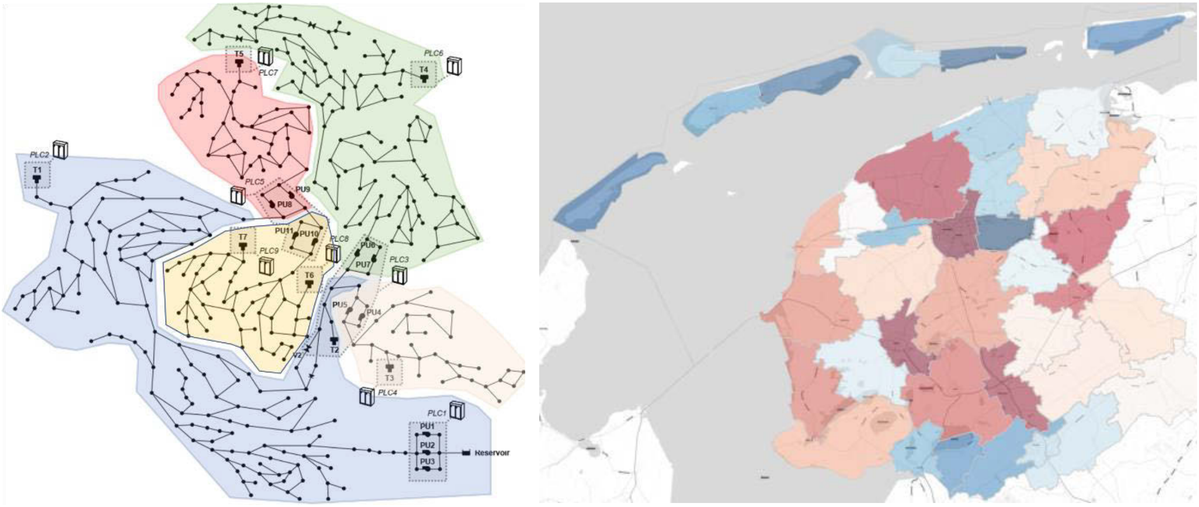
Then perform a sparse regression on $X = \Theta(X)\sum$. This is required to solve for all coefficients $\sum = [\sigma_1; \ldots ; \sigma_n]$, $\sigma_i \in R_p$. Then, let $\lambda > 0$ be the sparsity threshold and the following iterative procedure ensures that an sparse regression is obtained.

1. Initial guess: solve $X = \Theta(X)\sum$. via ordinary least squares
2. If $\sum(i,j) < \lambda$ set $\sum(i,j) = 0$
3. For $k = 1, 2, \ldots, n$

   solve $\dot{X}(:,\mathrm{k}) = \Theta(\mathrm{X})(:,\Sigma(:,\mathrm{k}) > \lambda)\Sigma(\Sigma(:,\mathrm{k}) > \lambda),\mathrm{k})$ via least squares
4. Repeat steps 2-3 until the coefficients do not change (or for a fixed number of iterations)

3 CASE STUDIES

Two different case studies are used. Firstly, is the simulated data from a benchmark WDN known as C-Town. Secondly, the water balance data of a province located in the northern part of the Netherlands and operated by Vitens.



(a) C-Town- with 5 DMA's. Shows also selected monitoring locations for tanks and pumping stations.

(b) Areas of water balance (large DMA's) in a province of the Netherlands operated by Vitens N.V.

Figure 1. Case studies

3.1 C-Town

It corresponds to a small WDN with 5 DMA's from which flows and pressures at particular locations can be fetched from the system. In this case study two different configurations are assumed (Fig 1a).

- First, a configuration in which all variables are observed. In this case a total of 724 variables is considered. Although it is unrealistic to collect all variables related to demands, pressures and flows within a WDN, the goal here is to determine whether or not SINDy is able to reconstruct both mass and energy balance according to the Global Gradient Algorithm (GGA) [11] [12], without prior knowledge of the equations. A dataset of 4 weeks (2688 timestamps every 15 minutes) is created.
- Second, a configuration of SINDy in which only a subset (43 variables) in the system are collected in the SCADA system for each DMA is presented. Variables which show no variation during the total length of the dataset where eliminated resulting in only 37 variables. The goal in this case is to be able to determine whether or not anomalies can be detected. Anomalies can represent multiple behaviours such as change of valve status, leakages, or even cyber-physical attacks. Here SINDy is compared to another surrogate. Two datasets are obtained from BATADAL (Battle of Attack Detection Algorithms) one of *normal* operation of the WDN and one *abnormal* (with anomalies). A SINDy model of the normal operation data is trained and subsequently tested on the abnormal data. The hypothesis is that SINDy is able to capture the system dynamics and will be able to identify the timestamps of anomalies as such.

3.2 Water balance areas of a province

Data collected from the last 4 full calendar years (2018-2021) of a northern province operated by Vitens are analysed. Data corresponds to the water balance in each of the Water Balance Areas (WBA) of the province. It is not possible to assess WBA as District Metered Areas (DMA's) due to the former being much larger. Fig. 1b, presents the localization of each of the WBA, while Fig. 2 represents the total water consumption pattern of the province within a 24 hour period at an hourly resolution. This water consumption is obtained by taking into account all the production locations of the province. The use of Automated Metered Readings (AMR's) is only available for a pilot area in the largest city of the province and for large customers, however given the low leakage percentage Fig. 2 is representative of the demand pattern. In this case the area is composed of 12 different DMA's where data is available between 2018 end 2021. The homogeneous period (Fig 2 below) where data is available for all variables is 11-Nov-2019 and 30-Oct-2020 (7840 timestamps) is highlighted.

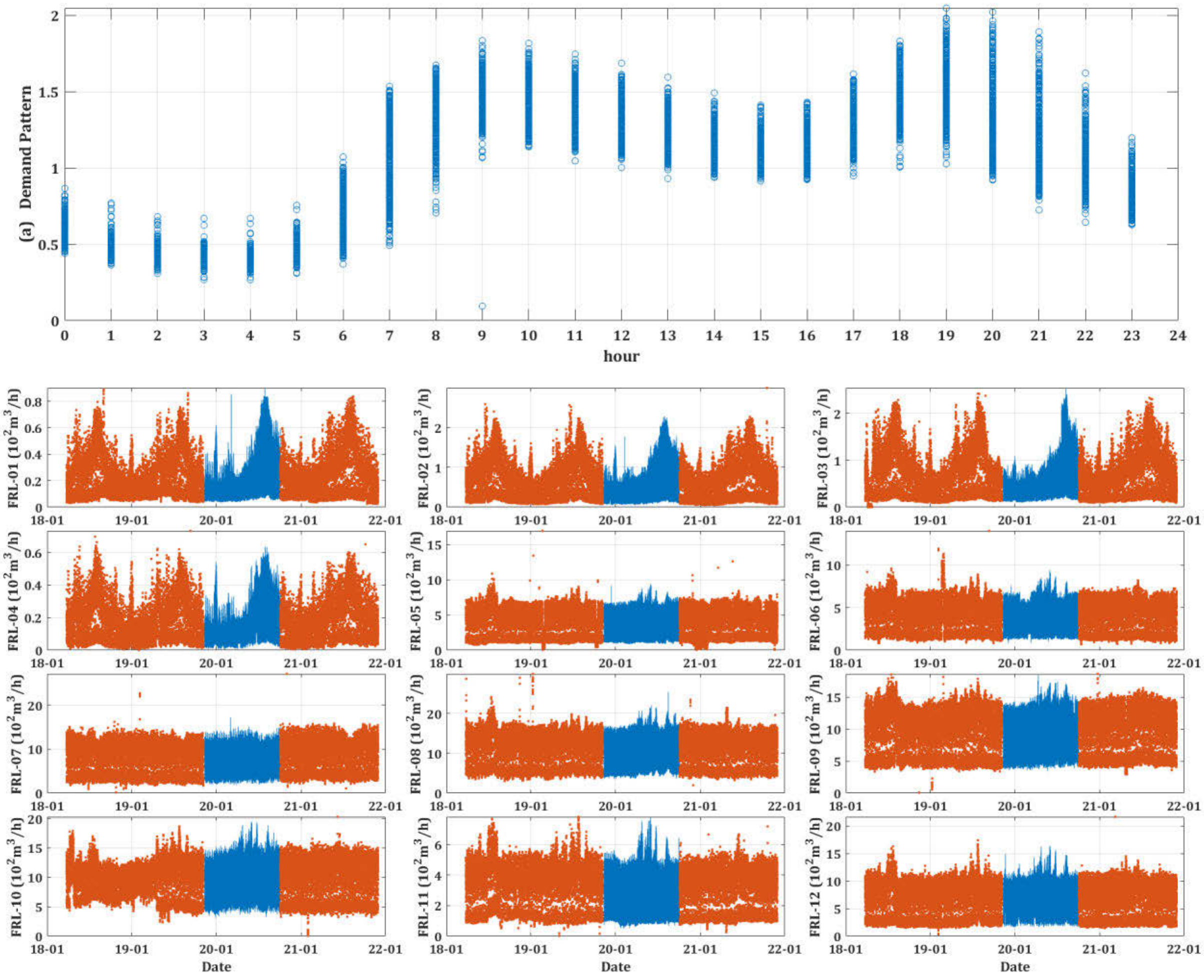


Figure 2. Water consumption province Period 2018-2021 in a North-Province of The Netherlands (a) daily water balance, (below) time series for each DMA's. Period in blue corresponds to a period of homogeneous data collection.

4 RESULTS

C-Town fully monitored

A fully monitored WDN implies installing volume meters, pressure sensors and AMR's on each location of the WDN. After training a SINDy model the results of RMSE for a full monitoring are presented in Fig. 3. Results have been ranked from higher to lower RMSE for pressure at nodes (A) and flows in links (B).

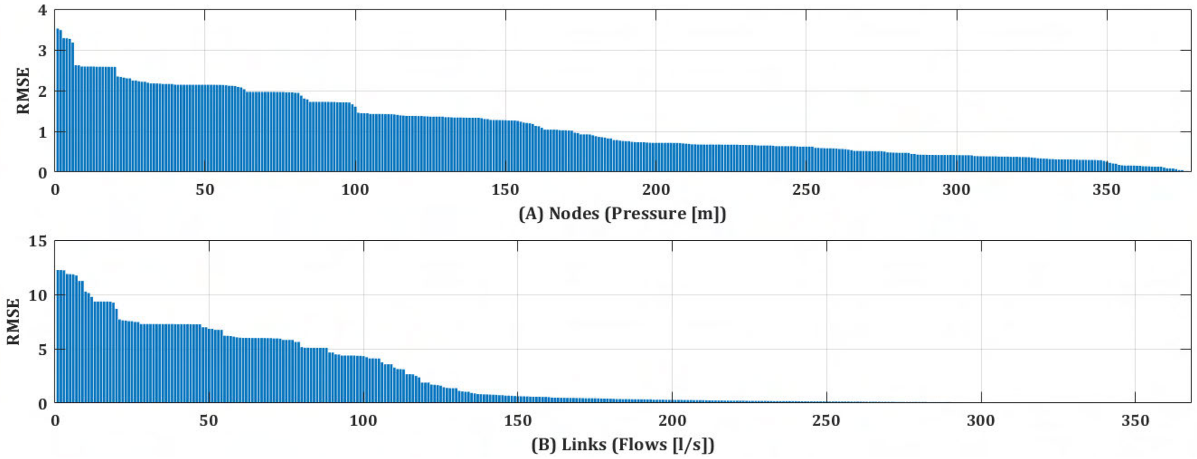


Figure 3. RMSE of all variables of C-Town. (A) Node pressures, (B) pipe flows.

This results indicate that SINDy is a suitable alternative for representation of a fully monitored WDN. To portrait the results obtained for each variable Fig. 4 presents the fitness of the time series for flow in pipe 11 (Q_{11}). This is a pipe where flows change direction throughout the simulation. The RMSE of this variable is high 11.86 l/s, however one can assess that most of the large errors occur during the change of trajectory of the flows in consecutive time stamps and the SINDy surrogate is able to return to the trend of the variable very fast. A similar behaviour of the application of SINDy has been obtained by other authors [4] for complex dynamical systems (such as Lorenz attractor). The errors are normally distributed as presented in Fig 4 (lower right).

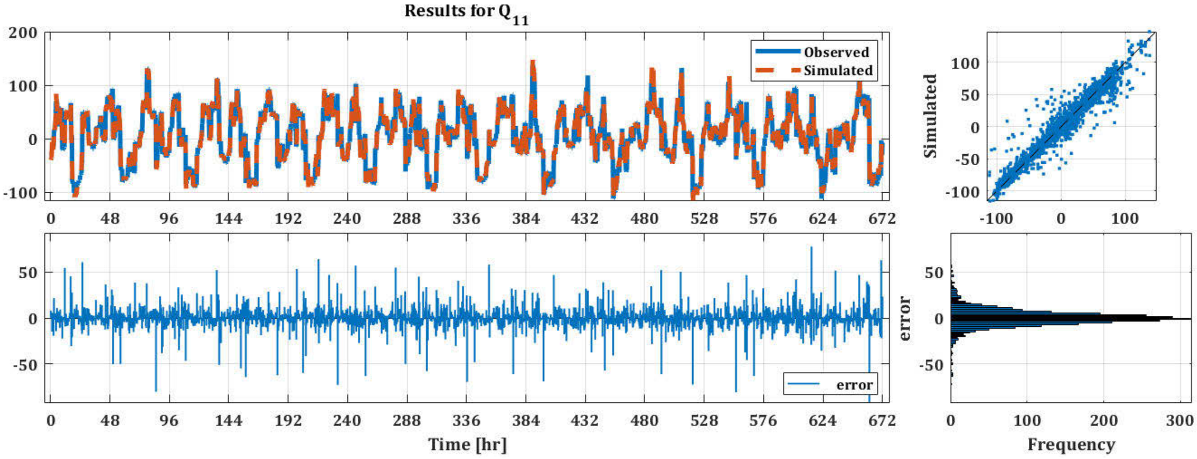


Figure 4. SINDy results for flows at pipe 11 (Q_{11}). Top left is the time series of Observed (EPANET) and Simulated (SINDy). Top right is the scatter of the time series Simulated vs Observed. Bottom left is the error obtained at each time stamp. Bottom right corresponds to the histogram of the errors obtained.

C-Town subset monitoring

On the second case, using a configuration with only 37 hydraulic variables the application of SINDy was not able to obtain similar results. The main issue on the formulation of the SINDy model is the fact that the data contains pump status as a variable. Such data was not used in the first case of analysis of the fully monitored network C-Town. Apparently, the inclusion of binary variables as independent variables in a SINDy model tends to create an overshoot in the behaviour of simulated dependent variables. Such behaviour is presented in Figure 5, where the time series shows that the status can have only values ϵ [0 or 1], while the estimation shows that the outcomes are real values in the range ϵ [-0.75, 1.30]. At this moment it is not known by the author whether or not there is a mechanism to handle binary variables within SINDy on the estimation of surrogates for WDN.

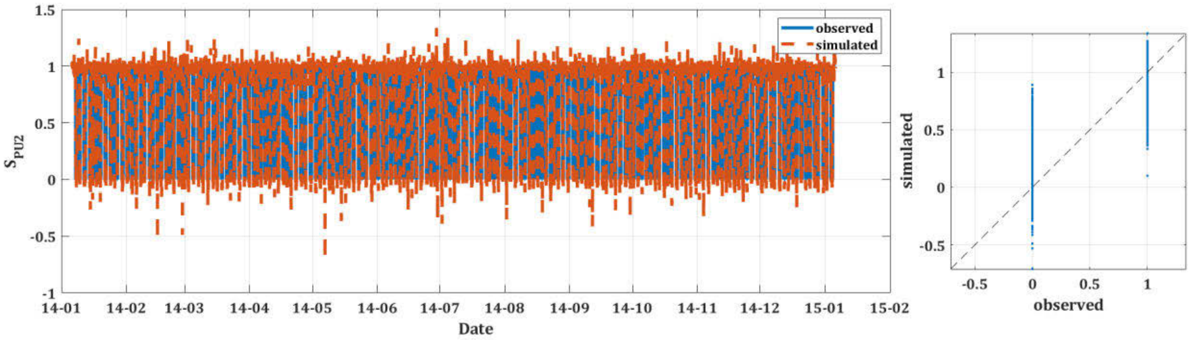


Figure 5. Results of time series for status of pump 2 and the corresponding error on the estimation.

Subsequently, the SINDy model obtained with normal data is applied to the data containing the anomalies (abnormal), however the results of the surrogate make it impossible to know whether the rapid variation of the time series is due to an event or due to the construct of the surrogate model itself.

Water balance of a province

Data of the water consumption is divided by 100 to obtain rescaled values. Subsequently a SINDy model is built. Lambda (λ) is used for sparsification and set as [0.01, 0.1, 1.0]. The maximum polynomial order is set to 2. The total number of variables considered in each case for Θ(X) is equal to 91 (i.e. order 0: 1; order 1: 12; and order 2: 78). Results of the RMSE for each value of λ are presented in Figure 6 where on the left are the sparse elements of $\Theta(X)^T$. Each row represents the coefficients which are active in the SINDy formulation for each DMA. Here *nnz* is equal to the total number of non-zero elements in each case. A higher value of λ will reduce the number of coefficients in Θ(X) which are non-zero from 954 to 131. On the right, the corresponding RMSE obtained after estimation of the consumption in each DMA. It needs to be mentioned that for values of $\lambda > 1.0$, the RMSEs increase continuously, while on the case that $\lambda < 0.01$ the additional gains on the error reduction are imperceptible. In addition, it is noticeable that the RMSEs are not linearly dependent on λ. This may lead to potentially select a different λ for the determination of the best surrogate of each area. Given the average consumption per area, it is expected that different areas with larger consumption will present larger errors.

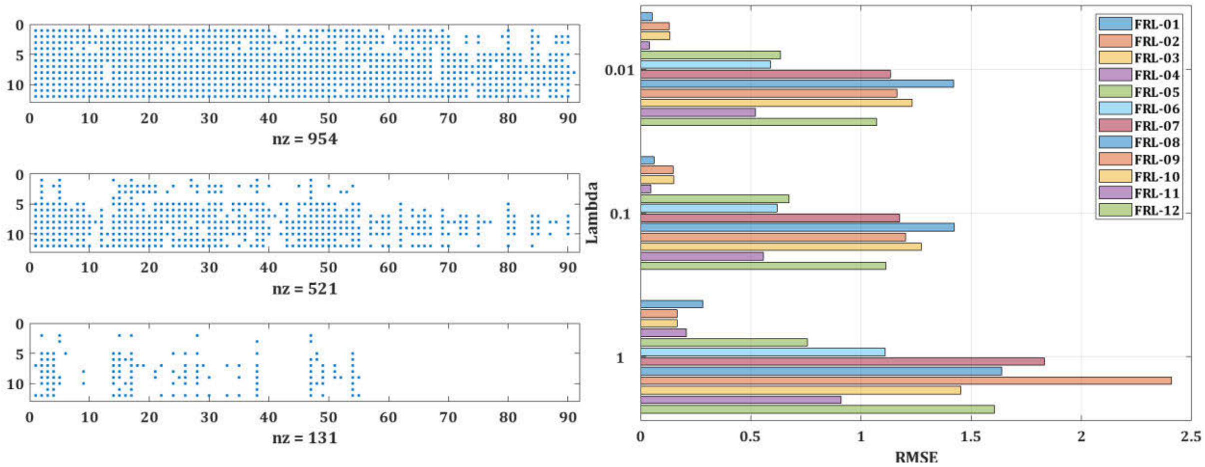


Figure 6. Left, matrix $\Theta(X)$ transposed. Right, the corresponding RMSE for each DMA.

Additional work is required to determine the minimum length of data required to generate comparable results. At this point 7,840 timestamps correspond to almost 1 year of continuous hourly data registration, this may not be possible for most utilities.

5 CONCLUSIONS

This article presents a new surrogate model for WDN. The method has been applied for both data of a benchmark WDN and for data of water consumption in a large province. In both cases SINDy was able to recreate the behaviour of the underlying system with low computational cost. The application for data of a fully monitored WDN shows the potential for the development of a very easy to setup surrogate model. Its application for a subset of monitoring variables of the same benchmark network were not able to reliably generate a surrogate model. In the case of the application of SINDy for the determination of a surrogate of water consumption in DMA's, once again the results show relatively good accuracy with respect to the observed values. Larger DMA's show larger RMSE and vice versa.

In addition, the possible application of the method as an anomaly detection algorithm for leakage detection or leakage localization are yet to be explored in a real system. Other aspect to consider as future work is the determination of the minimum length of the timeseries and the resolution required to build a trustworthy surrogate model.

6 REFERENCES

[1] F. Martinez-Alzamora, P. Conejos, M. E. Castro-Gama and I. Vertommen, "Digital Twins - A new paradigm for water supply and distribution networks," Hydrolink, pp. 48-54, 2021.

[2] Z. Rao and F. Alvarruiz, "Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model," Journal of Hydroinformatics, vol. 9, p. 15–24, January 2007.

[3] D. R. Broad, G. C. Dandy and H. R. Maier, "Water Distribution System Optimization Using Metamodels," Journal of Water Resources Planning and Management, vol. 131, p. 172–180, May 2005.

[4] F. Martínez, V. Hernández, J. M. Alonso, Z. Rao and S. Alvisi, "Optimizing the operation of the Valencia water-distribution network," Journal of Hydroinformatics, vol. 9, p. 65–78, January 2007.

[5] E. Salomons, A. Goryashko, U. Shamir, Z. Rao and S. Alvisi, "Optimizing the operation of the Haifa-A water-distribution network," Journal of Hydroinformatics, vol. 9, p. 51–64, January 2007.

[6] Z. Y. Wu and A. Rahman, "Optimized Deep Learning Framework for Water Distribution Data-Driven Modeling," Procedia Engineering, vol. 186, p. 261–268, 2017.

[7] N. Kutz, S. Brunton, J. Proctor and B. Brunton, Dynamic mode decomposition : data-driven modeling of complex systems, Philadelphia: Society for Industrial and Applied Mathematics, 2016.

[8] Z. Wu, S. L. Brunton and S. Revzen, "Challenges in Dynamic Mode Decomposition," September 2021.

[9] S. Brunton and N. Kutz, Data-driven science and engineering : machine learning, dynamical systems, and control, Cambridge, United Kingdom New York, NY: Cambridge University Press, 2019.

[10] S. L. Brunton, J. L. Proctor and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proceedings of the National Academy of Sciences, vol. 113, p. 3932–3937, March 2016.

[11] E. Todini and L. A. Rossman, "Unified Framework for Deriving Simultaneous Equation Algorithms for Water Distribution Networks," Journal of Hydraulic Engineering, vol. 139, p. 511–526, May 2013.

[12] E. Todini and S. Pilati, A Gradient Algorithm for the Analysis of Pipe Networks in: "Computer Applications in Water Supply. 1 - System Analysis and Simulation", B. Coulbeck and C. Orr, Eds., London, UK: Research Studies Press Ltd., 1988.

[13] B. Coulbeck, Computer applications in water supply, Letchworth, Hertfordshire, England New York: Research Studies Press Wiley, 1988.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14173

HOUSEHOLD BEHAVIOUR AND ENERGY LOSS IN INTERMITTENT WATER SUPPLY NETWORKS

Hamidreza Mohabbat[1] and David D. J. Meyer[2]

[1,2] Centre for Global Engineering, Department of Civil & Mineral Engineering, University of Toronto, Toronto, ON, Canada

[1] *H.Mohabbat@mail.utoronto.ca*, [2] *David.Meyer@utoronto.ca*

Abstract

Intermittent water supply (IWS) networks are a problematic reality for over one billion people. Despite their drawbacks, IWS networks persist, and projects to convert them to continuous supply often fail. Here we explore one reason such projects might fail: the energy loss associated with IWS. Under IWS, water is delivered over a shorter period, increasing flow rates and thereby increasing energy losses. Could this energy loss prevent utilities from increasing their supply durations?

To explore this question, we built two IWS versions of the Modena water network in EPANET. All households were assumed to withdraw water either i) as hastily as possible or ii) as patiently as possible. Artificial tanks and emitters modelled household storage and network leakage, respectively. Artificial tanks filled quickly, mimicking hasty withdrawals. To model patient withdrawals, a flow control valve was installed upstream of the tank, distributing withdrawals evenly throughout the duration of water supply.

Simulations showed that when households withdraw hastily, energy losses strictly increase as the supply continuity of the network increases. Conversely, when households withdraw water patiently, energy losses increase initially, reach at least one maximum, and then decrease as supply continuity increases. Our results suggest that since energy losses often increase as utilities increase continuity, energy loss could obstruct some utilities from increasing supply continuity and from achieving continuous supply. We also found clear evidence that network behaviour strongly depends on the hastiness of household withdrawals. We also found that when networks with patient households are supplied with ample continuity, leakage can substantially influence the energy loss. We recommend additional theoretical and field research on IWS investigate the pace at which household withdrawals occur.

Keywords
Intermittent water supplies, water distribution networks, Global South, energy loss, artificial string, EPANET.

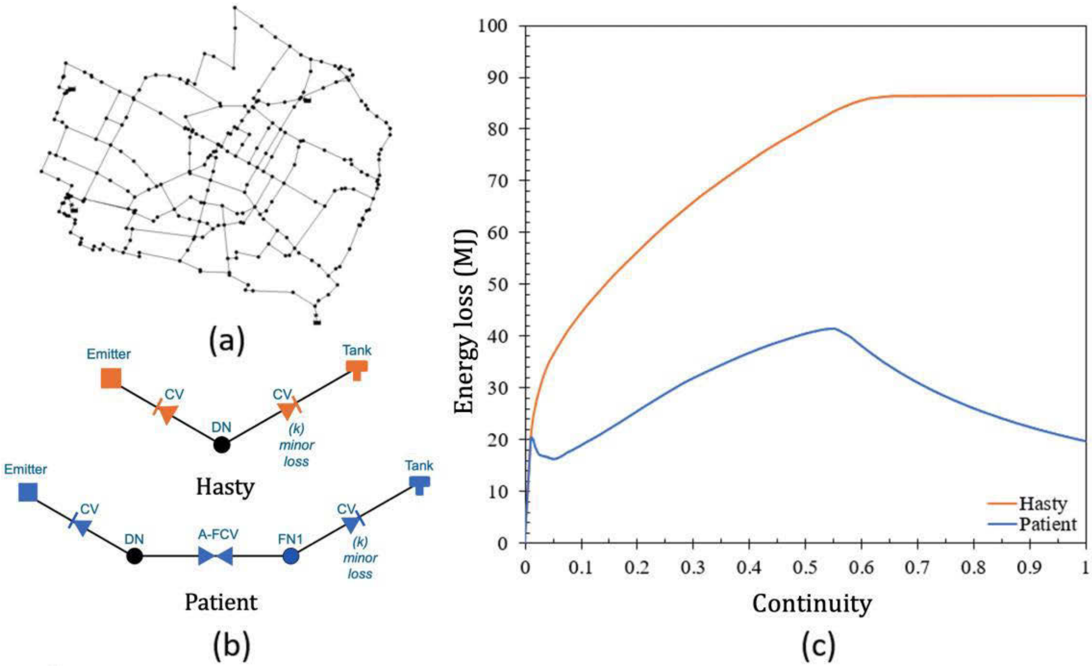


Figure 1 Visual Abstract, (a) Modena network before modification, (b) Artificial strings representing hasty and patient households in EPANET, (c) Energy loss versus water supply continuity in the modified Modena network, with hasty and patient households.

1 INTRODUCTION

Intermittent water supply (IWS) networks supply water for less than 24 hours per day, seven days a week [1]–[3]. Such networks serve over one billion people [4]. Many utilities resort to IWS when faced with water shortages [4] since intermittent operation can temporarily suppress leakage and consumer withdrawals [5]–[8]. Unfortunately, however, IWS worsens water quality and equality and complicates network hydraulics.

Given the drawbacks of IWS, many policies promote continuous water supply, or at least increasingly continuous supply (e.g., India launched new national 24x7 water policies and guidelines in 2021 and 2022 [9], [10]). But efforts to transition intermittent networks towards continuous supply often fail for varied and poorly understood reasons [11], [12]. To explain the persistence of IWS networks, some have suggested insufficient source water or treatment capacity, poor network management, and/or unsustainably high consumption [2], [13]. But to date, little attention has been given to energy losses in IWS networks.

We hypothesize, that energy losses are an important factor preventing some IWS networks from increasing their supply continuity. Under IWS, flow rates are high, which increases friction and induces a high rate of power loss in the network [14]. As an intermittent utility tries to increase its supply, flow rates and therefore power losses will decrease. But energy losses depend on the rate and duration of power loss. So as an IWS increases its supply duration, we expect power loss to decrease, but will that decrease outpace the increase in duration?

To evaluate potential barriers to 24x7 supply, previous research has often deployed hydraulic models of IWS [15]–[17]. These models included a variety of assumptions about how households respond to IWS [15]–[17]. Field research in IWS has consistently confirmed that households store water in response to IWS [2], [4], [18], but provides little detail on the rate at which household storage is filled. Without details on the rate of water withdrawals, flow and energy loss cannot be modelled.

This paper aims to provide quantitative evidence about the degree to which energy losses in IWS may thwart utility attempts to increase water supply duration. To do so we:

1. Propose two types of household withdrawal behaviour and construct EPANET models representing them;
2. Quantify how energy loss varies across simulations of these models for different supply durations;
3. Evaluate the degree to which energy losses may help explain the persistent of IWS; and
4. Assess the influence of the network leakage on energy loss.

2 DEFINITIONS AND MODELS

2.1 Water Supply Continuity Definitions

IWS networks vary in how often (frequency) they are pressurized and for how long (duration). We track the frequency of supply (e.g., every other day) by its inverse: the supply period (T). We define the average length of time during which a network is supplied with water as the supply duration (τ). Finally, we combine these metrics together to define the *supply continuity* (c) as the average percent of the time that a network supplies water: $c = \tau/T$. For instance, a network that provides water for a duration 4 hours ($\tau = 4\ hrs$) every four days would have a supply continuity of $c = 4\ hrs/4 * 24\ hrs = 4.2\%$, equivalent to the continuity of a network that runs for one hour ($\tau = 1\ hr$) every day.

2.2 Household Behaviour Types

Taylor et al. (2019) suggested that households in an IWS desire a certain volume of water per day, and as soon as they withdraw (i.e., satisfy) this demanded volume, they cease withdrawing water [8]. Through EPANET-based simulations, they showed that if an IWS network runs long enough, household demand becomes satisfied, and the qualitative behaviour of the network changes. In this paper, we denote the average percent of the time that an IWS must operate to satisfy households who withdraw water as fast as they can as the *satisfaction continuity*, c_s. For convenience, we similarly denote the supply duration required to satisfy such households as the *satisfaction duration*, $\tau_s = c_s T$.

Beyond the existence of a satisfaction threshold, little is known about the time-dependent and continuity-dependent behaviour of households in IWS. To explore the effects of household behaviour, we construct two models for how households withdraw water over time by imagining two types of household behaviour: hasty and patient.

***Hasty households* withdraw water as fast as possible**; the received flow rate for a hasty household can be modelled as:

$$Q_{\text{Hasty}}(t) = \begin{cases} \dfrac{V_s T}{\tau_s} = \mathrm{Q}_{\max} & \text{for } t < \tau_s \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where V_s is the household's desired water demand per day and τ_s is the satisfaction duration – the minimum required supply duration for the network to satisfy a household who hastily withdraws water, e.g., by leaving all taps and tanks open [6]. The satisfaction duration, τ_s, depends on the supply period (T) and the satisfaction continuity (c_s), while c_s depends only on physical network attributes and the household satisfaction volume (V_s). Hasty households, by definition, withdraw water as fast as hydraulically possible and so we sometimes denote it as $\mathrm{Q}_{\max}$.

Patient households **spread their demand over the supply duration;** hence, the received flow rate for a patient household depends on the continuity of the network. When the supply continuity is greater than the satisfaction continuity, a patient household utilizes the entire supply duration (τ) to withdraw its desired satisfaction volume: the daily desired volume, V_s, times the period of supply, T. When $c < c_s$, a patient household's withdrawal rate is equal to a hasty household's: both are limited by network hydraulics:

$$Q_{\text{Patient}}\ (t) = \begin{cases} \dfrac{V_sT}{\tau_s} = \mathrm{Q}_{\max} & \text{for } c < c_s \\ \dfrac{V_sT}{\tau} & \text{Otherwise} \end{cases} = \frac{V_sT}{\max(\tau_s, \tau)} \tag{2}$$

To explore the difference between these behaviours, consider a household supplied with water daily ($T = 1$), with a satisfaction continuity of 25% and in a network with a supply duration of 18 hours ($\tau = 18$ hrs, so $c = 75\%$). In this example, the satisfaction duration is $\tau_s = c_sT = 25\% * 24$ hours = 6 hors. During the first 6 hours of supply in this example, a hasty household would receive a flow rate of V_sT/τ_s; thereafter (the remainder of the supply duration, 12 hours) the hasty household will neither demand nor receive any water. Contrastingly, a patient household would utilize the entire supply duration (18 hours) to withdraw its desired volume V_sT. In this example both household types are ultimately satisfied and the initial flow rate to a hasty household is three times larger than to a patient household (Figure 2).

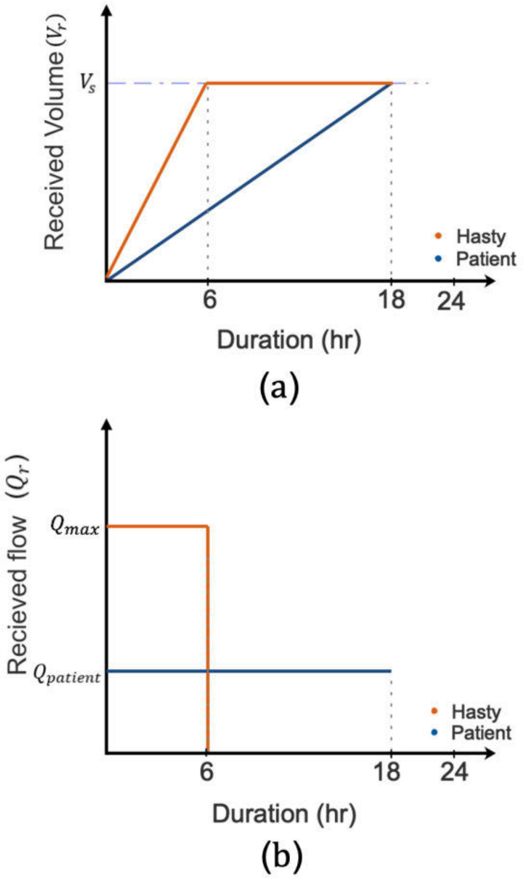


Figure 2 An example of how household type (hasty vs. patient) changes the received volume (a) and flow rate (b) over time in a network with daily supply of 18 hours/day and in which hasty households can be satisfied in only 6 hours.

If the network had operated for less than the required satisfaction continuity (c_s), neither of the households could have fulfilled their desired demand and both would withdraw water at maximum hydraulically feasible flow rate: $Q_{\max}$.

2.3 EPANET Construction

To explore the effect of hasty and patient households in a network, we converted a model of the Modena network (a continuous water supply, CWS) to behave like an intermittent one. Artificial tanks and emitters modelled household storage and network leakage, respectively. Artificial tanks filled quickly, mimicking hasty households. To model patient households, we installed artificial flow control valves upstream of the household tanks to distribute household withdrawals evenly throughout the supply duration.

2.3.1 Hasty Households in EPANET

Our EPANET model of hasty households is based on a method proposed by Batterman and Macke (2001), sometimes called the simple tank method (STM) [19]. Following the STM methodology, a CWS model can be modified to mimic an IWS by splitting demand at each of its nodes into hastily withdrawing households and leakage. Specifically, the STM adds two artificial strings per CWS demand node. Hasty households are represented by an artificial storage tank. Network leakage is represented by an artificial emitter. Both strings also include check valves to prevent reverse flow. The final composition of the strings is depicted in Figure 3.

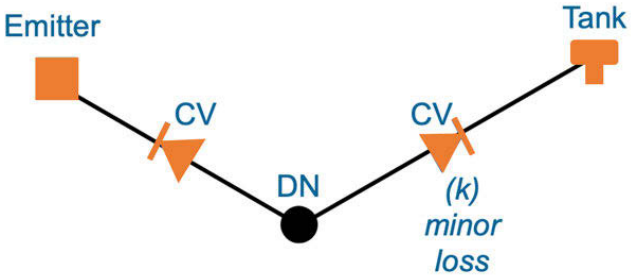


Figure 3 Schematic of the strings representing leakage and hasty households in EPANET, including the original demand node (DN) and two check valves (CV) connecting to an emitter and tank.

The parameters of the artificial strings depend on an assumed daily demand of an averageIWS household (V_s) and a network-wide leakage fraction [8]. The original demand of each CWS node is then split into leakage using an emitter and a tank-based representation of *N* hasty households, where *N* is determined based by the ratio of V_s to the demand at the CWS node (after subtracting leakage) [11]. By further assuming the hydraulic properties of a typical hasty household's service connection pipe and tank, the equivalent hydraulic attributes of *N* hasty households can be computed using formulations proposed by Batterman and Macke (2001) and summarized with increased specificity in Table A-2 [16]. The volume of each artificial tank is set to represent the combined demand of all *N* households represented by the original CWS node (detailed equations in Table A-2). The volume of these tanks corresponds to the households' combined daily demand, not the volume of their physical storage tanks.

Leakage is represented by an artificial string connecting an emitter to the original demand node (DN) via a check valve (Figure 3). The pipe is assumed to have a length of 1m, a roughness of 130 (Hazen-Williams), no minor losses, and a diameter equal to the average of connecting pipes at DN. The leakage flow through the emitter attached to demand node 'k' ($Q_{l,k}$) depends on the nodal pressure at the emitter (h_k), and the emitter discharge coefficient (C_k):

$$Q_{l,\mathrm{k}} = C_k h_k^{\alpha} \tag{3}$$

CC BY NC SA

where α is the emitter's pressure exponent in the absence of network data, we assume $\alpha = 1$ [20]. Here, we set the emitter coefficient by i) assuming leakage is a defined portion (β) of the original demand node in the leak-free CWS, and ii) assuming pressure head is 30 m. Hence:

$$C_k = \frac{Q_{l,k}}{h_k^{\alpha}} \approx \frac{\beta Q_{T,k}}{\overline{h_n}} \approx \frac{\beta Q_{T,\mathrm{k}}}{30} \tag{4}$$

where $\overline{h_n}$ is the average pressure head through the network, β is the leakage portion, and $Q_{T,k}$ is the original node demand in the CWS state, measured in m^3/s.

2.3.2 Patient Households in EPANET

Patient households usually satisfy their demands at a lower, more controlled flow rate compared to hasty households. To mimic this withdrawal patience, we added a flow control valve (FCV) upstream of the artificial tank, keeping all other elements equivalent to the STM (Figure 4).

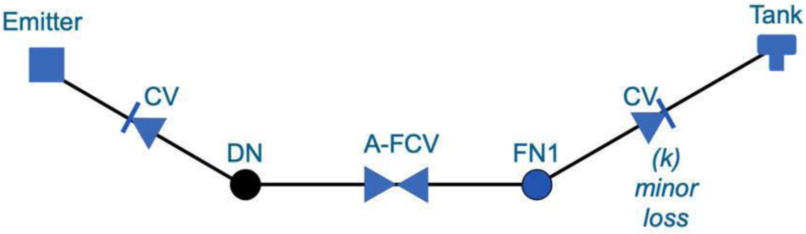


Figure 4 Schematic of the string representing leakage and patient households in EPANET, including the original demand node (DN), two check valves (CV), a fictitious node (FN) and an artificial flow control valve (A-FCV).

The flow setting for the described artificial valve is defined as:

$$Q_{FCV,k} = \frac{\overline{Q}_k T}{\tau} = \frac{0.85\, Q_{T,k} T}{\tau} \tag{5}$$

where $\overline{Q}_k$ is the household demand, here assumed as 85% of nodal demand in the leak-free CWS network ($Q_{T,k}$), and τ is the supply duration at which the network runs. As we are concerned with IWS networks that supply water at least occasionally, we assume $\tau>0$.

In our accounting of energy losses, we focus on energy lost in the network (rather than in households' taps). As such, we do not include the head loss values occurring inside the A-FCVs in our computation of total energy loss.

2.4 Modena Network

The unmodified Modena network comprises 268 demand nodes, 317 pipes and four reservoirs [21]. The spatial household demand distribution for the original leak-free Modena network is depicted in Figure 5. Its EPANET data are publicly accessible on the Water Distribution System Research Database of the University of Kentucky [22]. We adapt the Modena network to minimize inter-reservoir flows by changing the reservoir connection pipes into check valves. Using the Water Network Tool for Resilience (WNTR 0.4.0) package in Scientific Python Development Environment (Spyder 5.0.5), hasty and patient versions of the Modena networks were generated [23].

BY NC SA

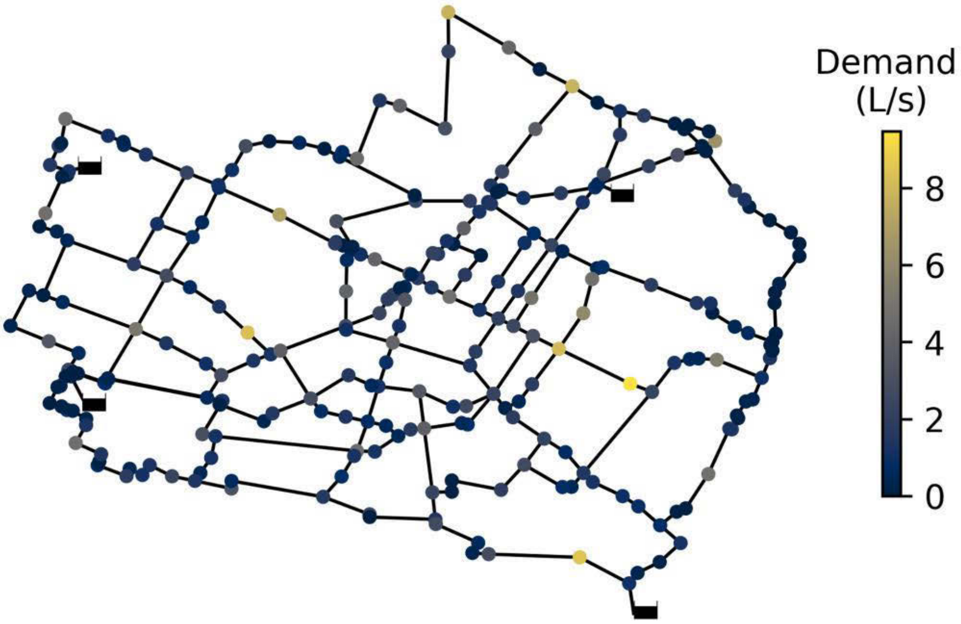


Figure 5 Spatial spread of Demand in Modena network (data originally published by [21])

2.5 Energy loss Calculation

To quantify the accumulated energy loss (E) during the supply duration (τ) in the hasty and patient Modena networks, energy loss was summed across all pipes (except the artificial flow control valves) and across all time steps:

$$E(\tau) = \gamma \sum_{t=0}^{\tau} \sum_{l=1}^{m} h_l Q_l \Delta t \tag{6}$$

where h_l, Q_l are the pipe-friction-induced head loss and flow rate in pipe l, Δt is the timestep (one minute), m is the number of pipes in the network, and τ is the supply duration (time during which the energy loss is accumulated). Note that E is the friction-induced energy loss and not the total energy supplied to the network.

The energy loss caused by the short (1 m) pipes connecting emitters to demand nodes is included in this sum, but its contribution is minimal (<0.001%).

3 RESULTS AND DISCUSSION

3.1 Energy loss vs. duration and continuity

To explore how household type affects energy loss, we first plot energy loss over time for two specific scenarios: when households in the hasty and patient Modena networks know the network will operate for 6 hours (Figure 6a) and for 18 hours (Figure 6b) per day, corresponding to continuity levels of 25% and 75%. The energy loss curves for the hasty Modena network under the continuity of 25% (Figure 6a) and 75% (Figure 6b) coincide because hasty households race to withdraw as fast as possible, regardless of the continuity, resulting in coincident and monotonically increasing energy losses. Contrastingly, the energy loss curves for the patient Modena network have slopes that decrease as the continuity of the network increases from 25% to 75% because patient households reduce their flow rates in response to having a greater supply continuity.

While energy losses always increase over time for a given supply continuity (e.g., Figure 6b), when the total energy loss is compared to continuity, more complex trends emerge. The total

BY NC SA

(accumulated) energy loss at the end of each supply is represented by the points $E_{H,A}$ and $E_{P,A}$ for the hasty and patient households in Figure 6a and $E_{H,B}$ and $E_{P,B}$ in Figure 6b. We aggregate points like $E_{H,A}$ and $E_{P,A}$ across a range of continuity levels to plot the relationship between total energy loss and continuity (Figure 6c).

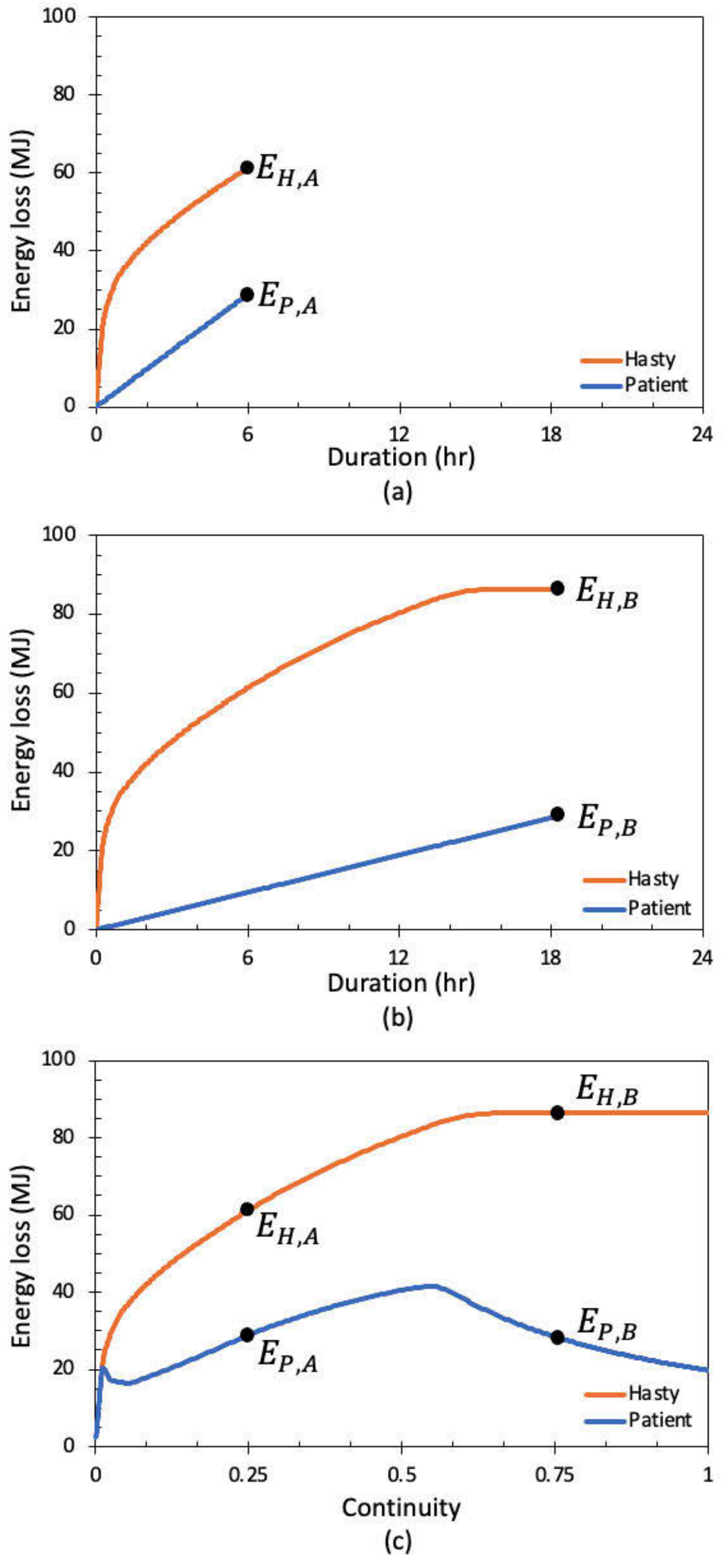


Figure 6 a, b: Cumulative energy loss against elapsed duration (time) in hasty and patient Modena networks when the network runs for 6 hours (a) and 18 hours (b) every day. c: Energy loss vs supply continuity for hasty and patient Modena networks. While energy loss increases monotonically with time for any given supply duration, patient households reduce flow rates at increased continuity levels, creating local maxima in the energy loss vs. continuity curve.

In both hasty and patient Modena networks, increasing continuity is associated with increasing energy losses when initial continuity values are <55% (Figure 6c). If the hasty Modena network began with 25% continuity, increasing to continuous supply (100% continuity) would increase total energy losses by 40% (Figure 6c). Similarly, if the patient Modena network increased from c=25%, energy loss would initially increase to a maximum of about 40% of their original value at their peak (at continuity≈55%; Figure 6c). However, after that peak energy loss, flow rates and energy losses would decrease with increasing continuity, ultimately reaching 30% less energy loss at c=100% in patient Modena as compared to c=25% (Figure 6c).

At any continuity, hasty households cause more energy loss than patient ones. Under continuous supply, hasty households induce 4.4x more energy losses than their patient counterparts. For hasty households, continuous water supply is the continuity with the highest energy losses. Contrastingly, the highest energy losses for patient households occurred when continuity was 55%, corresponding to the minimum continuity required to satisfy all households' demands.

Trends in energy loss versus continuity depend strongly on the assumption about how households will withdraw water (Figure 6). For hasty intermittent networks, monotonic increases in energy loss with continuity support the hypothesis that energy loss may be one barrier to increased (or continuous) supply continuity. Similarly, the existence of a peak energy loss in patient intermittent networks suggests where continuity is below this peak, energy losses may also be a barrier to increased supply continuity. Promisingly, for patient intermittent networks, the high energy losses associated with increased supply continuity are only temporary and the energy loss under CWS may be substantially lower than its value under IWS.

3.2 Effects of leakage on Energy loss

IWS networks are often described as having a high rate of leakage and a high rate of leakage is a commonly hypothesized barrier to increased supply continuity [18]. To explore the interaction of leakage and energy loss, we varied the assumed leakage share in our intermittent models (β in Equation (4)) from 5% to 45%, while holding demand constant (thereby varying the delivered water under continuous supply).

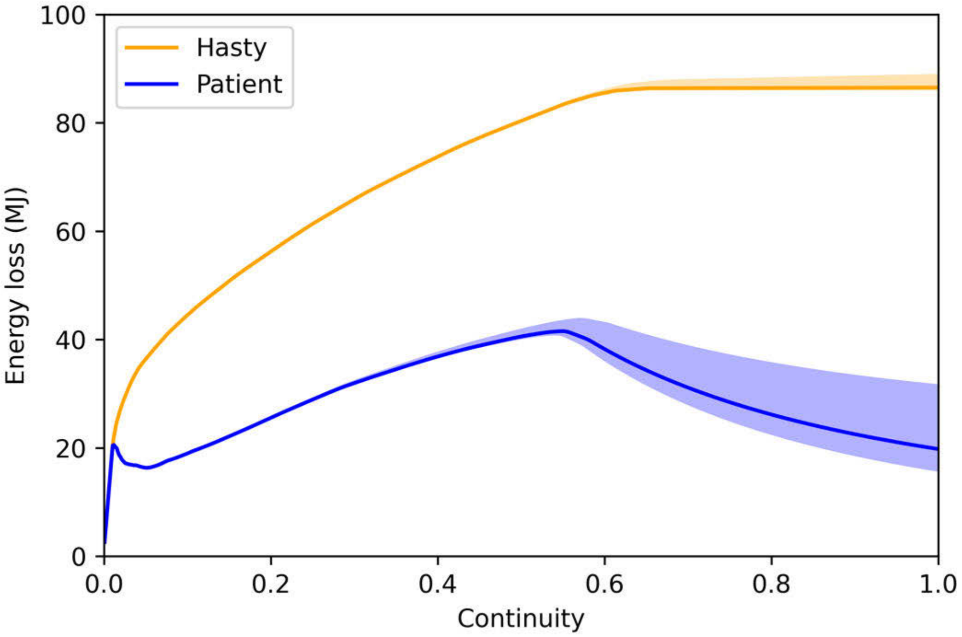


Figure 7 Energy loss vs. continuity plot in the intermittentized Modena network for hasty and patient networks under the range of leakage shares of 5% to 45% (the shaded areas) and 15% (as solid lines)

In both patient and hasty Modena networks, even a 9x increase in leakage coefficients had a much smaller effect on energy loss than changes in continuity (Figure 7). In the hasty Modena network, the 9x change in leakage coefficients changed the total energy loss by <3.6%. Conversely, leakage had a larger relative effect on energy loss in the patient Modena network (up to 103% increase) and was most notable when continuity levels were large enough to satisfy all households (>55%; Figure 7). While leakage may prevent some IWS from increasing their supply continuity, our results suggest that this barrier is not substantially compounded by leakage-induced energy losses.

4 CONCLUSION

Intermittent water supply networks persist despite their many drawbacks to utilities and households alike. This paper tested the hypothesis that energy loss might contribute to the persistence of IWS networks. To test this hypothesis, we imagined two extremes of household behaviour (hasty vs. patient) and implemented each in EPANET.

We found that network behaviour was drastically different when household withdrawals were hasty versus patient. Additional field research is needed (qualitative and/or quantitative) to help understand the rate at which consumers withdraw water from IWS and if it depends on the supply continuity. Are households in intermittent supplies predominantly hasty, patient, or something else entirely?

Our simulations of the hasty Modena network suggest that if every household behaved hastily, the energy loss would strictly increase with continuity, and energy loss could be one substantial reason why IWS are so difficult to convert to continuous water supplies. Conversely, energy loss in the patient Modena network initially increased, but once continuity was sufficient to satisfy all households, energy losses decreased with continuity. This may suggest that even where households withdraw patiently, energy losses may prove a barrier to increased continuity until continuity is long enough to satisfy all households.

At every supply continuity, energy losses were higher when households were hasty rather than patient. Where energy losses are limiting, slowing household withdrawals may prove key to achieving increased or continuous water supply. From the utility perspective, while limiting household withdrawal flows directly is not likely feasible, our results suggest that incentives and programs to reduce household withdrawal rates may prove highly beneficial.

5 APPENDIX

5.1. Household Connection Pipe Characteristics

The characteristics of the pipe connecting the storage tank to the demand node is determined based on the number, N, of hasty households represented by the node and based on the attributes of the pipe connecting a given hasty household to the network. Here, each household is assumed to have a demand of 1 m^3/day, which allows N to be calculated. Hence, the pipe characteristics are found using the equations in Table A-1.

Table A-2 Formulae to determine the connection characteristics for an artificial tank representing N households (taken from [16])

Connection Characteristic	**Equivalent connection**
Satisfaction consumption (m^3/day)	$\overline{Q_j} = V_s N = N$
Supply frequency ($1/day$)	$1/\mathrm{T} = 1$
Tank volume (m^3)	$V_j = V_s NT = N$
Tank height (m)	$H_j = H_i = 1$
Tank diameter (m)	$D_j = \sqrt{4V_i N}/\sqrt{\pi H_i}$
Length (m)	$L_j = 10$
Pipe friction, Hazen-Williams	$C_j = 130$
Pipe diameter (mm)	$d_j = d_i N^{0.380} = 15N^{0.380}$
Local loss coefficient	$k_j = k_j N^{-0.479} = 8N^{-0.479}$

Note: Subscript i denotes properties of an individual household, while j denotes the equivalent properties when N households are combined.

6 REFERENCES

[1] E. Kumpel and K. L. Nelson, "Intermittent Water Supply: Prevalence, Practice, and Microbial Water Quality," Environ. Sci. Technol., vol. 50, no. 2, pp. 542–553, Jan. 2016, doi: 10.1021/acs.est.5b03973.

[2] S. Galaitsi, R. Russell, A. Bishara, J. Durant, J. Bogle, and A. Huber-Lee, "Intermittent Domestic Water Supply: A Critical Review and Analysis of Causal-Consequential Pathways," Water, vol. 8, no. 7, p. 274, Jun. 2016, doi: 10.3390/w8070274.

[3] K. Simukonda, R. Farmani, and D. Butler, "Intermittent water supply systems: causal factors, problems and solution options," Urban Water Journal, vol. 15, no. 5, pp. 488–500, May 2018, doi: 10.1080/1573062X.2018.1483522.

[4] A. W. Bivins et al., "Estimating Infection Risks and the Global Burden of Diarrheal Disease Attributable to Intermittent Water Supply Using QMRA," ACS Publications, Jun. 20, 2017. https://pubs.acs.org/doi/full/10.1021/acs.est.7b01014 (accessed May 05, 2022).

[5] K. Vairavamoorthy, E. Akinpelu, Z. Lin, and M. Ali, "Design of Sustainable Water Distribution Systems in Developing Countries," pp. 1–10, Apr. 2012, doi: 10.1061/40569(2001)378.

[6] D. D. J. Taylor, A. H. Slocum, and A. J. Whittle, "Demand Satisfaction as a Framework for Understanding Intermittent Water Supply Systems," Water Resources Research, vol. 55, no. 7, pp. 5217–5237, 2019, doi: https://doi.org/10.1029/2018WR024124.

[7] S. Wunderlich, S. St. George Freeman, L. Galindo, C. Brown, and E. Kumpel, “Optimizing Household Water Decisions for Managing Intermittent Water Supply in Mexico City,” Environ. Sci. Technol., vol. 55, no. 12, pp. 8371–8381, Jun. 2021, doi: 10.1021/acs.est.0c08390.

[8] L. Santos Pereira, I. Cordery, and I. Iacovides, Coping with Water Scarcity. Dordrecht: Springer Netherlands, 2009. doi: 10.1007/978-1-4020-9579-5.

[9] Ministry of Housing and Urban Affairs, “Atal Mission for Rejuvenation and Urban Transformation 2.0.” Government of lndia, Oct. 2021. [Online]. Available: https://mohua.gov.in/upload/uploadfiles/files/AMRUT-Operational-Guidelines.pdf

[10] Ministry of Housing and Urban Affairs, “Guidelines for Planning, Design, and Implementation of 24x7 Water Supply Systems.” Government of lndia, Dec. 2021. [Online]. Available: https://mohua.gov.in/pdf/624eb498862a7Guidelines-for-Planning-Design-and-Implementation-of-24x7-Water-Supply-Systems.pdf

[11] S. P. Andey and P. S. Kelkar, “Performance of water distribution systems during intermittent versus continuous water supply,” Journal AWWA, vol. 99, no. 8, pp. 99–106, 2007, doi: 10.1002/j.1551-8833.2007.tb08011.x.

[12] L. Fan, G. Liu, F. Wang, C. J. Ritsema, and V. Geissen, “Domestic Water Consumption under Intermittent and Continuous Modes of Water Supply,” Water Resour Manage, vol. 28, no. 3, pp. 853–865, Feb. 2014, doi: 10.1007/s11269-014-0520-7.

[13] S. Hastak, P. Labhasetwar, P. Kundley, and R. Gupta, “Changing from intermittent to continuous water supply and its influence on service level benchmarks: a case study in the demonstration zone of Nagpur, India,” Urban Water Journal, vol. 14, no. 7, pp. 768–772, Aug. 2017, doi: 10.1080/1573062X.2016.1240808.

[14] E. Cabrera, E. Gomez, R. del Teso, and E. Estruch, “Energy Metrics to Assess the Efficiency of Pressurized Water Transport Systems:,” in WDSA / CCWI Joint Conference Proceedings, Jul. 2018, vol. 1. Accessed: Oct. 29, 2021. [Online]. Available: https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12069

[15] P. Sivakumar, N. B. Gorev, T. T. Tanyimboh, I. F. Kodzhespirova, C. R. Suribabu, and T. R. Neelakantan, “Dynamic Pressure-Dependent Simulation of Water Distribution Networks Considering Volume-Driven Demands Based on Noniterative Application of EPANET 2,” Journal of Water Resources Planning and Management, vol. 146, no. 6, p. 06020005, Jun. 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001220.

[16] D. Meyer, M. He, and J. Gibson, “Discussion of ‘Dynamic Pressure-Dependent Simulation of Water Distribution Networks Considering Volume-Driven Demands Based on Noniterative Application of EPANET 2’ by P. Sivakumar, Nikolai B. Gorev, Tiku T. Tanyimboh, Inna F. Kodzhespirova, C. R. Suribabu, and T. R. Neelakantan,” Journal of Water Resources Planning and Management, vol. 147, no. 8, p. 07021009, Aug. 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001428.

[17] H. Mohabbat and D. D. J. Meyer, “Energy Lost to Friction During Intermittent Water Supply,” in The virtual conference of AQUA≈360: Water for All - Emerging Issues and Innovations, United Kingdom, Aug. 2021, pp. 177–178.

[18] C. I. Gonzalez, J. Erickson, K. A. Chavarría, K. L. Nelson, and A. Goodridge, “Household stored water quality in an intermittent water supply network in Panama,” Journal of Water, Sanitation and Hygiene for Development, vol. 10, no. 2, pp. 298–308, Jun. 2020, doi: 10.2166/washdev.2020.156.

[19] A. Batterman and S. Macke, “A strategy to reduce technical water losses for intermittent water supply systems,” Ph.D. Thesis, 2001.

[20] L. A. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, EPANET 2.2 User Manual. Washington, US: U.S. Environmental Protection Agency, 2020.

[21] C. Bragalli, C. D’Ambrosio, J. Lee, A. Lodi, and P. Toth, “On the optimal design of water distribution networks: a practical MINLP approach,” Optimization and Engineering, vol. 13, no. 2, pp. 219–246, Jun. 2012, doi: 10.1007/s11081-011-9141-7.

[22] Kentucky Water Resources Research Institute, University of Kentucky, “Water Distribution System Research Database.” University of Kentucky Libraries. doi: 10.13023/KWRRI.WDSRD.

[23] K. Klise et al., Water Network Tool for Resilience WNTR User Manual. U.S. Environmental Protection Agency, 2020.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14176

ADVANCES IN PREMISE PLUMBING MODELING

William Platten[1], Regan Murray[2], Juneseok Lee[3], Robert Janke[4], Terra Haxton[5], Walter Grayman[6], Jonathan Burkhardt[7] and Steven Buchberger[8]

[1,2,4,5,7] U.S. Environmental Protection Agency, Cincinnati, OH, USA 45268
[3]Manhattan College, Riverdale, NY, USA 10471
[6]Walter M. Grayman Consulting Engineers, Oakland, CA, USA 94611
[8]University of Cincinnati, Cincinnati, OH, USA 45221

[8] *Steven.Buchberger@uc.edu*

Abstract

In summer 2019, a new task committee on Premise Plumbing Modeling (PPM) was approved under the auspices of the Water Distribution Systems Analysis Committee of ASCE EWRI. The primary purpose of the PPM task committee is to advance the science of the new field of premise plumbing modeling. In particular, the PPM committee intends to identify areas where methods and models developed over the past several decades to manage municipal water distribution systems can also be applied to solve vexing problems that arise in the premise plumbing systems of contemporary buildings. The ultimate goal of the PPM committee is to identify and develop water systems management tools for use by practicing engineers, as well as the larger water distribution research community, in support of the safe design and operation of indoor water distribution systems. Initially, three general areas were identified where current modeling approaches could be applied to premise plumbing: [1] hydraulic design and performance, [2] building water quality and age, and [3] water safety and security. Using these three core areas as a backdrop, this presentation will describe key accomplishments of the PPM task committee including: creation of a monthly webinar series reaching an international audience, distribution of a quarterly newsletter to the professional community, presentation of technical sessions at annual EWRI national Congress, and publication of a textbook style final report to archive the main topics covered in the PPM webinar series. Finally, this talk will conclude with a glimpse of the challenging problems posed by modern premise plumbing systems and the unique opportunities waiting for researchers working in the field of water distribution network analysis.

Keywords
ASCE Task Committee, premise plumbing, modelling.

1 BACKGROUND

Widespread provision of safe reliable drinking water to the general population ranks as one of the great engineering achievements of the 20th century [1]. A key component of contemporary water supply infrastructure is the “premise plumbing system”, representing the final leg of the journey in the delivery of water to the customer. Premise plumbing systems are an integral part of all modern buildings and encompass the complex collection of pipes, meters, heaters, tanks, valves, fittings, fixtures, and other appurtenances used for indoor water supply, heating, and sanitation.

Over the past half century, civil and environmental engineers have advanced the field of water distribution, with a focus mainly on municipal pipe networks. Current water distribution systems analysis generally stops at the curb. Historically, conventional analysis has not ventured into the building. Although some drinking water regulations extend to the point of consumption, there has been relatively little work related to the movement and transformation of water from the curb to the tap. Research is needed to bridge this gap and advance the science of water quality from source to tap. While similar in function, there are important operational distinctions between municipal water systems and premise plumbing systems (see Table 1).

Table 1. Some Salient Differences Between Water Distribution Systems and Premise Plumbing Systems.

Network Feature	Water Distribution System	Premise Plumbing System
Responsibility	Public water utility	Private building owner
Network topology	Looped pipes with some stubs	Branching pipes with many stubs
Pipe location	Below ground (in the soil)	Above ground (in the walls)
Flow regime	Mainly turbulent (except dead end)	Mixture of lamina and turbulent
Demands	Continuous random stream	Intermittent random pulses
Stagnation	Relatively rare	Very common
Design flow	Fire demand	99th percentile of peak period
Temperature	Ambient conditions	Hot and cold-water systems
Fixtures	Industry standards	Wide assortment of appliances
Modeling software	EPANET or similar	No comparable software

Three areas where established tools and techniques from analysis of municipal pipe networks could be applied to improve our understanding of premise plumbing systems are: (i) water quality within buildings—specifically, formation and delivery of opportunistic premise plumbing pathogens (OPPPs) and corrosion products; (ii) hydraulic design for improved efficiency, cost effectiveness and resource conservation; and (iii) water security.

1.1 Water quality

The United States has 140 million residential homes and nearly six million commercial buildings [2]. Maintaining a high standard of water quality at building fixtures is critical to the public health and safety of citizens. Water quality within buildings is a complex issue driven by plumbing materials, layout and usage patterns. Opportunistic premise plumbing pathogens that contaminate premise plumbing systems continue to cause outbreaks of waterborne illness such as *Legionnaires* disease, with serious consequences for human health. This is a problem particularly for systems with low-flow or infrequently used fixtures, and/or in-building storage tanks. Corrosion products also impact water quality during use, and reduction of exposure to lead within homes is a global priority. To minimize public health and safety issues, water system operators and managers recommend flushing contaminated water or decontaminating affected systems by installing in-building treatment facilities as reactive measures. However, *as yet there are no comprehensive information guidelines or models to help building managers, plumbing designers and health officials select and implement the optimum proactive and preventative actions in the facilities for which they are responsible.*

1.2 Hydraulic design

Recent efforts to improve water conservation/water efficiency and to comply with the requirements of the Energy Policy Act of 1992 follow recommendations made by the US Green

Building Council and the US EPA Water Sense initiative [3]. As a consequence, there has been a proliferation of low-flow premise plumbing water systems across the nation. While promoting conservation, such programs unintentionally raise the age of the water in the plumbing system, which can represent an emerging human health concern as OPPPs thrive under such low-flow/high water age conditions. Plumbing systems are not "right sized" to deal with this decrease in water demand, so water can dwell in pipes for lengthy periods before it is used with low or no disinfectant residual. To make matters worse, our existing water infrastructure is beginning to show its age and is starting to deteriorate, making it an ideal environment for OPPPs and corrosion to proliferate. It is critical to develop a better understanding of the potential impact of complex emerging situations on public health that will enable us to predict health risks that are likely to accompany continuing mandated efforts to conserve water in our drinking water systems.

1.3 System security

The security of our public water systems has long been a subject of concern. This was magnified in the aftermath of the September 11, 2001 attacks. Water utilities were required to perform vulnerability studies and premise plumbing was widely identified as a potential source for contaminant intrusion to the water supply system. Intrusions confined to a building would affect the customer directly within the particular premise; intrusions that escaped from the building (through pumpage back into the public system) could affect the entire downstream population. Several studies looked at this situation and applied simple standard distribution system models to the premise plumbing. Deficiencies in applying the general software under these situations were identified and included inadequate representation of dispersion, instantaneous demand patterns, pressure dependent demands, and some components (*e.g.,* hot water heaters, etc.). Whether a source of a contaminant is associated with the premise or came from the distribution system, the need to flush a system after contamination is important for the continued use of that building.

Generally, these problems have been approached from the narrow perspective of water chemistry/ microbial aspects in laboratory conditions or field sampling efforts. It is critical to approach these topics holistically because fate and transport of water contaminants are dictated by inherent hydraulics, demand, and systems characteristics within premise plumbing. This is further complicated by the lack of sampling protocols within premise plumbing regarding frequency, location, and timing. A modeling platform dedicated to premise plumbing issues has the potential to

- Improve our understanding of key phenomena that occur within premise plumbing systems,
- Provide a scientific basis for better planning, design, analysis, and operation of premise plumbing systems, and
- Identify new ways to optimize building water systems operations and management strategies.

2 PREMISE PLUMBING MODELING TASK COMMITTEE

To identify and address emerging challenges in the field of building water supply and to provide a forum for interested researchers in which to collaborate, a Premise Plumbing Modeling (PPM) Task Committee was launched during summer 2019 as part of the Water Distribution Systems Analysis (WDSA) committee. WDSA is a committee under the Environmental Water Resources Institute (EWRI) and is affiliated with the American Society of Civil Engineers (ASCE). EWRI has embraced that field through the Water Distribution System Analysis (WDSA) Committee, sponsorship of a track at their annual conferences, and publication of numerous papers in regular and special issues of the ASCE Journals for the past two decades.

2.1 PPM purpose and objectives

The purpose of the PPM task committee is to advance the science of the new field of premise plumbing modeling and to help develop building water systems management tools for use by practicing engineers, as well as the larger water distribution research community, in support of the safe design and operation of water distribution systems. The main objective is to develop and disseminate techniques and tools that can be used to improve planning, design, analysis, and operation of building water supply systems.

2.2 PPM research products

The Premise Plumbing Modeling committee identified several research milestones to achieve and end products to deliver over a three-year period, as highlighted in the list below:

- Sponsor technical sessions on premise plumbing and building water systems analysis at 2020-2022 EWRI Congresses and other venues.

- Organize on-line webinars with international experts to disseminate knowledge on the current state of building water systems modeling, highlight linkages to design, analysis, operational decision-making in practice, and identify gaps in knowledge, data or capabilities.

- Publish a final report, following a textbook format, that highlights the key findings of the international on-line workshops

- Publish a quarterly newsletter to keep the user community informed of latest developments in the PPM field.

- Publish an ASCE journal paper featuring a state-of-the-art review covering what has been done in the area of premise plumbing modeling and outlining what else needs to be done.

- Promote design goals for adoption in premise plumbing modeling software.

- Develop and distribute educational materials suitable for use in a Civil and/or Environmental Engineering curriculum, including a nomograph with public domain spreadsheet on estimating design water demands in premise plumbing systems.

- Promote interaction and collaboration with other professional organizations and agencies (AWWA, ASPE, IAPMO, NIST, USEPA, *etc.*) who work in the premise plumbing arena.

3 PPM COMMITTEE PROGRESS

Since its inception in 2019, the PPM task committee has made significant progress on many of its research milestones and end products, as discussed below:

3.1 Sessions at annual EWRI Congress and other venues

Table 2. Technical sessions on premise plumbing modeling at recent conferences.

Date	Location	Event	Status / Outcome
May 2020	Henderson, NV	ASCE EWRI Congress	Cancelled due to pandemic
June 2021	Milwaukee, WI	ASCE EWRI Congress	Sponsored virtual PPM session
June 2022	Atlanta, GA	ASCE EWRI Congress	Sponsored face-face PPM session
July 2022	Valencia, Spain	CCWI/WDSA Symposium	Presented PPM paper

3.2 International on-line webinars

Starting in April 2020, the PPM Task Committee organized a series of 13 on-line webinars to disseminate information on the current state of building water systems modeling, highlight linkages to design, analysis, operational decision-making in practice, and identify gaps in knowledge, data or capabilities. The schedule and theme of each webinar is listed in Table 3. As shown in Figure 1, 45 speakers participated in the on-line PPM webinars, representing a broad range of international expertise and work experience. All webinars were recorded and uploaded to a dedicated youtube channel accessible at:

(https://www.youtube.com/channel/UCRxdGu3wGpqRwez_VpoezCQ).

Table 3. Webinars hosted by the Premise Plumbing Modeling task committee.

Number	Month and Year	Theme
1	April 2020	Overview and Vision
2	June 2020	Water Age in Buildings
3	July 2020	Estimating Peak Water Demands in Buildings
4	August 2020	Hydraulics of Closed Conduits
5	September 2020	Water Quality and Contaminants
6	October 2020	System Operation
7	December 2020	Instrumentation for Monitoring and Measuring
8	January 2021	Hot Water Systems
9	February 2021	Design Considerations
10	March 2021	Integration Tools (BIM, CFD)
11	April 2021	Looking to the Future
12	October 2021	Committee Report
13	April 2022	Building Water Systems Data Repository

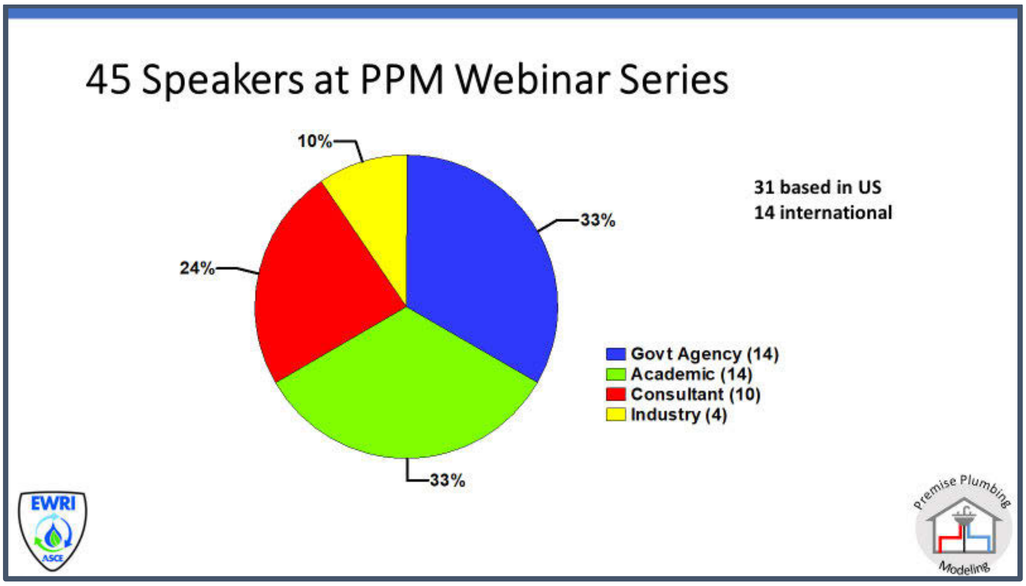


Figure 1. Speakers at the PPM webinar series represented a diverse mix of backgrounds and nationalities.

BY NC SA

3.3 Textbook style final report

The essence of the 13 PPM webinars were distilled to eight chapters in a text style book written to capture the current state of the science in premise plumbing modeling. The book covers critical aspects of premise plumbing modeling, including, but not limited to, hydraulics and water quality fundamentals, indoor water demand estimation (especially during peak conditions), current PPM modeling capabilities/limitations, design/code/standards, operations, and instrumentations.

In addition to highlighting current and past work, the objective of this effort is to identify limitations of current modeling, describe opportunities for future modeling, and documents needs related to data gaps or modeling tools for premise plumbing modeling. The book aims to disseminate knowledge regarding building water system modeling and its close connection to design, analysis, and operational decision-making.

The goal of this book is to discuss and examine the various aspects of premise plumbing systems and how they could be or have been modeled for purposes of engineering design and operation. Special emphasis is placed on highlighting weaknesses and limitations in current premise plumbing modeling efforts. Finally, promising opportunities for future premise plumbing modeling research are identified. The theme for each of the eight chapters and one appendix is listed in Table 4.

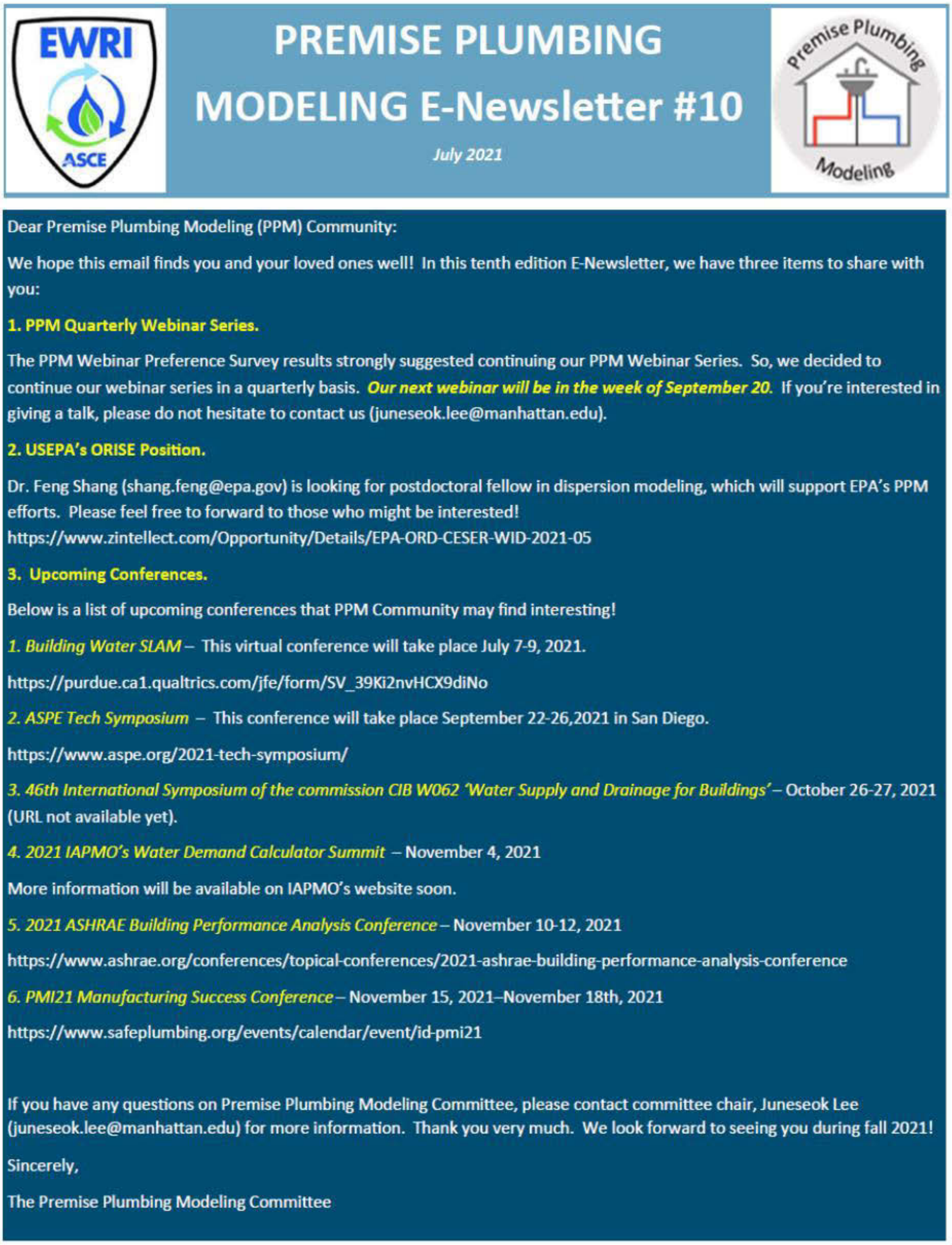
Table 4. Chapters in the Premise Plumbing Modeling textbook.

Chapter	Theme
1	Introduction and overview; the vital importance of premise plumbing systems in cities of the future.
2	Fundamental steady state conditions and transient hydraulic phenomena in premise plumbing systems.
3	Water quality concerns in premise plumbing; water age, and stagnation; chemical and biological issues.
4	Water demand in buildings: estimating instantaneous peak flows and simulating indoor diurnal water patterns.
5	Computer-based premise plumbing analysis, including tools for simulation and information management support.
6	Design and operation of premise plumbing systems in modern buildings; Management plans for premise plumbing systems.
7	Fundamentals of building hot water systems and interaction with building energy systems.
8	Future of premise plumbing modeling in a digital world; use of BIM models and CFD programs; importance of interagency collaboration and cooperation. Codes and standards.
Appendix A	Data requirements for modeling premise plumbing, with focus on selecting and collecting data using sensors.

The PPM textbook will be published by ASCE with an anticipated release date of summer 2023.

3.4 PPM e-newsletter

A periodic newsletter was distributed to keep the user community informed of latest developments in the PPM field. An example appears in Figure 2.


EWRI
ASCE
Premise Plumbing Modeling

PREMISE PLUMBING MODELING E-Newsletter #10

July 2021

Dear Premise Plumbing Modeling (PPM) Community:

We hope this email finds you and your loved ones well! In this tenth edition E-Newsletter, we have three items to share with you:

1. PPM Quarterly Webinar Series.

The PPM Webinar Preference Survey results strongly suggested continuing our PPM Webinar Series. So, we decided to continue our webinar series in a quarterly basis. ***Our next webinar will be in the week of September 20.*** If you're interested in giving a talk, please do not hesitate to contact us (juneseok.lee@manhattan.edu).

2. USEPA's ORISE Position.

Dr. Feng Shang (shang.feng@epa.gov) is looking for postdoctoral fellow in dispersion modeling, which will support EPA's PPM efforts. Please feel free to forward to those who might be interested!
https://www.zintellect.com/Opportunity/Details/EPA-ORD-CESER-WID-2021-05

3. Upcoming Conferences.

Below is a list of upcoming conferences that PPM Community may find interesting!

1. Building Water SLAM – This virtual conference will take place July 7-9, 2021.

https://purdue.ca1.qualtrics.com/jfe/form/SV_39Ki2nvHCX9diNo

2. ASPE Tech Symposium – This conference will take place September 22-26,2021 in San Diego.

https://www.aspe.org/2021-tech-symposium/

3. 46th International Symposium of the commission CIB W062 'Water Supply and Drainage for Buildings' – October 26-27, 2021 (URL not available yet).

4. 2021 IAPMO's Water Demand Calculator Summit – November 4, 2021

More information will be available on IAPMO's website soon.

5. 2021 ASHRAE Building Performance Analysis Conference – November 10-12, 2021

https://www.ashrae.org/conferences/topical-conferences/2021-ashrae-building-performance-analysis-conference

6. PMI21 Manufacturing Success Conference – November 15, 2021–November 18th, 2021

https://www.safeplumbing.org/events/calendar/event/id-pmi21

If you have any questions on Premise Plumbing Modeling Committee, please contact committee chair, Juneseok Lee (juneseok.lee@manhattan.edu) for more information. Thank you very much. We look forward to seeing you during fall 2021!

Sincerely,

The Premise Plumbing Modeling Committee

Figure 2. Sample PPM e-newsletter.

BY NC SA

3.5 Other PPM activities and products

Work continues on other PPM research products including preparation of a peer-reviewed manuscript for the ASCE *Journal of Water Resources Planning and Management.* The paper will present a state-of-the-art review of premise plumbing modeling with emphasis on identifying emerging PPM issues that can be investigated using tools already in the arsenal of engineers working in the field of water distribution systems analysis. In addition, the PPM task committee continues to facilitate and promote sustained productive collaboration among other professional organizations and agencies (AWWA, ASPE, IAPMO, NIST, USEPA, *etc.*) who work in the premise plumbing arena.

4 CONCLUSIONS

Premise plumbing systems deliver finished water from the municipal distribution network to individual consumers in a building. Premise plumbing is an integral part of every contemporary facility. The sophistication and complexity of premise plumbing systems have grown dramatically in recent decades. This has led to new operational challenges in achieving acceptable hydraulic performance and maintaining safe water quality in many buildings, large and small. These premise plumbing challenges will increase as the planet adds three billion people in the next three decades.

Many engineers working in the field of outdoor public water supply (especially water distribution systems) have expertise with tools and techniques that can be used to identify, analyze and solve the new problems emerging from indoor premise plumbing systems. The field is ripe for bold innovative modeling. Advanced building water systems modeling can create new knowledge that will lead to safer designs for low-flow premise water systems. This, in turn, will provide helpful “bi-directional feedback” to improve design, operation and maintenance strategies for conventional urban water distribution and supply systems.

Under the auspices of the EWRI WDSA committee, the PPM task committee has worked productively for three years to advance the state of practice of water resources and environmental engineering by providing membership with tools for use in the optimal design and safe operation of building water systems for public wellbeing. These initial efforts are the first steps in what promises to be a long and challenging but worthwhile journey for civil engineers into the field of premise plumbing modeling.

5 REFERENCES

[1] National Academy of Science, “Greatest Engineering Achievements of the Twentieth Century (greatachievements.org) June 2022.

[2] US Census Bureau, Census.gov June 2022.

[3] US Environmental Protection Agency, WaterSense | US EPA, June 2022.

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14181

DISINFECTION RESIDUAL BEHAVIOUR WITHIN DRINKING WATER DISTRIBUTION SYSTEMS

Jade Rogers[1], Katherine Fish[2], Vanessa Speight[3], Hunter Fairley[4], Graham Knowles[5] and Joby Boxall[6]

[1,2,3,6]Sheffield Water Centre, University of Sheffield. Sheffield. South Yorkshire. UK
[4,5]Scottish Water. Edinburgh. UK

[1] *jcrogers1@sheffield.ac.uk*, [2] *k.fish@sheffield.ac.uk*, [3] *v.speight@sheffield.ac.uk*,
[4] *Hunter.fairley@scottishwater.co.uk* , [5] *Graham.knowles@scottishwater.co.uk*,
[6] *J.B.Boxall@sheffield.ac.uk*

Abstract

Understanding the persistence of different disinfection residuals in drinking water distribution systems (DWDS) is critical to water quality safety and public health. Chloramines are expected to persist, due to their chemical composition, further into the extremities of DWDS providing reduced risks to water quality, however there is limited evidence of this from operational systems. Total chlorine data from regulatory sampling at WTW outlets and customer taps from multiple DWDS was collected and analysed. The first data set compares long term performance from equivalent systems at the same time. Results showed the decay of residuals was similar across all seasons in the DWDS analysed, showing little correlation with residual type. The second data set compares six systems that experienced a disinfection switch from chlorine to chloramine. Decay of total chlorine residuals was clearly reduced for three and marginally reduced for a fourth following the switch, suggesting these DWDS experienced increased persistence. However, two sites showed little change. This analysis highlights the uncertainties around residual persistence under different disinfection types, indicating the assumption that chloramine persists for longer than chlorine in every distribution system is not a given.

Keywords
Drinking Water Quality, Disinfection Residuals, Drinking Water Distribution System.

1 BACKGROUND

Disinfectant residuals (although not legally mandatory) are common practice across the majority of developed nations, implemented to control against DWDS microbial growth and/or regrowth, particularly in old or degrading networks. DWDS are complex interlinked systems composed of long connecting pipe sections, storage facilities and hydraulic components of which vary in terms of age and condition. These characteristics, alongside pipe-wall interactions and the biological (biofilm and planktonic), chemical and physical properties of the DWDS, drive water quality deterioration. Importantly, these complex interactions can impede disinfection residual maintenance, particularly at the network extremities – increasing the likelihood of microbial failures.

The most well-known and widely used disinfection residual by water utilities is free chlorine. Yet, in response to the regulation of chlorine-formed disinfection by-products (DBP), the use of chloramines, specifically monochloramine for secondary disinfection is increasing [1], [2]. This is particularly common in DWDS experiencing high organic loads. Additionally, the properties of chloramine mean it is less reactive in solution, showing decay rates two times slower than that of free chlorine [3]. For this reason, chloramine is sometimes used in DWDS with high water ages, where residual maintenance at the network extremities may present a challenge. However, the biological, chemical and physical characteristics of DWDS effect residual decay, predominantly pipe-wall interactions (with corrosion deposits and biofilms), reactions with naturally occurring compounds, such as total organic carbon (TOC) and elevated water temperatures [1], [4]–[8]. As a result, residuals can become depleted and drinking water quality may deteriorate during transit from the water treatment works (WTW) to customer taps [4]. TOC is of particular interest, as multiple studies show bulk water TOC to correlate with chlorine decay [9]–[12].

Previous research into chloramine and chlorine DWDS residual maintenance is heavily focused on reaction kinetics and decay models [13], [14], either computational or alongside bench top scale reactors [7], [8]. There seems to be limited studies considering the impact of both pipe wall and bulk water conditions on residual decay out in the field, whereby disinfection residuals are examined between two points within a DWDS [15]. Furthermore, there is limited direct comparisons of chloramine and chlorine residual maintenance in DWDS, likely due to DWDS complexity and variability. Despite this and the knowledge that residual decay is greatly influenced by network conditions and parameters, there still remains this idea that chloramine will persist for longer than chlorine in DWDS [16], [17]. Academic evidence in support of this statement is lacking - a concern when 90.6 % of U.S. utilities using chloramines stated that improved residual persistence was the main reason behind switching [2].

Improved understanding of disinfection residual persistence in both chlorine and chloramine systems is needed to protect against adverse water quality impacts and general water quality degradation. This is increasingly important as more and more water companies switch from chlorine to chloramine disinfection in response to DBP regulations and residual maintenance concerns. This paper compares the persistence of residuals in both long-term and newly switched chlorinated and chloraminated drinking water distribution systems (DWDS), to better understand residual behaviour. TOC (as a measure of organic load) is also considered, providing insight into potential reactions and bacterial food source availability.

2 METHODS

2.1 Case study introduction

In order to assess disinfection residual maintenance historic data was analysed. Two datasets were created from drinking water treatment works (WTW) outlet and tap water quality sample results. One data set allowed direct comparison, comparing different systems at the same time; while similar systems where compared there were differences. The second was for DWDS that experienced a switch in residual (from chlorine to chloramine), hence system characteristics were exactly the same but comparison is for sequential years. In combination the two datasets enable a more robust comparison.

2.2 Datasets

Dataset 1: TOC, total and free chlorine results from WTW outlets and distribution taps between 01.01.18 and 31.07.21, across 9 DWDS were selected. Drinking water travel distance impacts disinfection residual decay, for this reason systems within a 10% population size of 37,000 and 10,000 respectively were chosen.

To consider the effect of temperature, (shown to impact microbial activity and residual decay) dataset 1 was segregated by season. Due to the delayed response of water temperatures to air temperatures, seasons were as follows; "Winter" - January, February and March, "Spring" - April, May and June, "Summer" -July, August and September and "Autumn" - October, November and December. The specification for this dataset can be seen in Table 1.

TOC was analysed due to it being a nutrient used by planktonic and attached phase (biofilm) microorganisms. In addition, TOC indicates organic load within the bulk water, this threatens residual persistence due to direct reactions between organics and disinfection residuals. As a

BY NC SA

result TOC was also considered in this analysis. Annual average TOC from chloramine WTW outlets and taps for the period analysed was 1.59 mg/L and 1.57 mg/L, respectively. Annual average TOC from chlorine WTW outlets and taps was 1.08 mg/L and 1.07 mg/L, respectively.

Table 1 : Specifications of drinking water distribution systems used in dataset 1 and sample counts for total chlorine across each season. WTW data reported first followed by taps (n = "WTW", n ="Taps").

Disinfection Residual	No. of DWDS	DWDS Population Range	Data Points (WTW, Taps)
Chloramine	4	10% range of 37,000 people	Spring (n=729, n=395) Summer (n=537, n=312) Autumn (n=444, n=287) Winter (n=669, n=452)
Chlorine	5	10% range of 10,100 people	Spring (n=551, n=160) Summer (n=452, n=179) Autumn (n=431, n=178) Winter (n=545, n=192)

Dataset 2: TOC and Total chlorine results from WTW outlet and distribution taps of six DWDS (A-F), all which experienced a disinfection switch (chlorine to chloramine) between 2016 and 2021. Data from 1 year pre and 1 year post disinfection switch was included.

Table 2: Number of total chlorine data points per group in dataset 2, WTW data reported first followed by taps (n = "WTW", n ="Taps").

DWDS	Data points pre-switch (WTW, Taps)	Data points post-switch (WTW, Taps)
A	n=231, n=6	n=239, n=12
B	n=376, n=48	n=406, n=53
C	n=437, n=321	n=387, n=299
D	n=223, n=73	n=199, n=57
E	n=72, n=17	n=51, n=16
F	n=55, n=13	n=57, n=12

2.3 Analysis

Both datasets were analysed using the programming language and statistical software 'R'. In order to test the normality of the data Shapiro-Wilks was performed. As data was not normally distributed, box and whisker plots were used, with medians being used for % change calculations. Box plot thick horizontal bars specifies the median, the box itself represents the interquartile range (IQR) and the whiskers the maximum ($Q_3 + 1.5 \times IQR$) and minimum ($Q_1 - 1.5 \times IQR$) values, with dots indicating outliers beyond this range.

3 RESULTS

Figure 1 and Figure 2 show a greater decrease in total chlorine concentration (mg/L) between WTW outlet and taps, across all seasons in chloramine DWDS compared to chlorine DWDS. In contrast, the % difference in medians shown in Table 3 indicates a similar proportional decrease in total chlorine residual across both chloraminated and chlorinated DWDS. This is likely due to the higher chloramine concentrations at the WTW outlet than chlorine systems, which contributes to exponential decay. Figure 1 displays more visible outliers of <0.2 mg1/L than Figure 2 suggesting that in specific areas of the chloramine DWDS greater decay of total chlorine.

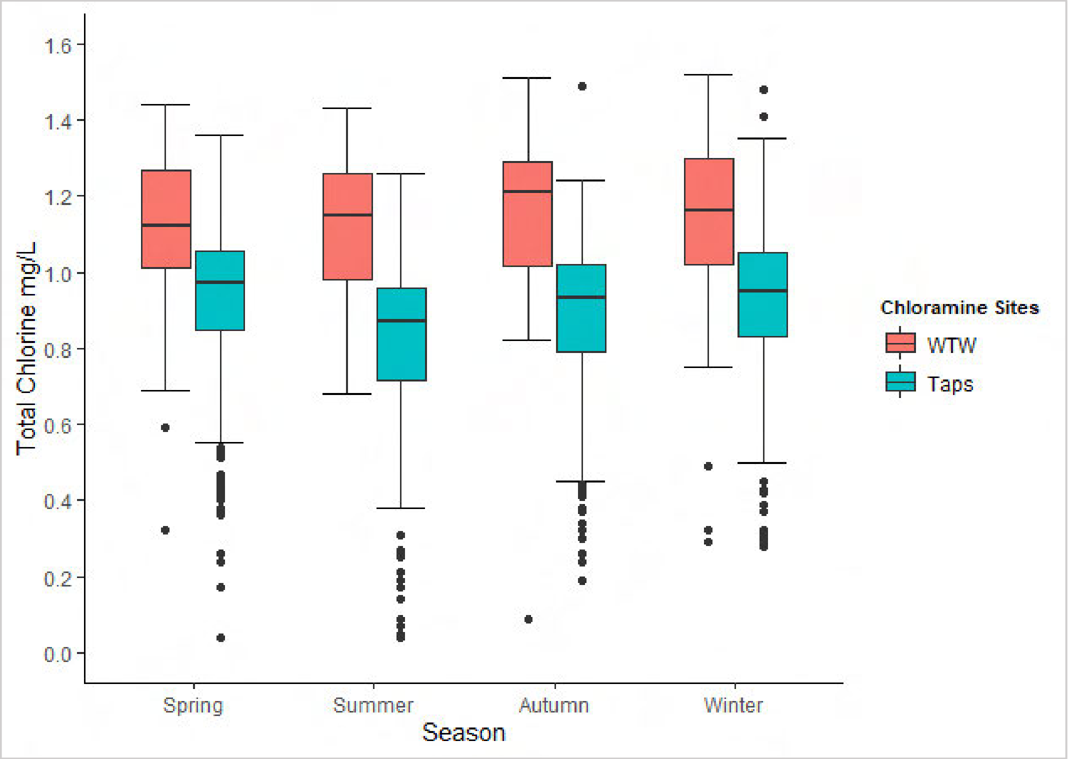


Figure 1: Impact of season on total chlorine concentrations at the water treatment works outlet and taps of four long-term chloraminated DWDS (Dataset 1)

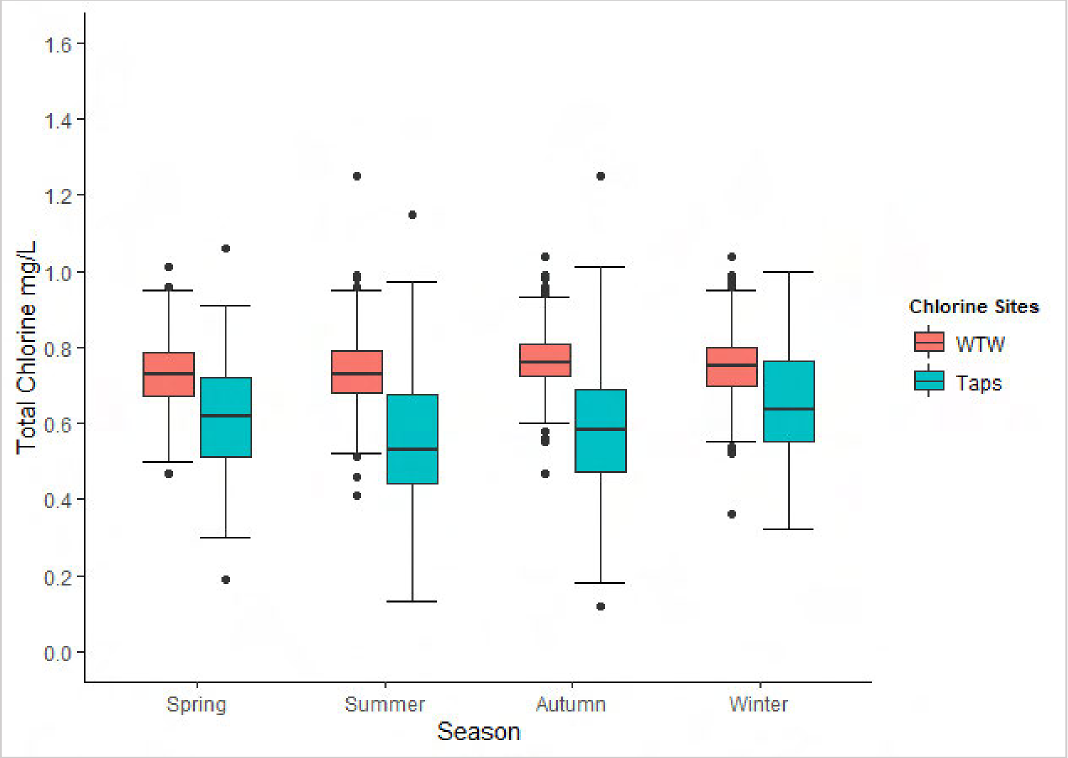


Figure 2: Impact of season on total chlorine concentrations at water treatment works outlet and taps of five long-term chlorinated DWDS (Dataset 1).

Table 3: Difference between the total chlorine median of WTW outlet and tap total chlorine as a percentage (%) across chlorine and chloramine DWDS, segregated by season, from dataset 1.

Chloramine DWDS	Chlorine DWDS	Season
18.1	15.13	Winter
13.39	14.47	Spring
24.35	26.32	Summer
23.14	23.03	Autumn

The largest % difference in total chlorine medians and thus decay is shown in summer of both chloraminated and chlorinated sites, 24.4% and 26.3% respectively. Indicating the importance of temperature in network residual persistence.

In order to consider the impact of water age on chloramine residual persistence, total chlorine data in dataset 2 was analysed (Figure 3). This dataset compares chlorine and chloramine persistence within the same DWDS (pre and post disinfection change), therefore network characteristics including water age remain the same. This is advantageous over Dataset 1 as this compares different DWDS experiencing different residuals and water ages. Dataset 2 analysis shows contrasting results to those presented in previous Figures 1 and 2, as higher total chlorine is seen at taps post-switch than pre-switch at DWDS A, C, D, E (Table 3).This suggests networks are experiencing less decay the year following the switch than the year before. However, low total chlorine concentrations are still recorded in some tap samples as shown by outliers, this is particularly true for DWDS A and C (Figure 3). DWDS B shows limited change between WTW and tap total chlorine concentration pre- and post-switch. Similarly, DWDS F shows a 12.5% reduction in residual through the network prior to the switch and a small increase 1.03% increase after (Table 3), indicating little change in residual persistence pre- and post- switch at these two sites.

The TOC concentration the year pre- and post-disinfection change from chlorine to chloramine is presented in Figure 4. DWDS A has been omitted due to a lack of data. TOC decreased between WTW and tap results prior to the switch in DWDS C, D and E. Post switch the opposite occurred, with TOC increasing from WTW to taps. Suggesting the switch to chloramine is contributing to increased TOC generation within the network. This trend is not seen in DWDS B and F, with TOC concentrations at DWDS B remaining low at both WTW and taps regardless of the disinfection switch and DWDS showing a similar reduction in TOC between WTW and taps pre and post switch, 18.75% and 22.73% respectively.

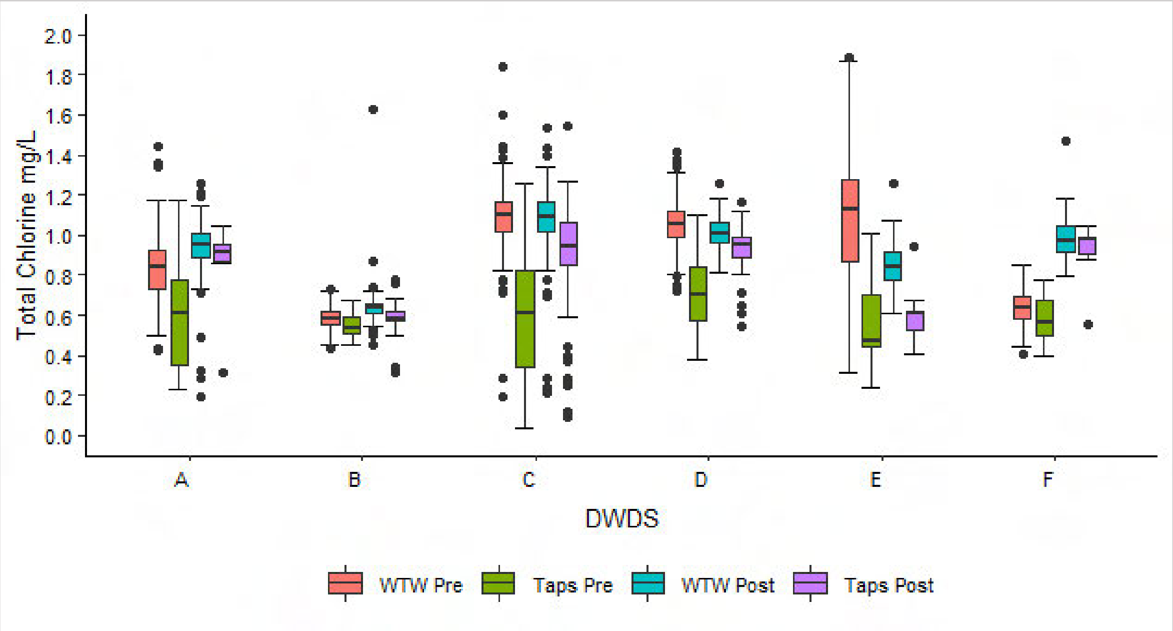


Figure 3: Total chlorine concentrations across six drinking water distribution systems (A-F), the year pre and post disinfection switch from chlorine to chloramine. Water treatment works concentrations before ("WTW Pre") and after ("WTW Post") and customer tap concentrations before ("Taps Pre") and after ("Taps Post") the disinfection switch are shown (Dataset 2).

BY NC SA

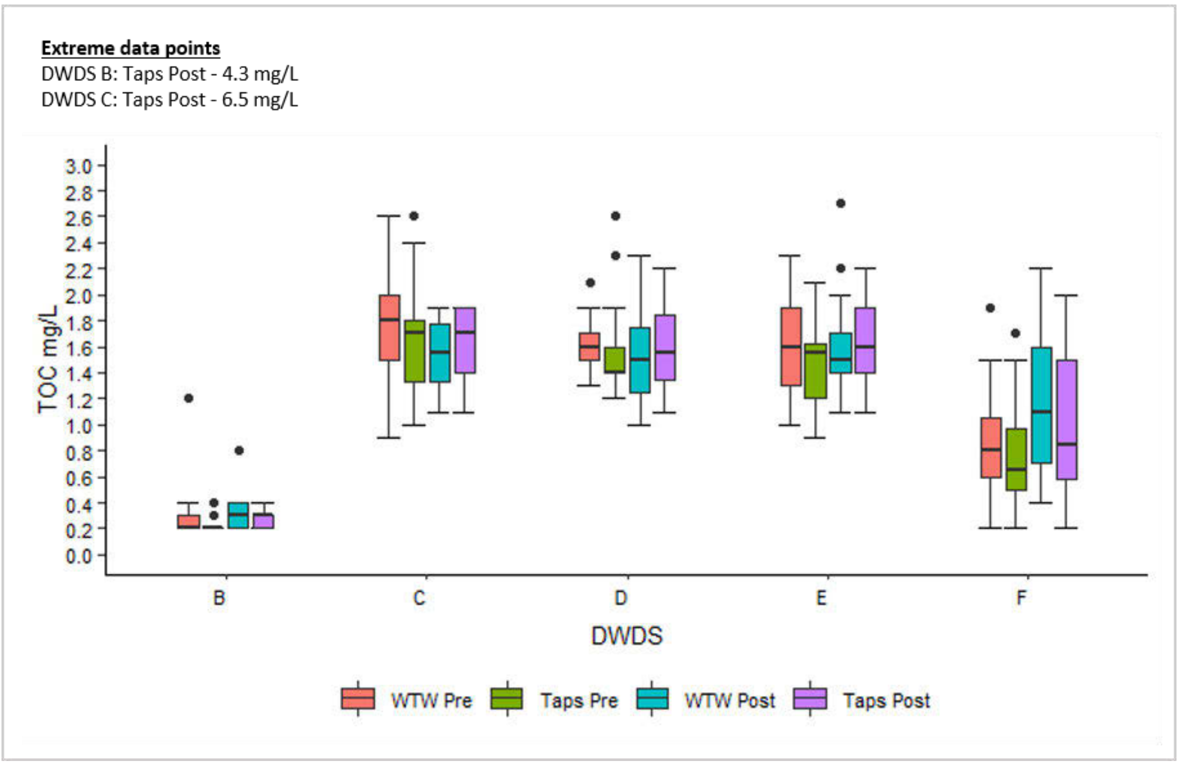


Figure 4: Total organic carbon (TOC) across five drinking water distribution systems (B-F), the year pre and post disinfection switch from chlorine to chloramine. Water treatment works concentrations before ("WTW Pre") and after ("WTW Post") and customer tap concentrations before ("Taps Pre") and after ("Taps Post") the disinfection switch are shown (Dataset 2).

Table 3: The % change in total chlorine and total organic carbon (TOC) medians between the WTW and taps pre and post disinfection switch. Increase in change (+) and decrease in change (-). N/A refers to sites that did not have sufficient data available.

DWDS	WTW Pre - Taps Pre		WTW Post - Taps Post	
	Total Chlorine	TOC	Total Chlorine	TOC
A	-27.38	N/A	-4.21	N/A
B	-7.76	0	-9.38	0
C	-44.55	-5.56	-13.76	+9.68
D	-33.33	-12.5	-5.94	+3.33
E	-58.41	-3.13	-27.98	+6.67
F	-12.50	-18.75	+1.03	-22.73

4 DISCUSSION

The first stage of analysis shows chloramine DWDS to experience similar if not greater decay in total chlorine throughout the network than chlorine. This contradicts the perception that chloramine's lower reactivity enables it to persist for longer within the DWDS than chlorine [16],

cc BY NC SA

[17]. However, this result should be treated with caution as dataset 1 doesn't consider all factors particularly that of water age – a result of chlorine and chloramine sites varying in size, which can influence residual decay. The construction of dataset 2, allowed for consideration of these factors and showed contradictory results, with four out of the six DWDS showing improved residual maintenance under chloramine disinfection than chlorine. Thus, supporting the belief that chloramine is less reactive and persists for longer within the network than chlorine. However, DWDS B and F indicate no benefit of chloramine residual persistence following the disinfection switch.

Post disinfection switch, DWDS F shows a 22.73% decrease in TOC, but maintains a comparable total chlorine residual throughout the network. In contrast, DWDS, C, D and E show TOC accumulation but also high residual decay the year post disinfection switch, suggesting that microbial activity may be contributing to the decay seen. However, this trend is tentative, due to the similarity of data ranges pre and post switch. The fact data sits well within the range of noise and the inability to perform statistical analysis to determine significance (due to unpaired data set) adds to this uncertainty. Additionally, it is difficult to infer biological activity based solely on TOC measurements. Assimilable organic carbon (AOC) would be a better parameter to determine microbial activity, particularly biofilm growth [18][18][18]. However, AOC is not routinely tested for and was not available for this historic data analysis.

It is important to note that the length of time that networks have been chloraminated differs between the two datasets. With dataset 1 being long-term chlorinated and chloraminated systems, and dataset 2, newly chloraminated systems. The length of time a network has been chloraminated is likely to influence the maintenance of residual. Thus, this could contribute to the contradictory results seen. Additionally, inferring residual loss from tap samples should be treated with caution due to the limited number of samples taken, which are unrepresentative of the vast complex DWDS that operate for 24 hours a day. The fact tap samples are collected randomly across various locations (unpaired data set), means statistical analysis to determine significance is not possible, this adds to the uncertainty of residual maintenance.

Dataset 1 shows the greatest decay in residual occurs in the summer of both chloraminated and chlorinated systems. Temperature is known to have a significant impact on chlorine decay rates, with decay approximately doubling for every 5 ˚C increase in water temperature [6] so this was expected. Whilst detailed water temperature data was not available for the datasets analysed, air temperature data was used to check summer months experienced higher temperatures range.

5 CONCLUSIONS

Analysis of data comparing long-term performance of chloramine and chlorine systems showed similar decay of residuals across all seasons, irrespective of residual type. The second dataset showed variable residual persistence across DWDS that underwent a disinfection switch from chlorine to chloramine. Specifically, three systems showed a clear reduction in residual decay and one a marginal reduction following the disinfection switch from chlorine to chloramine, indicative of improved residual persistence. Yet, two DWDS showed little change in residual persistence pre and post disinfection switch. This analysis highlighted the uncertainties around residual persistence under different disinfection residuals and suggests it would be wrong to assume that chloramine persists for longer than chlorine in every distribution system, meaning the common perception of this is not true.

6 REFERENCES

[1] P. Vikesland, N. G. Love, K. Chandran, R. Rebodos, and E. M. Fiss, "Seasonal Chlorination Practices and Impacts to Chloraminating Utilities," AWWA, vol. 2, pp. 1–95, 2007.

[2] C. J. Seidel, M. J. McGuire, R. S. Summers, and S. Via, "Have utilities switched to chloramines?," Journal / American Water Works Association, vol. 97, no. 10. pp. 87–97, 2005, doi: 10.1002/j.1551-8833.2005.tb07497.x.

[3] K. Chambers, J. Creasey, and L. Forbes, "Safe Piped Water: Managing Microbial Water Quality in Piped Distribution Systems," 2004. doi: 10.2166/9781780405841.

[4] K. Ma, J. Hu, H. Han, L. Zhao, R. Li, and X. Su, "Characters of chloramine decay in large looped water distribution system - The case of Tianjin, China," Water Sci. Technol. Water Supply, vol. 20, no. 4, pp. 1474–1483, 2020, doi: 10.2166/ws.2020.063.

[5] P. Keshvardoust, V. A. A. Huron, M. Clemson, N. Barraud, and S. A. Rice, "Nitrite production by ammonia-oxidizing bacteria mediates chloramine decay and resistance in a mixed-species community," Microb. Biotechnol., vol. 13, no. 6, pp. 1847–1859, 2020, doi: 10.1111/1751-7915.13628.

[6] M. Blokker, J. Vreeburg, and V. Speight, "Residual chlorine in the extremities of the drinking water distribution system: The influence of stochastic water demands," Procedia Eng., vol. 70, pp. 172–180, 2014, doi: 10.1016/j.proeng.2014.02.020.

[7] J. M. Arevalo, "Modeling Free Chlorine And Chloramine Decay In A Pilot Distribution System," (THESIS), 2007, [Online]. Available: http://library.ucf.edu.

[8] E. A. Curling, M. J. McKie, L. Meteer, B. Saunders, S. A. Andrews, and R. C. Andrews, "Estimation of chloramine decay in drinking water distribution systems," J. Water Process Eng., vol. 46, 2022, doi: 10.1016/j.jwpe.2022.102558.

[9] N. B. Hallam, F. Hua, J. R. West, C. F. Forster, and J. Simms, "Bulk Decay of Chlorine in Water Distribution Systems," J. Water Resour. Plan. Manag., vol. 129, no. 1, pp. 78–81, 2003, doi: 10.1061/(asce)0733-9496(2003)129:1(78).

[10] M. N. Saidan, K. Rawajfeh, S. Nasrallah, S. Meric, and A. Mashal, "Evaluation of factors affecting bulk Chlorine decay kinetics for the Zai water supply system in Jordan. Case study," Environ. Prot. Eng., vol. 43, no. 4, pp. 223–231, 2017, doi: 10.5277/epel70417.

[11] D. Nono, P. T. Odirile, and B. P. Parida, "Assessment of probable causes of chlorine decay in water distribution systems of gaborone city, Botswana," Online) = Water SA, vol. 45, no. 2, 2019, doi: 10.4314/wsa.v45i2.05.

[12] Y. Zhang, L. Zhou, G. Zeng, H. Deng, and G. Li, "Impact of total organic carbon and chlorine to ammonia ratio on nitrification in a bench-scale drinking water distribution system," Front. Environ. Sci. Eng. China, vol. 4, no. 4, pp. 430–437, 2010, doi: 10.1007/s11783-010-0247-5.

[13] R. Roy, A. Sathasivan, and G. Kastl, "Simplified chemical chloramine decay model for water distribution systems," Sci. Total Environ., vol. 741, Nov. 2020, doi: 10.1016/j.scitotenv.2020.140410.

[14] A. Sathasivan, J. Fisher, and G. Kastl, "Simple method for quantifying microbiologically assisted chloramine decay in drinking water," Environ. Sci. Technol., vol. 39, no. 14, pp. 5407–5413, Jul. 2005, doi: 10.1021/es048300u.

[15] N. B. Hallam, J. R. West, C. F. Forster, J. C. Powell, and I. Spencer, "The decay of chlorine associated with the pipe wall in water distribution systems," Water Res., vol. 36, no. 14, pp. 3479–3488, 2002, doi: 10.1016/S0043-1354(02)00056-8.

[16] US EPA, "Chloramines in Drinking Water ," EPA, Mar. 08, 2021. https://www.epa.gov/dwreginfo/chloramines-drinking-water (accessed Nov. 15, 2021).

[17] J. . Dyksen, J. Brandt-edwards, J. . Clement, R. Hoehn, and C. Spencer, "Long-term effects of disinfection changes on distribution system water quality," 2007.

[18] F. Pick, K. Fish, and J. Boxall, "Assimilable organic carbon cycling within drinking water distribution systems," Water Res., vol. 198, p. 117147, 2021, doi: 10.1016/j.watres.2021.117147.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14186

AN APPROACH TO IMPROVE DRAINAGE NETWORKS BASED ON THE STUDY OF FLOOD RISK

Leonardo Bayas-Jiménez[1], F. Javier Martínez-Solano[2], Pedro L. Iglesias-Rey[3] and Fulvio Boano[4]

[1,2,3] Department of Hydraulic Engineering and Environment, Universitat Politècnica de València, Valencia, Valencia, (Spain)
[1,4] Department of Hydraulics and Civil Infrastructures, Politecnico di Torino, Turin, Piedmont, (Italy)

[1] *leobaji@posgrado.upv.es*, [2] *jmsolano@upv.es*, [3] *piglesia@upv.es*, [4] *fulvio.boano@polito.it*

Abstract

In recent years, the number of pluvial floods in cities has increased. There are different reasons for this increase, but certainly the most important are the increase in the intensity of the rain and the growth of cities, which increases the impervious surface, decreases the concentration time, and consequently increases runoff. To determine the costs of flood damage, land use is very important in the estimation of the cost of flood damage because it defines the type, value, and vulnerability of the structures. On this wise, an approach of great interest to managers of drainage networks is the estimation of annual damages in the area under study, because by knowing these costs, corrective measures can be taken, and the investment needed to reduce flooding can be determined. This work aims to present a methodology to improve drainage networks with lack of capacity due to increased runoff with a focus on the study of the vulnerability to the risk of flooding. To improve the operation of the networks, the method considers the possibility of changing the existing pipes for others with greater capacity, the construction of storm tanks and the implementation of hydraulic controls. To find the best solutions, an optimization model was developed that uses a Pseudo Genetic Algorithm and the SWMM model to perform hydraulic simulations. To reduce calculation times, a search space reduction procedure is applied to identify the regions containing the best solutions. These actions are intended to improve the efficiency of the model. For the economic assessment of flood risks, the cost of flood damage is related to its probability of occurrence for different return periods. With these data, a curve with a log-linear relationship is constructed. The cost of the flood risk is obtained by integrating the area under the curve obtained. The methodology is applied in a drainage network with flooding problems called Balloon, located in located in northern Italy. The results are useful to demonstrate the benefit of risk cost analysis to make decisions and underline the relevance of including optimization models to prevent future damage in cities.

Keywords
Flood, Risk, Water, Drainage Networks, Optimization, Genetic Algorithms.

1 INTRODUCTION

Urban flooding is a phenomenon of growing importance, with the potential to affect the future of many cities around the world. There are two widely agreed reasons for the increase in the frequency of floods and the magnitude of flood damage. The first is urbanization that increases impervious surface and is often poorly planned, especially in developing countries. The second reason is climate change, which causes an increase in the frequency and intensity of precipitation events. Consequently, the risk of flooding has increased in recent years and is expected to worsen in subsequent years [1]. Urban floods occur as a result of the insufficiency of the urban drainage system to transport the runoff flows that are generated during episodes of high intensity rain. These floods can cause serious damage to people, infrastructure, and property, in addition to disrupting traffic and economic activity.

To face this problem, a frequently used tool is the optimization of drainage networks considering these new scenarios. Mathematical optimization is part of applied mathematics aimed at finding the best solutions to mathematically defined problems. The basic optimization problem consists of an objective function that will be minimized or maximized [2]. In this way, it can be mentioned the work carried out by Iglesias-Rey et al. [3] who used a Pseudo Genetic Algorithm (PGA) developed by Mora-Melia et al. [4] for the optimization of a drainage network considering the replacement of pipes, the installation of storm tanks, and the initial state of pumping units. In this work, the authors conclude that the combined use of pipe replacement and storm tank installation provides better results than the application of these actions separately. The importance of the different cost functions considered in the final optimization result was also highlighted. Continuing this line of research Ngamalieu-Nengoue et al. [5] confirm that the combined use of pipe replacement and storm tank installation provides the best results. In this work the authors warn that the increase in the number of Decision Variables (DVs) used when employing both measures together cause a significant increase in the size of the Search Space (SS). This problem in optimization with heuristic methods has been detailed by different authors [6], [7]. Mayer et al. [6] mention that this problem is called "curse of dimensionality" and not only does it make it more difficult to explore the SS, but there is also the risk that the algorithm fails to find the global solution. As a solution to this problem, Ngamalieu-Nengoue et al. [8] proposed a Search Space Reduction (SSR) method essentially based on two principles, the reduction of the resolution of the discretization of the DVs and the reduction of the number of DVs involved in the problem. In this work it was also evidenced that in the optimization process some solutions presented the reduction of pipes, with this reduction the water flow was slowed down. Based on this work Bayas-Jiménez et al. [9] consider the installation of hydraulic controls in conjunction with the installation of storm tanks and pipe replacement, conclude that the use of hydraulic controls reduces the volume of flooding, reducing the costs associated with the implementation of protection structures. In addition, the authors in their methodology implemented a convergence criterion to guarantee obtaining good solutions with a certain probability of success, adding robustness to the methodology. However, the use of this convergence criterion implies a significant increase in calculation times. To improve the performance of this optimization methodology Bayas-Jiménez et al. [10] propose a SSR method focused on the iterative reduction of DVs through three actions, the analysis of the topology of the network, the study of the discretization of the DVs and the application of a selection criterion of DVs. The authors point out that this methodology gives good results for the rain studied. However, it does not consider the effects of other rains that may occur within the design period. Although it is true that the most unfavorable situation is analyzed, rains may occur with less intensity and higher probability of occurrence that can cause flooding problems, as observed in their works by Freni, La Loggia and Notaro [11] and Olsen, et al. [12]. The main objective of this work is to study the impacts of these precipitations and identify the actions to be implemented to adapt the drainage networks to these conditions through the economic analysis of the risk of flooding.

2 THE RISK

The risk can be expressed as the product of the probability of danger and the consequences that are generated. The magnitude of the consequences will depend on the exposure and vulnerability of the affected area [13]. Flood risk management has become a central part of flood control policies throughout Europe [14]. This risk management approach is also increasingly adopted in the planning of urban drainage networks, so it is no longer only the probability of a storm event that is considered, but also the impact of these events [15]. If it is considered that the probability of flooding is usually expressed in years^{-1} and that the consequences of flooding can be expressed in monetary units, the cost of the risk of flooding can be expressed in euros/year. This cost is obtained by measuring the area under the curve drawn for rainfall with several return periods. [16]. As shown in Figure 1. It is important to keep in mind that flood risk assessment is not an easy task because it is a probabilistic variable that depends on stochastic rainfall events. It is impossible to predict exactly what the consequences of flooding will be in a specific area and time. But estimates can be made that are close to the expected scenarios.

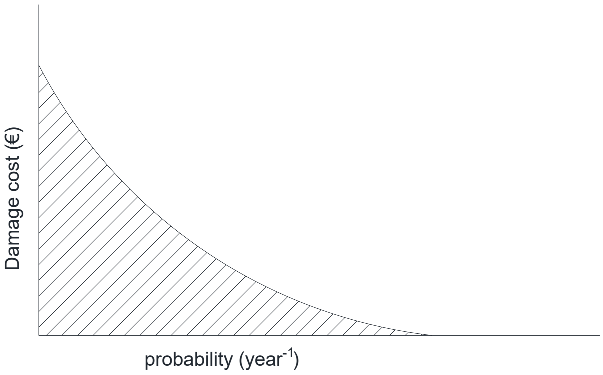


Figure 1. Annual cost of flood risk as the area under the curve.

3 OPTIMIZATION OF DRAINAGE NETWORKS

The optimization of the drainage networks is carried out considering the traditional method of substituting pipes for others with a larger diameter, the construction of storm tanks in the nodes that require it, and the installation of hydraulic control elements in the drainage network. In short, it is intended to rehabilitate drainage networks by finding the best solution to combine the installation of these infrastructures to reduce flooding. To find the best solutions, an optimization model is developed that uses the SWMM model and a Pseudo Genetic Algorithm connected by a toolkit. To compare the solutions found, an objective function is defined composed of the costs of these infrastructures and the cost implied by the risk of flooding. Thus, the model aims to minimize the objective function and find the best solutions to the analyzed problem. It is then necessary to define the DVs that the model must analyze. Specifically, there are three groups of DVs. The first group of decision variables are the diameters of the pipes in the network. The second group of DVs is the area of the storm tanks, the storm tank area is assumed as the DV because the depth of the existing manholes is set as the depth of the tank. The third group of DVs are the degrees of opening of the hydraulic control elements. When using a PGA as a search engine, it is understood that the values adopted by the DVs must be discretized. Pipe diameters are discrete while in the case of storm tanks and hydraulic controls, lists of options should be specified both for the areas of the tanks and for the degrees of opening of the hydraulic controls as proposed Bayas-Jiménez et al. [10].

Once the DVs and the values they can take have been defined, the SS that the model must explore to find the best solutions can be specified. The magnitude of the SS is calculated using equation (1).

$$\mathrm{S} = \mathrm{n_i}\ (\log \mathrm{NS}) + \mathrm{m_i}\ (\log \mathrm{ND}) + p_\mathrm{i}\ (\log \mathrm{N\theta}) \tag{1}$$

In the equation S represents the magnitude of the SS, $\mathrm{n_i}$ the nodes analyzed of the network, $\mathrm{m_i}$ the pipes analyzed of the network, ND represents the range of commercial diameters adopted, NS the list of areas defined, $\mathrm{p_i}$ the pipes analyzed of the network with the possibility to install hydraulic controls and Nθ represents the opening options defined. It can be inferred that the SS is quite large and that it potentially increases as the number of elements analyzed increases. The time and computational effort that the exploration of the total space would require would be very important, for this reason it is necessary to implement a SSR process. In the space reduction process as a first step, a first mapping of the entire SS of the algorithm is performed, each DV is analyzed by the algorithm using a list of options with a coarse discretization. Using a thick option list allows one to quickly identify DVs that can be included in the most promising region of the algorithm's exploration space. After performing a certain number of evaluations, the fifth percentile (P_5) of these evaluations is analyzed, selecting the candidate DVs for the new search region through a selection criterion based on repeatability. The process is repeated in the new defined region, eliminating the DVs that do not meet the selection criteria. In each iteration of the process, the convergence of the results towards certain DVs that increase the repeatability in the sampling, defining each time the region with the best solutions. The process ends when all DVs meet the selection criteria and the SS cannot be reduced any further. Once the final search region is defined, the final optimization is performed on this new scenario. This new optimization uses the refined option list. In order to perform a much more detailed exploration of the SS to identify the global optimum of the problem.

4 ANALYSIS OF THE COST OF RISK

To analyze the cost of the risk of flooding, it begins with the definition of the cost functions of both the infrastructure to be installed and the cost of the risk of flooding. The first infrastructure investment cost function is the cost of pipe renovation, for the case under study a second-degree polynomial equation is determined as shown in equation (2).

$$C_p(D_i) = (179.71\, D_i{}^2 + 281.32\, D_i - 14.139)\, L_i \tag{2}$$

In the equation $C_p(D_i)$ represents the cost of renewing the pipe in euros, D_i is the diameter of the pipe and Li represents the length of the pipe.

The second infrastructure investment cost function is the cost of storm tanks. To define this function, the cost of building tanks of different sizes has been analyzed. With these data, a cost function composed of two terms is determined and is shown in equation (3)

$$C_T(V_i) = 21220 + 9483.43 V_i{}^{0.65} \tag{3}$$

Where C_T (V_i) represents the cost in euros of the construction of a storm tank. and V_i is the flood volume that the tank must store.

The third structure investment cost function that has been determined for this project is the cost of hydraulic control. To define this function, the cost of acquiring and installing valves of different diameters has been analyzed. This analysis determines a second-degree polynomial function that is shown in equation (4).

$$C_v(D_i) = -271.53\, D_i{}^2 + 4401.70\, D_i + 148.32 \tag{4}$$

Where C_v (D_i) Is the cost of the hydraulic control, Di is the diameter of the pipe where the hydraulic control would be installed.

On the other hand, to account for the reduction in the initial investment in infrastructure, an annual amortization factor Λ is required that affects each cost function. In this work, the expression shown by Steiner (2007) and shown in equation (5) is used.

$$\Lambda = \frac{i}{1 - (1 + i)^{-n}} \tag{5}$$

In the equation i is the annual interest and n is the time in years in which it is proposed to recover the investment.

To obtain the annual cost of the flood risk, the cost of the flood damage must be calculated. One of the most used techniques to carry out the analysis of the damage of the flood is to use as a reference the depth or level of the flood. The analysis based on the flood level allows to reduce the flood and the cost of the flood depending on the flood zone. An expression that calculates the cost of flood damage as a function of flood depth and land use are the so-called damage-depth curves. In this sense, an expression that allows this calculation to be carried out is the one presented in equation (6).

$$C_y(y_i) = C_{max}\, A_i\ \left(1 - e^{-\lambda \frac{y_i}{y_{max}}}\right)^r \tag{6}$$

In equation, y_{max} is the maximum depth at which the flood reaches maximum damage. C_{max} is the maximum cost of flood damage obtained when y_{max} is reached A_i is the area of the analyzed sub catchment and y_i is the depth reached of the analyzed node. Once the cost of floods has been defined, in order to proceed to study the cost of the risk of flooding, the cost of the flood must be related to the return period of the analyzed scenario. For the different return periods, the respective calculations must be made to establish the corresponding cost of flooding. To obtain the curve, the cost obtained is related to the probability of exceedance of the event. In this sense, Olsen et al. (2015) [12] mention that these values can be represented as a linear-logarithmic relationship to extrapolate the cost of flood damage for intermediate return periods. Figure 2 shows this relationship, the area under the curve represents the annual cost of flood risk.

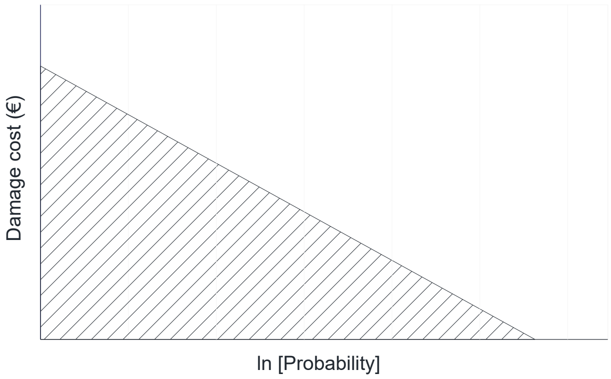


Figure 2. Annual cost of flood risk as the area under the curve in a log-linear relationship.

The curve can be expressed as shown in equation (7).

$$C_y = a \ln[p] + b \tag{7}$$

The limits of the integral are established as follows: a lower limit 0 that represents the probability of exceedance associated with an event with an infinite return period and an upper limit p_0 that represents the probability of exceedance for which the cost of flood damage. equation (8) shows the expression resulting from this integration and which is the equation used to calculate the cost of the annual risk of flooding.

$$C_F(p) = a\ (p_0 \ln[p_0] - p_0) + b\, p_0 \tag{8}$$

With these concepts defined, the Objective Function (Equation (9)) to be minimized by the model is obtained.

$$OF = \Lambda\,[C_p(D_i) + C_T(V_i) + C_v(D_i)\,] + C_F(p) \tag{9}$$

5 MODEL APPLICATION

The methodology is applied in the network called Balloon located in the north of Italy. The network is made up of 71 pipes and 70 nodes. 75 basins covering an area of 40.88 hectares are connected to the network. The network has a total length of 1848.60 meters. The elevation of the highest point is 229 meters above sea level and at the lowest point the elevation is 221 meters above sea level. These topographical characteristics make the area prone to flooding. The network works entirely in gravity. For its operation, the network has depths of up to 10 meters. Figure 3 shows the Balloon network and the sub catchments that make it up.

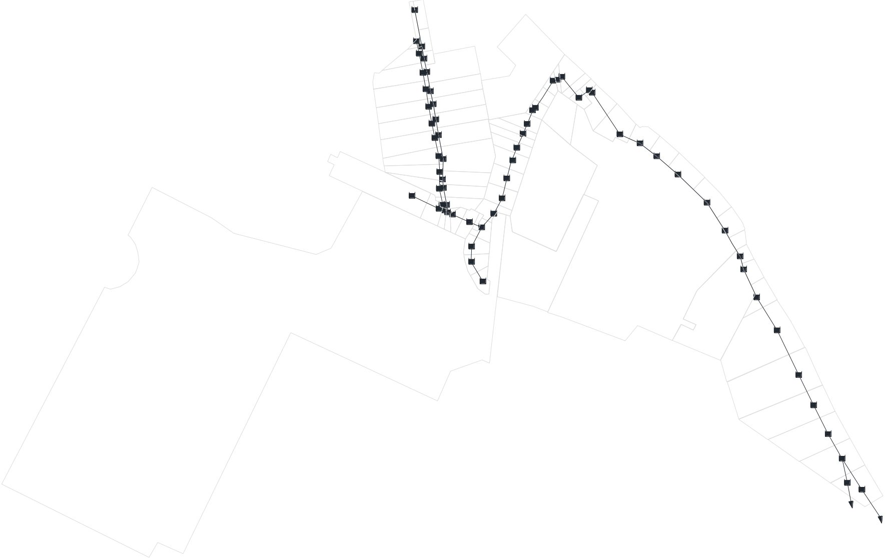


Figure 3. Balloon drainage network.

As for the area studied, it is characterized by being a large area with different land uses that are detailed in Table 1.

Table 1. Land uses of the studied area.

Land uses	Percentage
Streets and roads	28.97%
General trading	18.93%
Parks and playgrounds	10.10%
Education	7.87%
Restaurants	7.56%
Hotels	7.50%
Offices	6.71%
Car parks	3.47%

Churches	2.27%
Museums	1.75%
Health	1.70%
Warehouses	1.68%
Dwellings	1.51%

For the analysis of the risk in economic terms, the IDF curves are used for return periods of 2, 5, 10, 20, 50 and 100 years (Figure 4). With these curves storms of 1 hour duration with 5-minute intervals have been calculated. For the study of the risk of flooding, the 6 storms have been considered.

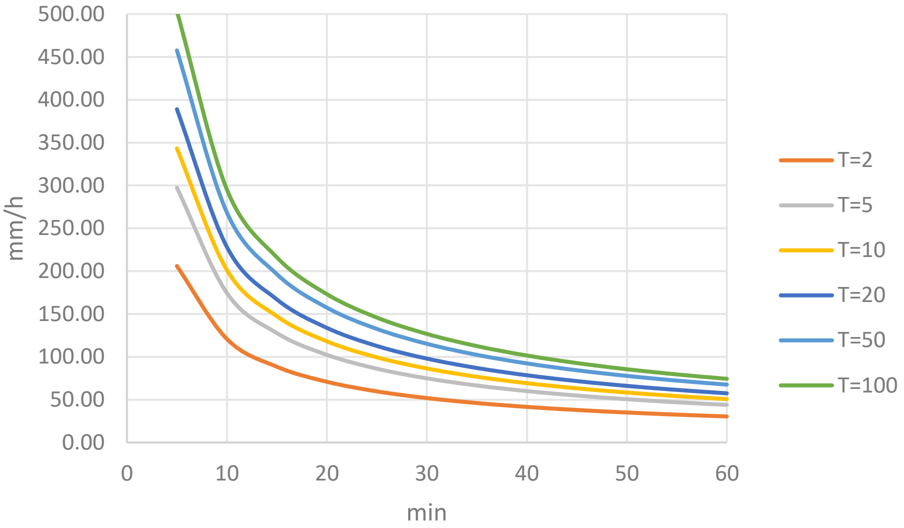


Figure 4. IDF curves for the study area.

In the analyzed scenario, the network in its current state generates economic losses related to flood damage of approximately €170,436.37 per year.

6 RESULTS

6.1 Application of the SSR method

The SSR method is applied in each secondary Branch of the network, three Branches that discharge the water in the main network were defined. Figure 5 shows the different branches that make up the Balloon network. When applying the method in each branch DVs that would not be part of the SS that contains the best solutions to the problem are eliminated. As a result, it is obtained that the elements of Branch L1 and Branch L3 are not part of the SS while only node N63 of Branch L2 is part of the final SS.

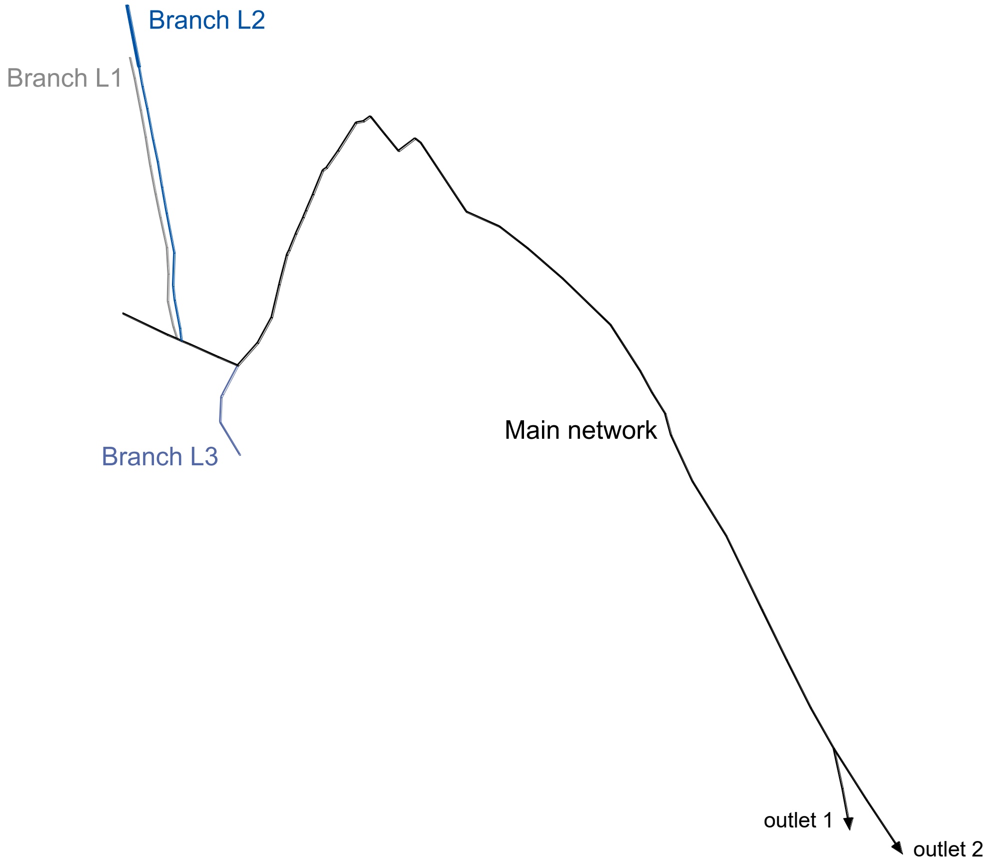


Figure 6. Branches of the Balloon network

Completed the reduction of the SS in each branch, the SSR method is applied in the total Balloon network, eliminating from the analysis the DVs discarded in the analysis of the branches. Applying the method in different iterations reduces the SS by limiting the search to the optimal region. The Table 2, Table 3 and Table 4 show the reduction in each interaction of the SSR method. Table 5 shows a summary of the SS reduction obtained by applying the SSR method compared to the total network without any previous reduction.

Table 2. Results of the first iteration

Pipes	~~C1~~	~~C2~~	~~C3~~	~~C4~~	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39
	C41	C42	C45	C46	~~C58~~	~~C59~~	~~C60~~	~~C61~~	~~C62~~	~~C63~~	~~C64~~	~~C65~~	~~C66~~	~~C67~~	~~C68~~	~~C69~~	~~C70~~	~~C71~~	~~C72~~
	~~C73~~	~~C74~~	~~C75~~	~~C76~~	~~C77~~	~~C78~~	~~C79~~	~~C80~~	C81	C82	C83	C84	C85	C86					
Nodes	~~N1~~	~~N2~~	~~N3~~	~~N4~~	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19
	N20	N21	N22	N23	N24	N25	N26	N27	N28	N29	N30	N31	N32	N33	N34	N35	N36	N37	N38
	N39	N41	N43	~~N45~~	~~N46~~	~~N47~~	~~N48~~	~~N49~~	~~N50~~	~~N51~~	~~N52~~	~~N53~~	~~N54~~	~~N55~~	~~N56~~	N57	~~N58~~	~~N59~~	~~N60~~
	~~N61~~	~~N62~~	N63	~~N64~~	~~N65~~	~~N66~~	~~N67~~	~~N68~~	N69	N70	N71	N72	N73						

Table 3. Results of the second iteration

Pipes	~~C1~~	~~C2~~	~~C3~~	~~C4~~	C5	~~C6~~	~~C7~~	~~C8~~	~~C9~~	~~C10~~	~~C11~~	~~C12~~	C13	~~C14~~	C15	~~C16~~	~~C17~~	C18	~~C19~~
	~~C21~~	C22	~~C23~~	~~C24~~	~~C25~~	~~C26~~	~~C27~~	~~C28~~	~~C29~~	~~C30~~	~~C31~~	~~C32~~	~~C33~~	~~C34~~	~~C35~~	~~C36~~	C37	C38	C39
	C41	C42	~~C45~~	~~C46~~	~~C58~~	~~C59~~	~~C60~~	~~C61~~	~~C62~~	~~C63~~	~~C64~~	~~C65~~	~~C66~~	~~C67~~	~~C68~~	~~C69~~	~~C70~~	~~C71~~	~~C72~~
	~~C73~~	~~C74~~	~~C75~~	~~C76~~	~~C77~~	~~C78~~	~~C79~~	~~C80~~	~~C81~~	~~C82~~	~~C83~~	C84	C85	C86					
Nodes	~~N1~~	~~N2~~	~~N3~~	~~N4~~	~~N5~~	~~N6~~	~~N7~~	N8	N9	~~N10~~	~~N11~~	N12	N13	~~N14~~	N15	N16	~~N17~~	~~N18~~	~~N19~~
	N20	~~N21~~	~~N22~~	~~N23~~	~~N24~~	~~N25~~	~~N26~~	~~N27~~	~~N28~~	~~N29~~	~~N30~~	~~N31~~	~~N32~~	~~N33~~	~~N34~~	~~N35~~	~~N36~~	~~N37~~	~~N38~~
	N39	~~N41~~	~~N43~~	~~N45~~	~~N46~~	~~N47~~	~~N48~~	~~N49~~	~~N50~~	~~N51~~	~~N52~~	~~N53~~	~~N54~~	~~N55~~	~~N56~~	~~N57~~	~~N58~~	~~N59~~	~~N60~~
	~~N61~~	~~N62~~	N63	~~N64~~	~~N65~~	~~N66~~	~~N67~~	~~N68~~	~~N69~~	~~N70~~	N71	~~N72~~	~~N73~~						

Table 4. Results of the third iteration

Pipes	~~C1~~	~~C2~~	~~C3~~	~~C4~~	C5	~~C6~~	~~C7~~	~~C8~~	~~C9~~	~~C10~~	~~C11~~	~~C12~~	~~C13~~	~~C14~~	C15	~~C16~~	~~C17~~	~~C18~~	~~C19~~
	~~C21~~	~~C22~~	~~C23~~	~~C24~~	~~C25~~	~~C26~~	~~C27~~	~~C28~~	~~C29~~	~~C30~~	~~C31~~	~~C32~~	~~C33~~	~~C34~~	~~C35~~	~~C36~~	~~C37~~	~~C38~~	~~C39~~
	~~C41~~	~~C42~~	~~C45~~	~~C46~~	~~C58~~	~~C59~~	~~C60~~	~~C61~~	~~C62~~	~~C63~~	~~C64~~	~~C65~~	~~C66~~	~~C67~~	~~C68~~	~~C69~~	~~C70~~	~~C71~~	~~C72~~
	~~C73~~	~~C74~~	~~C75~~	~~C76~~	~~C77~~	~~C78~~	~~C79~~	~~C80~~	~~C81~~	~~C82~~	~~C83~~	C84	C85	C86	-				
Nodes	~~N1~~	~~N2~~	~~N3~~	~~N4~~	~~N5~~	~~N6~~	~~N7~~	N8	~~N9~~	~~N10~~	~~N11~~	~~N12~~	~~N13~~	~~N14~~	N15	N16	~~N17~~	~~N18~~	~~N19~~
	~~N20~~	~~N21~~	~~N22~~	~~N23~~	~~N24~~	~~N25~~	~~N26~~	~~N27~~	~~N28~~	~~N29~~	~~N30~~	~~N31~~	~~N32~~	~~N33~~	~~N34~~	~~N35~~	~~N36~~	~~N37~~	~~N38~~
	~~N39~~	~~N41~~	~~N43~~	~~N45~~	~~N46~~	~~N47~~	~~N48~~	~~N49~~	~~N50~~	~~N51~~	~~N52~~	~~N53~~	~~N54~~	~~N55~~	~~N56~~	~~N57~~	~~N58~~	~~N59~~	~~N60~~
	~~N61~~	~~N62~~	N63	~~N64~~	~~N65~~	~~N66~~	~~N67~~	~~N68~~	~~N69~~	~~N70~~	~~N71~~	~~N72~~	~~N73~~	-					

Table 5. Magnitude of SS and number of DVs of the network without SSR and the final search space

Scenario	n_i	m_i	p_i	ND	NS	Nθ	DVs	S
Total network	70	71	71	40	5	10	212	233
Final SS	4	5	5	40	5	10	14	15

6.2 Final Optimization

Once the final search region has been delimited, a final optimization is applied using a refined optimization criteria and including the use of hydraulic controls. This optimization demands a greater computational effort, and it is for this reason that it becomes necessary to reduce the SS as a previous step.

Applying the optimization method, a solution is obtained, this solution is assumed to be very close to the optimal one. The proposed solution contemplates the substitution of the C84, C85 and C86 pipes. The installation of storm tanks in the nodes N8, N15, N16 and N63 and a Hydraulic Control element in the C75 pipeline. Figure 7 shows the infrastructure to be installed in the Balloon network. The characteristics of the elements to be installed are shown in Table 6.

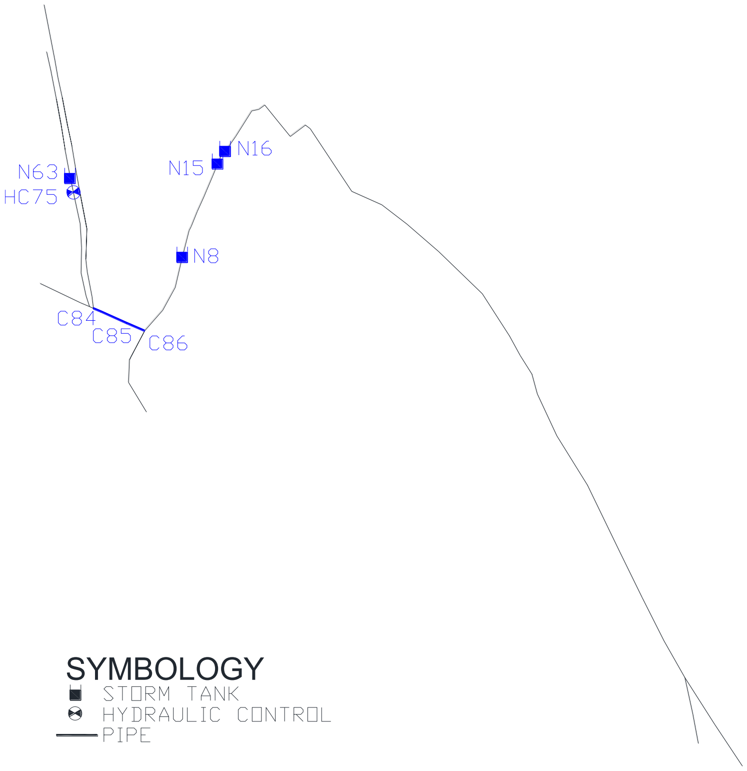


Figure 7. Elements to install in the Balloon

Table 5. Magnitude of SS and number of DVs of the network without SSR and the final search space

Characteristic	Elements								
	C5	C84	C85	C86	N8	N15	N16	N63	C75
Actual diameter (m)	0.70	0.70	0.70	0.70					
Optimized diameter (m)	1.00	1.10	1.00	1.00					
Volume (m^3)					2304	4914	4644	594	
Head-loss (m)									72.55

The costs of each term of the solution are shown in the Table 6. Finally, the Figure 8 shows the minimum value of the OF obtained in each iteration of the SSR and of the final solution.

Table 6. Term of objective function of the final solution

Terms in Objective Function (€/year)				
Flood	Storm Tank	Pipes	Hydraulic Control	Total
5,642.37 €	19,733.83	2,739.02	153.34	28,268.56

CC BY NC SA

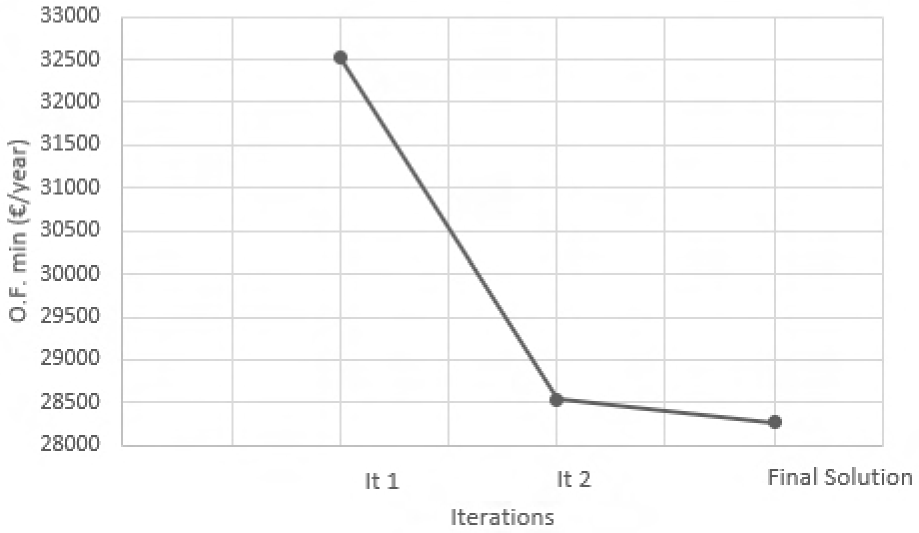


Figure 8. Minimum value objective function in each iteration

7 DISCUSSION

The result of the optimization process demonstrates the suitability of using the proposed method of changing pipes and installing tanks and hydraulic controls in the rehabilitation of the Balloon network. The cost of infrastructure also shows the suitability of the method with an annualized investment of €22,626.19, floods can be avoided that could generate very high costs of damage due to flooding. It should also be noted that the methodology allows for a certain level of flooding, so a cost associated with this damage appears in the objective function. This cost is 3.31% of the annual cost if measures were not taken to rehabilitate the drainage network. A strong point of the presented methodology is to establish the cost of flood damage according to the land uses of the cities. With the differentiation of land uses, the optimization of the network will allow more volume of flooding in areas where the flood does not cause damage or where the damage is minimal, such as in green areas and parking lots, etc. But it will greatly limit flooding in areas that, due to their activity, require greater protection (hospitals, dwelling, etc.). In this regard, it is important to point out that differentiating land uses in the studied area can be a starting point to link this methodology with the Low Impact Development technologies. These techniques can be installed in certain green areas of the study area.

On the other hand, the methodology based on risk analysis analyzes different rainfall events in order to obtain the curve of flood damage - probability. This action demands important calculation times, for this reason, the inclusion of an SSR method is an essential step. The results of applying the method are promising, reducing the SS by 93.56%, this reduction is produced meticulously to define the most promising region of the total SS. This can be seen in Figure 8 where it is observed that in each solution the minimum value of the objective function is reduced.

8 REFERENCES

[1] A. Jha et al., “Five Feet High and Rising Cities and Flooding in the 21 st Century,” Policy Research Working Paper, vol. 5648, May 2011.

[2] J. A. Snyman and D. N. Wilke, Practical mathematical optimization, 2nd ed., vol. 133, Springer Science+ Business Media, Incorporated.

[3] P. L. Iglesias-Rey, F. J. Martínez-Solano, J. G. Saldarriaga, and V. R. Navarro-Planas, “Pseudo-genetic Model Optimization for Rehabilitation of Urban Storm-water Drainage Networks,” in Procedia Engineering, vol. 186, 2017, pp. 617–625.

[4] D. Mora-Melia, P. L. Iglesias-Rey, F. J. Martinez-Solano, and V. S. Fuertes-Miquel, “Design of Water Distribution Networks using a Pseudo-Genetic Algorithm and Sensitivity of Genetic Operators,” Water Resour. Manag., vol. 27, no. 12, Sep. 2013, pp. 4149–4162.

[5] U. A. Ngamalieu-Nengoue, P. L. Iglesias-Rey, F. J. Martínez-Solano, D. Mora-Meliá, and J. G. S. Valderrama, “Urban drainage network rehabilitation considering storm tank installation and pipe substitution,” Water (Switzerland), vol. 11 , no. 3, 2019.

[6] H. R. Maier et al., “Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions,” Environ. Model. Softw., vol. 62, Dec. 2014, pp. 271–299.

[7] S. Sophocleous, D. Savić, and Z. Kapelan, “Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction,” J. Water Resour. Plan. Manag., vol. 145, no. 7, Jul. 2019, p. 04019024.

[8] U. A. Ngamalieu-Nengoue, P. L. Iglesias-Rey, and F. J. Martínez-Solano, “Urban drainage networks rehabilitation using multi-objective model and search space reduction methodology,” Infrastructures, vol. 4, no. 2, Jun. 2019, p. 35.

[9] L. Bayas-Jiménez, F. J. Martínez-Solano, P. L. Iglesias-Rey, D. Mora-Melia, and V. S. Fuertes-Miquel, “Inclusion of Hydraulic Controls in Rehabilitation Models of Drainage Networks to Control Floods,” Water, vol. 13, no. 4, Feb. 2021, p. 514.

[10] L. Bayas-Jiménez, F. J. Martínez-Solano, P. L. Iglesias-Rey, and D. Mora-Meliá, “Search space reduction for genetic algorithms applied to drainage network optimization problems,” Water (Switzerland), vol. 13, no. 15, Jul. 2021, p. 2008.

[11] G. Freni, G. La Loggia, and V. Notaro, “Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation,” Water Sci. Technol., vol. 61, no. 12, Jun. 2010, pp. 2979–2993.

[12] A. S. Olsen, Q. Zhou, J. J. Linde, and K. Arnbjerg-Nielsen, “Comparing Methods of Calculating Expected Annual Damage in Urban Pluvial Flood Risk Assessments,” Water, Vol. 7, vol. 7, no. 1, Jan. 2015, pp. 255–270.

[13] P. Samuels et al., Language of Risk – Project Definitions, 2nd ed., FLOODsite, Apr 2009.

[14] B. Merz, H. Kreibich, R. Schwarze, and A. Thieken, “Review article ‘assessment of economic flood damage,’” Natural Hazards and Earth System Science, vol. 10, no. 8, 2010, pp. 1697–1724.

[15] Butler and David, “Urban Drainage” , 4th ed., CRC Press, Apr. 2018.

[16] N. W. Arnell, “Expected Annual Damages and Uncertainties in Flood Frequency Estimation,” J. Water Resour. Plan. Manag., vol. 115, no. 1, pp. 94–107, Jan. 1989.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14191

SURVEY EXPLORING CUSTOMER COMPLAINT MANAGEMENT AND SMART WATER TECHNOLOGY ADOPTION ACROSS U.S. WATER SYSTEMS

Morgan DiCarlo[1], Emily Zechman Berglund[2], Nikhil Kaza[3], Andrew Grieshop[4], Luke Shealy[5] and Adam Behr[6]

[1,2,4,5,6] North Carolina State University, Department of Civil, Construction and Environmental Engineering
Raleigh, NC, USA
[3] University of North Carolina at Chapel Hill, Department of City and Regional Planning
Chapel Hill, NC, USA

[1] mdicarl@ncsu.edu, [2] emily_berglund@ncsu.edu, [3] nkaza@unc.edu, [4] apgriesh@ncsu.edu,
[5] ldshealy@ncsu.edu, [6] abehr@ncsu.edu

Abstract

High profile water quality events, including the Flint Lead in Water Crisis, have contributed to a decline in customer trust in their water utilities. For example, a recent study indicates as many as 60 million Americans do not drink tap water because they perceive risks in the cleanliness of water and do not trust water providers [1]. One way that utilities can build trust with customers is through improved management of customer complaints. Utilities can store, track, visualize, and share customer complaints to improve service and improve the way that customers interact with information about water quality. Smart water technologies, including Advanced Metering Infrastructure (AMI), data portals, and personal device applications (apps) can be utilized to better communicate with customers. Advanced data analytics can improve the insight that is gained about the source of water quality problems.

We surveyed utilities about their perspectives on trust and customer complaint management. This research explores the development, implementation, and results of a survey instrument distributed to water service providers across the United States. Survey questions explore the existing tools that utilities use to collect customer complaints, the adoption of smart technology by utilities, and characteristics of customer complaints. This research will assess capabilities to detect issues from customer complaints trends and the level of smart technology integration in United States water systems. We employ cross sectional analytical techniques to assess differences in complaint reporting and management system by utility size, urbanization, and socioeconomics of their service area. This research will develop new insight about the types of tools that utilities need and are willing to adopt to receive, analyse, and report customer complaints.

Keywords
Water Utilities, Smart Technology, Survey, Complaint Management, Smart water systems.

1 INTRODUCTION

Water utilities across the United States encounter many challenges in treating and delivering water [2]. Ageing infrastructure leads to pipe breaks and bursts, large volumes of lost water and revenue, and contamination of water supply [2]. Underfunded utilities are limited in their ability to expand water networks to meet growing demands, renew and restore aging pipes, update computational systems with advanced technology, and improve levels of service [2]. A recent study estimated that as many as 60 million Americans do not drink public water because they do not trust its safety [1]. Confidence in the quality of tap water declined at a particularly alarming rate from 2017-2018, after the Flint Water Crisis, in which toxic levels of Lead were discovered at multiple households connected to a water distribution system in Michigan [1]. The attention and timely response of utilities to address customer complaints can build public trust and track system wide quality trends [3]. Fostering communication with customers is one approach to improve customer confidence in water utilities [4]. Customers typically communicate with utilities to by reporting water quality issues, such as complaints about water discoloration, odor, or taste. In addition, analysis of trends in customer complaints can identify episodic water quality problems or water contamination events [5].

Communication between water utilities and customers can be enhanced through smart technologies such as social media, text alerts and data visualization platforms [4]. Water utilities are adopting smart technologies to track consumption, infrastructure performance, and operations[6]..Utilities can take advantage of real-time sensors, customer web portals, and social media to create transparency around water utility responses to complaints: social media expands the reach of utility communications [7]; Advanced Meter Infrastructure (AMI) provides subhourly insights about water consumption [8]; short message service (SMS) alert systems can quickly contact customers about a water quality incident [9]; data visualization platforms show household consumption trends to encourage conservation [10]; real time pressure sensors and machine learning models characterize water-end uses [11]; and Digital Twin, a computational replica of a real world pipe system that mimics its behavior can create new real-time insight around hydraulic performance of pipe systems [12]. There are different ways in which these smart technologies are useful, such as new avenues of data collection and enhancing information transfer between water service providers and consumers. However, it is not well understood how utilities are using smart technologies to build trust and improve customer satisfaction.

We conducted a survey to explore the processes and tools, including smart technologies, that utilities use to receive and store customer complaints. More than 500 respondents, representing community water systems (CWS) across the United States, participated. Data were analysed to evaluate how utilities communicate with customers, how they use smart technology, and how they use customer complaint data for operational insights. This research develops new information about the types of digital tools that utilities need and their approaches to receive, analyse, and report customer complaints, sharing an important perspective on some of the trust-building actions of water providers.

2 METHODS

2.1 Survey Instrument

The survey was developed and distributed to contact emails acquired using an internet scraper searching publicly available, online records of organizations that manage community water systems in the United States. All participants consented to sharing information about their water service operations, and the instrument was approved by an NC State University Internal Review Board. The survey was 46 questions in length and took approximately 15 minutes to complete. Respondents who indicated their consent and self-identified as a part of an organization that provides water services were subsequently shown the full survey. No incentives were offered. Responses were collected across the entire United States over a period of three weeks in September 2021. In total, there were 504 quality-controlled responses to the online survey distributed via Qualtrics surveying software. Qualtrics conducted a soft-launch survey pilot to evaluate the survey instrument and employed quality measures including flagging responses submitted in less than a third of the average response time, incoherent responses, and inconsistent responses. The topics of questions that were included in the survey are:

- Respondent job title
- Water system locale, size, and primary source water
- Types of smart technologies in use
- Budget for smart technology
- Barriers to smart technology adoption
- Number of staff in complaint management
- Frequency of and challenges in communicating with customers
- Method receiving and storing of customer complaints
- Frequency of complaints by complaint type (i.e., taste, odor, color etc.) and season
- Applications of complaint data in trend tracking and concerns about data storage

2.2 Sample Characteristics

Survey respondents varied widely in location, size, urbanization, and other operational details. More than a fourth of responses came from very small community water systems, defined as servicing fewer than 500 connections (26.5%). Most of the respondents (32.3%) were from small, community water systems, which serve populations from 501 to 3,300 households (Table 1). Another 17% of responses were from medium providers servicing 3,301 to 10,000 households, and 17.6% were large utilities with up to 100,000 service accounts. The remaining 4% of responses come from very large utilities with more than 100,000 water connections, including three respondents who reported serving more than 1 million people (Table 1). Table 1 shows data from the U.S. Environmental Protection Agency (EPA) about how many utilities are in the US overall, by size. This survey collected responses from water systems which service about 5.1% of the US population [13]. Small and very small utilities, serving less than 3,300 or 500 connections, are particularly well captured in this survey, compared to large utility surveys that were conducted recently [14]. The total number of responses (N=504) is a statistically representative sample size for total number community water systems (CWS) in the United States [15].

Respondents self-described their service areas as 51% rural, 13% suburban, 32% small city or city outskirts and 3% large cities. Most of the water providers (62%) reported groundwater as their region's primary water source, with 31% reporting surface water sources, and another 7% indicating other water sources such as bought from another supplier, spring water or cisterns. Eighty percent of the water providers surveyed are publicly owned, 12% were private companies, and the remaining responses indicated other ownership cases such as quasi-public, quasi-governmental, or a homeowner’s association. There were at least 30 responses received from each US Census Bureau region, with most responses (196) from the Southern United States.

Table 1 Sample Summary- Size Breakdown [a] 2020 U.S. population is 329 million [15]. Some estimates of utility populations are taken from previous census. [b] Data from U.S. EPA, 2021 [13]

Utility Size	Amount Served	Number of Responses	% of Responses	Total Population Served	% US population served	Number of CWS in the US [b]
Very Small	< 500	134	26.5	35,391	0.01%	26,963
Small	501-3,300	178	32.3	258,376	0.08%	13,334
Medium	3,301-10,000	86	17.0	514,099	0.16%	5,022
Large	10,001-100,000	89	17.6	2,788,882	0.85%	3,975
Very Large	>100,000	18	3.6	13,047,691	3.97%	446
	Total	504	100	16,644,439	5.06%[a]	49,740

3 SURVEY RESULTS AND DISCUSSION

3.1 Approaches to Customer Complaint Management

Responses to survey questions about approaches to manage customer complaints are explored by the size of the utility's service area. The ability of a utility to efficiently handle customer complaints is likely influenced by a broad range of factors, including size, number of staff, and importantly, the sheer volume of complaints received. Survey results indicate a wide variation in number of complaints received per week by utility size (Figure 3). Most respondents from very small (90.9%), small (91.1%) medium (80.2%), and even large (62.5%) utilities receive less than five complaints per week (Figure 1). The volume of complaints generally increases by size, and 42.1% of the very large utilities that responded receive 6 to 25 complaints per week, and 10.5% receiving more than 50. These findings demonstrate that given adequate maintenance and customer service staff, most utilities receive a small and manageable volume of complaints weekly. As utilities increase in size, growing numbers of complaints begets additional tools or smart technologies that can support response and resolution.

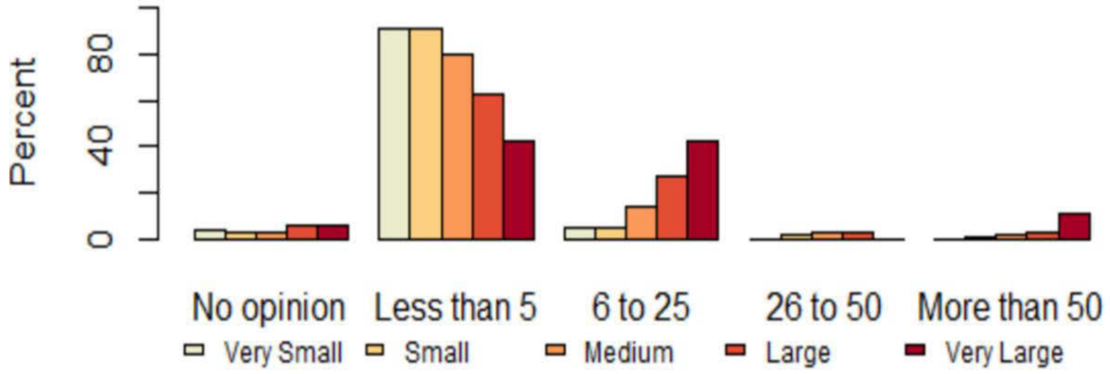


Figure 1. Percent of responses by size to "How many customer complaints are there per week?"

Analog records are less common as utility size increases (Table 2). Most large (59.1%) and very large (84.2%) utilities store records of customer complaints digitally, and all of the very large utilities are storing records in some capacity. Small (49.4%) and very small (33.3%) providers report mainly storing records on paper. Among very small providers serving less than 500 connections, 21.2% do not store records of customer complaints at all. Methods of complaint storage are important to gauge whether a utility can respond both to individual complaints that may reoccur, and clusters of complaints that can be indicative of system wide issues. For example, Gallagher and Dietrich implemented statistical analyses on the content of customer complaints from six utilities and found that incidences where there was both a high frequency of complaints and consistent descriptors in the data coincided with real episodic water quality problems [5]. With inaccessible or non-existent records, additional analyses of complaints that can yield operational insights are not possible.

The survey also explored whether complaint data is used to track system wide trends. Roughly half of very large utilities track system wide trends from complaint data (52.6%). All other utility size more often indicated that they do not use customer complaint data to track system wide trends. Table 3 presents a contingency table of how utilities store complaint data and their propensity to track system wide trends. Forty-two percent of utilities that mostly use digital storage methods of complaint data also reporting using that data to track system wide trends, while only 34.4% of those with paper storage and 6.7% of those not storing complaint records reporting trend tracking (Table 3). This supports the concept that digitalization and the integration of smart technologies is useful for utility efforts to make system wide improvements.

Table 2 Contingency table of the percentage of utilities that store complaint data on paper/digital by size

	Not stored	More on paper	Half and half	More digitally	No opinion
Very small	21.21	33.33	15.91	23.48	5.30
Small	12.78	49.44	11.67	23.89	2.22
Medium	12.79	29.07	13.95	39.54	4.65
Large	5.68	12.50	19.32	59.09	3.41
Very large	0.00	5.26	5.26	84.21	5.26

Table 3 Contingency table of utilities that store complaint data on paper/digital and their tracking of system wide trends

	Not stored	More on paper	Half and half	More digitally	No opinion
Yes, tracks trends	6.7%	34.4%	14.9%	42.1%	2.1%
No	18.7%	33.7%	11.9%	32.9%	2.8%
No opinion	12.3%	31.6%	22.8%	19.3%	14.0%

A set of open-ended questions further explored uses of customer complaint data. First, utilities were asked to indicate whether they felt their customer complaint data was utilized to its full potential. If yes, utilities were prompted to write about the ways they use their data at present; if no, utilities were asked about information they want to access from their data. A selection of free-form responses is reported in Table 4. Water providers that indicated they are using the data to its full potential frequently cited that there very few complaints, such that they can effectively solve customer issues from complaints on a case-by-case issue as soon as they arise. Additionally, responses described that complaint information is used to guide decisions about capital improvements and to identify areas where infrastructure repairs are needed. Utilities explained that they desire for methods or algorithms to improve understanding of system wide water quality, flag customers with payment issues, showcase the utility's efforts to help the public better understand water operations, and identify root causes of customer issues. These are fertile areas for future research and product development.

Table 4 Selection of free responses to questions about data utilization

You indicated more potential uses of customer complaint data. Please elaborate on potential applications of this information.	**You indicated that you are fully utilizing your customer complaint data. Please elaborate on your applications of this information.**
"To see trends or re-occurring issues"	"Track problems and make repairs"
"Could be used to identify constant problem customers and those who have payment issues"	"When problems come in, we investigate and address. If multiple complaints occur, we look for the cause"
"Flushing programs, line replacement priority"	"Customer complaints are used in the maintenance or replacement of infrastructure."
"We store the info in … monthly report. Really have no method to retrieve the data without reviewing e-mails or reading each report."	"Look for trends of pressure complaints"
"Should be able to group customer concerns & determine if there is an identifiable cause/solution over time."	"Complaints are rare and handled as they arise."
"Target certain neighborhoods to better help them… fix leaks at homes, address aging homes and repeated leaks, replace lead lines, help with affordability "	"We store and review complaints and responses and work to minimize issues in the future through capital projects and education and outreach."
"Seasonal trending of water quality; trending of complaints with ongoing upkeep (flushing and pipe pressure testing); identifying problem areas"	"We are a small rural system, so we rarely have any complaints, and when we do have complaints, we go straight to the complaint area and work directly with the customer"
"Identify recurring issues, show customers that these issues do occur, but can be - and are - fixed when we know about them."	"Not really relevant - we resolve issues as they infrequently arise"
"We could compile the data and release it regularly, rather than periodically. "	"We use the data to develop capital improvements and operation planning (flushing, sampling, disinfection)"

3.2 Smart Technology Adoption

Smart technologies integrate digital and internet-enabled tools to automate, streamline, and improve the operations of a system [6]. Figure 2 reports the percent of utilities by size that are already implementing various types of smart technologies. Social media can open a two-way dialogue between utility and customers and allow for more rapid communication transfer [7]. A large portion of utilities, regardless of size, indicated they already have a social media presence (Very small- 41.6%, Small- 57.8%, Medium- 69.7%, Large- 77.3% and Very large- 94.7%). Another commonly cited adoption was smart water meters. Smart meters were less common among very small providers (28.8%), but with increasing frequency by size, and 46.1% of small, 67.4% of medium, 77.3% of large and 78.9% of very large utilities have smart meters currently deployed.

Several technologies are well integrated at large utilities and more sparingly adopted among smaller providers. For example, hydraulic models are implemented by 89.5% of very large providers and 62.5% of large providers, but under 50% in other groupings; short message service (SMS) or text alert systems are in use by 63.2% of very large utilities but under 40% among smaller sizes; pressure sensors are adopted by 78.9% of very large utilities (Figure 2). Less than 40% of any utility grouping employs data dashboards that can visualize household consumption trends or allows consumers to regularly view their usage online. Implementation of data dashboards does steadily increase by utility size (Figure 2).

Only 15.8% of very large utilities and less than 5% of all smaller sizes reported using artificial intelligence or machine learning analytics, indicating these approaches are still cutting-edge and lack widespread uptake at present (Figure 2). Similarly, less than 2% of utilities of any size indicated using Digital Twins [12]. Results show differences in the current adoption of smart technologies by utility size, and, specifically, smaller utilities may have unequal access to tools that can improve infrastructure monitoring, operation, and management such as hydraulic models and real-time pressure sensors.

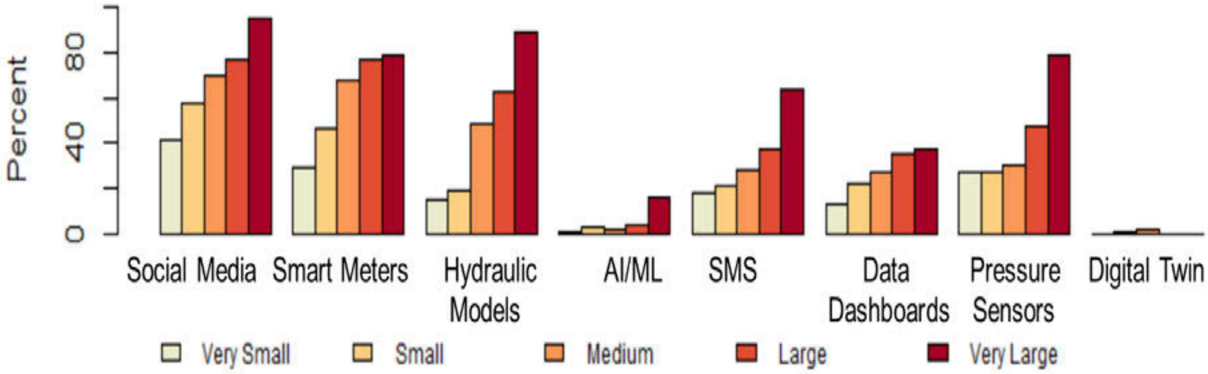


Figure 2. Percent of responses by size to "What smart technologies are currently in use?"

Most utilities, regardless of size, reported that it would be at least somewhat challenging to incorporate new smart technologies in their service area (Figure 3). Another question offered a list of possible barriers to the addition of smart technologies, and respondents indicated all which applied. Primary barriers to adoption of additional smart technologies across all respondents were finances (38%). Other barriers to the adoption of smart technologies included not enough personnel (23%), lack of staff training (22%), and that smart technologies should not be adopted because they add no value for the utility (11%). The survey also asked about smart technologies budgets, which is important because finances was the most frequently reported barrier to the adoption of smart technologies. As anticipated, budgets increased on average with utility size, and more than a quarter of very large utilities indicated at least $500,000 USD for smart technologies.

Six percent of respondents provided text responses to the question about barriers to the adoption of smart technologies and provide new insight into range of challenges utilities may face. Several cited lack of customer participation to implement new services, including a description of a "large senior population [with] low adoption of paperless billing and electronic form submission". Another theme of the text responses was trepidation about new techniques, such as responses stating "concerns about security", "personnel resistant to change", and "poorly established

BY NC SA

innovation goals". Finally, several respondents discussed difficulties managing rural and remote water systems that impede smart technology integration: "our service area is spread out", "terrain and geographic issues" and "limited internet and phone". These insights give important context for the bounds of digital transition in water systems, which relies on adequate funding, privacy and security measures, staff training, and the performance of interconnected public services such as internet and phone.

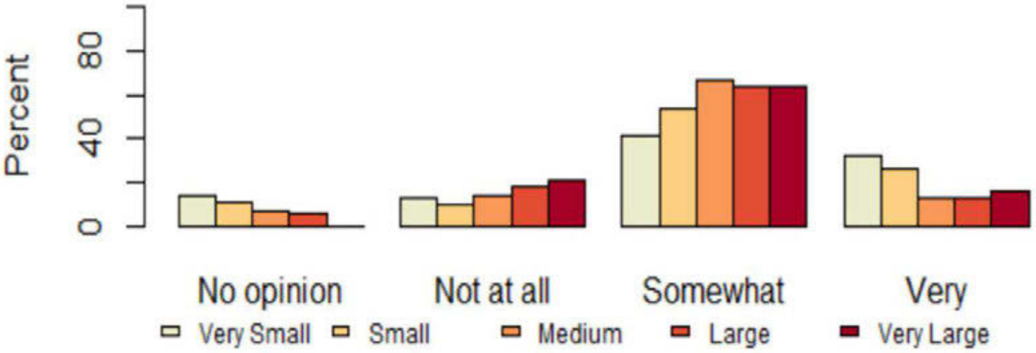


Figure 3. Percent of responses by size to "How challenging is it to adopt additional smart technology?"

4 CONCLUSION

In this research, a survey is developed and distributed to water providers across the United States to better characterize customer complaint management techniques and smart technology implementation. More than 500 water utilities responded, representing a statistically significant portion of the total number of community water systems in the United States. Respondents ranged from very small systems with less than 500 connections to urban centres with more than 100,000 households served. Responses demonstrated a range in attributes that affect system operations such as urbanization, geographic region, ownership, and primary drinking water sources.

Results reveal a dichotomy in funding, smart technology implementation, and customer complaint management approaches by utility size. Larger utilities tend to have digital storage of complaint records. Nearly a quarter of very small water providers do not store records customer complaint records at all. Some may view these records as unnecessary, given that results indicated most small providers see under five complaints per week. As a comparison, 10% of the largest utilities surveyed manage more than fifty complaints weekly. Utilities need information from data to make managerial decisions about infrastructure investment, in applications for additional funding, and to monitor system performance. Paper records or lack of records altogether are an area of improvement in which U.S. water utilities could use support. New complaint management systems can document this data automatically and securely over time. Further, less than half of all providers surveyed report applying complaint data to track system wide trends. Without collecting data and transforming it into actionable information, water systems may fail to detect systematic issues, such as the Flint lead crisis, while handling customer complaints in siloes.

Smart technologies including internet-enabled water meters, social media, hydraulic models, real-time pressure sensors and data visualization platforms help streamline utility operations, improve communication with customers, and offer structure for storing and processing customer complaint data. More than a quarter of respondents, regardless of size, already use social media and smart meters. Smart technology funding varies widely, with larger utilities being better funded, yet with many smaller utilities indicating having at least $50,000 USD to spend. Larger utilities indicated advanced capabilities more often than other groupings, including high instances of hydraulic models, text alert systems, and real time pressure sensors.

The survey collects free responses in addition to the measurable results, adding rich information about utility perceptions that can lead to new research directions. The water providers surveyed desire to learn more information for decision making from the data they collect. Particularly, utilities described want to use data to identify areas for repair and infrastructure improvements, to cluster problems for trend identification, and for education and outreach purposes. Finances and personnel issues are the main barriers to the digital transition of U.S. water systems, but other free responses about navigating rural terrain and lacking internet coverage are worthwhile areas of further investigation. Overall, this research identifies areas of improvement in customer complaint management and contribute to a more complete picture of current smart technology capabilities and barriers to further adoption in the United States.

5 REFERENCES

[1] Rosinger, A., Patel, A., & Weaks, F. (2021). Examining recent trends in the racial disparity gap in tap water consumption: NHANES 2011–2018. Public Health Nutrition, 1-7. pubmed.ncbi.nlm.nih.gov/34114536/

[2] American Society of Civil Engineers (ASCE). (2017). Water Infrastructure Report Card. Retrieved from https://www.infrastructurereportcard.org/cat-item/drinking_water/

[3] Grupper, M. A., M. G. Sorice, M. J. Stern, and M. E. Schreiber. 2021. Evaluating determinants of social trust in water utilities: implications for building resilient water systems. Ecology and Society 26(4):41. https://doi.org/10.5751/ES-12833-260441

[4] Junker, M. and Carpenter, A.T. (2021), Improving Public Perceptions of Tap Water. J Am Water Works Assoc, 113: 66-72. https://doi.org/10.1002/awwa.1692

[5] Gallagher, D. L., & Dietrich, A. M. (2014). Statistical approaches for analyzing customer complaint data to assess aesthetic episodes in drinking water. Journal of Water Supply: Research and Technology - AQUA, 63(5), 358-367. doi: http://dx.doi.org/10.2166/aqua.2014.144

[6] Berglund, E. Z., 2020. "Smart infrastructure: A vision for the role of the civil engineering profession in smart cities." J. Infrastruct. Syst. 26 (2): 03120001. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000549.

[7] Heath, A. (2020), COVID-19 Water Contamination Concerns Underscore Need to Engage With Consumers. J Am Water Works Assoc, 112: 20-25. https://doi.org/10.1002/awwa.1573

[8] Beal, C., Flynn, J. Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs, Utilities Policy, Volume 32, 2015, Pages 29-37, ISSN 0957-1787, https://doi.org/10.1016/j.jup.2014.12.006.

[9] Strickling, H., DiCarlo, M., Shafiee, E.M., and Berglund, E.Z. (2020). Simulation of Containment and Wireless Emergency Alerts within Targeted Pressure Zones for Water Contamination Management, Sustainable Cities and Society, 52, 101820.

[10] Novak, J. M. Melenhorst, I. Micheel, C. Pasini, P. Fraternali, and A.E. Rizzoli. (2018). Integrating behavioural change and gamified incentive modelling for stimulating water saving. Environmental Modelling Software, 102:120 – 137. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2017.11.038.

[11] Cominola, A., Giuliani,M., Piga, D., Castelletti, A., Rizzoli, A. (2015) Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review, Environmental Modelling & Software, Volume 72, Pages 198-214, ISSN 1364-8152

[12] Pesantez, J., Alghamdi, F., Sabu, S. Mahinthakumar, G., Berglund, E.Z. (2022) Using a digital twin to explore water infrastructure impacts during the COVID-19 pandemic, Sustainable Cities and Society, Volume 77, 103520, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2021.103520.

[13] U.S. Census Bureau 2020. QuickFacts-United States. Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045221

[14] Evans, J., Carpenter, A. (2019) Utility approaches to evaluating the effectiveness of consumer confidence reports, Utilities Policy, Volume 58, Pages 136-144, ISSN 0957-1787, https://doi.org/10.1016/j.jup.2019.05.004.

[15] U.S. Environmental Protection Agency (EPA) (2021) Government Performance and Results Act

(GPRA) Inventory Summary Report. Retrieved from https://www.epa.gov/ground-water-and-drinking-water/drinking-water-performance-and-results-report

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14200

ESTIMATION OF THE SPATIAL DISTRIBUTION OF SUBSTANCES IN ANAEROBIC DIGESTION TANKS WITH CFD

Soroush Dabiri[1], Prashant Kumar[2] and Wolfgang Rauch[3]

[1,2,3]Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria

[1] *soroush.dabiri@uibk.ac.at,* [2] *prashant.kumar@uibk.ac.at,* [3] *wolfgang.rauch@uibk.ac.at*

Abstract

Anaerobic digestion (AD) is used in wastewater treatment plants (WWTPs) to decompose organic matter. Highly efficient AD tanks are able to mix biochemical substances within the tank, and – in order to assess such spatial distribution of substances - a proper modelling method is necessary. Computational fluid dynamics (CFD) is a simulation tool for modelling fluid flows in WWTPs, but there is a lack of studies on CFD modelling of hydrodynamics together with an integrated modelling of bio-kinetics due to the complexity of biochemical reactions in AD processes. The current study aims at estimating the distribution of biochemical components within an AD tank. For doing this, an integrated modelling of hydrodynamics and bio-kinetics is conducted, in order to assess mixing quality. The novelty of the work is the derivation of a new AD-related solver in an open source CFD platform.

As the first step, the hydrodynamics of the fluid flow in our WWTP are estimated through iterative solution of fluid flow equations, and subsequently, by fixing the obtained velocity vector field, bio-kinetic equations are applied to the flow field. The bio-kinetics in our solver are based on ADM1, the state of the art transient model for AD process. The validation of our developed solver is done by comparing the results to a fully-mixed digester in an experimental setup. Then, the results of our hydrodynamic modelling are shown as the velocity profile and streamlines. Subsequently, regarding bio-kinetics, the concentrations of substances are estimated and plotted during the simulation time. In order to evaluate the distribution of the organic material within the tank, the uniformity index for the substances is analysed.

Keywords
wastewater treatment plant modelling, computational fluid dynamics, anaerobic digestion modelling, ADM1 modelling.

1 INTRODUCTION

Anaerobic digestion (AD) is often used in wastewater treatment plants (WWTPs) to decompose organic matter. Based on organic loading rate, mixing and the geometry of digesters, the distribution of biochemical components can vary. Mixing in anaerobic digestion (AD) tanks is of importance, to assure proper stabilizing organic wastes and efficient biogas production [1]. Computational fluid dynamics (CFD) are a valuable tool for investigating the flow field within digesters [2]. However, how to set proper initial and boundary conditions has always been an important question. e.g., the inlet configuration is an important issue in case of sludge recirculation [3]. Although many CFD investigations have been done to model the behaviour of mixers and gas injection in bioreactors, fluid recirculation specifications need to be investigated as well [4]. Additionally, mixing time should be analysed to give insight about the effect of the inlet configuration. Thus, we need to model a transient setup in real-scale to investigate sludge recirculation. In order to assure a homogeneous distribution of substrates, a proper modelling method is necessary.

Computational fluid dynamics (CFD) is a simulation tool for modelling fluid flows in AD tanks [5, 6]. CFD modelling of hydrodynamics together with an integrated modelling of bio-kinetics makes a recipe for a comprehensive modelling method in AD tanks. However, due to complexity of biochemical reactions in AD process, current CFD modelling studies usually focus mainly on the fluid flow within the digester, rather than the bio-kinetics [7, 8]. Thus, this study aims at estimating the distribution of biochemical components within an AD tank, which employs a sludge recirculation system, through an integrated modelling of hydrodynamics and bio-kinetics.

In the following, investigate the behaviour of the real world AIZ wastewater treatment plant in Tyrol, Austria represented by means of a full-scale model. The sludge is recirculated within the AD tank and through our two-step method the distribution of the biochemical material is estimated in 100 days, by using the CFD tool, OpenFOAM. The non-Newtonian characteristics of the sludge and proper turbulent enclosure are taken into account. The novelty of our project is applying ADM1 bio-kinetics to an open source CFD solver, which promotes the potential use of ADM1 in CFD platforms, in order to assess mixing quality in WWTPs.

2 METHOD

2.1 Procedure

In this study, initially, the hydrodynamics of the sludge recirculation is evaluated through an iterative solution of fluid flow equations (the geometry, meshing procedure, and the mixing simulation are also published in [9]), and subsequently, by fixing the obtained velocity field, bio-kinetic equations are applied to the velocity field. This is because there is a huge difference between of the required simulation time steps for fluid flows and in biochemical reactions.

2.2 Hydrodynamics modelling

Our case study is an egg-shaped AD tank, whose diameter reaches up to 15.4 m at the middle. The height of the sludge level is about 22.9 m. The outlet tube is 0.2 m in diameter, located at the bottom of the digester and extended from inside to the outside of the tank. According to [9], the splashing inlet configuration is investigated as emulating the sludge recirculation within the tank.

After designing the geometry a mesh analysis study is a necessity, in order to implement a proper mesh, which is explained in [9]. The final mesh network consists of 404878 elements.

For obtaining the hydrodynamics, a steady-state solver, simpleFoam, for incompressible flow is employed in OpenFOAM, which is an open source CFD software, based on finite volume method. This solver utilises the equations of conservation for mass (continuity) and momentum (Navier-Stokes).

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{V}) = 0 \tag{1}$$

$$\frac{\partial (\rho \vec{V})}{\partial t} + \vec{\nabla}(\rho \vec{V} \vec{V}) = -\vec{\nabla} \mathrm{p} + \vec{\nabla} \cdot (\bar{\bar{\tau}}) \tag{2}$$

where t is the time (for transient simulation), ρ is the density of the fluid and $\vec{V}$ is the fluid velocity, p represents the static pressure and $\bar{\bar{\tau}}$ is the stress tensor

In addition, since we have a high Reynolds number, we need to use a turbulent enclosure sub-model, which in this case we choose as k-ε model due to its robustness. Additionally, for modelling the non-Newtonian behaviour of the sludge, the power law equation is used, taking into account the total solids (TS) concentration of 5% (see [9]).

2.3 Bio-kinetics modelling

The bio-kinetics within an AD tank are based on ADM1, a comprehensive dynamic model for AD processes [10]. The concentration of ADM1 components are estimated through scalar transport equations [11]. ADM1 equations are classified in the following. With the start of the digestion process, the biomass (X_c) is disintegrated:

$$\frac{dX_i}{dt} = k_{dis} X_c \tag{3}$$

where X_i is the either carbohydrates, proteins or lipids, and k_{dis} is disintegration rate. Then, the first step in AD is the hydrolysis of the material, which is calculated as follows:

$$\frac{dS_{sub}}{dt} = k_{hyd,i} X_i \tag{4}$$

where S_{sub} is the concentration of produced substrate and $k_{hyd,i}$ is the hydrolysis rate. Afterwards, the uptake rate of each soluble substrate depends on its own concentration:

$$\frac{dS_{sub}}{dt} = k_{m,i} \frac{S_{sub}}{K_{S,i} + S_{sub}} X_{bac} I \tag{5}$$

where $k_{m,i}$ and $K_{S,i}$ are the Monod maximum rate of uptake and half saturation value for the process *i*. X_{bac} is the bacteria and *I* represent the inhibitions, which are classified as pH, nitrogen, ammonia and hydrogen inhibitions. The decay rate of bacteria should be considered as follows:

$$\frac{dX_i}{dt} = k_{dec,i} X_i \tag{6}$$

where $k_{dec,i}$ is the decay rate for each degrader bacteria.

The hydrodynamic equations are solved through an OpenFOAM solver, simpleFoam in steady-state, and the bio-kinetic equations are solved via a developed solver that we denote as passivScalarADMFoam.

3 VALIDATION

3.1 Hydrodynamics validation

As we have conducted our model in 2D, a two-step simulation set should be conducted for comparing the 2D and the 3D results, which has been done and explained in [9].

Moreover, regarding validation of hydrodynamics, the velocity profile at the centre of the AD tank as computed by OpenFOAM is compared to a previously validated ANSYS Fluent simulation results – see [9]. Both OpenFOAM and ANSYS results depict a good match, with differences less than 1%.

3.2 Bio-kinetics validation

With regard to modelling the bio-kinetics within the AD tank, the model is validated by a previous experimental set in a fully-mixed lab-scale digester [12] as suggested by Wu [13]. They conducted a set of experiments at mesophilic temperature in a 25-day period. The digester had 0.06 m3 volume and was fed daily with 0.63 kg COD/m^3 of biomass.

As the model is fully-mixed - for validation purposes - we can remove the spatial variation in our calculation domain and reduce the mesh network to a single cell with an inlet and an outlet. A single vector represents the velocity field within the cell, the amount of which corresponds to the 25-day HRT in the tank. *Figure 1* shows the results of our passiveScalarADMFoam solver, compared to Fatolahi experiments [12], where only 9% of difference is observed between the biogas yield in our solver and Fatolahi experiments, which validates the efficiency of our developed solver.

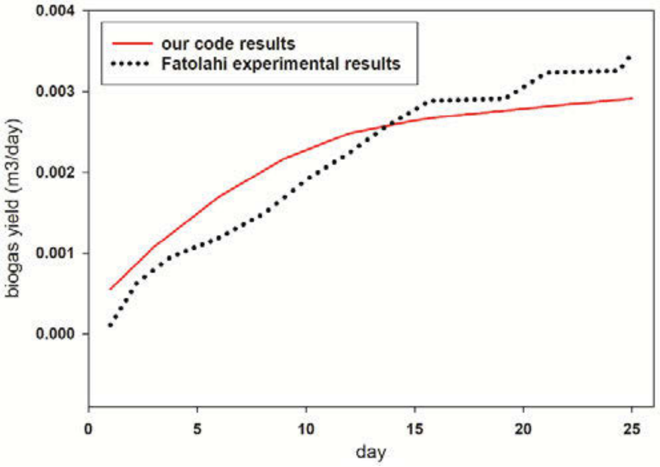
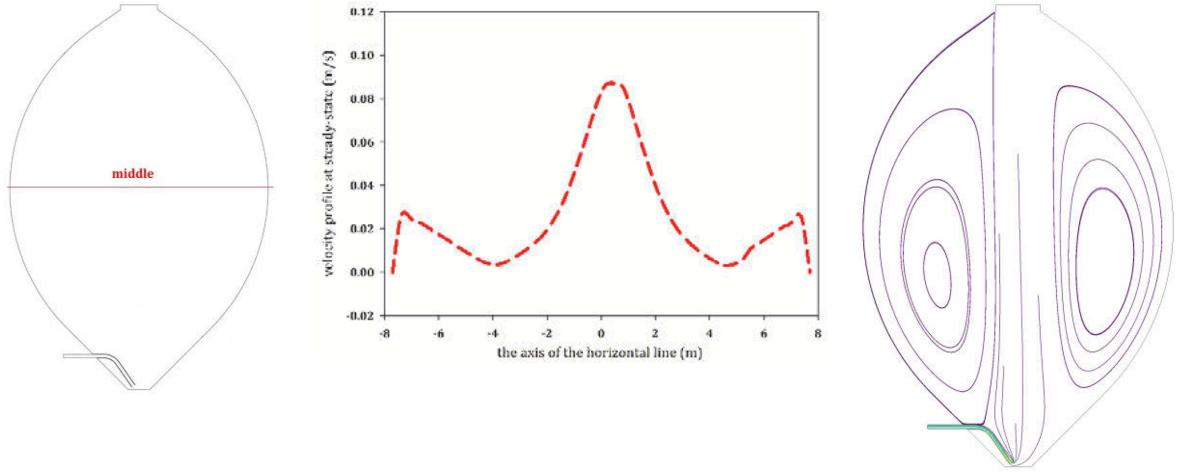


Figure 1 comparing the results of our passiveScalarADMFoam solver to Fatolahi experiments

4 RESULTS AND DISCUSSION

4.1 Hydrodynamics results

Regarding the hydrodynamics of the model, the results of the steady-state simulation are depicted as velocity profiles at the central line at the middle of the digester, 11 m from the bottom. Moreover, the streamlines of the fluid flow are displayed, in order to give an insight on the behaviour of the fluid flow within the tank. *Figure 2*(a) shows the velocity profiles at the central horizontal line and the streamlines.



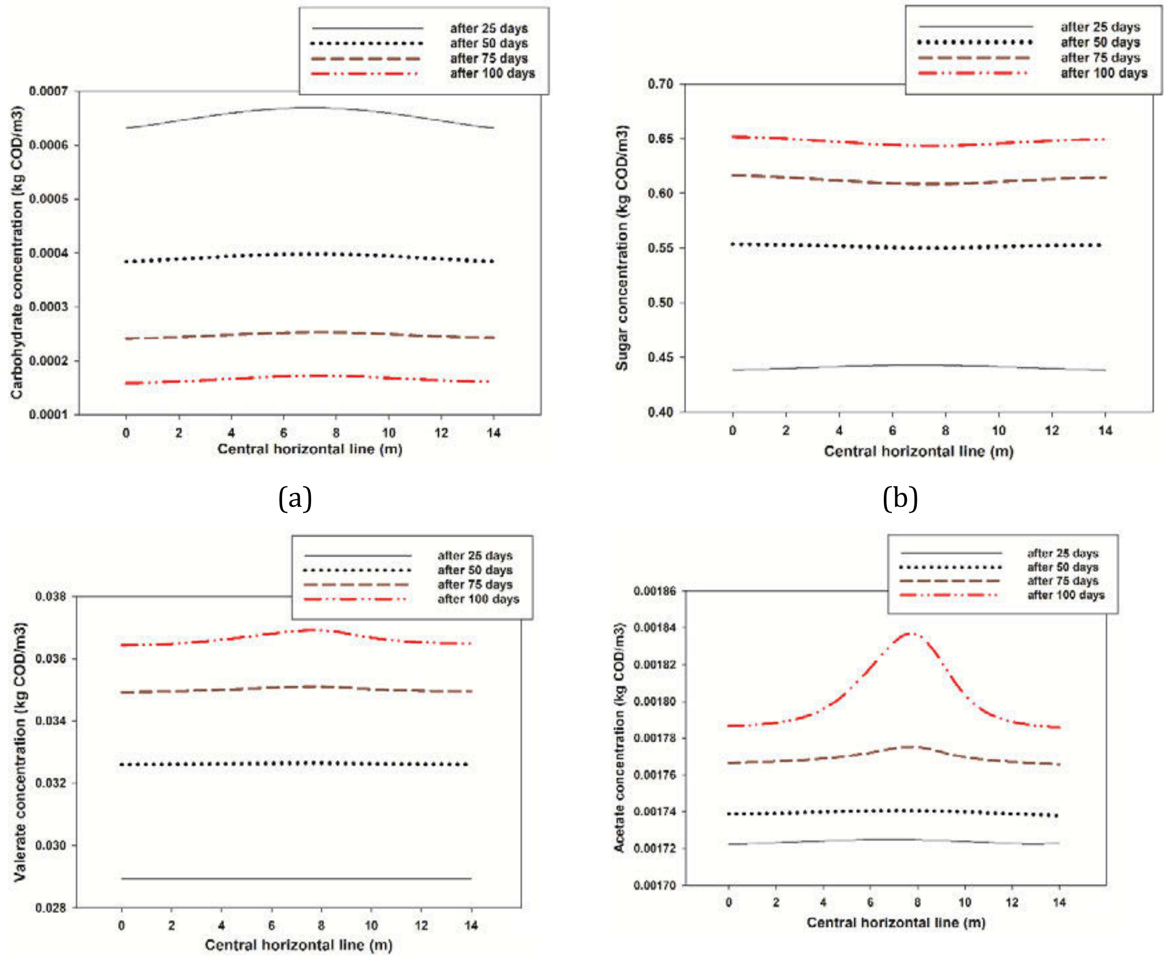
(a) (b) (c)

Figure 2 the central horizontal line (a), the velocity profile at the central horizontal line (b), the streamlines within the tank (c)

Figure 2(a) shows the location of the central line, which later is used for showing the concentration of bio-kinetic substances. The velocity profile along the central line in *Figure 2*(b) depicts that the maximum velocity at the centre is about 0.085 m/s. The plot also shows that a higher velocity range is formed at the center of the tank, while a smaller increase in at the regions close to the wall represent the upward flow regime at the two lateral sides. This is better shown in *Figure 2*(c), where the streamlines illustrate the creation of two circular flow regime within the tank. Still, the downward flow in the tank is more dominant than the upward flow near the walls.

4.2 Bio-kinetic results

The concentration of the substances is plotted for each component along the central line of the digester (located 11 m from the bottom). Here, we display only the concentration of the substances that are representative of the various steps of the AD process. The first step is the disintegration of the biomass, from which hydrocarbons are produced. Thus, *Figure 3*(a) shows the amount of hydrocarbons concentration along with the central line during a constant distributed times within 100 days of ADM1 simulation, that is, after 25, 50, 75 and 100 days. The next step is hydrolysis, and one of its products is sugar. Hence, sugar is also plotted in *Figure 3*(b). Similarly, the concentrations of valerate, as the product of acidogenesis, and the concentration of acetate, as the product of acetogenesis is plotted within the mentioned times, and shown in *Figure 3*(c) and *Figure 3*(d), respectively.



(a) (b)

(c) (d)

Figure 3 the substrates concentration along with the central horizontal line in the tank

From the four substance concentration profiles at the horizontal line located at the centre of the tank it is clear that there is little variation along the central line, that is, the distribution of the materials are almost evenly within the tank. However, after 100 days the distribution of the acetate is less homogenous as compared to other plotted substances. *Figure 3*(a) shows that the concentration of carbohydrates decreases during the simulation time. After 25 days, it is more than $6*10^{-4}$ kg COD/m^3, while it decreases to less than $6*10^{-4}$ kg COD/m^3 after 100 days. Carbohydrate, as the product of disintegration, is the only substance which decreases along the time. Sugar (from 0.44 to 0.65 kg COD/m^3), Valerate (from less than $2.9*10^{-2}$ to more than $3.6*10^{-}$

^{2} kg COD/m^3) and Acetate (from 1.72*10^{-3} to more than 1.78*10^{-3} kg COD/m^3) increase during the time.

Regarding the homogeneity of the substances in the tank, the uniformity index is suggested in [14], which is calculated as follows:

$$\gamma = (\,1 - \frac{\sum_{i=1}^{n}(|\varphi_i - \bar{\varphi}|A_i)}{2\bar{\varphi}\sum_{i=1}^{n}(A_i)}) * 100 \qquad (7)$$

where A_i and φ_i are the area of the ith cell and its corresponding scalar variable, respectively, and $\bar{\varphi}$ is the mean value for each variable on the surface.

In order to have a good mixing quality, the uniformity index of all the substances should be high enough. This is true for most of the components. *Figure 4* shows the uniformity index of the analysed material during the 100 days of the simulation time. Since it is assumed that the materials initially are distributed equally at the beginning of the simulation, the uniformity index is 100% at the first day. Among the plotted substances, the uniformity index of sugar and Valerate stays above 95%, however the uniformity index of carbohydrates reaches only above 80%. The uniformity index for acetate does not goes below 70%. These values prove that the recirculation system of the AD tank accounts for an almost even distribution of the material i.e., for a proper mixing of tank.

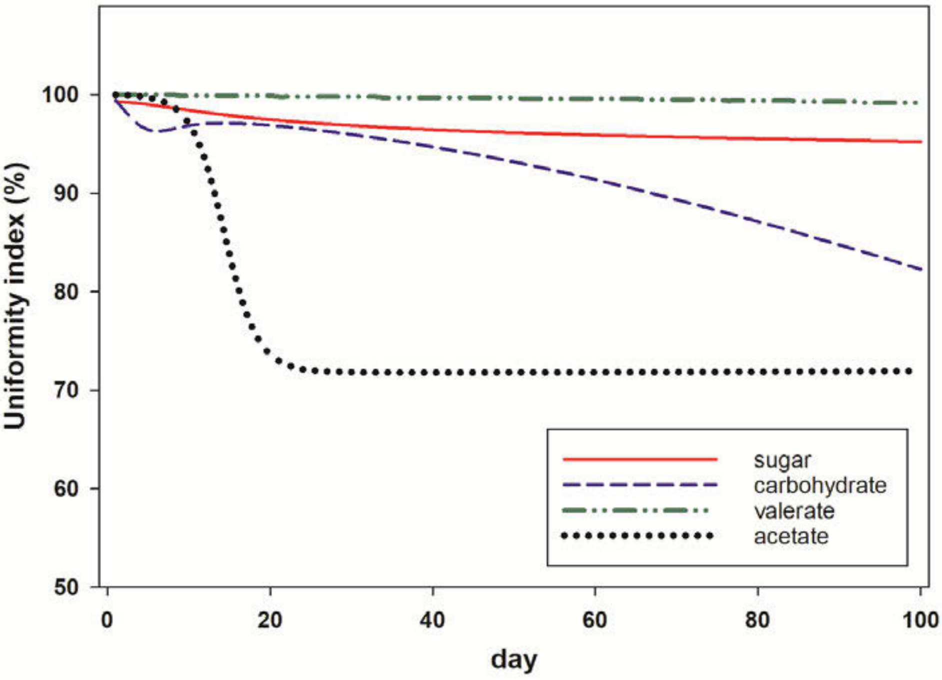


Figure 4 the uniformity index of the analyzed substances during the 100-day period

5 CONCLUSIONS

This paper promotes the potential use of ADM1 in CFD platforms, by applying ADM1 to an open source CFD platform (OpenFOAM). The developed solver is called passiveScalarADMFoam and is employed exemplarily for assessing the mixing quality of sludge recirculation system in the real world case study AIZ WWTP, located in Austria. After calculating the hydrodynamics of mixing and the validation of the solver via experimental case-studies, the following points are concluded:

- The velocity at the centre of the investigated tank does not exceed 0.085 m/s, and it is higher at the center than the regions near the walls.
- Two circular flow regimes are formed within the tank, due to the sludge recirculation (see *Figure 2*(c)).
- The results of passiveScalarADMFoam solver are compared to Fatolahi experiments. Due to low difference in the resulting biogas yield, we conclude the efficiency of our developed solver.
- Carbohydrate, as the product of disintegration, decreases along the time, but other investigated substances, i.e. sugar, valerate and acetate increase.
- The uniformity index of the substances within the tank is not lower than 70%, which proves the efficiency of the recirculation mixing strategy in our studies WWTP.

6 REFERENCES

[1] H. Wang, R. A. Larson, M. Borchardt, and S. Spencer, "Effect of mixing duration on biogas production and methanogen distribution in an anaerobic digester," Environmental technology, 2019.

[2] S. Xie et al., "Anaerobic co-digestion: A critical review of mathematical modelling for performance optimization," Bioresource Technology, vol. 222, pp. 498-512, 2016.

[3] E. Hernandez-Aguilar, A. Alvarado-Lassman, A. Osorio-Mirón, and J. M. Méndez-Contreras, "Development of energy efficient mixing strategies in egg-shaped anaerobic reactors through 3D CFD simulation," Journal of Environmental Science and Health, Part A, vol. 51, no. 7, pp. 536-543, 2016.

[4] D. Dapelo, F. Alberini, and J. Bridgeman, "Euler-Lagrange CFD modelling of unconfined gas mixing in anaerobic digestion," Water research, vol. 85, pp. 497-511, 2015.

[5] Y. Kamla, H. Ameur, A. Karas, and M. I. Arab, "Performance of new designed anchor impellers in stirred tanks," Chemical Papers, vol. 74, no. 3, pp. 779-785, 2020.

[6] C. Sadino-Riquelme, R. E. Hayes, D. Jeison, and A. Donoso-Bravo, "Computational fluid dynamic (CFD) modelling in anaerobic digestion: General application and recent advances," Critical Reviews in Environmental Science and Technology, vol. 48, no. 1, pp. 39-76, 2018.

[7] C. C. Chang, C. Kuo-Dahab, T. Chapman, and Y. Mei, "Anaerobic digestion, mixing, environmental fate, and transport," Water Environment Research, vol. 91, no. 10, pp. 1210-1222, 2019.

[8] H. Che, D. Liu, W. Tian, S. Gao, J. Sun, and L. Xu, "CFD-DEM study of gas-solid flow regimes in a Wurster type fluidized bed with experimental validation by electrical capacitance tomography," Chemical Engineering Journal, vol. 389, p. 124280, 2020.

[9] S. Dabiri, J. Sappl, P. Kumar, M. Meister, and W. Rauch, "On the effect of the inlet configuration for anaerobic digester mixing," Bioprocess and biosystems engineering, vol. 44, no. 12, pp. 2455-2468, 2021.

[10] D. J. Batstone et al., "The IWA anaerobic digestion model no 1 (ADM1)," Water Science and technology, vol. 45, no. 10, pp. 65-73, 2002.

[11] Y. M. Tobo, U. Rehman, J. Bartacek, and I. Nopens, "Partial integration of ADM1 into CFD: understanding the impact of diffusion on anaerobic digestion mixing," Water Science and Technology, vol. 81, no. 8, pp. 1658-1667, 2020.

[12] Z. Fatolahi, G. Arab, and V. Razaviarani, "Calibration of the Anaerobic Digestion Model No. 1 for anaerobic digestion of organic fraction of municipal solid waste under mesophilic condition," Biomass and Bioenergy, vol. 139, p. 105661, 2020.

[13] B. Wu, "Integration of mixing, heat transfer, and biochemical reaction kinetics in anaerobic methane fermentation," Biotechnology and bioengineering, vol. 109, no. 11, pp. 2864-2874, 2012.

[14] D. Dapelo and J. Bridgeman, "A CFD strategy to retrofit an anaerobic digester to improve mixing performance in wastewater treatment," Water Science and Technology, vol. 81, no. 8, pp. 1646-1657, 2020.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14201

OPTIMAL DESIGN OF SEWER NETWORKS INCLUDING TOPOGRAPHIC CRITERIA AND DROP MANHOLES

Juan Saldarriaga[1], Juana Herrán[2] and Pedro L. Iglesias-Rey[3]

[1,2]Department of Civil and Environmental Engineering, Water Distribution and Sewerage Systems Research Center, Universidad de los Andes, Bogotá, (Colombia)
[3]Department of Hydraulic Engineering and Environment, Universitat Politècnica de València, Valencia (Spain)46022 Valencia, Spain; piglesia@upv.es

[1] *jsaldarr@uniandes.edu.co,* [2] *jm.herran10@uniandes.edu.co,* [3] *piglesia@upv.es*

Abstract

Optimal sewer network design is a complex problem that has been widely studied in literature. A methodology published in the literature that has shown great potential is the proposed by Duque, Duque, Aguilar, Saldarriaga. This methodology uses Mixed Integer Programming to select the layout of the network and Dynamic Programming to select the optimal combination of diameters and depths of pipes. Although the methodology showed good performance, two improvements have been implemented in later works. The first one is to consider topographic criteria when selecting the layout. This demonstrated to reduce the sewer network design's costs. The other improvement is the addition of drop manholes in the optimization process. This allowed the methodology to comply with maximum velocity constrains in hilly regions. The present work tested the three methodologies mentioned before in a sewer network located in Bogotá, Colombia to evaluate the contributions and disadvantages of each methodology considering the cost and the hydraulic features of the resulting designs, and the computational resources required with each methodology.

Keywords
Optimal sewer network design, layout selection, hydraulic design, topographic criteria, drop manholes.

1 INTRODUCTION

Sewer network design is a problem that is divided into two subproblems: the layout selection and the hydraulic design. In the layout selection, the structure of the network is defined; this includes the flow direction, flow rate and type of connection of pipes. In the hydraulic design the diameter and upstream and downstream invert elevation of pipes is defined.

Due to the complexity of both subproblems and the immense number of alternatives that exists to design a sewer network, finding the optimal design it's been a challenge for the researchers in this field. For this reason, many approaches have been proposed to try to find the optimal or a near-optimal design of a sewer network. For example, Li and Matthew [1] proposed the searching direction method for the optimal selection of the layout and used discrete dynamic programming (DDDP) for the hydraulic design. Haghighi [2] proposed an adaptative method entitled loop-by-loop cutting algorithm to solve the layout selection and used a discrete differential dynamic programming model for sizing the pipes. Later, Haghighi and Bakhshipour [3] used the loop-by loop cutting algorithm for the layout selection and tabu search (TS) for the hydraulic design. After that, Duque et al. [4] used mixed-integer programming (MIP) for the layout selection and Dynamic Programming (DP) for the hydraulic design.

This last approach achieved the designs with the lowest cost at the time when it was published in two sewer network benchmarks studied in the literature. In later works, some extensions have been added to improve the approach of Duque et al. For example, Saldarriaga et al. [5] proposed a methodology to include topographic criteria in the layout selection model which achieved designs with lower costs than the work of Duque et al. Also, in a methodology that will be described later in this work, drop manholes are included in the optimization process. This is especially important in hilly regions where drop manholes are necessary to comply with maximum velocity constrains and to dissipate flow energy in the system.

In the present work, a comparison between the three methodologies mentioned before is done in order to identify their advantages and disadvantages related with the construction cost and the hydraulic features of the resulting designs. The three methodologies were tested using a sewer network located in Bogotá, Colombia, with a modified ground elevation to simulate a hilly region and in this way, to include the possibility of having drop manholes in the designs.

2 DESIGN METHODOLOGIES

2.1 Sewer network design methodology proposed by Duque et al. [4]

Duque et al. proposed an iterative scheme to find a near-optimal solution for the sewer network design problem. As mentioned before, this problem is composed by two subproblems: the layout selection and the hydraulic design. Each iteration of the methodology is composed by the solution of the layout selection and the hydraulic design, and with each iteration the construction cost of the network is refined.

With this methodology, the layout selection problem is modelled using graph theory. The input of the problem is an undirected graph composed by nodes and arcs. The nodes represent each manhole of the sewer network with its respective x and y coordinates, ground elevation, and inflow. The arcs represent the connection between two manholes that depend on the direction and the type of connection. There are two types of connection in pipes, outer and inner branch pipes. Outer branch pipes are the first pipe in a series, while the inner branch pipes are the rest of the pipes. These types of connection are represented as t_1 and t_2, respectively.

Since the connection between two manholes can have to possible directions (i.e., $i \rightarrow j$ or $j \rightarrow i$) and two possible types of connection (i.e., t_1 or t_2), there are four possible arcs between a couple of manholes as is shown in Figure 1.

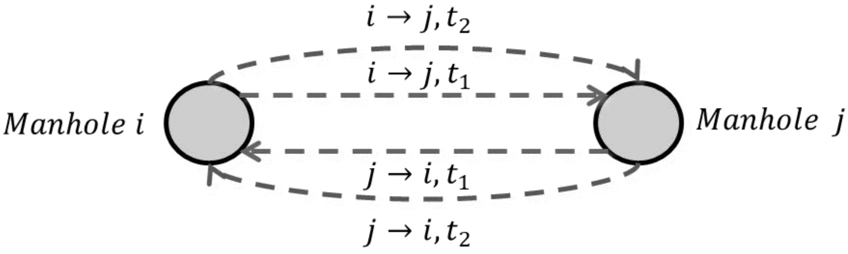


Figure 1. Arcs between two manholes.

With the given undirected graph, linear optimization is used to find the layout of the sewer network. This is, to establish the flow direction, type of pipe connection, and flow rate in pipes. This is done using two decision variables: x_{ijt} and q_{ijt}. x_{ijt} is a binary variable that takes the value of one (1)f the pipe from manhole *i* to manhole *j* and type *t* is part of the layout selection. q_{ijt} is a continuous variable that represent the flow rate that is transport in the pipe from manhole *i* to manhole *j*. The linear optimization model also considers the constrains that guarantee a feasible layout and the objective function.

The objective function of the model should be the minimization of an equation that represents the construction cost of the network. However, the construction cost depends on the diameters and excavation depths of pipes, which are unknown in the layout selection subproblem. For this reason, the objective function that Duque et al. proposed for this subproblem is an approximation of the construction cost of the sewer network. This approximation is done with a linear regression between the flow rate and cost of pipes. Nevertheless, in the first iteration the cost of pipes is also unknown, therefore, in the first iteration the linear regression coefficients are random values, but with the following iterations their values are refined. Equation (1) presents the objective function where c_{ij} and a_{ij} are the linear regression coefficients, and x_{ijt} and $q_{ijt}$are the decision variables of the model.

$$min\left(\sum_{t\in\mathcal{T}} \sum_{(i,j)\in\mathcal{A}_L} c_{ij}q_{ijt} + \sum_{t\in\mathcal{T}} \sum_{(i,j)\in\mathcal{A}_L} a_{ij}x_{ijt}\right) \quad (1)$$

After selecting the layout of the network, the hydraulic design is performed. In this subproblem the diameters and depths of the pipes are defined. This was done by Duque et al. using a Shortest Path Algorithm. In this kind of algorithm, the problem is modeled as a directed graph where each arc has a cost associated and the objective is to find the path of arcs with the minimum cost.

To model the hydraulic design as a Shortest Path Algorithm, Duque et al. proposed a graph composed by nodes and arcs. The nodes represent the possible combinations of diameters and depths, while the arcs represent the diameter and depth of a specific pipe. Each manhole of the network has associated a group of nodes. The arcs connect two nodes from two different manholes representing a specific pipe. Figure 2 shows an example of the arcs and nodes that represent a pipe between two manholes. In this example, each manhole has 9 nodes that correspond to the possible combinations of 3 diameters and 3 depths. d_1, d_2 and d_3 are different diameters, and Z_1, Z_2 and Z_3 are different depths. The arrows represent the feasible arcs that model the pipe. Note that the upstream node can never be connected with a downstream node that has a smaller diameter or a shallower depth. Also, in the example there are only 3 diameters and 3 depths, but, in the methodology, there could be as many diameters and depths as desired.

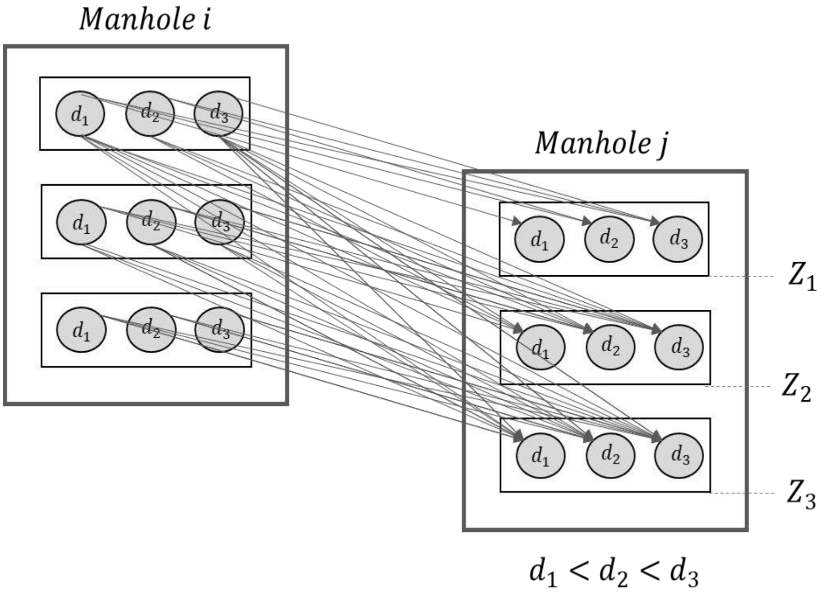


Figure 2. Example of nodes and arcs that represent a pipe.

Each arc of the model has a cost that depend on the diameter and depth of its downstream node. Considering this cost, the Shortest Path Algorithm selects the path of minimum cost, this is, the optimal combination of diameters and depths of the sewer network.

In the hydraulic design problem, the real cost of the pipes is known because the diameters and invert elevations are known. With this information, the linear regression of the objective function of the layout selection is done and with each iteration, the objective function is improved. As a consequence, with each iteration the real cost of the network is reduced until the methodology converges. For this reason, around 10 to 30 iterations must be performed to find the solution of the sewer network design problem with this methodology.

2.2 Layout selection including topographic criteria proposed by Saldarriaga et al. [5]

Since the methodology proposed by Duque et al. requires a random initialization of the objective function in the layout selection, Saldarriaga et al. proposed a modification to improve this equation. The improvement consists in the addition of the parameter b_{ijt} in the objective function as is shown in Equation (2). The purpose of this parameter is to consider the topography of the land in the layout selection and, in this way, that the direction of the pipes follows the land's slope, since this reduces the required excavation depth, and therefore, reduces the cost of excavation.

$$min\left(\sum_{t\in\mathcal{T}} \sum_{(i,j)\in\mathcal{A}_L} c_{ij}q_{ijt} + \sum_{t\in\mathcal{T}} \sum_{(i,j)\in\mathcal{A}_L} a_{ij}x_{ijt} + \sum_{t\in T} \sum_{(i,j,t)\in A_L} b_{ijt}x_{ijt}\right) \tag{2}$$

The parameter b_{ijt} is calculated for each arc in the network (recall the existing arcs in Figure 1). This means that for a pipe, four parameters b_{ijt} are calculated depending on the direction and the type of connection of the pipe. From these four values of b_{ijt}, the linear optimization model selects the one that has the minimum value, and in this manner, the direction and type connection of the pipe is stablished.

As mentioned before, it is preferred that the pipes follow the land's slope. For this reason, the arcs that follow this direction have a b_{ijt} with lesser value. Regarding the type of connection, it was found that trying to minimize the outer branch pipes let designs with lower costs, so the parameter b_{ijt} tries to give priority to the inner branch pipes. The value of b_{ijt} is established using three different topographic criteria.

2.2.1 Topographic criterion 1

This criterion is defined by the multiplication of the land slope where the pipe is located and -1. In this way, the arcs that follow the land slope will be prioritized because its parameter b_{ijt} will have a negative value and this is preferred by the objective function since it is a minimization function.

Also, to give priority to the inner branch pipes, the outer branch pipes are multiplied by a penalty that increases the value of b_{ijt}. A sensitivity analysis was done to stablish that 0.65 is a good value to use as a penalty when the slope is positive, and 1.65 when the slope is negative. Note that the penalty must be different depending if the slope is negative or positive because, in the first case, b_{ijt} will be a negative value, so to increase this value it must be multiplied by a number between 0 and 1. In the second case, b_{ijt} will be a positive value so to increase this value, it must be multiplied by any number greater than 1.

To resume, Equation (3) and (4) presents how b_{ijt} is calculated using criterion 1 in inner and outer branch pipes, respectively, where s is the land slope and μ is the penalty for outer branch pipes.

$$b_{ijt_2} = s_{ijt_2} * (-1) \tag{3}$$

$$b_{ijt_1} = s_{ijt_1} * (-1) * \mu \tag{4}$$

2.2.2 Topographic criterion 2

This is a power-based criterion that is calculated the same way as criterion 1 but the land slope is also multiplied by the length of the pipe. The penalty for the outer branch pipes is the same as in criterion 1. Equation (5) and (6) presents how b_{ijt} is calculated using criterion 2, where L is the length of the pipe.

$$b_{ijt_2} = s_{ijt_2} * (-1) * L_{ij} \tag{5}$$

$$b_{ijt_1} = s_{ijt_1} * (-1) * L_{ij} * \mu \tag{6}$$

2.2.3 Topographic criterion 3:

This criterion was proposed specially for flat regions since it does not depend on the land slope. With this criterion b_{ijt} is calculated as the distance between the downstream manhole of the pipe and the outfall of the sewer network. With this criterion there is not a penalization in the outer branch pipes.

The layout selection model must be executed using each of the three topographic criteria at the time. This means that three iterations must be done and in each one of them a different layout will be obtained. With each layout, a different hydraulic design is achieved. The design with the lowest cost is used in a fourth iteration where the excavation depth is penalized to try to reduce the cost of the design even more.

2.2.4 Iteration with penalization in excavation depth

In this iteration, two cost are considered: the cost per unit length of pipes and the cost of the extra excavation that must be done if an arc is chosen compared to the average installation slope of the design with the lowest cost obtained with the topographic criteria.

For the inner branch pipes, the cost per unit length is considered in the linear regression coefficients (c_{ij} and a_{ij}) of the objective function. Regarding the cost of the extra excavation, if the land slope where the pipe is located is smaller than the average installation slope, b_{ijt} is the extra excavation cost. In the opposite case, b_{ijt} is calculated as the extra excavation cost multiplied by -1.

In the case of the outer branch pipes, the cost per unit length is almost always the same because they tend to have the smallest diameter and depth. For this reason, the cost per unit length is not included in the linear regression and it must be included in the value of b_{ijt}. In outer branch pipes, when the slope of the pipe is smaller than the average installation slope, b_{ijt} is the average cost per unit length multiplied by the length of the pipe. In the opposite case, b_{ijt} is the sum of the extra excavation cost and the cost of the pipe.

With this method to penalize the excavation depth, the layout selection model tries to find a solution that reduces the extra excavation. However, the hydraulic design obtained from the resulting layout of this iteration has not always a lower cost then the designs from the topographic criteria. For this reason, from the four iterations done with the methodology, the one with the lowest design should be selected as the solution of the sewer network design problem.

2.3 Hydraulic design including drop manholes

Designing a sewer system in hilly regions is a challenge because high slopes induce to high velocities that exceed the maximum velocity constrain. If this constrain is no complied, pipes and other sewer structures could be damaged. To comply with the mentioned constrain and to dissipate the flow energy in sewer networks, drop manholes are used. These structures are vertical manholes where the upstream pipe is in a higher level that the downstream pipe. The difference between the level of those pipes is the drop height that helps with the dissipation of the flow energy.

As part of this work, a methodology to include drop manholes in the sewer network design problem was introduced. This methodology consists in the addition of a new type of arc that represent the possibility of having drop manholes in the network. These arcs have a vertical direction and their downstream node have a greater depth than their upstream node. Also, the length of the arcs represents the drop height of the drop manhole. Figure 3 presents an example of an arc that represent a drop manhole where its upstream node has a diameter d_2 and an excavation depth Z_1, and its downstream node has the same diameter but an excavation depth of Z_2.

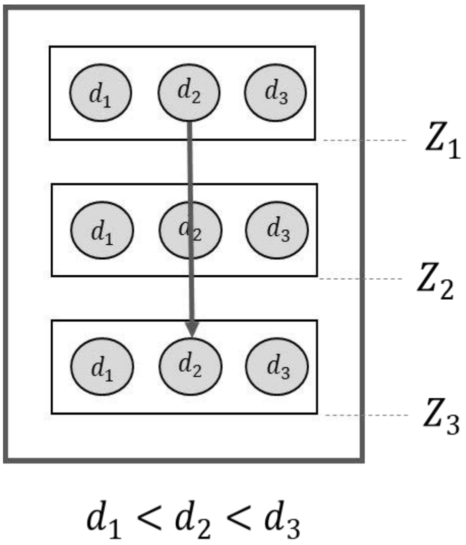


Figure 3. Example of an arc that represents a drop manhole.

Since the new types of arcs represent drop manholes and the original arcs of the methodology represent pipes, the cost function is different for each type of arc. In the case of the latter, their cost function depends on the diameter and depth of the pipe, but in the case of the arcs that represent drop manholes, their cost function must represent the cost of installing a drop manhole in the network.

The Shortest Path Algorithm works the same way as in the original methodology, it finds the path of arcs with the minimum cost that represent the hydraulic design of the network. With the addition of the new type of arcs, the algorithm needs to evaluate a greater number of arcs to find the optimal solution.

3 CASE STUDY

To test the three methodologies, a sewer network labelled Chicó was used. This network is located in Bogotá, Colombia and is composed of 109 manholes and 160 pipes that transport a total flow rate of 1.525 m^3/s. The ground elevations of the manholes were modified to have a slope of 5%. This was done to simulate a hilly region where drop manholes can improve the cost of the sewer network designs. Figure 4 illustrates the structure of the Chicó sewer network with the respective location of the outfall.

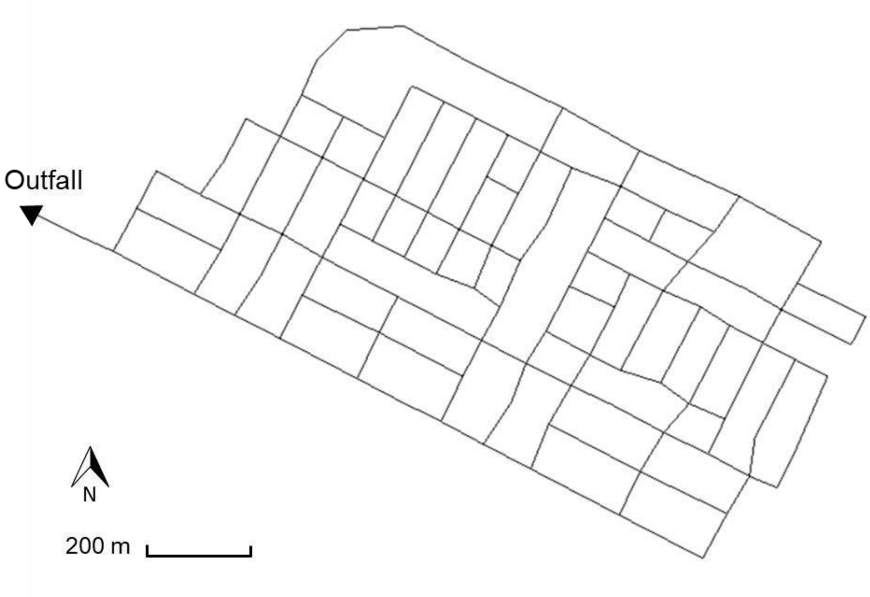


Figure 4. Chicó sewer network.

All the designs obtained with the three tested methodologies comply with hydraulic constrains that allow an adequate operation of the sewer network. For this work, the constrains proposed by Li and Matthew [1] were used. This constrains are presented in Table 1.

Table 1. Hydraulic constrains for the sewer network design

Constrain	Value	Condition
Minimum diameter	0.2 m	Always
Maximum filling ratio	0.6	$d \leq 0.3$ m
	0.7	0.35 m $\leq d \leq 0.45\ m$
	0.75	0.5 m $\leq d \leq 0.9\ m$
	0.8	$d \geq 1\ m$
Minimum velocity	0.7 m/s	$d \leq 0.5$ m and Flow rate $> 0.015\ \mathrm{m}^3/s$
	0.8 m/s	$d > 0.5$ m and Flow rate $> 0.015\ \mathrm{m}^3/s$
Maximum velocity	5 m/s	Always
Minimum gradient	0.003	Flow rate $< 0.015\ m^3/s$
Minimum depth	1 m	Always

Also, the list of available diameters used is: {0.2, 0.25, 0.3, 0.35, 0.38, 0.4, 0.45, 0.5, 0.53, 0.6, 0.7, 0.8, 0.9, 1.0, 1.05, 1.20, 1.35, 1.4, 1.5, 1.6, 1.8, 2, 2.2, 2.4} in meters. And the material of the pipes used was concrete with a Manning's n of 0.014.

To measure the construction cost of the sewer networks, two cost function were used, one for the pipes and another one for the drop manholes. The cost function used for the pipes was proposed by Maurer et al. [6] and is presented in Equation (7), where C is the construction cost of a pipe in USD, h is the average excavation depth of the pipe in meters, l is the length of the pipe in meters, and α and β are coefficients that depend on the diameter of the pipe.

$$C = (\alpha h + \beta) * l \tag{7}$$

For the cost of the drop manholes, Equation (8) was used, where C_{dmh} is the cost of a drop manhole in USD, h is the excavation depth of the drop manhole in meters, and d is the diameter of the downstream pipe of the drop manhole.

$$C_{dmh} = 4354.38 - 776.76h + 5404.52d - 6370.59hd + 870.05h^2 + 12820.76d^2 \tag{8}$$

4 RESULTS AND DISCUSSION

Table 2 presents the results of the designs achieved with each methodology. The table includes the construction cost of the sewer network, the most important hydraulic constrains in hilly regions, the number of drop manholes in the design and the number of iterations required in each methodology.

Table 2. Results of the design of the Chicó sewer network.

Design	Methodology	Construction cost (USD x10^6)	Maximum filling ratio (%)	Maximum velocity (m/s)	Maximum depth (m)	N° of DM	N° of iterations
1	MIP and DP	$ 8,96	74.75	4.97	7.80	0	23
2	MIP with TC and DP	$ 7,67	74.77	4.94	4.20	0	4
3	MIP with TC and DP with DM	$ 7,62	74.77	4.98	3.00	2	4

MIP: Mixed Integer Programming.
DP: Dynamic Programming.
TC: Topographic criteria.
DM: Drop manholes.

Regarding the construction cost, the design with the highest cost is the one obtained with the methodology of Duque et al., which is the expected result because it is the original methodology without any improvement. When including the topographic criteria in the layout selection model, the cost of the design is reduced in 14.39%, which is equivalent to $ 1.29 x 10^6 USD. Also, when including the possibility of designing with drop manholes, the cost is reduced even more. It is reduced in 14.89% compared with the original design.

The three designs comply with all hydraulic constrains, which are important to guarantee an adequate operation of the sewer network. Regarding the maximum filling ratio and the maximum velocity, the three designs achieved values that are close to the maximum allowed, that is, 75% for the maximum filling ratio, and 5 m/s for the maximum velocity.

In contrast, the maximum depth is very different in each design. The design with the highest depth is also the one with the highest cost. The use of the topographic criteria to try to select a layout that follows the land slope reduces the maximum depth in 3.6 meters. In addition, the use of drop

BY NC SA

manholes can also let to a reduction of the maximum depth, although it is not as significant as the reduction achieved with the topographic criteria.

It is important to note that the only design that includes the possibility of having drop manholes is the third one. Two drop manholes were obtained in the design with this methodology. Both are located just before the outfall, and both have 1.1 meters of drop height.

Lastly, including topographic criteria in the layout selection model allows a significant reduction of the required number of iterations. In the original methodology, 23 iterations were needed, while with the other methodologies, only 4 iterations were required. Figure 5 illustrates the cost achieved in each iteration of the three evaluated methodologies.

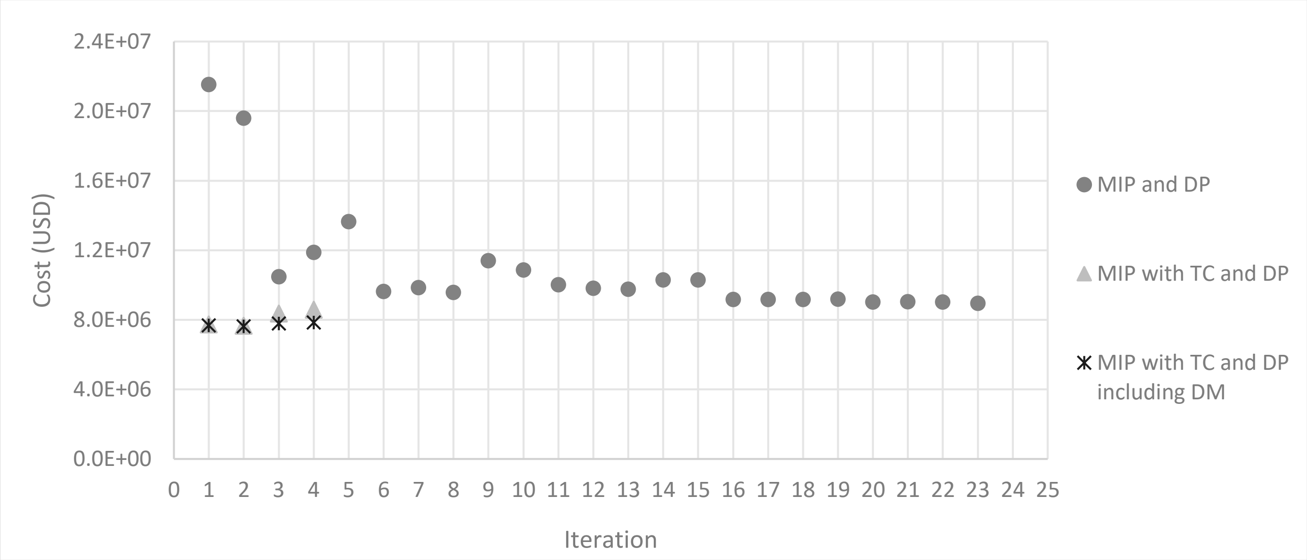


Figure 5. Cost of the designs in each iteration of the methodologies.

Figure 5 shows that the MIP and DP methodology converges, this means that the cost of the designs is refined with each iteration. This is not the case with the methodologies that uses topographic criteria, since with these methodologies each iteration uses a different strategy to follow the land slope. As it can be seen, considering topographic criteria in the model not only reduces the iterations required to find a near-optimal solution, but also the 4 iterations achieved with this methodology have a lower cost than the best solution from the original methodology.

Also, when analysing the iterations with the methodology that include drop manholes it can be seen that the cost of the designs is lower, but the reduction is not very significant. Furthermore, the addition of drop manholes to the model does not increment the number of iterations, but it does increment the computational time in each iteration because the Shortest Path Algorithm needs to evaluate a greater number of arcs, since the drop manholes are modelled as arcs.

5 CONCLUSIONS

In the present work three methodologies for the optimal design of sewer networks were compared using as a case study a real sewer network located in Bogotá, Colombia. The first methodology consists in an iterative scheme were the layout selection and hydraulic design models are embedded. With each iteration the cost of the design is reduced. The layout selection model is solved using MIP, while the hydraulic design model is solved using DP, or more specifically, a Shortest Path Algorithm that guarantees to select the optimal combination of diameters and excavation depths for a given layout.

The second methodology is an extension of the methodology previously described. The extension consists in the addition of topographic criteria to the layout selection model to try to follow the land slope. Using these criteria, the methodology does not require an iterative scheme because the layouts obtained following the land slope let to great hydraulic designs. This means that the number of iterations is reduce using topographic criteria.

The third methodology is also an extension of the original methodology, and it also includes topographic criteria. The contribution of this methodology is that it includes the possibility of having drop manholes in the hydraulic design. These structures are important in hilly regions where it is difficult to comply with maximum velocity constrains.

To compare the three methodologies, the Chicó sewer network was used. This network has been used as a case study in previous works in the literature, but in the present work it was important to simulate a hilly region to study the effect of drop manholes in the designs. For this reason, the ground elevation of the network was modified to have a land slope of 5%.

From the results, the main conclusion is that considering topographic criteria in the model reduces the construction cost of the network in a significative way and also reduces the required amount of iterations to find a near-optimal solution. This is because the topographic criteria try to select the direction of pipes in the layout in order to follow the land slope, this allows a reduction of the maximum depth required by the pipes, which also reduces the construction costs.

When using the third methodology, two drop manholes were obtained in the design. These structures allowed a reduction of the maximum depth required in the network, for this reason, the cost was also reduced when using the methodology with drop manholes, although the reduction of costs was not as significant. Also, the addition of drop manholes in the model does not increment the number of iterations, but it increments the computational time in each iteration.

In the case study used, the three methodologies achieved to comply with all hydraulic constrains that are important to guaranty a satisfactory operation of the sewer network. The designs obtained had similar maximum filling ratios and maximum velocities, which means that, for the case study used, the drop manholes did not have an important impact on the compliance of constrains. However, the drop manholes did help to reduce the excavation depth and, therefore, the construction cost of the network.

6 REFERENCES

[1] G. Li and Matthew Robert G. S., “New Approach for Optimization of Urban Drainage Systems,” J. Environ. Eng., vol. 116, no. 5, pp. 927–944, Sep. 1990, doi: 10.1061/(ASCE)0733-9372(1990)116:5(927).

[2] A. Haghighi, “Loop-by-Loop Cutting Algorithm to Generate Layouts for Urban Drainage Systems,” J. Water Resour. Plan. Manag., vol. 139, no. 6, pp. 693–703, Nov. 2013, doi: 10.1061/(ASCE)WR.1943-5452.0000294.

[3] A. Haghighi and A. E. Bakhshipour, “Deterministic Integrated Optimization Model for Sewage Collection Networks Using Tabu Search,” J. Water Resour. Plan. Manag., vol. 141, no. 1, p. 4014045, Jan. 2015, doi: 10.1061/(ASCE)WR.1943-5452.0000435.

[4] N. Duque, D. Duque, A. Aguilar, and J. Saldarriaga, “Sewer network layout selection and hydraulic design using a mathematical optimization framework,” Water (Switzerland), vol. 12, no. 12, 2020, doi: 10.3390/w12123337.

[5] J. Saldarriaga, J. Zambrano, J. Herrán, and P. L. Iglesias-Rey, “Layout selection for an optimal sewer network design based on land topography, streets network topology, and inflows,” Water (Switzerland), vol. 13, no. 18, pp. 1–20, 2021, doi: 10.3390/w13182491.

[6] M. Maurer, M. Wolfram, and H. Anja, “Factors affecting economies of scale in combined sewer systems,” Water Sci. Technol., vol. 62, no. 1, pp. 36–41, 2010, doi: 10.2166/wst.2010.241.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14247

ROADMAP TOWARDS SMART WASTEWATER TREATMENT FACILITIES

Daniel Aguado[1], Henri Haimi[2], Michela Mulas[3] and Francesco Corona[4]

[1] CALAGUA – Unidad Mixta UV-UPV, Institut Universitari d'Investigació d'Enginyeria de l'Aigua i Medi Ambient – IIAMA, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain.
[2] FCG Finnish Consulting Group Ltd, Osmontie 34, P.O. Box 950, FI-00601 Helsinki, Finland
[3] Department of Teleinformatics Engineering, Federal University of Ceará, Brazil
[4] School of Chemical Engineering, Aalto University, Finland.

[1] daaggar@hma.upv.es, [2] Henri.Haimi@fcg.fi, [3] michela.mulas@ufc.br, [4] francesco.corona@aalto.fi.

Abstract

To protect human health and natural ecosystems, wastewater treatment plants (WWTPs) have been traditionally designed to remove pollutants from wastewater. With remarkable success, WWTPs continuously adapt to increasingly stringent discharge limits. Nowadays, municipal wastewater treatment facilities are facing a double transition and new challenges: On the one hand, the transition towards a sustainable and circular water economy, in which resource recovery from wastewater (water, energy, and nutrient recovery) plays a fundamental role for its effective implementation. On the other hand, the digital transition, which aims at making the operation of these facilities smart, will undoubtedly have a synergistic effect together with the paradigm shift towards the effective implementation of a circular water economy.

To make our current facilities smart, there is a growing interest in finding the way to convert the collected process data into intelligent actions for improving their operation. This is not an easy task for many reasons:

- the harsh environment in which the instrumentation must work (corrosive, sludgy, biofilm formation with biological activity…),
- almost complete absence of metadata that would make it easy the interpretation of the process data that it is being collected and that would enable its future use,
- the almost complete absence of automated data quality assurance, required to avoid "garbage in – garbage out"
- the ever-increasing number of available process sensors (data overload), that must be properly processed and made easily available for further use to make them useful
- large amounts of data are collected and stored in databases but not wisely used, thus, resulting in data graveyards,
- the excessive cost of nutrient and organic matter sensors/analysers which moreover are labour maintenance intensive, fact that restrict their availability to the range of large facilities, thus, they are not usually available for small size facilities (which are the vast majority),
- the intelligent sensors and data-driven models must be maintainable by the plant workers (not by Data scientists),
- the lack of process expertise in the development of the artificial intelligent tools,
- plant operators are often accustomed to their operational routines and, therefore, cultural change is needed in the organization for successful digital transition and adopting new intelligent tools.

The progress in computing capabilities together with the large amount of collected process data in WWTPs have created the perfect storm for the machine learning boom we are observing, but all the aforementioned issues can make the incredible digital transition opportunity that exists today completely lost. In an attempt to avoid this disaster, this paper tries to shed light on the path towards increasing the value of the large amount of data that nowadays are being collected in WWTPs and WWRFs. Thus, digital transition could be safely embraced and the enormous potential of data analytics fully exploited, enabling it to play an essential role in the future automation and operation of our municipal facilities.

Keywords
Control, data analytics, digitalization, metadata, wastewater treatment plant, water resource recovery facility.

1 INTRODUCTION

The presence of organic compounds, nutrients, solids, pathogens, and other pollutants in wastewater made it to be traditionally considered an undesired waste. Due to the impact of these pollutants on the environment, wastewater must undergo energy-intensive treatment to remove them, prior to its discharge into natural aquatic environments. To protect human health and natural ecosystems, Wastewater Treatment Plants (WWTPs) have been designed to remove the pollutants contained in wastewater (Metcalf & Eddy, 2013).

With remarkable success WWTPs have fulfilled the tasks and, over the years, they relentlessly evolved to adapt to the increasingly stringent discharge limits. In the last decade, the transition towards a sustainable and circular water economy (CE) has stimulated a paradigm shift that is transforming the perception of sewage- and wastewater from an undesirable waste into a product that is rich in valuable resources to be recovered (Guest *et al.*, 2009), such as reusable water itself, nutrients and energy. From this perspective, a CE process aims at improving productivity of resources by keeping products, materials and infrastructure in use for longer than in the traditional linear 'take-make-consume-waste' economic model. CE has been promoted by policymakers (e.g., European Commission 2020) and adopted by many industries (Mhatre *et al.*, 2021). To reflect the increased focus on resource recovery in wastewater treatment (MacDonald & Crawford 2017) many WWTPs have been rebranded as Water Resource Recovery Facilities (WRRFs) as long as they incorporate some type of resource recovery process. However, the introduction of new process units for recovery and additional operational goals resulting from the shift from WWTPs to WRRFs renders the operation of the facilities more challenging.

The structural changes and new goals emphasize the need of efficient process monitoring and control tools, as well as multi-objective process optimization strategies (Arnell *et al.*, 2017; Solon *et al.*, 2019). At the same time, these technologies are key enablers of a second transition that municipal WWTPs are facing today: the digital transition. This transition aims at making WWTPs and WRRFs smart and their operations intelligent and efficient. According to Ingildsen and Olsson (2016) a smart water utility operates according to an optimal decision-making management, deployed at all process levels. Primary enablers of smart operations are online water quality and quantity sensors and process actuators and control levers. The use of these devices and their inclusion in computational solutions for designing operational strategies that account for the full water cycle, from water intake to water effluent, must be systematic and pervasive. The ultimate goal is a system-level management of the operations that is autonomously able of ensuring adequate water quality and quantity, with a minimum consumption of energy and materials, and minimum environmental impact.

Yet, today, most of wastewater treatment facilities operate using only basic sensor arrangements and the coupling between sensors and actuators is limited to simplified control schemes, if not ad hoc rules. Already at this level, the potential benefits offered by instrumentation, control, and automation (ICA) technologies remains largely under-utilized. Ingildsen and Olsson (2016) estimate that the use and exploitation of ICA technologies could, however, improve the capacity of a biological nutrient removal WWTP by 10–30% in the short term, and by 20–50% in the mid- and long-term (10–20 years from now). Moreover, the opportunities resulting from the availability of modern sensor technologies and more instrumented facilities are still to be discovered and used to support plant operation and management. Efficient statistical and data-driven models can be used to explore and model the wealth of information available in process data. This could lead to the discovery of new phenomena and would enable the development of process models and control strategies.

Although recommended and beneficial, it is essential to understand that having many installed sensors is not sufficient for a WWTPs and WRRFs to be considered smart facilities. To benefit from process data, raw measurements must be firstly processed and then made easily available for further use at different levels of granularity by process operators, engineers, and plant managers. The data collected from sensors must be (a) digested into process insight, the resulting knowledge is then (b) used to develop predictive models that can help characterise the state of the plant and its units, before it is eventually (c) embedded in automatic control structures where it is transformed into optimal control actions aiming at driving plant operations: The technology involved along this workflow can be understood as a combination of statistics, optimisation, and control theory or, as the general public oftentimes denotes it, machine learning and artificial intelligence.

Designing, developing, and then effectively using these technologies for operating wastewater treatment facilities is not a straightforward task. To start with, it is well-known that the quality and reliability of sensors' signals is affected by the harsh environment in which the instruments operate (corrosive and sludgy environment, as well as biofilm formation with biological activity are commonplace). In addition, the biochemical processes and the hydraulics occurring in wastewater treatment plants are complicated, highly nonlinear, and only partially understood from a mechanistic point of view. The portrait is completed when adding the challenges of controlling monolithic facilities perpetually operated in transient conditions. With such complex systems, when large amounts of data are collected, but simply stored in databases and not modelled to achieve specific process monitoring and control tasks, the risk of forming data graveyards is not negligible. This is a risk that must be confronted, mitigating actions must be taken to minimise it, and a clear roadmap to harness the full potential made available by modern data-based technologies laid down.

Our starting point is the recognition that, regretfully, in most wastewater treatment facilities, data quality and data analysis procedures are still rudimental, if existent at all.

To reverse this situation, several limitations need to be overcome and counteractions taken:

- the almost complete absence of automated data quality assurance, required to avoid "garbage in – garbage out"
- the almost complete absence of metadata collection that would make it easy the interpretation of the process data that it is being collected and that would enable its future use,
- the excessive cost of nutrient and organic matter sensors/analysers which moreover are labour maintenance intensive, fact that restrict their availability to the range of large facilities,
- the inclusion of process expertise in the development of the artificial intelligent tools and make them understandable by WWTPs' personnel to favour their short-term adoption.

Importantly, it is believed that a successful transition in the direction of digitalization and the adoption of new computational tools must be achieved through a cultural change in the organization and the management of the facilities and the resources available for their operation. This has been recognized by the many water utilities that have accepted the challenge and developed explicit roadmaps to be implemented step by step. In this regard, the human factor plays a central role in catalysing the digital transition. Because the primary users of smart applications are frontline staff responsible for daily operational decisions, it is important to involve them in developing these tools and to ensure that new information is presented in a user-friendly format and is actionable for their needs (Torfs *et al.*, 2022).

Adequate training of end-users is also a necessity to enable the successful adoption of new tools. Eerikäinen *et al.* (2020) found that employees of WWTPs are expecting next generation of digital tools for process data analysis. They emphasize that those tools should combine competences of both automation providers and wastewater process experts with a thorough understanding of treatment phenomena. The challenge for this is that the number of experts who have adequate skills and experience in both data techniques and treatment processes is limited. Therefore, it is expected that universities also enable and encourage learning this kind of mix of technologies in their curricula. That would promote new business opportunities in the form of, for instance, machine learning applications tailored for WWTPs, as well as services that facilitate the introduction of new tools with in-house data methodology understanding of new generation of workforce members. In fact, it also is of primary importance that universities provide students with sufficient mathematical education as that is the backbone required for learning advanced data analysis methods.

Even though the majority of smart water products that have been traditionally available on the market are targeted for water and wastewater network operations, some companies have started offering data quality management solutions and advanced modelling services which are specifically tailored for WWTPs (Corominas *et al.*, 2018). In addition, a large number of the computational tools that could be used to support the operations of WWTPs have been designed and developed in the academia. While many of these tools have mainly kept their research-oriented nature (Haimi *et al.*, 2013; Corominas *et al.*, 2018; Newhart *et al.*, 2019), it is important to note that successful smart water companies have often strong connections with university and routinely adopt ideas from novel academic research to add functionalities to their products.

Academia also plays a highly important role in showcasing the benefits of advanced data mining, modelling, and control systems to WWTP decision-makers and other stakeholders. To reach the multiple operational goals set when operating a modern facility, it is crucial that end-users are offered the opportunity to clearly appreciate the potential of advanced monitoring and control systems from an economic and safety of people, the environment as well as the equipment: As a driver of the cultural change towards efficient utilization of measured data for improved operation, that would actuate inclusion, for instance, of advanced monitoring and control systems in public procurements of water utilities when upgrading plants. In addition, procurements of smart systems for WWTPs are challenging to master, for example, because individual application solutions should be integrated into existing and future software platform solutions (Müller-Czygan 2020). Nevertheless, procuring advanced systems would also act as a driver for an increased competence of automation companies and consultants providing services for water utilities.

To avoid that the extraordinary digital transition opportunity that exists today be completely lost due to the aforementioned issues, this paper tries to shed light on the path towards increasing the value of the large amount of data that are being collected in current WWTPs and WRRFs making it possible to leverage the machine learning boom. It serves also as a roadmap to ease the intelligent automation of these facilities, thus paving the way to their digitization. Thus, digital transition could be safely embraced and the enormous potential of data analytics fully exploited, enabling it to play an essential role in the future automation and operation of our municipal facilities.

2 MAKING WASTEWATER TREATMENT FACILITIES INTELLIGENT

2.1 The digital transition in WWTPs

The digital transition aims at making the operation of wastewater facilities intelligent, being on-line monitoring, real-time process control and automation essential parts of the digitalization of wastewater facilities. Thus, on-line measurement data from the process are the foundation of a smart facility. To be profitable and allow process control and decision-making, the frequency of measurement of a variable (i.e., the temporal resolution of the sensor) should make it possible to capture its dynamics, i.e., the phenomena of interest is captured by the sensor (e.g., from lower to higher frequency needed: suspended solids in the reactor, influent flow rate, dissolved oxygen concentration in the aerobic reactor).

A wide range of physical and chemical parameters relevant to operation of WWTPs can be measured continuously or semi-continuously with commonly used sensors and analysers (Ingildsen & Olsson 2016). Development of novel instruments is still taking place, for instance, for measurement of organic carbon, metals, and emerging contaminants in wastewaters, for more affordable and improved nutrient sensors (Zhang *et al.*, 2020), and for enabling continuous monitoring of some parameters relevant to optimization of anaerobic digestion processes (e.g., measuring individual volatile fatty acid species) (Jimenez *et al.*, 2015).

Figure 1 shows a scheme of the water line and the sludge line including the typical sensors that could usually be deployed within a standard wastewater facility. Please note that mainly process variables (also known as secondary variables) are recorded which are easy-to-measure with relatively cheap sensors. Less frequently available (in small and medium-size WWTPs) quality variables (also known as primary variables) - like nutrients and organic matter which are measured with expensive sensors/analyzers.

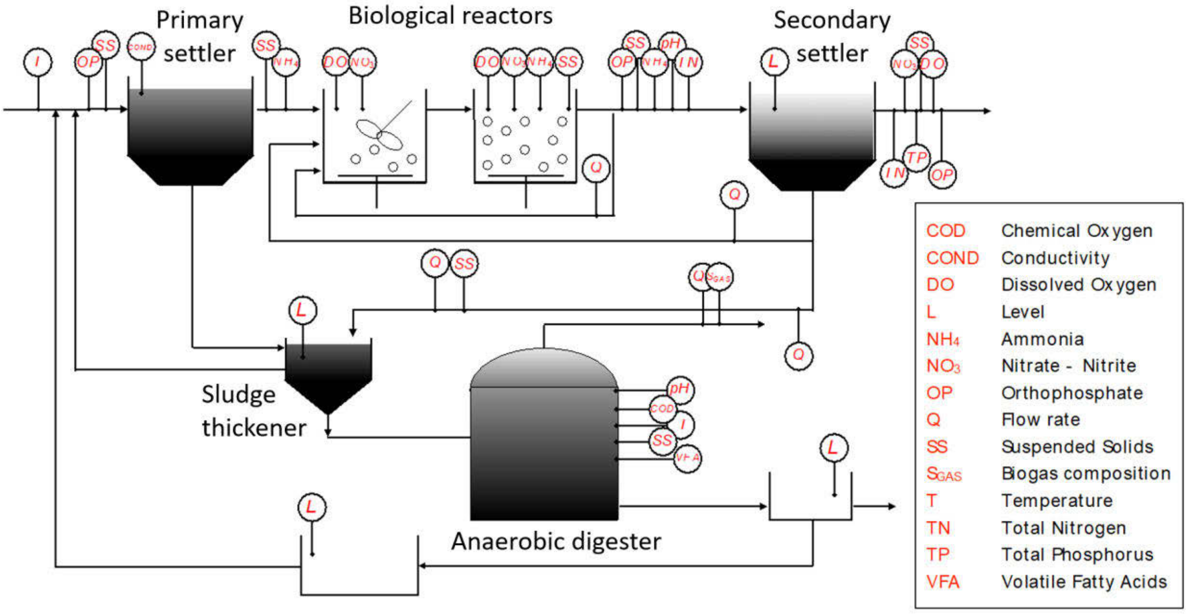

Primary settler
Biological reactors
Secondary settler
Sludge thickener
Anaerobic digester
T
OP
SS
COND
SS
NH4
DO
NO3
DO
NO3
NH4
SS
SS
pH
OP
NH4
TN
L
SS
NO3
DO
TN
TP
OP
Q
Q
Q
SS
Q
SGAS
L
Q
pH
COD
T
SS
VFA
L
L
COD Chemical Oxygen
COND Conductivity
DO Dissolved Oxygen
L Level
NH4 Ammonia
NO3 Nitrate - Nitrite
OP Orthophosphate
Q Flow rate
SS Suspended Solids
SGAS Biogas composition
T Temperature
TN Total Nitrogen
TP Total Phosphorus
VFA Volatile Fatty Acids

Figure 1. Typical sensors found in the water and sludge lines of a conventional wastewater treatment facility.

Although there is a wide variety of sensors available, practical challenges still exist: operators of treatment facilities find, for instance, quality of measured data and laborious maintenance needs of instrumentation as barriers for efficient use (Eerikäinen *et al.*, 2020). Particularly, there are simpler sensors (e.g. dissolved oxygen, pH, flow, level) that have been proven to be robust, sufficiently accurate and need minimum maintenance, but challenges concern more advanced sensors (e.g. ammonium, nitrate, nitrite and phosphate) that have been found to be less reliable, preventing a wider application of the advanced control algorithms that are dependent on these sensors (Yuan *et al.*, 2019). Even there are a number of undesired sensor states: excessive drift, shift, fixed value, complete failure, wrong gain and isolated fault (a single incorrect value) (Therrien *et al.*, 2020).

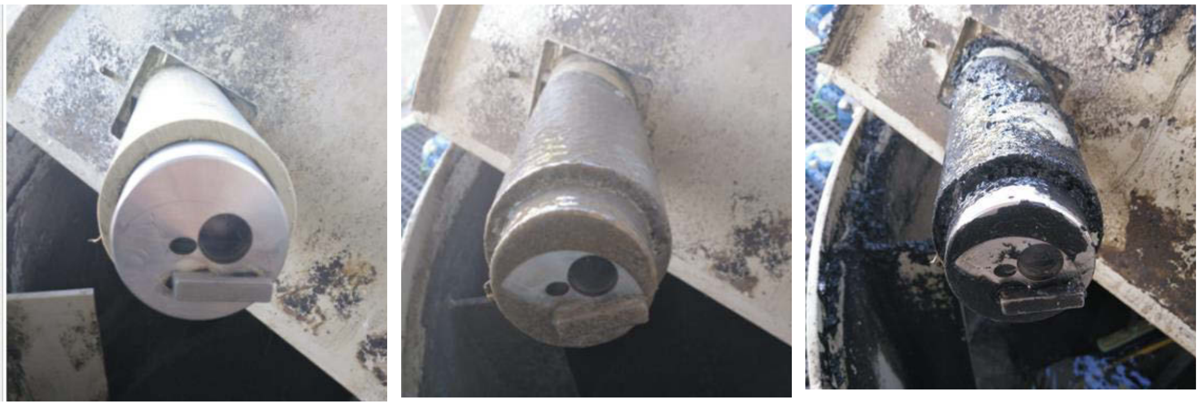
Recent instrumentation surveys in Swedish and Danish WWTPs (Åmand *et al.*, 2017; Nilsson & Andersson 2018, respectively) indicate that less than 70 % of investigated facilities in both countries had written instructions for quality control of at least part of the instruments. However, because large amounts of measured data are generated in modern WWTPs, in addition to conventional sensor maintenance efforts, incorporation of techniques for fast detection and diagnosis of faults are needed to guarantee sufficient data quality (Corominas *et al.*, 2018). Another challenge that is also emphasized by increasing amounts of measured data in WWTPs, is that stored sensor data is often not augmented with adequate meta-data i.e. descriptive information, which hinders the use of historical data to address future problems (IWA 2021a).

In our experience, efforts for keeping instrumentation operational varies quite much between facilities even if their size and personnel numbers would be similar. Also other reasons than laborious maintenance and incomplete technology have a crucial impact on this: if operators do not understand the value of certain measurements for process operation, they easily loose interest in those sensors (Olsson & Ingildsen 2018). Therefore, implementation of each sensor should be justifiable, and their purpose and received benefits need to be thoroughly explained to employees who operate the process. This is of key importance for successful instrumentation and control design projects. Moreover, adequately maintained sensors are vital for smart wastewater treatment facilities.

2.2 Stages to transform raw data into actionable insight

2.2.1 From raw data to quality checked data fitted for purpose

As can be seen in Figure 2, the first tough challenge for ICA in a wastewater facility is the harsh environment in which the instrumentation has to work (corrosive, sludgy, biofilm formation with biological activity…), which directly impacts the quality and reliability of a sensors' signal. Surprisingly, despite this is evident and known, data quality and data analysis are essentially non-existent in most wastewater facilities around the world. Therefore, nowadays with numerous sensors installed in many wastewater facilities, a huge amount of data is being collected that is neither analysed nor utilized, resulting in data graveyards (Corominas *et al.*, 2018). To make the situation even worse, the almost complete absence of metadata (that would enable the correct interpretation of the collected data) prevents its future use.



(a) **(b)** **(c)**

Figure 2. Pictures of a suspended solids probe installed in the buffer tank of a WWTP: (a) new probe (b) freshly removed from the buffer tank (c) after cleaning with water jet. This illustrates the harsh environment in which the instrumentation has to work (corrosive, sludgy, biofilm formation with biological activity…).

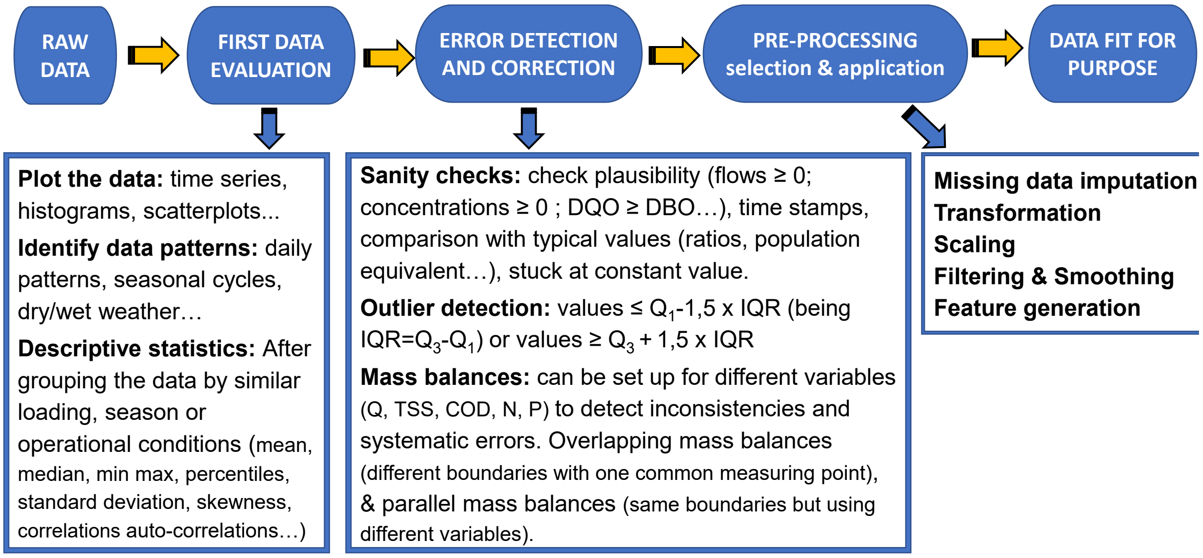
BY NC SA

Since the on-line measurements are the basis of a smart facility, appropriate data quality check is vital to make it possible the exploitation of the ICA. It is evident and of paramount importance for process operation that every on-line measurement should have to be quality checked prior to be used in a control loop.

The raw collected data must be properly processed and made easily available for further use to make them truly useful. However, there is no general standard on how the on-line quality check should be done (Ingildsen & Olsson 2016). Figure 3, shows different pre-processing steps that can be applied to the collected data (raw data) to check its quality and improve its information content. Depending on each particular case, some steps or others can be applied. For example, for automatic control purposes human intervention should be not required, while for process modelling purposes it will be a quite important component: starting from the initial visualization of the data by the wastewater treatment expert to gain an overview and feeling of the plotted data to the application of mass balances to detect inconsistencies in the collected data.

A relatively easy to implement automated data quality assurance workflow would include the following stages: **raw data -> sanity checks -> outlier detection -> pre-processing (missing data imputation, scaling and filtering) -> Data fit for purpose**. As can be seen in Figure 3, the end of the pre-processing steps results in "data fit for purpose", which is quality-checked data that has been reshaped into a better form for further analysis by the methods shown in Figure 4.

The pre-processed on-line measurement can be used in a control loop (i.e., used as input of the control algorithm), either directly (e.g., the oxygen concentration that is used to regulate the amount of air supplied by the blowers to the aerobic reactor) or indirectly via the on-line estimation of another variable using for example a data-driven model (e.g. a soft-sensor based on an artificial neural network or on a support vector machine or on a partial least squares model....) or simply multiplied by another variable. There are several signals of interest in the wastewater treatment context that are obtained by a simple combination of multiple signals (e.g., the organic mass flow (Q x COD), the solids mass flow (SS x Q),...).


RAW DATA
FIRST DATA EVALUATION
ERROR DETECTION AND CORRECTION
PRE-PROCESSING selection & application
DATA FIT FOR PURPOSE

Plot the data: time series, histograms, scatterplots...

Identify data patterns: daily patterns, seasonal cycles, dry/wet weather…

Descriptive statistics: After grouping the data by similar loading, season or operational conditions (mean, median, min max, percentiles, standard deviation, skewness, correlations auto-correlations…)

Sanity checks: check plausibility (flows ≥ 0; concentrations ≥ 0 ; DQO ≥ DBO…), time stamps, comparison with typical values (ratios, population equivalent…), stuck at constant value.

Outlier detection: values ≤ Q_1-1,5 x IQR (being IQR=Q_3-Q_1) or values ≥ Q_3 + 1,5 x IQR

Mass balances: can be set up for different variables (Q, TSS, COD, N, P) to detect inconsistencies and systematic errors. Overlapping mass balances (different boundaries with one common measuring point), & parallel mass balances (same boundaries but using different variables).

Missing data imputation
Transformation
Scaling
Filtering & Smoothing
Feature generation

Figure 3. Pre-processing steps that can be applied on raw data to check its quality and improve its information content. Selecting appropriate steps, a data pipeline can be developed for each case.

2.2.2 From good quality data to data-driven models

From the perspective of data processing, the challenge is to effectively use the data that can be acquired for a modern wastewater treatment plant for developing models useful for achieving reuse, resource and energy recovery, and minimal carbon and greenhouse gas footprint. Within

BY NC SA

machine learning and artificial intelligence, the task is approached mainly as supervised and unsupervised learning.

In a typical supervised learning scenario, the output, called response or dependent variable, is a quantitative (such as "difficult-to-measure-variables") or categorical (such as effluent quality or process indicator) that we wish to predict. The output type leads to a naming convention for the prediction tasks: regression (to predict quantitative outputs) and classification (to predict qualitative outputs). In both situations, a set of measurements data from the sensors and the laboratory/experimental analysis is available. These are the features, also called predictors, inputs variables or independent variables. We have a training set of data, in which we observe the outcome and feature measurements for a set of samples. Using these data a prediction model is built which will enable to predict the outcome for new unseen test samples. Unsupervised learning, often performed as part of the exploratory data analysis, refers to a situation in which for every feature, we observe a vector of measurements but no associated response. The goal is to directly infer some interesting properties of the process from the available features without the help of an associated response variable providing correct answers or degree-of-error for each observation (Hastie *et al.*, 2017).

Figure 4 schematically summarises the methodologies applied to the pre-processed data within the wastewater treatment applications. Most studies focused on regression problems to predict and monitor the output variables based on a given number of historical observations, as summarized by Haimi *et al.* (2013) and lately by Ching *et al.* (2021). Recent machine learning algorithms for WWTP classification problems include random forest, tree-based algorithms, support vector machine and the comparison of different methods in various applications (as for instance in Guo *et al.*, 2015, Nourani *et al.*, 2018, Wang *et al.*, 2021 and 2022). Unsupervised methods have been utilized for determining changes in process variables and for anomaly detection (Corominas *et al.*, 2018). A recent example of clustering application is Xu *et al.* (2021) for optimizing the processes configuration of full-scale WWTP predesign through an integrated strategy consisting of t-distributed stochastic neighbour embedding (t-SNE) and deep neural networks (DNNs).

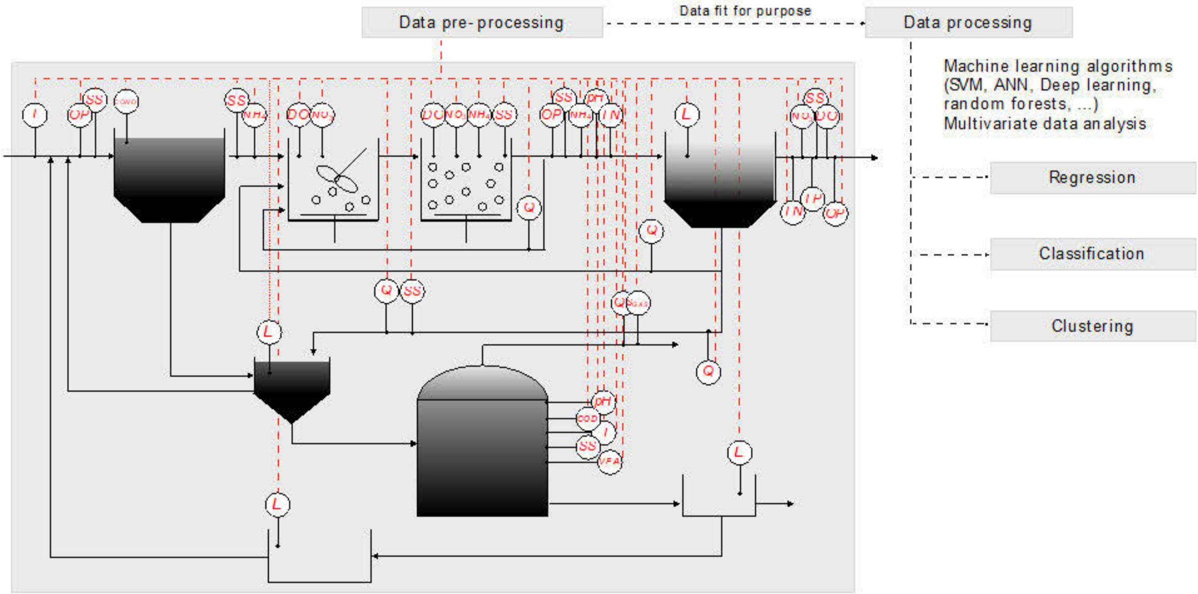


Figure 4. Data-driven techniques that can be applied to the quality-checked data that is fit for purpose.

BY NC SA

2.2.3 From good quality data to WWTP process control

The development and implementation of data-driven models and automatic process controllers will make the collected data truly useful. Real-time process control and automation can significantly contribute to the optimization of different processes (chemical dosing, pumping, aeration, energy consumption) and can take care of the repetitive low-level tasks necessary to keep the facility running. It should be highlighted that sound control actions require good quality data. The increasingly complex control strategies (required to efficiently operate the increasingly complex WWTPs) will fail more often the lower the quality of the measured process data. There are control loops based on variables that are directly measured from the process (e.g., dissolved oxygen concentration) or on the output of a data-driven model (e.g., soft-sensor, or a data-driven predictive model).

From the perspective of automatic control, the challenge is a system-level technology to optimally operate WWTPs in a management of wastewater that includes reuse, resource and energy recovery, and minimal carbon and greenhouse gas footprint. Ideally, facilities and recipients of recovered resources must be operated as interacting entities, to satisfy operational objectives that aim at matching demand and resources, while always ensuring safety of people and equipment, environmental permits, and sustainability boundaries.

Nowadays, there is no standardised control solution for WWTPs that aim at being operated as WRRFs. General solutions for control algorithms are useless because they are case-specific, and the algorithms require tunning on a regular basis. To tune them, skilled personnel are needed. It is still rare to encounter WWTPs in which downstream operations are accounted for to define the planning over the recovery of energy and materials: When the operational goal is dictated only by disposal permits that are unaware of the fate of recoverable resources, plant management cannot be expected run WWTPs as bio-refineries within the water chain. Yet, there exists practical evidence that wastewater treatment facilities often times have the technological capability and flexibility to be operated towards these objectives.

In a less heuristic approach, the goals of resource recovery, subject to neutral- or positive-energy constraints and minimal air emission footprint, must be formulated as explicit control objectives to be achieved by manipulating material and energy fluxes across WWTPs, in response to downstream needs and upstream conditions. By coordinating the right synergy between treatment plants on the one hand, and the receivers of recovered resources on the other, flexibility between treatment, reuse, and resource recovery can be largely achieved by making the best use of existing facilities, with little-to-none capital investments for upgrading or retrofitting. In this framework, plant-wide planning must be designed around controllers that must be capable to (a) deal with complex and uncertain unit- and plant-wide dynamics; (b) satisfy the constraints given by current and forthcoming permits; (c) manage the production of reuse water and recovered resources; (d) enforce energetic neutrality, if not positivity; and (e) minimise the environmental impact of the plant. Moreover, controllers must be capable to determine an optimal equilibrium when conflicting objectives are at stake.

The importance of optimising the operations of a WWTP using automatic control is largely recognised, both from an environmental and an economical viewpoint (Ingildsen & Olsson 2016). Efforts made to reduce energy use by replacing specific devices (Daw *et al.*, 2012) focused on the control of aerobic processes (Mulas *et al.*, 2015, Stentoff, 2020], but also on structural changes in process configuration (Sarpong and Gude, 2020). Recently, Neto et al. (2022) have shown that conventional activated sludge processes can be optimised with respect to non-conventional objectives, like quality and quantity of released water. While these results are rooted on the availability of dynamic models (Henze 2020) that enable the definition of advanced control strategies, progress remains to be made to integrate the mechanistic models in the determination of the technological margins for controlling WWTPs as WRRFs, and on how to safely complement them with empirical counterparts. There exist critical knowledge gaps and open questions also regarding the feasibility of full energy and nutrient recovery, and on the thresholds as of when these practices are safe. It is also unclear what are the conditions for environmental neutrality. These questions are at the core of the research.

A key pathway towards a smart management of wastewater treatment systems and the foundation for a sustainable water management is built upon the development of information and decision-support systems. In a model-based approach, state-space process models derived from first-principles, either mechanistic or statistical, are learn using process measurements, then analysed and finally embedded in receding-horizon controllers for optimal decision-making and planning. The integration of explicit operational policies for supervising smart operations, on a high-level, and the deployment of regulatory actions with low-level controllers can be developed according to a general control architecture consisting, in its basic formulation, of a dynamic process model, a state estimator, and a predictive controller.

In Figure 5, the integration of an optimal controller is illustrated on a conventional activated sludge process consisting of a certain number of actuators (in this case, 13), operated using low-level PI controllers whose set-points are dynamically determined using a predictive controller. The controller, in turn, determines the set-points as the decision variables that optimize a user-defined operational objective over a fixed time-horizon, subject to the dynamics of the plant (a process model) and to a number of technological and operational constraints. As the controller requires knowledge of the current state of the process, this information is reconstructed for sensor measurements (here, 14) by a state estimator, again based on a dynamical model of the process. In the example, the process model used to represent the plant is the BSM1.

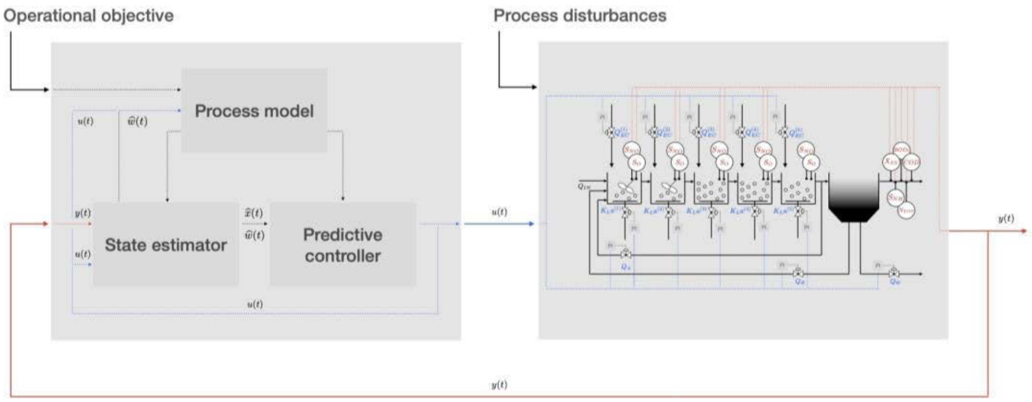


Figure 5. A schematic of a model-based predictive controller for a conventional activated sludge process. The actions of the actuators (in blue) are defined using PI controllers whose setpoints are determined using a predictive controller. Sensor measurements (in red) are used by a state estimator to determine the current state of the process. Both the controller and the state estimated are based on a mechanistic process model.

In Figure 6, the architecture is instantiated to use a model-predictive controller (MPC) and a moving-horizon estimator (MHE). The process model is explicitly described by a state-space formulation, with a set of dynamic models (the 145 differential equations in the BSM1) and measurement models (14 algebraic equations defining the measurements as function of the process state, again based on the BSM1). From a process perspective the actuators allow for the control of aeration and the addition of external carbon sources to the biological reactors, and the control of sludge recycles and removal in the secondary settler.

BY NC SA

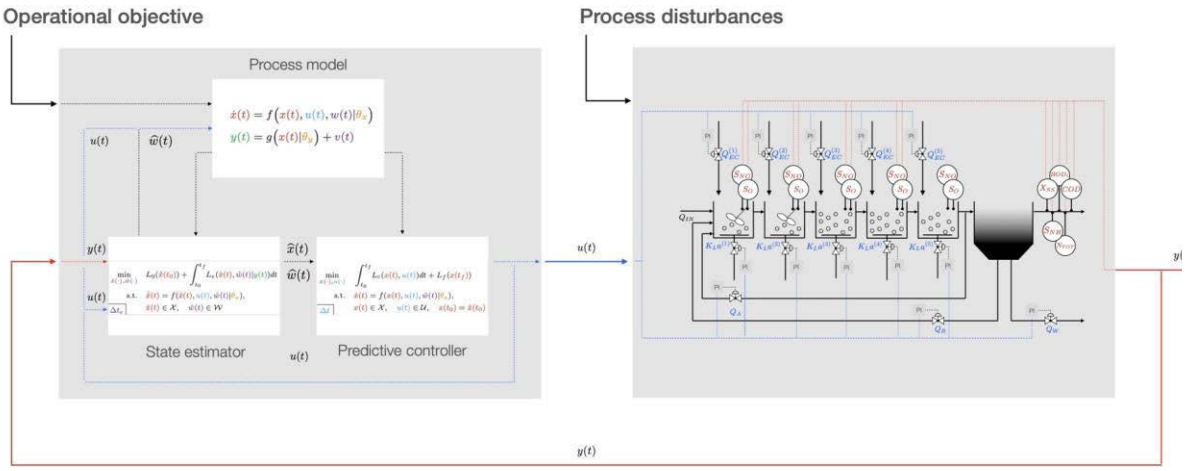


Figure 6. General structure of a model-based predictive control (MPC) and moving-horizon state estimator (MHE). Both the MPC and the MHC solve a nonlinear optimization problem, subject to the process dynamics, encoded by the process model, and a number of technological and operational constraints after

When compared to simpler error-feedback control strategies that require an ad-hoc pairing between controlled and manipulated variables, a model-based control architecture defines the control actions over the entire set of actuators comprehensively, at the plant-level, according to a state-feedback principle. It is also important to note that additional modules can be integrated into this general structure. Typically, it is expected to include a fault-analysis module that certifies the normality of the operations and health status of the equipment: This module is necessary to authorize the deployment of the supervisory actions to the regulatory layer. Moreover, it is often beneficial to add a module that support planning by predicting the future evolution of the disturbances. More advanced modules with data analysis capabilities can be included to support monitoring, from high-level KPIs to instrumentation, at different time scales and process levels.

To simulate real-world situations in a more realistic way allowing to test control actions and the staff to experience what different situations would be like in real life (without the associated costs and dangers), there is a current trend to develop digital twins (DT). DTs are virtual representations that serve (near) real-time counterparts of physical objects (twins). The core of DTs of wastewater treatment processes are process models that often are mechanistic models (for instance, ASM and ADM models), but hybrid models have also been found promising approaches (Torfs *et al.,* 2022). According to these authors, three key features that separate DTs of wastewater treatment processes from off-line models are that (a) a physical counterpart for model must exist, (b) there is an automated data connection to the physical twin, and (c) there needs to be means to continuously update the process model according to evolution of the physical process over time. Because an essential property of a DT is use of near live data, appropriate automated data management is crucial for successful implementations. DTs have been used, for instance, for evaluating current process status and for performing automated scenario analysis in the Singapore PUB Changi WRRF (Johnson *et al.*, 2021), for process monitoring and operational advice with focus on improving resource recovery and reducing energy footprint in Egå WRRF (Denmark) (Polesel *et al.,* 2021) and for predictive control of influent flow in Kolding WRRF (Denmark) (Stentoft *et al.,* 2020). Other potential applications of DTs in treatment plants include operator training, failure analysis, and asset management and predictive maintenance (Torfs *et al.*, 2022).

BY NC SA

3 CONCLUSIONS

Collected data in WWTPs and WRRFs can be utilized for many useful purposes: process monitoring, real-time process control and automation, increase process knowledge, process optimization, data-driven modelling, data mining, better understanding of process state, fault prediction, decision-making. Currently a lot of data is collected in WWTPs and WRRFs, but most of them is not used and the almost complete absence of metadata prevents its future use not having the necessary knowledge of its context to allow their interpretation. In this paper it has been shown how to increase the value of the large amount of data that nowadays are being collected in WWTPs and WWRFs. The main take-home messages are the following:

- Nowadays, many facilities around the world are not exploiting the value of the available data, being data rich yet information poor a frequent situation that results in data graveyards.
- It has been shown that it is possible to take advantage of the data that nowadays it has been collected with the already deployed instrumentation in each facility, by assuring its quality (an automatic quality workflow has been proposed) and storing its context and related information (metadata) to enable its interpretation and future use.
- The deployment of real-time automatic process control algorithms and data-driven models can avoid some repetitive low-level tasks to process operators while keeping the WWTP running in a cost-effective way, thus making the collected data truly valuable and useful.
- The human factor is vital for a successful digital transition: the staff should be involved in the development of the artificial intelligent tools (e.g., taking advantage of their process expertise and know-how), as well as be trained on these tools (so the staff can understand the new smart-tools and some of them even update or tuning them).
- More information from the process can be obtained deploying new sensors and quality probes, which will allow the development of more complex control strategies, as well as a more detailed and in-depth knowledge of the monitored processes, which is always valuable for an informed decision-making.

4 REFERENCES

[1] Arnell M., Rahmberg M., Oliveira F, Jeppsson U. (2017): Multi-objective performance assessment of wastewater treatment plants combining plant-wide process models and life cycle assessment. Journal of Water and Climate Change, 8(4), pp. 715-729.

[2] Ching P.M.L, So R.H.Y., Morck T. (2021) Advances in soft sensors for wastewater treatment plants: A systematic review. Journal of Water Process Engineering, 44.

[3] Corominas Ll., Garrido-Baserba M., Villez K., Olsson G., Cortes U., Poch, M. (2018). Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. Environ. Model. Software, 106, 2018, 89–103.

[4] Daw J., Hallet K., DeWoldfe J., Venner I. (2012) Energy efficiency strategies for municipal wastewater treatment facilities. Technical Report. Department of Energy Office of Scientific and Technical Information.

[5] Eerikäinen S., Haimi H., Mikola, A., Vahala R. (2020) Data analytics in control and operation of municipal wastewater treatment plants: qualitative analysis of needs and barriers. Water Sci. Technol. 82(12), 2681–2690.

[6] Elsayed A., Siam A., El-Dakhakhni W. (2022) Machine learning classification algorithms for inadequate wastewater treatment risk mitigation, Process Safety and Environmental Protection, 159, pp.1224-1235.

[7] European Commission (2020) A new circular economy action plan – for a cleaner and more competitive Europe. Brussels, Belgium.

[8] Guest J.S., Skerlos S.J., Barnard J.L., Beck M.B., Daigger G.T., Hilger H., Jackson S.J., Karvazy K., Kelly L., Macpherson L., Mihelcic J.R., Pramanik A., Raskin L., van Loosdrecht M.C.M., Yeh D. and Love N.G. (2009) A New Planning and Design Paradigm to Achieve Sustainable Resource Recovery from Wastewater, Environmental Science and Technology, 43, 6126-6130

[9] Guo H., Jeong K., Lim J., Jo J., Kim Y. M., Park J., Kim J.H., Cho K. H. (2015) Prediction of effluent concentration in a wastewater treatment plant using machine learning models, Journal of Environmental Sciences, 32, pp. 90-101.

[10] Haimi H., Mulas M., Corona F., Vahala R. (2013) Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environ. Model. Softw. 47, 88–107.

[11] Hastie T., Tibshirani R., Friedman J. (2017) The Elements of Statistical Learning - Data Mining, Inference, and Prediction. Second Edition, Springer.

[12] Henze (2000). Activated sludge models ASM1, ASM2 and ASM3 IWA Publishing. ISBN:9781900222242

[13] Ingildsen P., Olsson G. (2016) Smart Water Utilities: Complexity made simple. IWA Publishing, London.

[14] IWA (2021a) Digital Water: The Value of Meta-Data for Water Resource Recovery Facilities. White Paper. London, UK, p. 15.

[15] Jimenez J., Latrille E., Harmand J., Robles A., Ferrer J., Gaida D., Wolf C., Mairet F., Bernard O., Alcaraz-Gonzalez V., Mendez-Acosta H., Zitomer D., Totzke D., Spanjers H., Jacobi F., Guwy A., Dinsdale R., Premier G., Mazhegrane S., Ruiz-Filippi G., Seco A., Ribeiro T., Pauss A., Steyer J.-P. (2015) Instrumentation and control of anaerobic digestion processes: a review and some research challenges. Rev. Environ. Sci. Bio/Technology 14, 615–648.

[16] Johnson B.R., Kadiyala R., Owens G., Mak Y.P., Grace P., Newbery C., Sing S., Saxena A., Green J. (2021) Water reuse and recovery facility connected digital twin case study: Singapore PUB's Changi WRP process, control, and hydraulics digital twin. Proceedings of WEFTEC 2021, Chicago, USA.

[17] McDonald G., Crawford J. (2017). Energy neutral WWTPs - A dream or a reality for New Zealand? The Water New Zealand conference & Expo 2017.

[18] Mhatre P., Panchal R., Singh A., Bibyan S. (2021): A systematic literature review on the circular economy initiatives in the European Union. Sustainable Production and Consumption, 26, pp. 187-202

[19] Metcalf & Eddy (2013). Wastewater Engineering: Treatment and Resource Recovery. McGraw Hill; 5th edition; ISBN-13 : 978-0073401188.

[20] Mulas M., Tronci S., Corona F., Haimi H., Lindell P., Heinonen M., Vahala R., Baratti R. (2015). Predictive control of an activated sludge process: An application to the Viikinmäki wastewater treatment plant. Journal of Process Control, 35:89-100

[21] Müller-Czygan, G. (2020) Smart Water–How to Master the Future Challenges of Water Management; Chandrasekaran P.T., Javaid M.S., Sadiq A., Eds.; Resources of Water; IntechOpen: London, UK, 19–33.

[22] Neto O.B.L., Mulas M., Corona F. (2022). About the classical and structural controllability and observability of a common class of activated sludge plants. Journal of Process Control, 11, 8-26.

[23] Newhart K.B., Holloway R.W., Hering A.S., Cath T.Y. (2019) Data-driven performance analyses of wastewater treatment plants: A review. Water Research, 157(6), pp. 498-513

[24] Nilsson S., Andersson S. (2018) Summary of instrumentation survey Denmark. IVL Swedish Environmental Research Institute Ltd., Stockholm, Sweden.

[25] Nourani V., Elkiran G., Abba, S. I. (2018) Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach, Water Science and Technology, 78 (10), pp. 2064-2076.

[26] Olsson G ., Ingildsen P. (2018) Making water operations smarter. In Advances in Wastewater Treatment (Mannina, G., Ekama, G., Ødegaard, H. and Olsson, G., eds). IWA Publishing, London.

[27] Polesel, F., Remigi, E.U., Sanford, R.K.E., Merrild, M.H.B., Nissen, Thage, R.K., Lynggaard-Jensen, Ramin, A.P., Drejer, L.H., Andreasen, P., Sørensen, H.R. and Dalkvist, T. (2021) Digital Twinning of Water Resource Recovery Facilities – From concept to implementation. Proceedings of 7th IWA Water Resource Recovery Modelling Seminar WRRMod2021, virtual conference.

[28] Sarpong G., Gude V.G. (2020). Near future energy self-sufficient wastewater treatment schemes. Int. Journal of Environmental Research, 14:479-488.

[29] Solon K., Volcke E.I.P., Sperandio M., van Loosdrecht M.C.M. (2019) Resource recovery and wastewater treatment modelling. Environmental Science: Water Research and Technology, Royal Society of Chemistry, 2019, 5 (4), pp.631-642.

[30] Stentoff P.A. (2020). Stochastic modelling and predictive control of wastewater treatment processes. PhD thesis, Technical University of Denmark.

[31] Stentoft P.A., Vezzaro L., Mikkelsen P.S., Grum M., Munk-Nielsen T., Tychsen P., Madsen H., Halvgaard R. (2020) Integrated model predictive control of water resource recovery facilities and sewer systems in a smart grid: example of full-scale implementation in Kolding. Water Science and Technology, 81, 1766–1777.

[32] Therrien J.-D., Nicolaï N. Vanrolleghem P.A. (2020) A critical review of the data pipeline: How wastewater system operation flows from data to intelligence. Water Sci. Technol. 82 (12), 2613–2634.

[33] Torfs E., Nicolaï N., Daneshgar S., Copp J.B., Haimi H., Ikumi D., Johnson B., Plosz B.B., Snowling S., Townley L.R., Valverde-Pérez B., Vanrolleghem P.A., Vezzaro L. Nopens I. (2022). The transition of WRRF models to Digital Twin applications. Water Science & Technology, accepted for publication.

[34] Xu R., Cao J., Fang F., Feng Q., Yang E., Luo J. (2021) Integrated data-driven strategy to optimize the processes configuration for full-scale wastewater treatment plant predesign, Science of The Total Environment, 785, 147356,

[35] Yuan Z., Olsson G., Cardell-Oliver R., van Schagen K., Marchi A., Deletic A., Urich C., Rauch W., Liu Y. Jiang G. (2019) Sweating the assets – the role of instrumentation, control, and automation in urban water systems. Water Res., 155 (2019), pp. 381-402

[36] Zhang W., Tooker N.B., Mueller A.V. (2020) Enabling wastewater treatment process automation: leveraging innovations in real-time sensing, data analysis, and online controls. Environ. Sci. Water Res. Technol., 6, pp. 2973-2992

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14249

METADATA: A MUST FOR THE DIGITAL TRANSITION OF WASTEWATER TREATMENT PLANTS

Daniel Aguado García[1], Frank Blumensaat[2], Juan Antonio Baeza[3], Kris Villez[4], María Victoria Ruano[5], Oscar Samuelsson[6], Queralt Plana[7] and Janelcy Alferez[8]

[1] CALAGUA – Unidad Mixta UV-UPV, Institut Universitari d'Investigació d'Enginyeria de l'Aigua i Medi Ambient – IIAMA, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain.
[2] ETH Zurich/Eawag, Institute of Environmental Engineering, Chair of Urban Water Management Systems, Zurich, Switzerland
[3] Department of Chemical, Biological and Environmental Engineering, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain
[4] Oak Ridge National Laboratory, Oak Ridge, TN, USA
[5] Chemical Engineering Department, Universitat de València, Burjassot, Spain
[6] IVL Swedish Environmental Research Institute, Sweden
[7] modelEAU, Université Laval, Québec, Canada
[8] VITO - Vision on technology, Boeretang 200, BE-2400 MOL, Belgium.

[1] *daaggar@hma.upv.es*, [2] *Frank.Blumensaat@eawag.ch*, [3] *JuanAntonio.Baeza@uab.cat*,
[4] *villezk@ornl.gov*, [5] *m.victoria.ruano@uv.es*, [6] *oscar.samuelsson@ivl.se*,
[7] *queralt.plana@gmail.com*, [8] *janelcy.alferes@gmail.com*

Abstract

The increment in the number and diversity of available (and affordable) sensors together with the advances in information and communications technologies have made it possible to routinely measure and collect large amounts of data at wastewater treatment plants (WWTPs). This enormous amount of available data has boosted the interest in applying sound data-driven solutions to improve the current normal daily operation of these facilities. However, to have a real impact in current operation practices, useful information from the massive amount of data available should be extracted and turned into actionable knowledge.

Machine learning (ML) techniques can search into large amounts of data to reveal patterns that a priori are not evident. ML can be applied to develop high-performance algorithms useful for different tasks such as pattern recognition, anomaly detection, clustering, visualization, classification, and regression. These ML algorithms are very good for data interpolation, but its extrapolation capabilities are low. Hence, the data available for training these data-driven models require data covering the complete space for the independent variables. A significant amount of data is required for this purpose, but data of good quality.

To transform big data into smart data, giving value to the massive amount of data collected, it is of paramount importance to guarantee data quality to avoid "garbage in – garbage out". The reliability of on-line measurements is a hard challenge in the wastewater sector. Wastewater is a harsh environment and poses a significant challenge to achieve sensor accuracy, precision, and responsiveness during long-term use. Despite the huge amount of data that are currently being recorded at WWTPs, in many cases nothing is yet being done with them (resulting in data graveyards). Moreover, the use of the data collected is indeed very limited due to the lack of documentation of the data generation process and the lack of data quality assessment.

Metadata is descriptive information of the collected data, such as the original purpose, the data-generating devices, the quality, and the context. Metadata is needed to clearly identify the data that should be used for the development of data-driven models. These data should be selected from the same category. If we include data that shouldn't be in the same data set because they were obtained under different operational conditions, this would lead to unreliable model predictions. ML algorithms learn from data, thus to be useful tools and to really improve the decision-making process in WWTP operation and control, representative, reliable, annotated and high-quality data are needed.

Effective digitalization requires the cultivation of good meta-data management practices. Unfortunately, there are no wastewater-specific guidelines available to the production, selection, prioritization, and management of meta-data. To address this challenge, the IWA Task Group on Meta-Data Collection and Organisation (MetaCO TG to which the authors of this paper belong) which has been supported by the International Water Association since 2020 will soon finish the scientific and technical report containing such guidelines specifically for WWTPs. This paper highlights why meta-data should be considered when collecting data as part of good digitalisation practices.

Keywords
Digitalization, metadata, wastewater treatment plant, water resource recovery facility.

1 INTRODUCTION

Water is a scarce natural resource, essential for life and for the exercise of the vast majority of economic activities. After its use, whether for human or industrial activities, its composition and quality notably degrade, and it becomes wastewater. Treatment of urban wastewater is essential to protect the aquatic environment as well as the human health [1]. The importance of access to clean water and sanitation as well as clean water at sea is embedded in goals 6 and 14 of the United Nations' Sustainable Development Goals (SGDs). Wastewater treatment is essential to reduce marine pollution (SDG 14); the development of energy-efficient treatment and control solutions for pollutants' removal (or resource recovery) and the production of renewable energy from the organic matter, can contribute to the affordable and clean energy goal (SDG 7) as well as to achieve more sustainable cities (SDG 11). Water scarcity, poor water quality and inadequate sanitation negatively impact food security.

Wastewater treatment is currently undergoing a paradigm shift as a result of the transition to the circular economy. This paradigm change consists in ceasing to consider wastewater as a mere waste, to consider it a source of resources and energy [2]. Wastewater Treatment Plants (WWTPs) can contribute to the circular economy in different ways, such as producing a water effluent that can be reused, resource recovery (e.g. phosphorus is a macro-nutrient essential for live and a non-renewable resource that can be recovered in WWTPs) and clean energy production through anaerobic digestion of the organic matter contained in the influent wastewater.

Traditionally, wastewater treatment has been centralized, due to the economy of scale associated with the construction of WWTPs. However, in less populated areas, other technologies have been implemented due to different reasons: less inversion is possible, the lack of suitable qualified personal to operate the WWTP, and the need to treat very low flow rates as well as to deal with strong fluctuations in influent flow and composition. As a consequence, many different treatment schemes and configurations for wastewater treatment can be found today, even in the same city. Moreover, the type and level of instrumentation, control and automation varies significantly from one WWTP to another. There are analysers whose installation and maintenance costs (require frequent maintenance by skilled staff) may not be covered by the energy cost savings on small or remotely located WWTPs. All these factors make that the number of analogical and digital signals registered vary widely from less than 500 to more than 30,000 from one WWTP to other [3].

Recent advances in information and communications technologies (ICT), online sensors, and autonomous energy supplies make ubiquitous sensing of WWTP viable today, even in remote locations. The monitoring capabilities offered by the ICT will greatly improve the decision-making process for design, operation, and control, but this will only be realized if the data produced by and sent to devices can be trusted with very high reliability [4].

The amount and diversity of available sensors in WWTPs and other data from the processes that take place in these facilities has increased massively in the last decades [5]. Thanks to increased efficiency in communication networks and extreme reductions in data storage costs, data collection is extremely scalable which means today's WWTPs have entered the era of big data by covering the three V's introduced by Doug Laney in 2001 [6]. Volume (a high amount of data is gathered from the instrumentation deployed at WWTPs), velocity (data values are produced at high speed, with sensors recording data every second or less) and variety (data from WWTPs is heterogeneous: different types of sensors, data from laboratory analysis, spectrophotometry, omics, images…). More recently other V's have been added like veracity (data and sources must be trusted) and value (to refer that big data only is only of value when presents high veracity with low vulnerability).

Variety and veracity represent a global and big challenge for making the most of the available data in WWTPs. The heterogeneity of data types, including traditional univariate time series (e.g., dissolved oxygen, temperature…) structured multivariate data (e.g., from spectrophotometry) as well as unstructured data (e.g., omics, images), makes the creation of a data fusion processing pipeline to extract and synthesize valuable information from all available data sources challenging and extremely difficult to apply in a large-scale fashion. This, in turn, affects how data quality assessment and quality control is ensured in the WWTP sector. The combination of data variety together with the exposition of the instrumentation deployed in the facilities to the hostile environment that WWTP represent (wastewater, particles, sludge, corrosive gases…) means that ensuring fit-for-purpose data quality remains a tough challenge today. The hostile environment in which these sensors are installed make the instrumentation prone to malfunction (bias, drift, precision degradation, missing data….) leading to poor data quality and reduced sensor accuracy and reliability [7].

Since external data is now less costly and ubiquitously available more data is still available at WWTPs. Consequently, information on important boundary conditions such as population estimates/ behaviour, weather ..., can be fed into supervisory control and data acquisition (SCADA) systems and wastewater services. Therefore, there is a realistic possibility to literally "drown" in data, especially if data management tools are not up to the task regarding quality checks and filtering out unreliable information. Many methods are available from data science [8], and the community seems to be familiar with them on a theoretical level. However, in practice, the developed capacities are not sufficient to harvest their full potential [4]. Transparent and standardized data treatment protocols may help to achieve this. Extracting information from the available data is not only important for the facilities themselves but also for other entities that can also be interested in WWTP data because they contain valuable information on societal behaviour (eg. spatial and temporal variations on drugs use as well as viruses tracking in wastewater-based epidemiology) or the expected quality of aquatic ecosystems.

Effective digitalization requires the cultivation of good meta-data management practices. Unfortunately, there are no wastewater-specific guidelines available to the production, selection, prioritization, and management of meta-data. To address this challenge, the IWA Task Group on Meta-Data Collection and Organisation (MetaCO TG to which the authors of this paper belong) which has been supported by the International Water Association since 2020 will soon finish the scientific and technical report containing such guidelines specifically for WWTPs. This paper highlights why meta-data should be considered when collecting data as part of good digitalisation practices.

2 META-DATA

Facilities usually feature many sensors that generate a large amount of data from the process (see Figure 1). It is important to highlight that to obtain accurate and useful data, every sensor:

- should be correctly installed (considering that it has also to be accessible for calibration, maintenance, and replacement when necessary),

- have a purpose (a reason to be there)
- and be adequately maintained by process operators according to manufacturer's instructions.

If only the numerical values of the variables are stored (which is the standard practice in most WWTPs), after a certain time, it will not be possible to know if the sensor was properly maintained, how many times it was calibrated, if it was replaced, the exact position where it was located or if it was moved temporarily to another position of the facility to gather data from another stream (e.g., influent, anoxic reactor, aerobic reactor, effluent of the treatment chain….) which could exhibit a completely different range of recorded values. All this information is of paramount importance to enable the current use of the data by an expert in data science as well as the future use of the data.

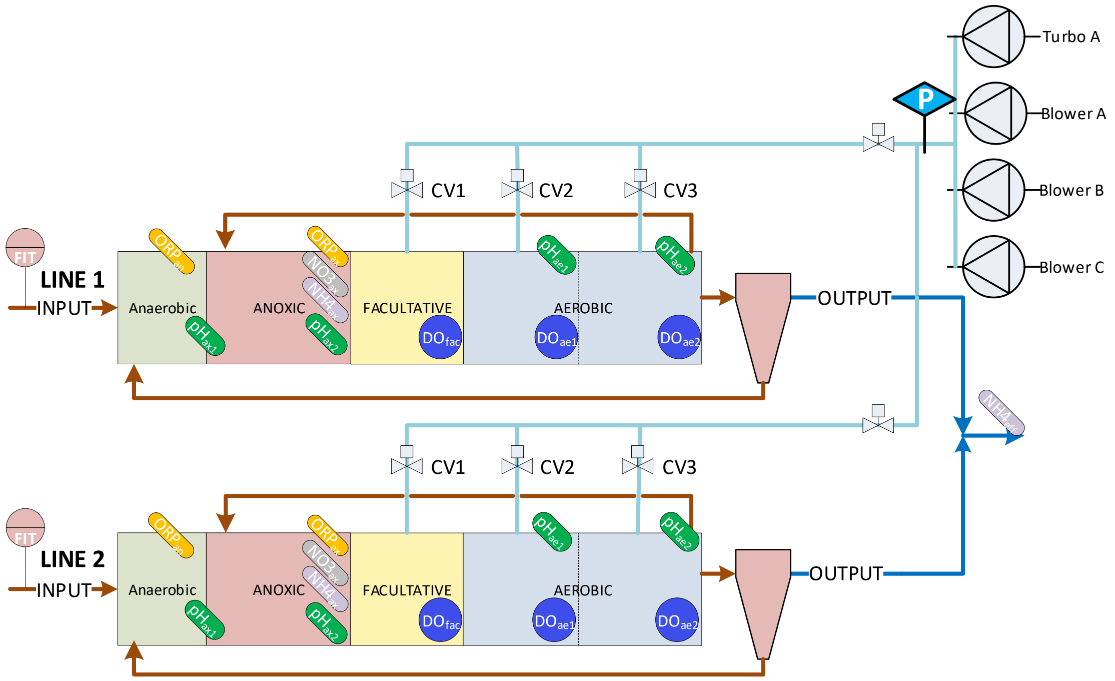


Figure 1. Layout of two parallel lines of a large WWTP featuring biological nutrient removal showing where each sensor is located within the biological reactor.

As can be seen in Figure 1, the same type of sensor has been installed in different locations along the water-line of the facility. For instance, in this case, six dissolved oxygen (DO) sensors have been located in different parts of the biological reactor, eight pH sensors, and so on. The sensors have been installed in different positions because variations in the recorded values of the variable are expected. The locations have been selected to gather information from the biological processes taking place as well as for control purposes. Note that if the location of each sensor would not be stored together with the recorded data, how would a data scientist be able to extract valuable information from the recorded data? In contrast, sometimes for variables that are critical for the process as well as for cheap and ubiquitous sensors such as pressure and temperature, the sensor is duplicated to obtain a reference measurement. Thus, in these cases two sensors measuring the same variable are located in the same reactor close to each other (and similar recorded values from both sensors are expected). The reason is to achieve hardware redundancy for data quality assessment. Thus, to extract valuable information from the data and to enable the use and reuse and future use of the gathered data for multiple purposes, additional information describing the context of the data should be also stored. This additional descriptive information is what is known as meta-data.

Meta-data is any descriptive information that is required and/or useful to interpret the data from a sensor or from the laboratory. There are several pieces of basic information that should be collected as much as possible, like the following aspects (full-details and extended information will be available soon in the scientific and technical report specifically targeted for WWTPs):

- Unit of measurement
- Measurement range
- Sensor location
- Sensor ID
- Temporal resolution
- If any change of units is applied
- If any type of missing data imputation is applied
- Operational state (operational, calibration, validation, maintenance)
- Purpose of the sensor
- Operational condition of the plant (normal conditions, toxic spill, dry-weather conditions, type of mechanical failure,…)

The first and foremost evident piece of information required to interpret the numerical value of a variable is the measuring unit which allows the correct interpretation of the values recorded.

The measuring range describes the interval along which the value of the measured variable must be situated so that the technical specifications of the sensor are met and thus they are the possible values that can be generated by a properly functioning sensor. Knowing the lowest and highest value that can be produced by a sensor that is functioning properly, can be useful to detect a malfunctioning sensor.

Maintenance actions often require a change in the exact location of the sensor, and occasionally a sensor can be temporary moved to another position within the facility for a given period of time to measure the variable of interest in another stream. These changes in position can lead to a complete different range of values (e.g., an expensive ammonia sensor moved from the aerobic reactor to the anoxic reactor, or to the influent or effluent of the facility), making the recorded data not always representative of the original measurement location. Therefore, logging the exact location of sensors is important being ideal that they would incorporate a position measurement system that enable automatic logging. The location is highly important because it can provide the context for the values interpretation, in addition to being also useful for the detection, isolation and diagnosis of anomalous values.

Knowing the purpose of a sensor can assist on to taking decisions on the prioritization of maintenance actions and it can also help a data scientist in deciding whether some signals could be useful to develop a data-driven model.

The operational condition of the plant (normal conditions, toxic spill, dry-weather conditions, type of mechanical failure, jammed valve…) can be considered contextual information that can be added through annotation. Thus, it is normally a manual process executed by a human expert. The information provided is of great help for the later use of the data. Normally this information is provided in a freeform text since the type and number of events that can occur is very large. However, free text allows different terms used for the same event, for instance if a different person is annotating. This can be avoided by providing a set of labels with an accepted description of most of the possible events and let a final option for including a free text in case the expert considers that none of the labels describe what it is wanted to reflect on the observed data.

It should be noted that the more unstructured would be the meta-data it would also be more difficult to managed and analysed in a conventional way.

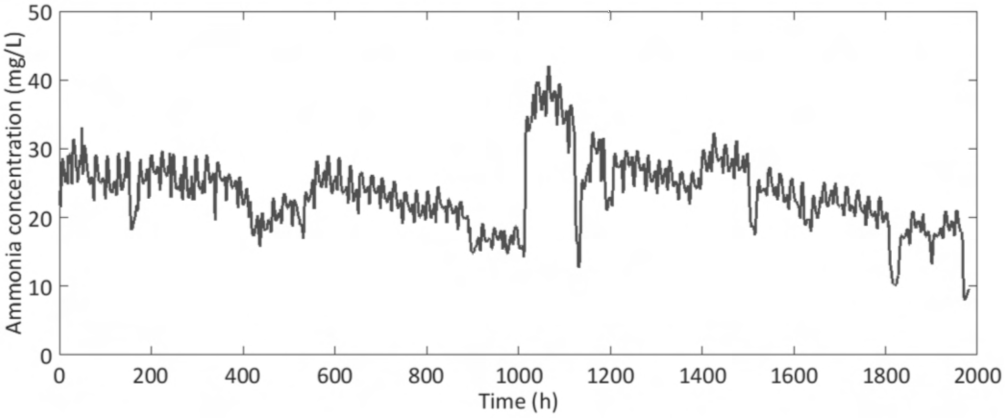
3 ILLUSTRATIVE EXAMPLE

Figure 2a shows the temporal evolution of ammonium concentration at the influent of the biological process of a large WWTP along nearly three months of operation. In this figure it can be seen noticeable variations in the ammonia concentration. Apart from a daily pattern, a temporal trend can be seen together with several relative abrupt changes in the recorded values of ammonia concentration.

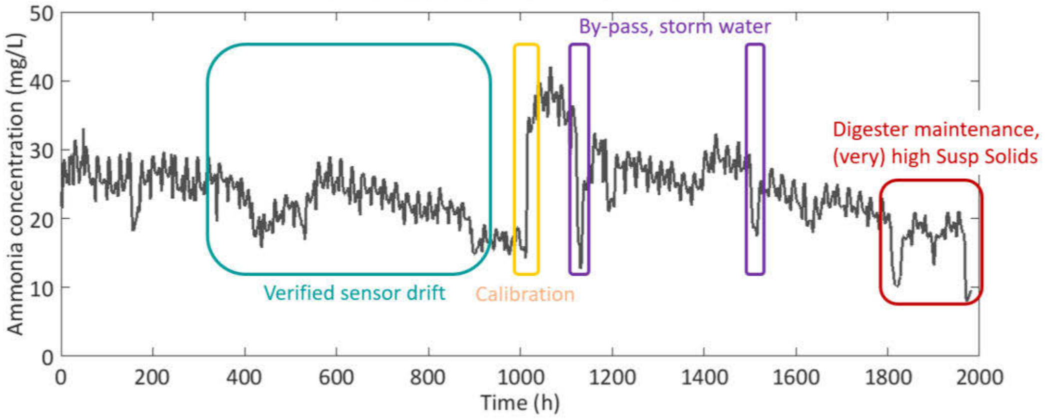
In Figure 2b, the same variable and period is depicted, together with information of the observed changes by the process expert. As can be seen, Figure 2b is much more informative and useful, from the point of view of a later and future use of the data, for instance by a Data Scientist (with little knowledge on wastewater treatment processes). It should be highlighted how these pieces of information provided as meta-data can assist in an automation of the data triage to a high degree, and how the need for subjective assessment of the recorded time series has been significantly reduced, fact that will boost the trust in the data-driven predictions and decision-making.

The supervised approach in machine learning applications need a labelled dataset with the real anomalies and the normal-operating conditions data (i.e., clean data) for training and testing the data-driven models used for classification purposes as well as those used prediction applications. Thus, meta-data will play a central role to implement supervised machine learning models and intelligent tools in practice.

The most recent developments in data mining, machine learning, and artificial intelligence promise to extract the maximum of information from historical data. However, not understanding the process and the conditions under which data is collected makes even the best of algorithms fail. The use of WWTP data after they has been collected is typically limited due to a lack of documentation on the data generating process and a lack of data quality assessment. The scalability of the AI methods is expected to enable an effective use of data from multiple, uncoordinated sources, including data sources outside of the wastewater resource recovery plant’s fence (e.g., weather forecasting data) to provide answers to questions through understanding new relationships that are ever-larger in scope. Unfortunately, this will remain a daydream unless the raw data becomes traceable.



(a)



(b)

Figure 2. Time evolution of influent ammonium concentration along almost three months of operation: (a) only recorded values of the sensor including the units of measurement (b) adding process expert annotation.

4 CONCLUSIONS

This paper has illustrated the importance of meta-data to make it possible to transform raw data into intelligent actions for the operation of WWTPs and WRRFs, and it has been highlighted why meta-data should be considered when collecting data as part of good digitalisation practices. The main take-home messages are the following:

- Nowadays, a large amount of data is collected in WWTPs and WRRFs, but the almost complete absence of meta-data is compromising their potential use as well as their future use of the data, allowing their loss as data graveyards.
- There are no wastewater-specific guidelines available to the production, selection, prioritization, and management of meta-data.
- Meta-data greatly facilitates the interpretation of the collected data, helping to avoid the saying "data rich but information poor" that characterises many wastewater treatment data sets around the world.
- Meta-data plays a central role to implement useful supervised machine learning models and intelligent tools in practice.
- Systematic meta-data management will make it possible the future leverage of sensor and laboratory data, increasing its value.
- To really improve the decision-making process in the operation and real-time control of WWTPs and WRRFs, based the on-line collected data, there is need of annotated, representative, reliable and high-quality data.
- The IWA Task Group on Meta-Data Collection and Organisation (MetaCO TG to which the authors of this paper belong) will soon finish the scientific and technical report containing such guidelines specifically for WWTPs.

5 COPYRIGHT NOTICE

This research is sponsored by the US Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Advanced Manufacturing Office, under contract DE-AC05-00OR22725 with UT-Battelle LLC. This manuscript has been authored by UTBattelle LLC under contract DE-AC05-00OR22725 with DOE. The US government retains—and the publisher, by accepting the article

for publication, acknowledges that the US government retains—a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript or allow others to do so for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

6 REFERENCES

[1] Metcalf & Eddy, G. Tchobanoglous, H.D. Stensel, R. Ryujiro Tsuchihashi, F. Burton (2013). Wastewater Engineering: Treatment and Resource Recovery. McGraw-Hill Education. ISBN: 978-0073401188.

[2] E. Neczaj, A. Grosser (2018). Circular Economy in Wastewater Treatment Plant–Challenges and Barriers. Proceedings.; 2(11):614. https://doi.org/10.3390/proceedings2110614.

[3] Ll. Corominas, M. Garrido-Baserba, K. Villez, G. Olsson, U. Cortés, M. Poch (2018). Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques, Environmental Modelling & Software, 106, 89-103, doi: 10.1016/j.envsoft.2017.11.023

[4] F. Blumensaat, J.P. Leitão, C. Ort, J. Rieckermann, A. Scheidegger, P.A. Vanrolleghem, K. Villez (2019). How Urban Storm- and Wastewater Management Prepares for Emerging Opportunities and Threats: Digital Transformation, Ubiquitous Sensing, New Data Sources, and Beyond - A Horizon Scan. Environmental Science & Technology 2019 53 (15), 8488-8498 doi:10.1021/acs.est.8b06481.

[5] G. Olsson (2012). ICA and me — A subjective review. Water Research, 46 (6), pp. 1585-1624, 10.1016/j.watres.2011.12.054

[6] Doug Laney, 3d data management: Controlling data volume, velocity and variety, META Group Research Note 6 (2001).

[7] O. Samuelsson, A. Björk, J. Zambrano, B. Carlsson (2018). Fault signatures and bias progression in dissolved oxygen sensors. Water Science & Technology 78 (5), 1034-1044. doi: 10.2166/wst.2018.3501979.

[8] K.B. Newhart, R. W. Holloway, A.S. Hering, T.Y. Cath (2019). Data-driven performance analyses of wastewater treatment plants: A review. Water Research, 157, 498-513. doi: 10.1016/j.watres.2019.03.030.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14267

OPTIMAL DESIGN-FOR-CONTROL OF WATER DISTRIBUTION NETWORKS VIA CONVEX RELAXATION

Filippo Pecci[1], Ivan Stoianov[2] and Avi Ostfeld[3]

[1,2]Department of Civil and Environmental Engineering, Imperial College London, London (UK)
[3]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa (Israel)

[1] *f.pecci14@imperial.ac.uk*, [2] *ivan.stoianov@imperial.ac.uk*, [3] *ostfeld@cv.technion.ac.il*

Abstract

This paper considers joint design-for-control problems in water distribution networks (WDNs), where locations and operational settings of control actuators are simultaneously optimized. We study two classes of optimal design-for-control problems, with the objectives of controlling pressure and managing drinking-water quality. First, we formulate the problem of optimal placement and operation of valves in water networks with the objective of minimizing average zone pressure, while satisfying minimum service requirements. The resulting mixed-integer non-linear optimization problem includes binary variables representing the unknown valve locations, and continuous variables modelling the valves' operational settings. In addition, water utilities aim to maintain optimal target chlorine concentrations, sufficient to prevent microbial contamination, without affecting water taste and odour, or causing growth of disinfectant by-products. We consider the problem of optimal placement and operation of chlorine booster stations, which reapply disinfectant at selected locations within WDNs. The objective is to minimize deviations from target chlorine concentrations, while satisfying lower and upper bounds on the levels of chlorine residuals. The problem formulation includes discretized linear PDEs modelling advective transport of chlorine concentrations along network pipes. Moreover, binary variables model the placement of chlorine boosters, while continuous variables include the boosters' operational settings.

Computing an exact solution for the considered mixed-integer optimization problems can be computationally impractical when large water network models are considered. We investigate scalable heuristic methods to enable the solution of optimal design-for-control problems in large WDNs. As a first step, we solve a convex relaxation of the considered mixed-integer optimization problem. Then, starting from the relaxed solution, we implement randomization and local search to generate candidate design configurations. Each configuration is evaluated by implementing continuous optimization methods to optimize the actuators' control settings and compute feasible solutions for the mixed-integer optimization problem. Moreover, the solution of the convex relaxation yields a lower bound to the optimal value of the original problem, resulting in worst-case estimates on the level of sub-optimality of the computed solutions

We evaluate the considered heuristics to solve problems of optimal placement and operation of valves and chlorine boosters in water networks. As case study, we utilize an operational water network from the UK, with varying sizes and levels of connectivity and complexity. The convex heuristics are shown to generate good-quality feasible solutions in all problem instances with bounds on the optimality gap comparable to the level of uncertainty inherent in hydraulic and water quality models. Future work should investigate the formulation and solution of multiobjective optimization problems for the optimization of pressure and water quality, to evaluate the trade-offs between these two objectives. Moreover, the formulation and solution of robust optimization problems for the design of water networks under uncertainty is the subject of future work.

Keywords
Pressure management, water quality, optimization.

1 BACKGROUND AND MOTIVATION

Water distribution networks (WDNs) are critical infrastructure, facing unprecedented challenges due to increasing water demand, climate change, and more stringent economic and environmental constraints. The rising demand for intelligent water systems requires analytical and technological innovations to support design and automatic control of such complex networks. In particular, the COVID-19 pandemic has highlighted the need for near real-time monitoring and control of water quality to protect public health and provide assurance to customers [1]. The reduction in human resources caused by the pandemic is increasing the demand for automation and reliability of WDNs. The efficient management of WDNs requires the satisfaction of multiple objectives, including optimal control of pressure and water quality. Pressure control allows network operators to minimize leakage [2] and reduce probability of pipe failures [3]. Furthermore, disinfectant residuals in WDNs are critical control variables to preserve water quality and eliminate the risks of pathogen contamination, with chlorine being a commonly used water disinfectant – see for example [4], [5]. Pressure control schemes are implemented by installing and operating pressure reducing valves (PRVs), which control pressure at their downstream node. Optimal pressure control problems in WDNs aim at minimizing average zone pressure (AZP) in WDNs, while satisfying regulatory requirements on minimum pressure at demand nodes [6]. At the same time, water utilities aim to maintain optimal target chlorine concentrations, sufficient to prevent microbial contamination, without affecting water taste and odour, or causing growth of disinfectant by-products [7]–[9]. Moreover, optimized chlorine dosage should avoid spatial and temporal variations, which are perceived as water quality problems by customers. This is achieved by the optimal placement and operation of chlorine booster stations, which re-apply disinfectant at selected locations within WDNs to maintain target chlorine concentrations [10]–[12].

We consider two classes of optimal design-for-control problems, with the objectives of controlling pressure and managing drinking-water quality. First, we formulate the problem of optimal placement and operation of valves in water networks with the objective of minimizing average zone pressure, while satisfying minimum service requirements. Then, we consider the problem of optimal placement and operation of chlorine booster stations, which reapply disinfectant at selected locations within WDNs. In this case, the objective is to minimize deviations from target chlorine concentrations, while satisfying bounds on the levels of chlorine residuals. These problems result in mixed integer optimization problems, combining continuous variables modelling the operational settings of the control actuators, with discrete variables representing their locations.

As shown by the numerical experience reported in [13], [14], off-the-shelf global optimization solvers fail to solve these mixed integer optimization problems, even for small-scale water network models. In [14], the authors proposed new heuristics based on convex optimization to generate good quality feasible solutions when large WDN models are considered. The developed convex heuristic relies on randomization to generate candidate locations for valves and chlorine boosters. Moreover, the heuristic generates lower bounds to the optimal value of the original problem, resulting in estimates on the optimality gap of the computed solutions. This allows a worst-case estimate on the level of sub-optimality of the computed solutions. In this manuscript, we investigate the implementation of the method proposed in [14], using as case study an operational water network from the UK. We evaluate the optimal operation of valves and chlorine boosters over extended time simulations, and compare network pressures, flows, and chlorine concentrations computed by the optimization process, with those obtained by the widely used hydraulic and water quality simulation tool EPANET [15].

2 PROBLEM FORMULATION

In this study, we consider two design-for-control problems in water networks: optimal valve placement and operation, and optimal booster placement and operation. Firstly, we formulate the problem of computing optimal locations of pressure control valves, and their operational settings, while satisfying hydraulic constraints, i.e. mass and energy conservation laws. The objective is to minimize Average Zone Pressure (AZP), while enforcing lower and upper bounds on hydraulic variables. Binary variables are used to model the placement of valves on network links. In our formulation, we consider unidirectional pressure control valves, assuming that the direction of operation of the valves is kept constant for the whole time period. These binary variables are subject to physical and economical constraints, limiting the number of valves considered for installation, and ensuring that only one valve is installed on each link. A detailed description of the problem formulation can be found in [16] .

In comparison, the problem for optimal placement and operation of chlorine boosters is formulated as a mixed integer quadratic program, where decision variables include locations and operational settings of chlorine boosters, and chlorine concentrations at water sources – see as examples [10], [11], [17], [18]. As done in [11], we aim to minimize the deviation from target chlorine concentrations at demand nodes, weighted by nodal demands - we refer to this as Average Target Deviation (ATD). We also enforce lower and upper bounds on water quality variables. In particular, these include maximum allowed chlorine concentrations at network nodes. We include binary variables to model the placement of boosters, with the total number of boosters considered for installation being modelled as optimization constraint. The operational settings of chlorine boosters are defined as the chlorine concentrations at the node where the booster is installed. The transport of chlorine residuals through network pipes is governed by a linear advection reaction partial differential equation (PDE) [19], which we approximate within optimization constraints using an upwind implicit discretization scheme [20]. A complete description of the problem formulation can be found in [21].

3 SOLUTION METHOD

Both design-for-control problems for optimal placement and operation of valves and boosters can be represented as:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax + g(x) = 0 \\ & Bx + Dz \leq b \\ & z \in \{0,1\}^m. \end{array} \tag{1}$$

The objective $f(x)$ is a (possibly non-linear) convex function representing either AZP or ATD. In the case of optimal valve placement, $g(x)$ models the frictional energy losses across network pipes as non-linear function of the flow rates. In contrast, in the case of optimal booster placement, the discretized PDEs result in $g(x) = Tx + c$, for opportunely define matrix T and vector c. Problem (1) belongs to the class of mixed integer non-linear programming (MINLP). As shown in [13], off-the-shelf MINLP solvers can fail to compute feasible solutions for (1), even when modest size water network models are considered. In [14], the authors have proposed convex heuristics to compute feasible solutions for mixed-integer programs like (1), together with bounds on the level of sub-optimality. For a detailed description of the methods, the reader is referred to [14]. Here, we provide a summary of the proposed convex heuristics.

Step 1. Solve a convex relaxation of (1), where non-convex constraints (if any) are substituted by their polyhedral relaxations:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax + Rx \leq r \\ & Bx + Dz \leq b \\ & z \in [0,1]^m, \end{array} \tag{2}$$

for opportunely defined matrix R and vector r – see [16]. The convex problem in (2) can be efficiently solved by convex optimization solvers [22], including GUROBI [23] when $f(x)$ is either linear or quadratic. By solving Problem (2), we obtain a lower bound to the optimal value of Problem (1), denoted by f^* , as well as a vector of fractional values $z^* \in \mathbb{R}^m$.

Step 2. Implement the randomized rounding heuristic described in [14], where the fractional values in z^* are used to defined a discrete probability distribution over the set $\{1,\ldots,m\}$, to generate vectors of binary variables $\hat{z} \in \{0,1\}^m$. For each $\hat{z}$, solve the following optimization problem using a non-linear programming solver like IPOPT [24] to compute a locally optimal solution:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax + g(x) = 0 \\ & Bx \leq b - D\hat{z}\,. \end{array} \tag{3}$$

Solving Problem (3), we obtain a feasible vector of continuous variables $\hat{x}$, and the corresponding objective function value $f(\hat{x})$. Let $(z^{\text{best}}, x^{\text{best}})$ be the feasible solution corresponding to the lowest objective function value computed by the randomized rounding heuristic. The value $f^{\text{best}} = f(x^{\text{best}})$ is an upper bound to the optimal value of Problem (1). A worst-case estimate on the level of sub-optimality of the solution $(z^{\text{best}}, x^{\text{best}})$ is given by:

$$\text{Gap} = \frac{f^{\text{best}} - f^*}{f^{\text{best}}}. \tag{4}$$

4 CASE STUDY

We evaluate the performance of the considered methods using BWFLnet, the hydraulic model of a water distribution network from the UK – the EPANET model of BWFLnet is given in [25]. BWFLnet has two inlets and it includes 2281 links and 2221 nodes. The network layout is presented in Figure 1. We consider 24 different time steps, one for each hour of the day.

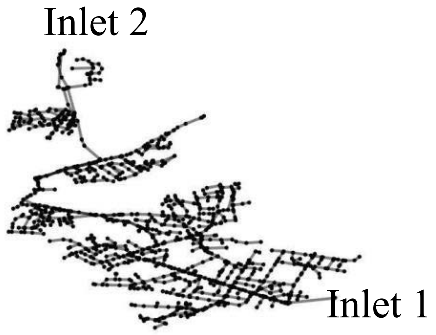


Figure 1. Layout of BWFLnet

4.1 Optimal valve placement and operation in BWFLnet

We formulate the problem of computing optimal locations and operational settings for 3 pressure control valves to be placed in BWFLnet. The convex heuristic method described in Section 3 is implemented to compute a feasible solution with bounds on its level of sub-optimality. The optimization process terminates after 174 seconds, returning a feasible solution corresponding to an AZP value of 33.90 m , and an estimated optimality Gap of 12%. The locations of the three pressure control valves are shown in Figure 2.

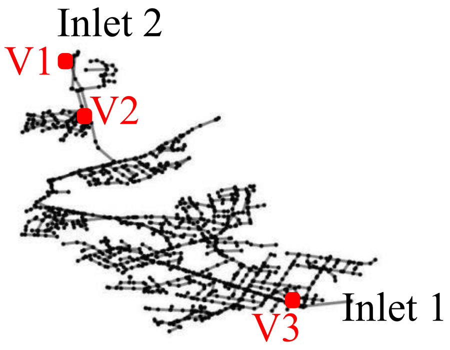


Figure 2. Optimised locations of pressure control valves computed solving Problem (1).

In order to validate the results, we implement the optimised control settings for the three pressure control valves in EPANET. The simulated AZP is very close to the value obtained solving Problem (1), and it is equal to 33.89 m. The largest difference between optimised and simulated pressures is equal to 0.02 m. The accuracy of the predictions made by the optimization model is also illustrated by Figure 3, which reports the temporal average of the errors in nodal pressures:

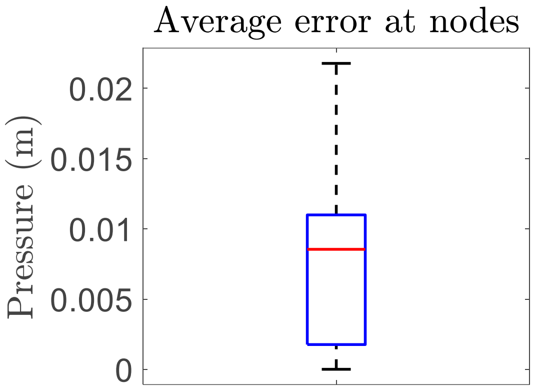


Figure 3. Temporal average of the differences in nodal pressures between EPANET simulations and optimization results.

4.2 Optimal booster placement and operation in BWFLnet

We formulate the problem of optimal placement and operation of 3 chlorine boosters in BWFLnet. We assume a target chlorine concentration of 1 mg/l. The hydraulic head and flows have been fixed based on the solution computed in 4.1, where 3 pressure control valves are optimally operated to minimize AZP. In this case, Problem (1) results in a mixed integer quadratic program (MIQP). We have implemented the state-of-the-art MIP solvers GUROBI and CPLEX but they both

BY NC SA

failed to compute a feasible solution after two hours of computation. In comparison, our convex heuristic has resulted in a feasible solution with an ATD value of 27.35, and an optimality Gap of 24%. The computational time required by the convex heuristic to converge was 212 seconds. The optimised locations of chlorine boosters are shown in Figure 4.

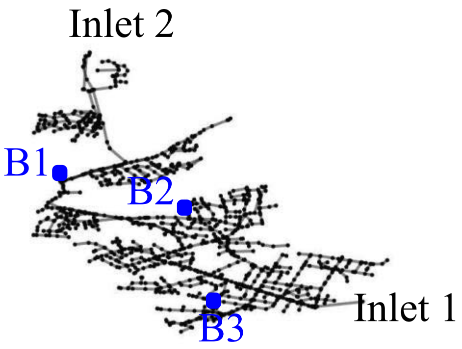


Figure 4. Optimised locations of chlorine boosters computed solving Problem (1).

As done in Section 4.1, we validate our results comparing the optimised chlorine concentrations with those obtained by simulating the boosters' operation in EPANET. As shown in Figure 5, the temporal average differences between simulated and optimised concentrations are smaller than 0.15 mg/l in 99% of nodes. This is comparable to the level of uncertainty inherent in modelling of chlorine residuals in operational water networks.

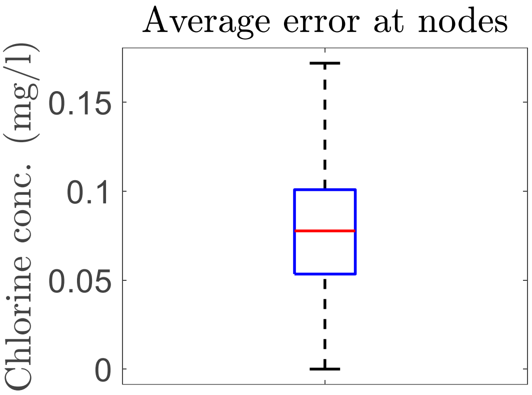


Figure 5. Temporal average of the differences in chlorine concentrations between EPANET simulations and optimization results.

Finally, Figure 6 reports the distribution of simulated chlorine concentrations in BWFLnet at 20:00, where three boosters are optimally operated to minimize ATD. Observe that the vast majority of nodes experience chlorine concentrations close to the target value of 1 mg/l. The few nodes with lower chlorine concentrations have either zero or low demand. Hence, they do not have a significant impact on ATD, whose weights are nodal demands.

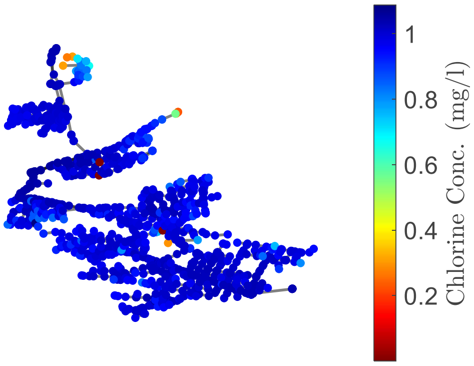


Figure 6. Chlorine concentrations at 20:00 in BWFLnet, with optimized boosters.

5 CONCLUSIONS

The convex heuristics are shown to generate good-quality feasible solutions in for the considered problem instances with bounds on the optimality gap comparable to the level of uncertainty inherent in hydraulic and water quality models. Future work should investigate the formulation and solution of multiobjective optimization problems for the optimization of pressure and water quality, to evaluate the trade-offs between these two objectives. Moreover, the formulation and solution of robust optimization problems for the design of water networks under uncertainty is the subject of future work.

6 REFERENCES

[1] R. S. Quilliam, M. Weidmann, V. Moresco, H. Purshouse, Z. O'Hara, and D. M. Oliver, "COVID-19: The environmental implications of shedding SARS-CoV-2 in human faeces," Environ. Int., vol. 140, no. April, p. 105790, 2020, doi: 10.1016/j.envint.2020.105790.

[2] B. Ulanicki, P. Bounds, J. P. Rance, and L. Reynolds, "Open and closed loop pressure control for leakage reduction," Urban Water, vol. 2, no. 2, pp. 105–114, 2000, doi: 10.1016/S1462-0758(00)00048-0.

[3] A. Lambert and J. Thornton, "The relationships between pressure and bursts – a state-of-the-art update," Water21 - Magazine of the International Water Association, no. 21, pp. 37–38, 2011.

[4] M. M. Polycarpou, J. G. Uber, Z. Wang, F. Shang, and M. Brdys, "Feedback control of water quality," IEEE Control Syst. Mag., vol. 22, no. 3, pp. 68–87, 2002, doi: 10.1109/MCS.2002.1004013.

[5] T. Sakomoto, M. Lutaaya, and E. Abraham, "Managing water quality in intermittent supply systems: The Case of Mukono Town, Uganda," Water (Switzerland), vol. 12, no. 3, 2020, doi: 10.3390/w12030806.

[6] R. Wright, E. Abraham, P. Parpas, and I. Stoianov, "Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming," Water Resour. Res., vol. 51, no. 12, pp. 9925–9941, 2015, doi: 10.1002/2015WR017466.

[7] B. T. Carrico and P. C. Singer, "Impact of booster chlorination on THM production: A simulated analysis," J. Environ. Eng., vol. 135, no. 10, pp. 928–935, 2009, doi: 10.1061/40792(173)126.

[8] T. Walski, "Raising the Bar on Disinfectant Residuals," World Water Mag. Water Environ. Fed., vol. 42, no. 4, pp. 34–35, 2019.

[9] S. Avvedimento, S. Todeschini, C. Giudicianni, A. Di Nardo, T. Walski, and E. Creaco, "Modulating Nodal Outflows to Guarantee Sufficient Disinfectant Residuals in Water Distribution Networks," J. Water Resour. Plan. Manag., vol. 146, no. 8, p. 04020066, 2020, doi: 10.1061/(asce)wr.1943-5452.0001254.

[10] M. Propato and G. Uber, James, "Linear Least-Squares Formulation for Operation of Booster Disinfection Systems," J. Water Resour. Plan. Manag., vol. 130, no. 1, pp. 53–62, 2004, doi: 10.1061/(ASCE)0733-9496(2004)130:1(53).

[11] M. Propato and J. G. Uber, "Booster System Design Using Mixed-Integer Quadratic Programming," J. Water Resour. Plan. Manag., vol. 130, no. 4, pp. 348–352, 2004, doi: 10.1061/(ASCE)0733-9496(2004)130:4(348).

[12] D. Kang and K. Lansey, "Real-Time optimal valve operation and booster disinfection for water quality in water distribution systems," J. Water Resour. Plan. Manag., vol. 136, no. 4, pp. 463–473, 2010, doi: 10.1061/(ASCE)WR.1943-5452.0000056.

[13] F. Pecci, I. Stoianov, and A. Ostfeld, "Relax-Tighten-Round Algorithm for Optimal Placement and Control of Valves and Chlorine Boosters in Water Networks," Eur. J. Oper. Res., vol. 295, no. 2, pp. 690–698, 2021, doi: 10.1016/j.ejor.2021.03.004.

[14] F. Pecci, I. Stoianov, and A. Ostfeld, "Convex Heuristics for Optimal Placement and Operation of Valves and Chlorine Boosters in Water Networks," J. Water Resour. Plan. Manag., vol. 148, no. 2, pp. 1–14, 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001509.

[15] L. A. Rossman, "EPANET 2," U.S. Environmental Protection Agency, Cincinnati, 2000. doi: 10.1177/0306312708089715.

[16] F. Pecci, E. Abraham, and I. Stoianov, "Global optimality bounds for the placement of control valves in water supply networks," Optim. Eng., vol. 20, no. 2, pp. 457–495, 2019, doi: 10.1007/s11081-018-9412-7.

[17] M. E. Tryby, D. L. Boccelli, J. G. Uber, and L. A. Rossman, "Facility location model for booster disinfection of water supply networks," J. Water Resour. Plan. Manag., vol. 128, no. 5, pp. 322–333, 2002, doi: 10.1061/(ASCE)0733-9496(2002)128:5(322).

[18] K. Lansey, F. Pasha, S. Pool, W. Elshorbagy, and J. Uber, "Locating satellite booster disinfectant stations," J. Water Resour. Plan. Manag., vol. 133, no. 4, pp. 372–376, 2007, doi: 10.1061/(ASCE)0733-9496(2007)133:4(372).

[19] L. A. Rossman and P. F. Boulos, "Numerical methods for modeling water quality in distribution systems: A comparison," J. Water Resour. Plan. Manag., vol. 122, no. 2, pp. 137–146, 1996, doi: 10.1061/(ASCE)0733-9496(1996)122:2(137).

[20] M. R. Islam, M. H. Chaudhry, and R. M. Clark, "Inverse modeling of chlorine concentration in pipe networks under dynamic condition," J. Environ. Eng., vol. 125, no. 3, pp. 296–298, 1999, doi: 10.1061/(ASCE)0733-9372(1999)125:3(296).

[21] F. Pecci, I. Stoianov, and A. Ostfeld, "Optimal Design-for-Control of Chlorine Booster Systems in Water Networks via Convex Optimization," in Proceedings of the European Control Conference (ECC) 2022, 2022.

[22] S. Boyd and L. Vandenberghe, Convex optimization. Cambridge University Press, 2004.

[23] Gurobi Optimization, "Gurobi Optimizer 9.1 Reference Manual." 2021, [Online]. Available: https://www.gurobi.com/documentation/9.1/refman/index.html.

[24] A. Waechter and L. T. Biegler, "On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming," Math. Program., vol. 106, no. 1, pp. 25–57, 2006, doi: 10.1007/s10107-004-0559-y.

[25] F. Pecci, I. Stoianov, and A. Ostfeld, "Case Studies for Optimal Placement and Control of Valves and Chlorine Boosters in Water Networks." Mendeley Data, 2021, doi: 10.17632/ws9pwxkbb2.4.

[26] B. J. Eck and M. Mevissen, "Non-Linear Optimization with Quadratic Pipe Friction," 2012.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14437

AN OPTIMIZATION FRAMWORK FOR LARGE WATER DISTRIBUTION SYSTEMS BASED ON COMPLEX NETWORK ANALYSIS

Robert Sitzenfrei [1]

[1]Environmental Engineering, Faculty of Civil Engineering, Institute of Infrastructure, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria

[1] *robert.sitzenfrei@uibk.ac.at*

Abstract

The major task of water distribution networks (WDNs) is to reliably supply water in sufficient quantity and quality. Due to the complexity in design and operation of WDNs, and to ensure a reliable level of service with minimum costs, multi-objective design approaches are used which are usually rely on evolutionary algorithms. However, for large WDNs the decision variable space increases exponentially. When considering multiple objectives (e.g., resilience, costs, water quality), for complex, large (real) WDNs with several thousand decision variables, evolutionary algorithms are practically infeasible to apply. With complex network analysis mathematical graphs of WDNs can be analysed very computationally efficient and therefore such an approach is especially suitable for analysing large spatial transport networks. Recently, based on complex network, a highly efficient approach for Pareto-optimal design of WDNs was developed. Based on topological features and a customized graph measure for the demand distribution (demand edge betweenness centrality), a graph-based multi-objective design approach was developed, which outperformed the results of an evolutionary algorithm regarding the quality of solutions and computation time (factor 10^5 faster). Further, also based on complex network analysis, a highly efficient surrogate method for assessing water quality in large WDNs was developed ($2.4 \cdot 10^5$ times faster than extended period simulation Epanet2). In this paper, these two approaches based on complex network analysis: (1) two objective optimization model and (2) the graph-based water quality model, are combined in a novel graph optimization framework which is especially suitable for complex, large (real) WDNs. The applicability of this very computationally efficient, novel approach is shown on a real case studies with 4,000 decision variables for which the results are be obtained within 18.5 seconds of computation time, while with a state-of-the-art evolutionary algorithm it took more than 8 weeks.

Keywords
Multi-objective optimization, Demand edge betweenness centrality, Hydraulically informed Graph analysis.

1 INTRODUCTION

The major task of water distribution networks (WDNs) is to reliably supply water in sufficient quantity and quality [1]. Such systems are usually grown over decades and have an organic and complex structure. The complex hydraulic interactions of the different components, is no trivial problem to solve. Due to the complexity in design and operation of WDNs, and to ensure a reliable and resilient level of service with, at the same time, minimum of costs, multi-objective design approaches are used which are usually solved with evolutionary algorithms [2]. These approaches are well investigated and described in literature [3, 4]. However, research on real all-pipe models is rare in this regard due to the large decision space as for large WDNs the decision space increases exponentially. When considering multiple objectives (e.g., resilience, costs, water quality), for complex, large (real) WDNs with several thousand decision variables, evolutionary algorithms are practically infeasible to apply due to the limitations in computation time. However, due to data availability (geographic information system, digital twins), the trend in modelling is going

CC BY NC SA

towards all-pipe models. Therefore, there is a need for fast approximations for solving such multi-objective design problems.

The topology of WDNs can be modelled as mathematical graphs. With complex network analysis such mathematical graphs can be analysed very computationally efficient and therefore such approaches are especially suitable for analysing large spatial transport networks [5]. Recently, based on complex network analysis, a highly efficient approach for Pareto-optimal design of WDNs was developed [6]. Based on topological features and a customized graph measure for the demand distribution (demand edge betweenness centrality), a graph-based multi-objective design approach was developed, which outperformed the results of an evolutionary algorithm in some areas of the Pareto-front regarding the quality of solutions and especially computation time (factor 10^5 faster). That approach is particularly suitable to solve problems with an extremely large decision space (>100,000) in acceptable time [6]. However, when comparing the design solutions from the graph-based optimization with those from evolutionary algorithm, for design solutions with a high level of redundancy, the graph-based method is outperformed. However, in that region of the Pareto-front (high resilience and high costs), water quality issues arise due to the low flow velocities. While water quality analysis is also computationally intensive for large scale WDS, also based on complex network analysis, a highly efficient surrogate method for assessing water quality in large WDNs was developed ($2.4 \cdot 10^5$ times faster than extended period simulation Epanet2) [7]. Therein, the edges in the graph of the WDS are weighted based on the residence time in the pipes. Based on shortest path analysis, pattern correction and topological correction functions, estimates for water age values were obtained.

In this paper, the two approaches based on complex network analysis: (1) two objective optimization model [6] and (2) the graph-based water quality model [7], are combined in a novel framework for optimization of WDNs which is especially suitable for complex, large (real) WDNs. Pareto optimal design solutions (minimum costs versus maximum resilience) are evaluated regarding exceedance of a threshold for water age. Therewith, technically unsuitable solutions are excluded from the Pareto-front. The applicability of this very computationally efficient, novel approach is shown on a real case studies with 4,000 decision variables for which the results can be obtained within a few seconds of computation time.

2 MATERIALS AND METHODS

In this work, a large real case study is optimized with a graph-based approach (2.1) with two contradicting objectives (minimal costs versus maximum resilience) and subsequent, from that Pareto-front of optimal solutions, design solutions exceeding a water quality threshold (maximum water age) are identified with a graph-based water quality approach (2.2). Finally, the obtained solutions are compared with design solution based on an evolutionary algorithm [8] from literature and extended period simulations with Epanet2 for water quality assessment are performed for the final design solutions.

2.1 Graph-based multi-objective design with demand edge betweenness centrality

A WDN can be modelled as mathematical graph, which can be analysed computationally efficient with complex network analysis (CNA). A graph consists of a set of vertices (e.g., nodes) which can be interlinked with links (e.g., pipes). The adjacency matrix, which is a symmetric matrix of the size of the number of nodes, describes if there is a link between two vertices. If there is a connection between vertices i and j, the matrix element $a_{ij}=1$, otherwise $a_{ij}=0$. Each link/edge k in the graph can have a weight w_k. Often unweighted graphs ($w_k=1$), or the Euclidean distance (i.e., pipe length L_k) $w_k=L_k$ is used. But also, other weights can be used e.g., mimicking hydraulic or water quality characteristics such as friction losses or residence time in an edge. The shortest path $\sigma_{i,j}$, is the path between two vertices i and j, where the path length (i.e., the sum of edge weights) is minimal.

The edge betweenness centrality (EBC) counts how often an edge is part of $\sigma_{i,j}$ when connecting all possible node pairs. For a WDN, each node has to be connected to at least one source node, therefore an EBC value, counting shortest paths from all demand nodes to the source node can be an indicator for required transport capacity. Instead of counting the number of shortest paths, the nodal demand can be added to the EBC values of an edge, and it gives an estimate of optimal/design flows. This customized EBC measure is denoted 'demand edge betweenness centrality' EBC^Q [6]. When EBC^Q is used for design of a WDN, only the pipe length L_k is available as edge weight, as no other hydraulic characteristics are known before the design process. In a WDN, where there are multiple redundant flow paths, a disadvantage of EBC^Q with edge weights L_k is that flow is concentrated in a few edges. When we consider all edges which are part of a $\sigma_{i,j}$ from all demand nodes to the source (all edges with $EBC^Q>0$, respectively), this collection of shortest paths is called the shortest path tree.

In real hydraulics of WDNs, loops are formed by two or more flow paths. The flow division in that loop adjusts it self according to the energy and mass balance. However, EBC^Q, as describe before, concentrates all the demands (flows) in just one of these flow paths, i.e., the shorter one (even if it is only slightly shorter than the alternate path). To overcome this shortcoming in the EBC^Q evaluations, the edge weights can be changed iteratively to achieve a more realistic flow division. This flow division can be achieved, by e.g., after identifying a shortest path for a demand parcel with edge weights $L_{k,0}$, artificially lengthening the edge weights in that path to $L_{k,1}$ for the next demand parcel (iteration). This lengthening of the edge weights is denoted dynamic weights (in contrast to static weights). The lengthening can be a function of the size of the demand parcel Q_i (e.g., $L_{k,1}=(1+Q_i(L/s)^2)\cdot L_{k,0}$). Small toy examples of determining EBC^Q with static and dynamic weights can be found in Figure 1.

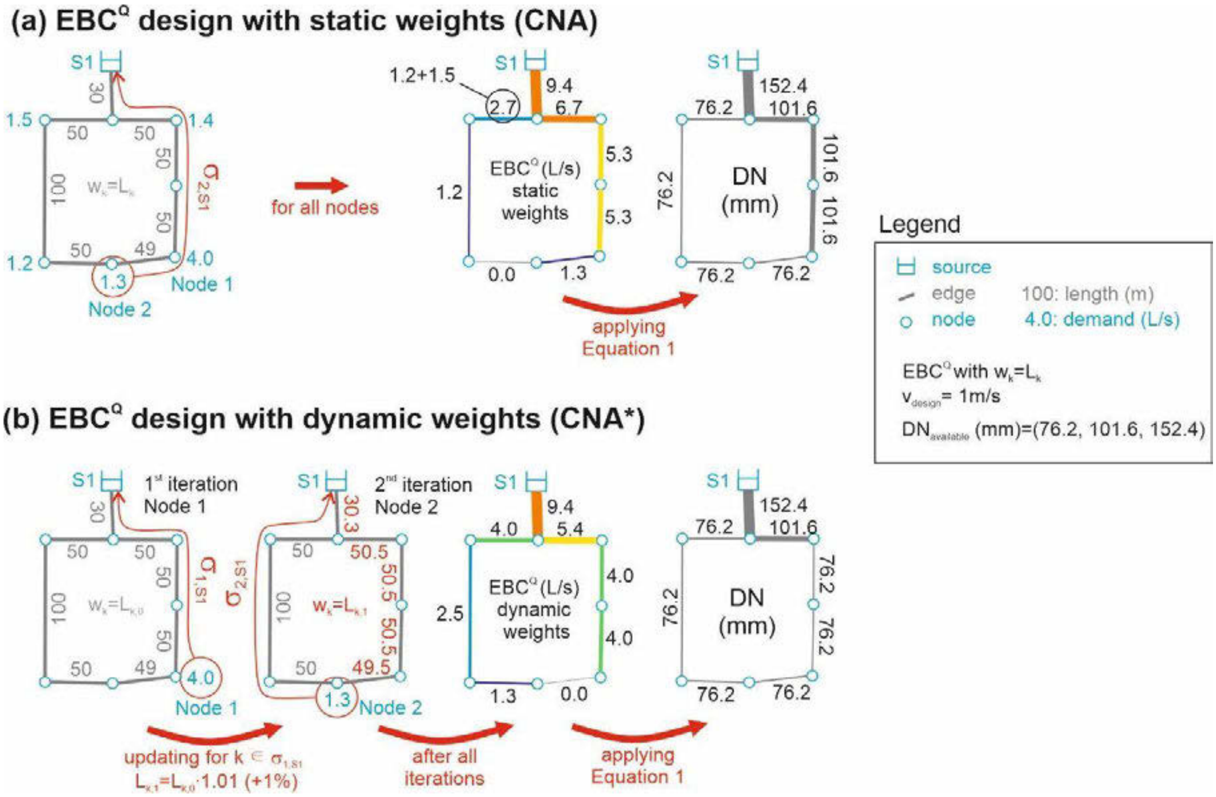

(a) EBC^Q design with static weights (CNA)
S1
30
1.5
50
50
1.4
100
w_k=L_k
50
50
σ_2,S1
1.2
50
49
4.0
1.3
Node 1
Node 2
for all nodes
1.2+1.5
2.7
9.4
6.7
5.3
1.2
EBC^Q (L/s) static weights
5.3
0.0
1.3
applying Equation 1
152.4
76.2
101.6
76.2
DN (mm)
101.6
101.6
76.2
76.2
Legend
source
edge
100: length (m)
node
4.0: demand (L/s)
EBC^Q with w_k=L_k
v_design= 1m/s
DN_available (mm)=(76.2, 101.6, 152.4)
(b) EBC^Q design with dynamic weights (CNA*)
S1
1st iteration Node 1
30
50
50
100
w_k=L_k,0
50
σ_1,S1
50
49
4.0
Node 1
S1
2nd iteration Node 2
30.3
σ_2,S1
50
50.5
100
w_k=L_k,1
50.5
50.5
50
49.5
1.3
Node 2
9.4
4.0
5.4
4.0
2.5
EBC^Q (L/s) dynamic weights
4.0
1.3
0.0
152.4
76.2
101.6
76.2
DN (mm)
76.2
76.2
76.2
76.2
updating for k ∈ σ_1,S1
L_k,1=L_k,0·1.01 (+1%)
after all iterations
applying Equation 1

Figure 1. Multi-objective design procedure based on CNA [6](licensed under CC-BY 4.0)

BY NC SA

Based on the EBC^Q values (static or dynamic weights), the diameters DN_k for each pipe k can be determined with continuity equation, an assumed flow velocity v_{design}, and commercially available diameter classes ($DN_{available}$):

$$DN_k = \lceil \sqrt{\frac{4}{\pi} \cdot \frac{EBC^Q(k)}{v_{design}}} \rceil \in DN_{available} \tag{1}$$

A summary of the design process can be found in Figure 1. Sitzenfrei, et al. [6] have shown, that the when considering different values for v_{design} in the described design process, Pareto optimal design solutions can be obtained (minimal pipe costs versus maximum resilience according to [9]) which partly outperforms designs from optimization with an evolutionary algorithm. For the proposed CNA design procedure itself, no hydraulic simulations are required. Only to check the pressure threshold afterwards (e.g., minimal pressure under design load of 30m), an simulation with Epanet2 is needed [10].

2.2 Water quality assessment with complex network analysis

For water quality assessment with CNA, again shortest path analysis $\sigma_{i,j}$ is used. This time, as an edge weights, residence times in the edges are considered. The shortest paths can then be interpreted as minimal residence time of water during the transport from a source to a demand node (i.e., water age or travel time in the WDN). In order to avoid additional hydraulic simulations, the hydraulic simulation results from the resilience assessment (see 2.1) with the design load are used. To account that for water age simulation the average demand load is decisive, the flow velocities from the resilience assessment are reduced by a factor describing the ration between design and average demand. Sitzenfrei [7] showed that with that assumption, only marginal errors regarding the flow velocities are obtained.

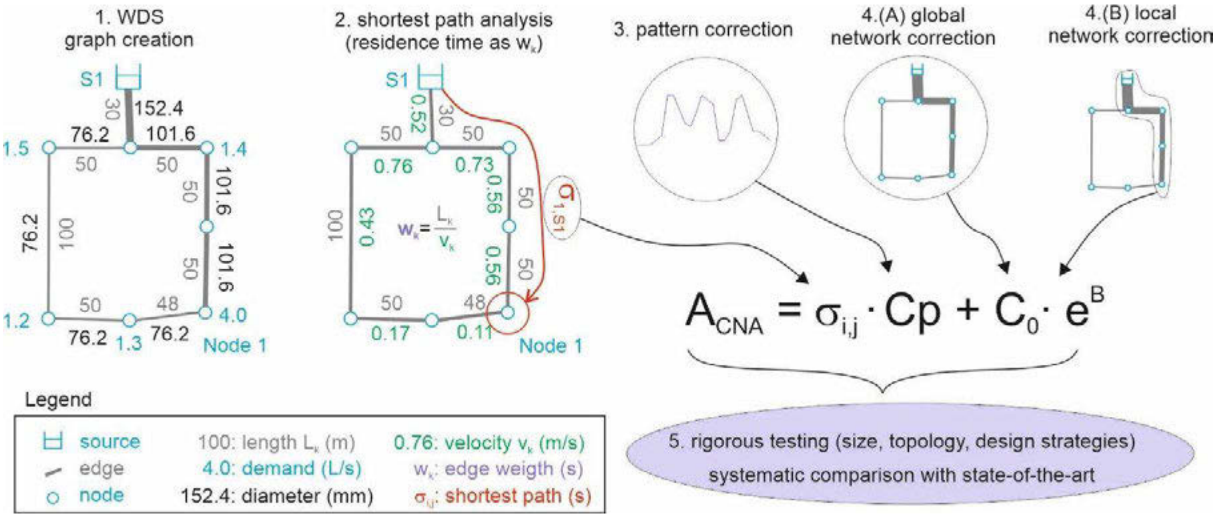

1. WDS graph creation
2. shortest path analysis (residence time as w_k)
3. pattern correction
4.(A) global network correction
4.(B) local network correction
S1
30
152.4
76.2
101.6
1.5
1.4
50
50
50
101.6
76.2
100
50
101.6
50
48
4.0
1.2
76.2
1.3
76.2
Node 1
S1
0.52
30
50
50
0.76
0.73
0.56
50
$\sigma_{1,S1}$
100
0.43
$w_k = \frac{L_k}{v_k}$
0.56
50
50
48
0.17
0.11
Node 1
$A_{CNA} = \sigma_{i,j} \cdot Cp + C_0 \cdot e^B$
5. rigorous testing (size, topology, design strategies) systematic comparison with state-of-the-art
Legend
source
edge
node
100: length L_k (m)
4.0: demand (L/s)
152.4: diameter (mm)
0.76: velocity v_k (m/s)
w_k: edge weigth (s)
σ_{ij}: shortest path (s)

Figure 2. Water quality assessment with CNA [7] (licensed under CC-BY 4.0)

The determined values $\sigma_{i,j}$ for water age are basically similar to a hydraulic snapshot simulation, where only edges in the shortest path tree are considered in the model. To account for extended period simulations (i.e., 24 different hourly values for diurnal demand pattern for multiple days), a simple way time consideration can be assumed [7] and a factor for pattern correction Cp can be determined. To consider that the flow is not only in the shortest path tree, corrections based on the network topology can be applied. As global network correction, the fraction of flow in alternative flow paths ($0 \leq Q_{alt} \leq 1$) and the mean node degree (mD) of the WDN are identified to describe the global network dispersion process regarding water age [7]. The node degree describes the number of edges connected to a node and mD is the average number of edges connected to a node in the network. Q_{alt} is the sum of flows in edges outside the shortest path tree divided by the total flow. Q_{alt} can also be determined based on hydraulic results from the resilience assessment (see 2.1). As local network correction, the flow lengths from the source to each demand node are determined (denoted mixL) and used. For a better understanding, also the number of mixing nodes in this flow paths are determined (denoted number mixing nodes) with CNA. Based on these different correction terms, from the shortest path values $\sigma_{i,j}$ the water age based on CNA can be determined (A_{CNA}). The entire procedure, systematic testing with different network topologies and validating the graph-based water quality model can be found in [7]. A summary of the procedure can be found in Figure 2 and the developed model in Equation 2.

$$A_{CNA} = \sigma_{i,j} \cdot C_P + 0.1 \cdot \mathrm{e}^{mD^2 \cdot Q_{alt} \cdot mixL} \tag{2}$$

2.3 Case study and design solutions from literature

As a case study, a large real case study is used. Due to data protection issues, the real layout cannot be shown here. However, an anonymized layout is shown in Figure 3. The hydraulics are fully preserved (same height differences and same pipe lengths), only the visual representation is changed. The case study serves approximately 100,000 inhabitants with one single source. The WDN model has 3,558 nodes and 4,021 edges and therefore decision variables. The Pareto front in Figure 3 (b) is taken from Sitzenfrei, et al. [11] and was determined GALAXY [8] which is based on an state-of-the-art NSGA-II algorithm for two-objective design. In general, the extended period simulations for water quality assessment are conducted with a water quality time step Δt= 1 min and a total simulation time of 10 days.

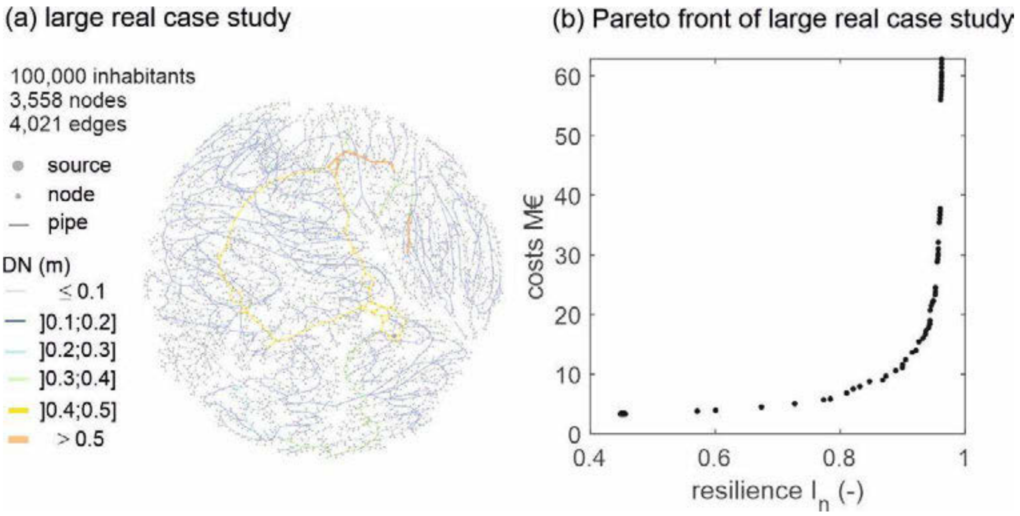


Figure 3. Anonymized layout of large real WDS (a) and Pareto front of optimal design solutions from meta-heuristic optimization

3 RESULTS AND DISCUSSIONS

The proposed graph-based water quality model, has the advantage of very fast execution time compared to extended period simulations (EPS) with Epanet2 but with less accuracy. In general, depending on the smallest network elements (pipes), the water quality time step Δt should be

BY NC SA

chosen in Epanet2. One could now argue, that a larger Δt could be used in order to ensure a short computation time with a potentially acceptable loss of accuracy. Therefore, for one design solutions from the EBC^{Q} designs, the impact of Δt sizes is investigated. The potential loss of accuracy with larger Δt, can be interpreted as numerical dispersion.

In Figure 4 (a), the water ages obtained for different water quality Δt (60min, 15min, 5 min and 1 min) are shown. Based on all demand nodes in the system, the four boxplots are created for the calculated nodal water ages. One can already see, how Δt sizes have an impact on the results of the water age simulations. However, the smallest nodal values are for all solutions are close to zero, while there is a great difference in the maximum water ages. Therefore, in Figure 4 (b), the nodal water ages for the different Δt sizes are evaluated depending on the number of mixing nodes from the source to the demand node. In Epanet2, at each node, the inflows are flow weighted averaged. If there is water age between two certain time steps, the next larger water age is used based on Δt. The maximum number of mixing nodes in the investigated WDN is 51, meaning for the furthest distant node, a water parcel in $\sigma_{i,j}$ with the shortest residence time passes 51 nodes, where it is potentially mixed with flows from outside the shortest path tree. Due to the rounding up to the next Δt, the water age is overestimated with increasing Δt. That effect is clearly shown in Figure 4 (b) where on the x-axis the number of mixing nodes for each node is shown and, on the y-axis, the ΔAge as difference of the current to the water age determined with Δt = 1min is shown. For the case study, the over-estimation of water age for 1h water quality time step is up to 100h, 15min up to 20 hours and for 5 minutes up to 7 hours, respectively. This analysis outlines the importance of using a small enough Δt for water quality analysis with Epanet2.

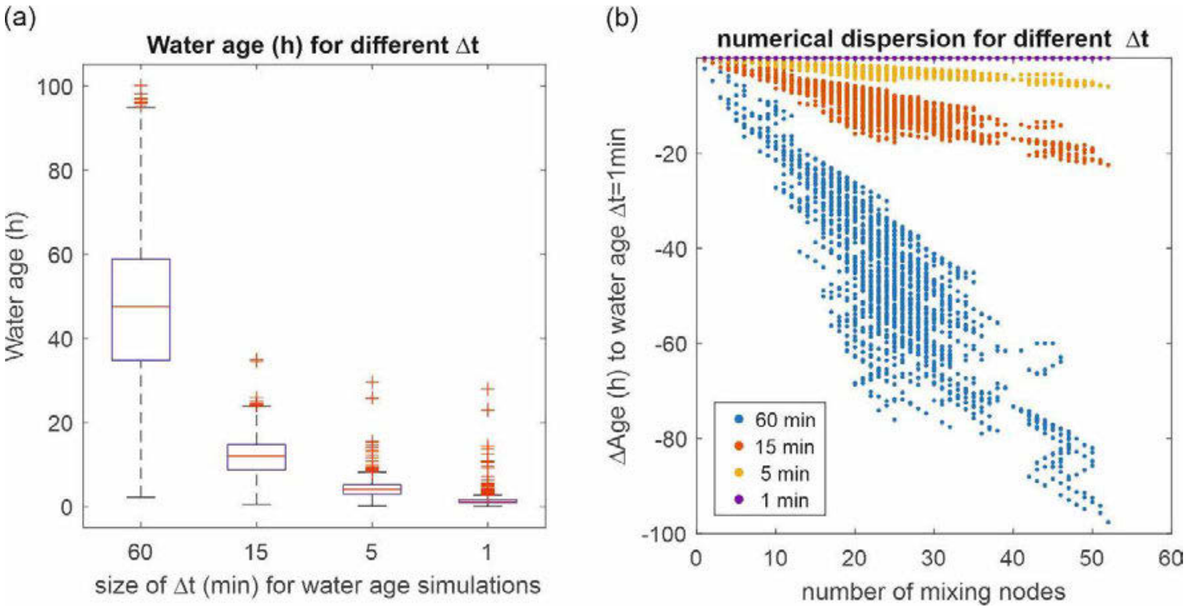


Figure 4. Impact of different Δt on the water age (a); impact of number of mixing nodes on the water age difference ΔAge to the results with Δt =1min (b)

In Table 1, for one design solution from EBC^{Q} design (see also Figure 4), the statistical differences in nodal water age and the computation times are shown. The graph-based method has by far the shortest computation time, and compared to the results with Δt=1 min, the best accuracy of the solution (median difference -0.19h)

Table 1. Mean differences in water age with different methods for one design solution and computation time

Δt / method	median difference (h) and standard deviation to Δt=1min	Computational time for water quality assessment of one design solution
60 min	-46.18 h ± 17.22h	0.09 min = 5.5 sec
15 min	- 10.57 h ± 3.97 h	0.13 min = 7.9 sec
5 min	- 2.74 h ± 1.06 h	0.88 min = 53.0 sec
1 min	-	30.61 min = 1,836.7 sec
A_{CNA}	-0.19 h± 0.27 h	0.00038 min = 0.023 sec

For the graph-based design, v_{design} is varied from 0.25 m/s to 2.0 m/s in 0.025 m/s steps. This results in 71 design solutions. When checking these 71 design solutions regarding the minimum pressure requirement, 55 of them fulfil the criterion pressure ≥ 30 m. The largest design value without pressure violation is v_{design} = 1.6 m/s. In Figure 5, for these 55 design solutions, the water age is determined with the graph-based water quality method (a) and an EPS with Epanet2 (b) with Δt=1min. Although there are also some discrepancies between (a) and (b), the overall trend is very well represented by the graph-based water quality method.

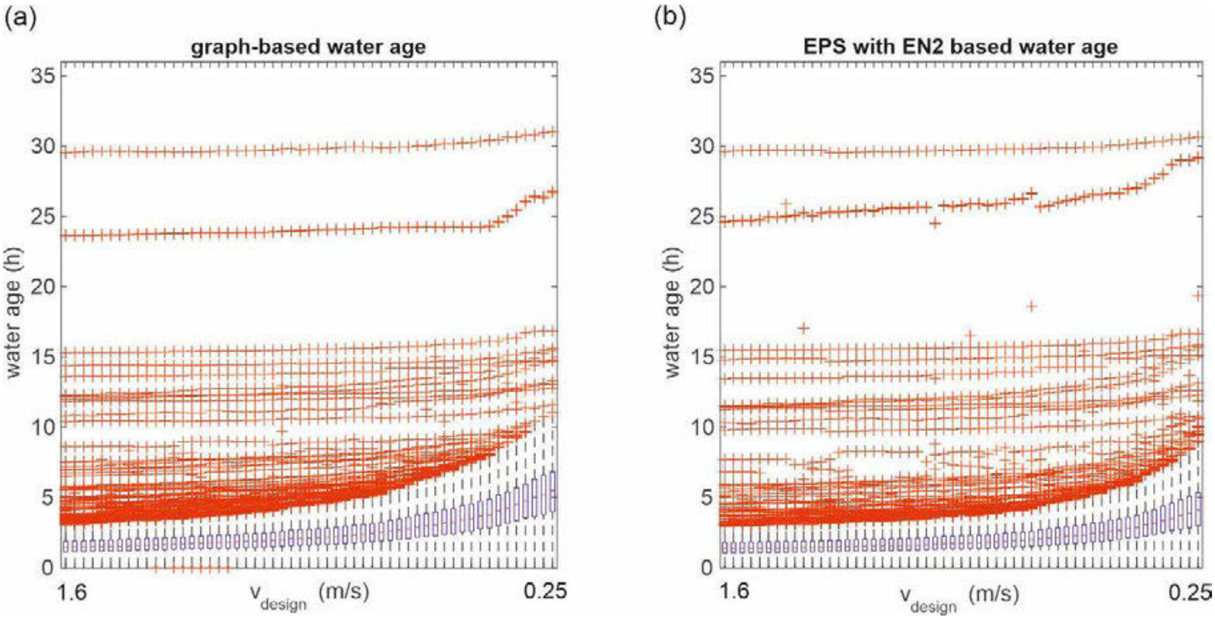


Figure 5. Results for graph-based water age (a) in comparison with results from EPS with Epanet2 (EN2) (b) for 55 graph-based design solutions

In a next step, the obtained Pareto-front for minimizing costs and maximizing resilience from the graph-based design method is compared with results obtained with GALAXY. For the graph-based designs, the water quality is also assessed with the graph-based water quality model and design solutions with a nodal water age exceeding 30 hours are identified. For the design solutions obtained with GALAXY (population size 100 and 300,000 generations results in 30 Mio. function evaluations), an EPS over 10 days with a water quality time step Δt=1min is performed. Note that in Sitzenfrei [7] it was shown, that when checking the water quality constraints for design solutions from GALAXY with the graph-based water quality model, produces almost identical results as with EPS with Epanet2.

When we compare the obtained design solutions from the graph-based design procedure (coloured dots), with design solutions form GALAXY (black and grey dots) in Figure 6, it can be observed, that for some regions of the Pareto-front, similar results can be obtained with both methods. The least cost solutions cannot be reproduced with the graph-based method and also the high resilient solutions are not competitive. However, when looking at the water quality performance, one can see that the solutions with a high resilience value from GALAXY, all have a very large water age (>168 hours) (grey dots). When using the resilience metric according to Prasad and Park [9], the high resilient solutions are also driven by the uniformity factor, which favours similar sized pipe diameters connected to a node. This results in large capacities all over the WDN. Such design solutions with almost uniformly distributed high capacity cannot be obtained with the graph-based design (so far, but will be tackled in future work). However, when looking at the graph-based design solutions with costs above 12 Mio. €, the maximum water age is still close to 30 hours (see also Figure 5) while for the design solution from GALAXY above 12 Mio. €, more than 168 hours and more are obtained.

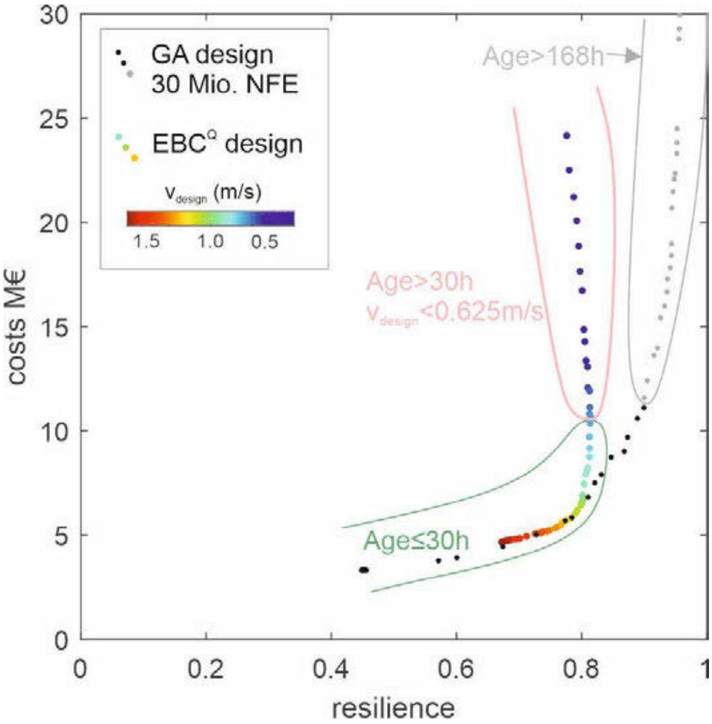


Figure 6. Results from graph-based design (coloured dots) in comparison with results from GALAXY Genetic Algorithm (GA) with number of function evaluations NFE = 30 Mio.

In summary with the combination of the two graph-based methods for design and water quality assessment, a relevant part of the Pareto front of optimal designs, considering water age exceedances as constraint, can be obtained. For a certain part (costs between 8 Mio € and 10 Mio €) the graph-based method does not provide competitive results compared to GALAXY. However, great potential is seen to further improve the graph-based design [12] and to also cover that part of the Pareto-front in future.

When we now investigate the computational time, for the graph-based design, for 55 EBC^Q design solutions it took in total 7.01 seconds to determine the diameters. Determining resilience (with Epanet2) and costs takes 10.45 seconds. So, in total for 55 design solutions, the entire procedure took 17.46 seconds. The graph-based water quality model took for all 55 design solutions in total 1.04 seconds. The entire graph analysis including design and water quality assessment took therefore 18.50 seconds. The water quality analysis with Epanet2 and EPS for the 55 graph-based design solutions with Δt=1min took 28.06 hours. The optimization of that case study with GALAXY with a population size of 100 and 300,000 generations took in total 8 weeks and the EPS of the 100 design in the final generation took 154h. So, in summary, the GALAXY solutions took more than 8 weeks and the graph-based solutions 18.50 seconds. Such a fast surrogate method enables a multitude of further applications, where a lot of evaluations are required (e.g., deep uncertainty analysis, scenario analysis, etc.) or fast solutions are expected (e.g., in a pre-design software).

4 SUMMARY AND CONCLUSIONS

In this work, a large real case study was optimized with a graph-based design approach with two contradicting objectives (minimal costs versus maximum resilience). Subsequent, from that Pareto-front, design solutions exceeding a water quality threshold (maximum water age) are identified with a graph-based water quality model and for validation purposes also extended period simulation with Epanet2 are performed for identification. By that the importance of an adequate (small) time steps (Δt=1min) was highlighted and the error associated with too large time steps was quantified. With the graph-based water quality method, significantly better results were obtained as for e.g., with Δt=5 min but with tremendously less computational efforts.

In a last step, the obtained design solutions from the graph-based design method are compared with design solution from an evolutionary algorithm. These design solutions were also assessed with extended period simulations with Epanet2 to check if the water quality constraints are met. It was shown, that with the combination of the two graph-based methods for design and water quality assessment, a relevant part of the Pareto front of optimal designs, considering water age exceedances as constraint, were obtained. For a certain part of the Pareto front with solutions with medium resilience, the graph-based method does not provide competitive results compared to the evolutionary algorithm. However, great potential is seen to further improve the graph-based design [12] and to also cover that part of the Pareto-front in future.

With the proposed graph-based methods, very fast approximate results for multi-objective optimization can be obtained. The fast execution times (18.5 seconds compared to 8 weeks) enables an implementation in e.g., hydraulic simulation software, where a fast and user-friendly pre-assessment is needed or any kind of application where many scenarios need to be investigated (e.g., deep uncertainty analysis, future developments, etc.). Further, the graph-design procedure can serve as a global search method for multi-objective optimization, which is then further improved by a local search algorithm based on an evolutionary algorithm.

The graph-based water quality model enables also other further applications. For example, in course of multi-objective optimization (e.g., with GALAXY), an additional constraint could be implemented excluding design solutions which are (by far) exceeding a water age threshold. For the 30 Mio. function evaluations with GALAXY (which required 8 weeks of computation time), the graph-based water quality assessment of all 30 Mio. designs would only require additional 64.1h. By that, the search space is also reduced, potentially improving also the computation time of the evolutionary algorithm.

5 REFERENCES

[1] M. Möderl, R. Sitzenfrei, and W. Rauch, "How Many Network Sources are Enough?," in World Environmental & Water Resources Congress, Challenges of Change, Providence, Rhode Island, USA, 2010.

[2] H. R. Maier, Z. Kapelan, J. Kasprzyk, J. Kollat, L. S. Matott, M. C. Cunha, et al., "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions," Environmental Modelling & Software, vol. 62, pp. 271-299, 2014/12/01/ 2014.

[3] A. De Corte and K. Sörensen, "Optimisation of gravity-fed water distribution network design: A critical review," European Journal of Operational Research, vol. 228, pp. 1-10, 2013.

[4] H. Mala-Jetmarova, N. Sultanova, and D. Savic, "Lost in optimisation of water distribution systems? A literature review of system operation," Environmental Modelling & Software, vol. 93, pp. 209-254, 7// 2017.

[5] M. Möderl, R. Sitzenfrei, T. Fetz, E. Fleischhacker, and W. Rauch, "Systematic generation of virtual networks for water supply," Water Resources Research, vol. 47, Feb 1 2011.

[6] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savić, "Using Complex Network Analysis for Optimization of Water Distribution Networks," Water Resources Research, vol. 56, p. e2020WR027929, 2020.

[7] R. Sitzenfrei, "Using complex network analysis for water quality assessment in large water distribution systems," Water Research, vol. 201, p. 117359, 2021/08/01/ 2021.

[8] Q. Wang, D. A. Savić, and Z. Kapelan, "GALAXY: A new hybrid MOEA for the optimal design of Water Distribution Systems," Water Resources Research, vol. 53, pp. 1997-2015, 2017.

[9] T. D. Prasad and N.-S. Park, "Multiobjective Genetic Algorithms for Design of Water Distribution Networks," Journal of Water Resources Planning and Management, vol. 130, pp. 73-82, 2004.

[10] L. A. Rossman, "EPANET 2 user manual," National Risk Management Research Laboratory - U.S. Environmental Protection Agency, Cincinnati, Ohio2000.

[11] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savic, "A complex network approach for Pareto-optimal design of water distribution networks," in World Environmental and Water Resources Congress 2021, ed, 2021.

[12] M. Hajibabaei, S. Hesarkazzazi, A. Minaei, D. Savic, and R. Sitzenfrei, "Dynamic Edge Betweeness Centrality and Optimal Design of Water Distribution Networks," presented at the 2nd International Joint Conference on Water Distribution Analysis & Computing and Control in the Water Industry 2022, Valencia, Spain, 2022.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14441

LONG-TERM TRANSITIONING OF WATER DISTRIBUTION SYSTEMS: ERC WATER-FUTURES PROJECT

Dragan Savić[1], Barbara Hammer[2], Phoebe Koundouri[3], Marios Polycarpou[4]

[1]KWR Water Research Institute, Nieuwegein, The Netherlands
[1]Centre for Water Systems, University of Exeter, Exeter, United Kingdom
[2]Faculty of Technology, Bielefeld University, Germany
[3]School of Economics, Athens University of Economics and Business, Greece
[3]Sustainable Development Unit, ATHENA RC, Greece
[3]Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[4]KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
[4]Dept of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus

[1] *Dragan.Savic@kwrwater.nl*, [2] *bhammer@techfak.uni-bielefeld.de*, [3] *pkoundouri*@aueb.gr,
[4] *mpolycar@ucy.ac.cy*

Abstract

The percentage of the world population living in urban settlements is expected to increase to 70% of 9.7 billion by 2050. Historically, as cities grew, the development of new water infrastructures followed as needed. However, these developments had less to do with real planning than with reacting to crisis situations and urgent needs, due to the inability of urban water planners to consider long-term, deeply uncertain and ambiguous factors affecting urban development and water demand. The "Smart Water Futures: Designing the Next Generation of Urban Drinking Water Systems" or "Water-Futures" project, which was funded by the European Research Council (ERC), aims to develop a new theoretical framework for the allocation and development decisions on drinking water infrastructure systems so that they are: (i) socially equitable, (ii) economically efficient, and (iii) environmentally resilient, as advocated by the UN Agenda 2030, Sustainable Development Goals. The ERC Synergy grant project tackles the "wicked problem" of transitioning water distribution systems in a holistic manner, involving civil engineering, control engineering, machine learning, decision theory and environmental economics expertise. Developing a theoretical foundation for designing smart water systems that can deliver optimally robust and resilient decisions for short/long-term planning is one of the biggest challenges that future cities will be facing. This paper presents an overview of related past research on this topic, the knowledge gaps in terms of investigating the problem in a holistic manner, and the key early outcomes of the project.

Keywords
Drinking water networks, transitioning, real-time monitoring, long-term design, sustainability, deep uncertainty.

1 INTRODUCTION

Nearly 80% of the world's population is exposed to high levels of threat to water security due to anthropogenic climate change 0. Latest studies confirm that considerable changes in freshwater resources have been occurring across the globe, indicating a future in which already limited water resources will become even more precious 0. On the other hand, the continuous expansion of the urban footprint means that an estimated 70% of the world's population will live in urban areas by 2050 0. The dramatically increased water demands resulting from this unprecedented urbanization, together with increasingly uncertain climate conditions indicate the need for a holistic, intelligent decision-making framework for managing water infrastructures in the cities of the future. This framework needs to ensure that allocation and development decisions on water infrastructure systems will be made in a socially equitable, economically efficient and environmentally resilient way as advocated by the UN Agenda 2030, Sustainable Development Goals, SDGs 0. Consequently, there is a need for a new approach to designing the next generation of urban drinking water systems that applies not only to the planning and management of mature water infrastructure systems such as those found in developed countries, but also to developing countries where the fastest population growth is predicted over the next 50 years. We need a framework that can: (i) adapt to evolving urban water networks, new sensing technologies and consumer behaviours; (ii) integrate real-time monitoring and control with long-term planning and policy-making; (iii) assimilate water quality issues with water supply problems; and (iv) incorporate economic, social, ethical and environmental considerations. Moreover, the new framework needs to be human-centric so that intelligent algorithmic solutions are explainable and acceptable by human policymakers, managers, operators and consumers.

The "Smart Water Futures: Designing the Next Generation of Urban Drinking Water Systems" or "Water-Futures" project, which was funded by the European Research Council (ERC), aims to develop a new theoretical framework for the allocation and development decisions on drinking water infrastructure systems 0. We consider four challenges for the development of the framework for drinking water systems: scenario-based staged development, real-time smart operation, explainable machine learning and economic considerations. This paper presents an overview of related past research on this topic, the knowledge gaps in terms of investigating the problem in a holistic manner, and the key early outcomes of the project.

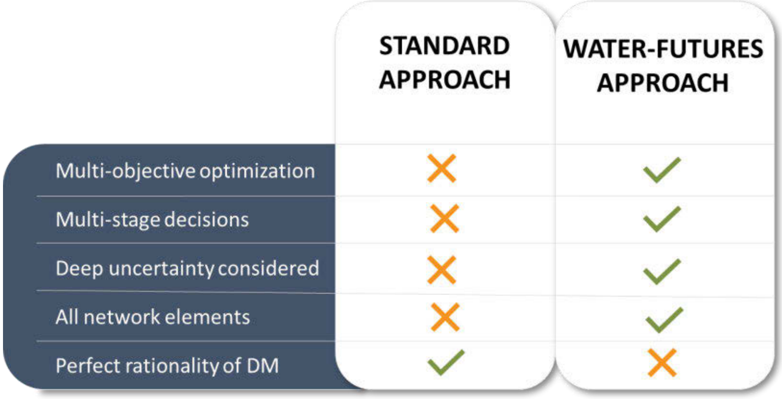
2 SCENARIO-BASED STAGED DEVELOPMENT OF DRINKING WATER SYSTEMS

2.1 Staged development

World cities are facing crucial policy decisions about how to achieve long-term water security considering their ageing water infrastructures. The key challenges are based on whether cities: (i) can anticipate the future growth trends; (ii) will adopt a long-term perspective; (iii) will take into account decision robustness; (iv) will consider policy flexibility; and (v) are keen to develop strategic visions to support water infrastructure planning decisions 0. Most of the early planning models were based on financial criteria and considered only pipes as network elements for replacement or renewal at a particular future time, i.e., a static, one-stage problem of deciding when it is economic to replace a pipe 0,0,0. Engelhardt et al. 0reviewed long-term planning strategies for water infrastructures and identified key advantages of implementing an optimisation-based, holistic approach to infrastructure planning. Several studies have since optimised mainly small benchmark systems considering one, or at most a limited number of the following issues: (i) a single-objective economic criterion; (ii) a static, one-stage problem; (iii) ideal foresight; i.e., the correct prediction of future; (iv) only pipes as decision variables; and (v) perfect rationality of decision-makers 0, as seen in Figure 1. What is currently missing, is extending the problem of planning and management into the more generalized problem of sustainable transitioning of urban water systems, considering multiple objectives and decision variables (e.g., pricing, investment, number of sensors to install), while considering both hydraulic and quality dynamics as well as deep uncertainties.

Staged design of drinking water networks may be defined as the problem of identifying a sequence of design decisions that need to be taken over several consecutive stages during the planning horizon, to optimize benefits and costs, subject to specific constraints at each stage 0. Optimization tools are often used to seek the best sequence of interrelated designs, which cannot be optimized individually as each of them encompasses the solutions from previous stages. The main advantage of staged development is the ability to use *an adaptive design strategy* to make decisions that adapt over time to changing circumstances rather than base them on a fixed design over the entire planning horizon. This is particularly important when decisions have to be made under deep uncertainty with multiple possible futures, such as in the case of increased climate variation.

BY NC SA



	STANDARD APPROACH	WATER-FUTURES APPROACH
Multi-objective optimization	✗	✓
Multi-stage decisions	✗	✓
Deep uncertainty considered	✗	✓
All network elements	✗	✓
Perfect rationality of DM	✓	✗

Figure 1. Approaches to long-term planning strategies for water infrastructure, the standard approach vs. the approach taken by the Water-Futures project

With the realisation that climate change poses fundamental threats to infrastructures, people and urbanised areas, the last 20 years have witnessed an increase in research related to the impact of deep uncertainty on the long-term security of water infrastructure systems 0. Combined with that, two key considerations have to be incorporated into the adaptive design of drinking water networks:

1) *Design flexibility* - the ability to implement the first-stage design while keeping a view of the long-term system development, and
2) *Design robustness* - the ability of a design to satisfy as many future scenarios as possible.

2.2 Long-term scenarios

While several researchers have developed approaches to introduce flexibility and/or robustness to infrastructure planning 0, 0, 0 0, the ultimate goal of integrating short-term and long-term planning activities in a holistic theoretical decision-making framework under deep uncertainty (e.g., climate, demographic or economic projections) has not been addressed.

Engineers are accustomed to using historical data in the design of infrastructure. For example, demand projections based on past realisations are used to assess the need for new infrastructure. However, predicting the long-term future is extremely difficult and often ends up with wrong predictions. For example, when twenty-seven top US scientists in the sixties predicted what the world would look like in 20 years; out of 335 predictions, nearly all were wrong 0. Multiple scenarios are used instead to provide a better understanding of the range of possible environments the water infrastructure system must contend with in the future. Using multiple scenarios avoids the situation when building a water infrastructure project specifically for one (wrong) scenario, but uses a diversity of scenarios over which the project should perform robustly and satisfactorily 0.

3 REAL-TIME SMART MANAGEMENT OF WATER DISTRIBUTION SYSTEMS

Decision making in urban water distribution networks can be viewed in terms of the long-term planning and management, discussed in the previous section, as well as in terms of the operational management of the water system, which is characterized by real-time and short-term decisions. Real-time decision making is typically implemented by automated systems or algorithms and supervised by water operators during the everyday monitoring and control of the system. In this section, we focus on some of the key issues that need to be addressed in real-time monitoring and control of urban drinking water distribution systems, especially in view of new technological advances that are starting to become widely available for the management of water distribution systems.

CC BY NC SA

The objectives of monitoring and control of water systems include detecting and resolving unanticipated events in the system, controlling the system parameters to minimize losses and safeguard quality, optimising resource allocation and making sure the system works in the most efficient way (e.g., energy efficiency). On the other hand, real-time monitoring and control must be aligned with the long-term management and planning decisions, as well as the general high-level policies. This includes monitoring how the risk of abnormal events changes in time and how to reconfigure the system in order to mitigate extreme events, which may be of low probability but high impact. At the same time, it is crucial that future technology developments (e.g., new embedded sensors and actuators) will not increase risks to urban water systems (e.g., malicious attacks).

Various aspects of the real-time monitoring and control problem have been investigated in water systems 0. In practice, usually, the system parameters are unknown, therefore the models need to be calibrated using an optimization method 0,0. Since water consumption affects the flows in a network, consumers play an important role in the dynamics of drinking water distribution systems, and their behaviour has been studied and modelled 000. Moreover, placement problems have been studied, to determine where to optimally install water quality and hydraulic sensors, to improve event detectability 0,0. Using sensor information, various methods have been proposed for contamination diagnosis 0,0,0, as well as leakage detection and localization 0. In parallel to monitoring, feedback control methods have been proposed for water quality control 0,0, pump scheduling 0 and pressure valve control 00.

From a monitoring and control perspective, water distribution systems are cyber-physical-social systems 0 with multiple interactive dynamics and feedback loops:

- *Physical part* – this includes all the physical components that are required for the normal operation of a water distribution system, such as pipes, tanks, valves, etc
- *Cyber part* – this includes all the algorithms that are required for monitoring the operation of the water system and for autonomously controlling the behaviour of the physical system.
- *Social part* – this includes the human behaviour that influences the operation of the water system, such as the water demand of consumers, behaviour of operators and policy makers, etc.

The cyber-physical-social framework for the design and analysis of real-time monitoring and control methods for water distribution systems facilitates a holistic approach that takes into consideration the embedded nature of the cyber-physicals parts of the system, while at the same time incorporating human behaviour (social part), which is an integral part of the operation of water systems. The dynamics of drinking water distribution networks include the hydraulic dynamics (e.g., flows and pressures) and the water quality dynamics (concentration of various chemical substances and biological species). Due to the conventional decay of chemicals over time, the hydraulic dynamics affect the water quality dynamics.

During the last few years, there have been significant advances in the development of new sensing devices that measure various hydraulic and quality parameters of water systems. These sensing devices are often deployed in an Internet-of-Things (IoT) setting, with the capability to be embedded and integrated with real-time decision and control algorithms.

It is anticipated that this technological trend will continue in the future with further proliferation of IoT technology and the advancement of information and communication technologies and data analytics. This will result in novel sensing devices for measuring water quality and hydraulic parameters, which are expected to be smaller, cheaper and possibly easier to install and maintain. Moreover, these devices will be connected to the internet, so the information will be available for processing in real-time, both to human operators as well as to algorithms for automated decision

BY NC SA

making. Other advances that are affecting the operation and smart management of water distribution systems are the development of virtual sensors, which are based on algorithmic methods for measuring parameters at certain locations (in contrast to hardware devices)0, and IoT actuators, which are internet-enabled devices that can enhance the automation and supervision processes of water distribution systems.

While these technological advances provide the potential for significantly enhancing the capability for real-time smart management of water distribution systems, they also pose some key challenges and risks. For example, with the wide deployment of internet of things devices, there are risks associated with privacy issues, which need to be seriously addressed. Moreover, there are significant risks associated with the potential of malicious cyber-physical attacks. Since water systems are critical infrastructures (similar to energy and power grids, telecommunication networks, transportation systems, etc), it is crucial that they are protected against any potential cyber-physical attack that may compromise its smooth operation, or even worse, cause contamination of the drinking water. Therefore, in addition to handling normal operation, real-time monitoring and control algorithms are required to be able to detect malicious attacks and to be able to distinguish between normal or accidental faults (e.g., sensor faults, actuator faults) and malicious attacks (e.g., replay attacks).

4 EXPLAINABLE MACHINE LEARNING

Although hydraulic equations of water distribution systems (WDS) are well understood, and powerful simulation technologies of WDS exist 0, modelling of real networks is subject to severe uncertainties: More than 16% of water pipes have surpassed their useful lives and face serious ageing and deterioration challenges, where the exact state is usually unknown 0. Moreover, optimal planning and control depend on expected future demand, a widely unknown quantity in particular in the light of yet unclear effects caused by global warming or growth of cities 0. In such settings, data-driven modelling and prediction constitute one possibility to match formal models to reality.

The increase of digital information including historical data on water demand as well as real-time sensor information, which mirrors the current state of the network, has led to a rise of data-driven methods, in particular machine learning models, in WDS. These offer crucial technologies, which are capable of enhancing physical simulation and control by information relevant for decisions which depend on the specific network state **¡Error! No se encuentra el origen de la referencia.**. Tasks which have been addressed by machine learning in this context are widespread: estimating the condition of water pipelines, leakage detection and localization, prediction and management of pipe failures, modelling of water quality, demand prediction, optimization of water treatment plans, early warning systems, or efficient data-driven optimization, to name just a few 0,0,0,0,0,0. Besides classical machine learning methods such as Bayesian modelling, random forests, or kernel methods, recent approaches often rely on deep learning 0,0,0.

In the context of WDS, a number of specific challenges arise, which cause the need for adaptations of common workflows in machine learning:

1) Heterogeneous data format – while most machine learning technologies have been designed for homogeneous vectorial data, measurements in the domain of WDS include heterogeneous sensor data, which are subject to spatial and temporal characteristics. Mixed real-valued and discrete representations can render optimization difficult, and data heterogeneity causes the need for data assimilation 0. In recent years, dedicated models which can directly deal with temporal or spatial data such as deep recurrent and graph neural networks have led to promising results 0,0, but it is yet unclear how to best represent digital information in the light of underlying domain knowledge in WDS 0.

BY NC SA

2) Imbalanced data and changing distribution – the frequency of observed phenomena does not necessarily scale with their relevance in WDS. As an example in the context of prediction and management, pipe failures are observed much less frequently than normal behaviour, hence the data distribution is skewed. As a consequence, machine learning models need to correct for such biases and deal with imbalanced data 0. In real-time dynamic systems, another challenge is given by data drift, i.e., the fact that the data-generating process might change over time, caused by sensor fatigue or changing demands in developing cities, for example. Such phenomena lead to a violation of one of the fundamental assumptions of classical machine learning, the assumption that data are identically distributed and representative of the underlying regularity. Here special care has to be taken to continuously adapt machine learning models to possibly changing demands using online learning technologies 000.
3) Necessity of human-centred design – WDS as critical infrastructure directly affect humans in their daily life. Thereby, humans have different roles: (i) as customers to whom service is provided, (ii) as actors who determine the development of WDS via their behaviour (e.g. exhaustive consumption of water) and decisions (e.g. price policy), and (iii) as engineers who need to guarantee a sustainable quality of service. Human-centred design in WDS needs to take these roles into account to achieve robust functionality and sustainability of WDS 0. While machine learning technologies can help in short-term control and long-term planning of WDS, the black-box nature of modern technologies such as deep networks adds a possible complication here: humans might be incapable of understanding the rationale behind decisions made by ML technologies, and human intention and objectives implemented in ML systems might be severely misaligned 0. Hence human-centred design in WDS faces the challenge to make ML technologies transparent to humans.

4.1 Explainable AI (XAI)

Explainable AI (XAI) or, more specifically, explainable machine learning refers to methods which substantiate black box technologies with components which can be understood by humans 0. Commonly, one distinguished global XAI methods, which provide insight into the global function of a model (such as the most relevant rules which characterize a leakage), and local XAI methods, which explain a single decision only (such as an explanation of why a specific sensor signal should be interpreted as a leakage rather than normal behaviour). In recent years, a variety of different technologies have been proposed, whereby they differ w.r.t the form of explanation (such as feature-based versus exemplar-based methods), the algorithmic choices used to compute the explanation (such as post-hoc methods versus embedded methods), and the objective of the explanation (such as proposing actions how to repair an observed fault versus explanations which specify who should be held liable for an observed failure).

Most local XAI technologies for deep models have been proposed in the last few years only; hence, in WDS, existing XAI approaches mostly focus on global models: as an example, natively interpretable global XAI methods have been proposed in the form of neuro-fuzzy-systems 0, i.e., extensions of logical rules to continuous measurements, which characterize conditions of water pipelines. Global hybrid explanation methods are presented in the work 0: more specifically, leakage detection methods combine hydraulic transient modelling and machine learning technologies to extract the most relevant features based on which to design the model. Post-hoc technologies, which determine the most relevant features to decide the water quality in the context of Algal bloom for trained machine learning models, have been proposed in the work 0. These latter methodologies belong to the class of feature-relevance-determination methods, using different principles to account for possible redundancies and correlations of the information which is contained in diverse sets of features.

In a recent approach, local explanation technologies have successfully been used to explain sensor failures in spatio-temporal networked data in WDS 0: Here time series models predict local sensor values based on the neighbourhood, and a threshold strategy is used to indicate deviations of sensor measurements from expected behaviour. Afterwards, so-called counterfactual explanations are used to explain why the deviation takes place. Counterfactual explanations provide the information on what needs to be changed in the input to obtain the desired output change. They can be computed particularly efficient for specific models 0. In the work 0, local counterfactual explanations are coordinated within the WDS in such a way that it becomes possible to identify the global source of the sensor fault as a `consensus' of all local explanations. The results demonstrate the benefit which arises when harvesting on the network structure in WDS for XAI methods, yet it still deals with a comparably simple setup. Recent advances in XAI technologies for deep graph neural networks 0 or distributional changes 0 offer promising starting points to explore the capabilities offered by XAI technologies for complex spatio-temporal systems as present in WDS.

4.2 Fairness and trust

Since XAI technologies provide insight into the mechanism based on which automatic decisions are taken, they offer a convenient possibility to inspect the objectives implemented by an AI model in an explicit form. In particularly, XAI methodologies enable humans to identify deviations of an AI model from the desired functional and non-functional goals set by a human partner. Here, two crucial objectives are fairness and trust.

The notion of `fairness' formalizes the intuition that an AI model treats individuals or groups similarly unless there are valid reasons not to do so. As an example, the outcome of a recidivism decision should be independent of a person's ethnicity, hence the latter should constitute a `protected' feature which does not influence the final outcome 0. XAI methods can uncover violations of this objective, since they are capable of explaining the dependency of decision outcomes and such protected attributes. Indeed, it has been shown that there exist popular AI models used in practice which display a severe bias in diverse areas including automated hiring, recidivism assessment, or language models 0,0. In WDS, however, the notion of fairness is yet widely unexplored, albeit highly relevant. Here fairness refers to the question of whether access and quality of WDS are evenly distributed among all customers of a WDS unless prohibited by unavoidable physical constraints. In the light of long-term developments, fairness in WDS also refers to differences in costs, quality, and services of WDS over several generations.

The notion of`trust' summarizes the prerequisites which are required such that a human is willing to use an offered AI system. In the first place, this notion refers to the trust that the objectives aimed for by a human are met by the system, including non-functional ones such as fairness; beyond this alignment, it includes the trust that the AI model does so robustly and in possibly changing or adversarial realistic environments. Hence a second crucial demand to establish trust are guarantees for the security and safety of WDS in realistic and possibly changing environments 0. Partially, robustness against attacks can be guaranteed by mathematical properties 0. XAI technologies offer an additional avenue based on which to inspect, which attacks and changes can be harmful to a system – an opportunity which is yet widely unexplored in the domain of WDS.

5 ECONOMIC CONSIDERATIONS FOR LONG-TERM TRANSITIONING OF WATER DISTRIBUTION SYSTEMS

The Water-Futures human-centric approach aspires to develop a methodology that integrates economic, social, ethical and environmental considerations, with direct relevance to UN Agenda 2030, into an interdisciplinary decision-support framework that will allow agent-based societal welfare maximization in the short, medium and long-run, under deep uncertainty. In doing so our research focuses on five central unresolved scientific questions, challenging the traditional paradigm of Neoclassical Economics. These questions will drive our investigations for an alternative, deeper, more mature understanding of the structure of human preferences and the decision-making process.

The traditional paradigm in Neoclassical Economics toward welfare maximization passes through rationality. A rational agent is assumed to seek to maximize utility given information and geographical boundaries. However, research has shown that rationality is a situation that is relative and under (deep) uncertainty it is violated. Moreover, different definitions of rationality exist and each produces different results. Time and uncertainty are correlated, while uncertainty often takes the form of ambiguity (when probabilities of uncertain events are unknown). The term risk refers to situations in which the probabilities of events' occurrence are known, while the notion of uncertainty is broader and refers to situations in which this may not be the case 0. Most decisions indeed must be made in situations in which some events do not have an obvious, unanimously agreed-on probability assignment. This might be because too little information is available or because different predictions exist, resulting from different models or datasets or different experts' opinions. Currently, the evaluation of climate policy is generally performed using models that do not distinguish between risk and uncertainty but actually reduce any kind of uncertainty to risk. In this task we will augment the mathematical decision-making framework towards treating deep uncertainty, enabling robust decision making with regard to the short, medium and long-run development of urban water systems under climate change, as short and long-run decisions should be dynamically consistent and integrated into a unifying framework.

People do not only differ in their tastes for goods and services, but also with regard to how selfish or fair-minded they are, which has important economic consequences. This also highlights the ethical dimension of welfare maximization, the concept of eudaimonia which has been neglected so far 0. We will develop a mathematical decision framework for the allocation of urban water (over time and space, and between societal layers) and the development of the systems technology and infrastructure that supports this allocation, which will augment the current neoclassical paradigm to internalize the ethical dimension in decision making that lead societies to eudaimonia. The suggested augmentation builds on the literature on “Subjective Well-Being”, which has produced remarkable results over the last years on determinants of happiness, consequences of happiness, causality, integration into standard economics and policy consequences.

“Subjective Well-Being” entails a deeper understanding of human preferences with regard to public goods, such as water infrastructure and related environmental concerns. These preferences are not documented in any markets and as a result, we need to infer them from other choices people make, or to directly elicit them via field or laboratory experiments. Water-Futures will use non-market valuation experimental methods 0 aided by virtual reality experiences (to the best of our knowledge this is the first attempt for this integration) to elicit willingness to pay and welfare benefits for exciting, or planned water infrastructures. People's preferences and valuation are dynamic and are shaped by available information and people’s ability to understand this information -which is multidisciplinary and science hectic. Water-Futures search for responses to these challenges will build on a novel combination of the literature on “Subjective Well-Being” and “Experimental Behavioural Economics” for developing a new mathematical decision-making framework, integrated with the water systems engineering, optimal control and machine learning algorithms, that can support the design of smart urban water systems in a way that leads societies to eudaimonia (happiness via preference satisfaction). As far as implementation is concerned Water-Futures aims to use the systems innovation approach to co-design the future vision for urban water systems and co-develop the technological, policy, and financial pathways towards achieving this future vision, by engaging through living labs, all relevant stakeholders from different countries across the world.

In a response to United Nations Secretary-General Antonio Guterres' call for action: "Today, Sustainable Development Goal 6 is badly off track" and it "is hindering progress on the 2030 Agenda, the realization of human rights and the achievement of peace and security around the world", our endeavour is to support and accelerate the implementation of SDG6 for the people, the planet and their prosperity.

6 SUMMARY AND CONCLUSIONS

The dramatic rise in water demand resulting from unprecedented urbanization, together with increasingly uncertain climate conditions indicates the need for a holistic, intelligent decision-making framework for managing water infrastructures in the cities of the future. This framework needs to ensure that allocation and development decisions on water infrastructure systems will be made in a socially equitable, economically efficient and environmentally resilient way as advocated by the UN Agenda 2030, Sustainable Development Goals. Consequently, there is a need for a new approach to designing the next generation of urban drinking water systems that applies not only to the planning and management of mature water infrastructure systems such as those found in developed countries but also to developing countries where the fastest population growth is predicted over the next 50 years.

The new design approach needs to: (i) be adaptable to evolving urban water networks (in stages), new sensing technologies and consumer behaviours; (ii) be able to integrate real-time monitoring and control with long-term planning and policy making; (iii) be able to assimilate water quality issues with water supply problems; and (iv) incorporate economic, social, ethical and environmental considerations. Moreover, the new framework needs to be human-centric so that intelligent algorithmic solutions are explainable and acceptable by human policymakers, managers, operators and consumers. Due to its complexity and many interdependent factors, this challenge is a typical 'wicked problem', which seems impossible to solve 0.

The success of the Water-Futures project depends on inter- and trans-disciplinary synergies, which combine and transcend the different expertise and methodologies (Sections 2-5) into a holistic design framework. As an example, scenario generation constitutes a key factor to enable decision-making under deep uncertainties, yet it requires socio-economic insights into the rationality of human decision making and machine technologies to uncover homogeneous clusters and critical transitions in long-term dynamics. Conversely, a valid quantitative evaluation of the explainability of machine learning models is impossible without a reference to expert knowledge and human perception, hence vocabulary and validity need insights from control theory and socio-economics. Short- and long-term objectives of control and decision making of smart water systems are closely interrelated, yielding a strong interdependency of technical as well as societal modelling objectives, at the same time requiring machine learning technologies to tame the involved combinatorial complexity. Integrating ethics and fairness in control systems requires socio-economics expertise. Conversely, integrating physical feedback loops and the different social dynamics requires deep collaboration between systems and control theory as well as water engineering. Although only in its first year this six-year project has already identified the key areas of research and addressed some of the fundamental elements of the Water-Futures framework for long-term transitioning of water distribution systems.

7 ACKNOWLEDGEMENTS

This paper is a result of the project Water-Futures which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 951424).

8 REFERENCES

[1] Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. and Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature, 467(7315), p.555.

[2] Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoing, H.K., Landerer, F.W. and Lo, M.H., 2018. Emerging trends in global freshwater availability. Nature, 557(7707), p.651.

[3] United Nations (UN), the Department of Economic and Social Affairs, 2018. World Urbanization Prospects: The 2018 Revision, Key Facts. United Nations Department of Economics and Social Affairs, Population Division: New York, NY, USA.

[4] United Nations (UN), 2015. Transforming Our World, the 2030 Agenda for Sustainable Development. General Assembly Resolution A/RES/70/1.

[5] The website of the Water-Futures ERC Synergy Project: https://waterfutures.eu/ (last accessed on 02/05/2022).

[6] Pot, W., 2019. Anticipating the Future in Urban Water Management: an Assessment of Municipal Investment Decisions. Water Resources Management, 33(4), pp.1297-1313.

[7] Shamir, U. and Howard, C.D., 1979. An analytic approach to scheduling pipe replacement. Journal AWWA, 71(5), pp.248-258.

[8] Walski, T.M., 1982. Economic analysis of rehabilitation of water mains. Journal of the Water Resources Planning and Management Division, 108(3), pp.296-308.

[9] Walski, T.M. and Pelliccia, A., 1982. Economic analysis of water main breaks. Journal AWWA, 74(3), pp.140-147.

[10] Engelhardt, M.O., Skipworth, P.J., Savic, D.A., Saul, A.J. and Walters, G.A., 2000. Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. Urban Water, 2(2), pp.153-170.

[11] Mala-Jetmarova, H., Sultanova, N. and Savic, D., 2018. Lost in optimisation of water distribution systems? A literature review of system design. Water, 10(3), p.307.

[12] Tsiami, L. et al., 2022, A review of staged design of water distribution networks. 2nd WDSA/CCWI Joint Conference, Valencia, Spain.

[13] Cunha, M., Marques, J., Creaco, E. and Savić, D., 2019. A Dynamic Adaptive Approach for Water Distribution Network Design. Journal of Water Resources Planning and Management, 145(7), p.04019026.

[14] Lansey, K.E., Basnet, C., Mays, L.W. and Woodburn, J., 1992. Optimal maintenance scheduling for water distribution systems. Civil Engineering Systems, 9(3), pp.211-226.

[15] Basupi, I. and Kapelan, Z., 2015. Flexible water distribution system design under future demand uncertainty. Journal of Water Resources Planning and Management, 141(4), p.04014067.

[16] Creaco, E., Franchini, M. and Walski, T.M., 2014. Accounting for phasing of construction within the design of water distribution networks. Journal of Water Resources Planning and Management, 140(5), pp.598-606.

[17] Marques, J., Cunha, M. and Savić, D.A., 2015. Multi-objective optimization of water distribution systems based on a real options approach. Environmental Modelling & Software, 63, pp.1-13.

[18] Schnaars, S.P., 1989. Megamistakes. Collier Macmillan.

[19] Cantoni, M., Weyer, E., Li, Y., Ooi, S.K., Mareels, I. and Ryan, M., 2007. Control of large-scale irrigation networks. Proceedings of the IEEE, 95(1), pp.75-91.

[20] Savic, D.A., Kapelan, Z.S. and Jonkergouw, P.M., 2009. Quo vadis water distribution model calibration?. Urban Water Journal, 6(1), pp.3-22.

[21] Zierolf, M.L., Polycarpou, M.M. and Uber, J.G., 1998. Development and autocalibration of an input-output model of chlorine transport in drinking water distribution systems. IEEE Transactions on Control Systems Technology, 6(4), pp.543-553.

[22] Alvisi, S., Franchini, M. and Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. Journal of Hydroinformatics, 9(1), pp.39-50.

[23] Blokker, E.J.M., Vreeburg, J.H.G. and Van Dijk, J.C., 2009. Simulating residential water demand with a stochastic end-use model. Journal of Water Resources Planning and Management, 136(1), pp.19-26.

[24] Donkor, E.A., Mazzuchi, T.A., Soyer, R. and Alan Roberson, J., 2012. Urban water demand forecasting: review of methods and models. Journal of Water Resources Planning and Management, 140(2), pp.146-159.

[25] Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. and Faloutsos, C., 2008. Efficient sensor placement optimization for securing large water distribution networks. Journal of Water Resources Planning and Management, 134(6), pp.516-526.

[26] Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., Pierro, F. d., Khu, S.-T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y. & Walski, T., 2008. The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. Journal of Water Resources Planning and Management, 134(6), pp.556-568.

[27] Perelman, L., Arad, J., Housh, M. and Ostfeld, A., 2012. Event detection in water distribution systems from multivariate water quality time series. Environmental Science & Technology, 46(15), pp.8212-8219.

[28] Eliades, D.G., Stavrou, D., Vrachimis, S.G., Panayiotou, C.G. and Polycarpou, M.M., 2015. Contamination event detection using multi-level thresholds. In Proc. of Computing and Control for the Water Industry (CCWI), 119, pp.1429-1438.

[29] Afshar, A. and Najafi, E., 2014. Consequence management of chemical intrusion in water distribution networks under inexact scenarios. Journal of Hydroinformatics, 16(1), pp.178-188.

[30] Mounce, S.R., Mounce, R.B. and Boxall, J.B., 2011. Novelty detection for time series data analysis in water distribution systems using support vector machines. Journal of Hydroinformatics, 13(4), pp.672-686.

[31] Polycarpou, M.M., Uber, J.G., Wang, Z., Shang, F. and Brdys, M., 2002. Feedback control of water quality. IEEE Control Systems Magazine, 22(3), pp.68-87.

[32] Wang, Z., Polycarpou, M.M., Uber, J.G. and Shang, F., 2005. Adaptive control of water quality in water distribution networks. IEEE Transactions on Control Systems Technology, 14(1), pp.149-156.

[33] Wang, Y., Ocampo-Martinez, C. and Puig, V., 2016. Stochastic model predictive control based on Gaussian processes applied to drinking water networks. IET Control Theory & Applications, 10(8), pp.947-955.

[34] Creaco, E. and Franchini, M., 2013. A new algorithm for real-time pressure control in water distribution networks. Water Science and Technology: Water Supply, 13(4), pp.875-882.

[35] Wright, R., Stoianov, I., Parpas, P., Henderson, K. and King, J., 2014. Adaptive water distribution networks with dynamically reconfigurable topology. Journal of Hydroinformatics, 16(6), pp.1280-1301.

[36] Zhou, Y., Yu, Richard F., Chen, J., and Kuo, Y., 2020. Cyber-physical-social systems: a state-of-the-art survey, challenges and opportunities. IEEE Communications Surveys & Tutorials, 22(1), pp. 389-425.

[37] Vrachimis, S.G., Timotheou, S., Eliades, D.G., and Polycarpou, M., 2019. Iterative hydraulic interval state estimation for water distribution networks. *Journal of Water Resources Planning and Management*, 145(1), p. 04018087.

[38] Paluszczyszyn, D., Skworcow, P., & Ulanicki, B. (2015). Modelling and Simulation of Water Distribution Systems with Quantised State System Methods. *Procedia Engineering*, *119*, 554–563.

[39] Fan, X., Zhang, X., & Yu, X. (2021). Machine learning model and strategy for fast and accurate detection of leaks in water supply network. *Journal of Infrastructure Preservation and Resilience*, 2(1), 10.

[40] Elshaboury, N., & Marzouk, M. (2022). Prioritizing water distribution pipelines rehabilitation using machine learning algorithms. *Soft Computing*. doi:10.1007/s00500-022-06970-8

[41] Savić, D. ed., 2013. *Machine Learning in Water Systems: Proceedings of the AISB Convention 2013, Exeter, UK, 3.-5. April 2013*. University of Exeter.

[42] Zukang Hu, Beiqing Chen, Wenlong Chen, Debao Tan, Dingtao Shen; Review of model-based and data-driven approaches for leak detection and location in water distribution systems. *Water Supply* 1 November 2021; 21 (7): 3282–3306.

[43] Robles Velasco, A., Muñuzuri, J., Onieva, L., & Rodríguez Palero, M. (2021). Trends and applications of machine learning in water supply networks management. *Journal of Industrial Engineering and Management, 14*(1), 45-54.

[44] Candelieri A, Ponti A, Giordani I, Archetti F. Lost in Optimization of Water Distribution Systems: Better Call Bayes. *Water*. 2022; 14(5):800.

[45] Wu, Z. Y., & Rahman, A. (2017). Optimized Deep Learning Framework for Water Distribution Data-Driven Modeling. *Procedia Engineering*, *186*, 261–268.

[46] Qian, K, Jiang, J., Ding, Y., Yang, S. "Deep Learning Based Anomaly Detection in Water Distribution Systems," *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 2020, pp. 1-6.

[47] Hajgató Gergely, Paál György, & Gyires-Tóth Bálint. (2020). Deep Reinforcement Learning for Real-Time Optimization of Pumps in Water Distribution Systems. *Journal of Water Resources Planning and Management*, *146*(11), 04020079.

[48] Bragalli, C., Fortini, M., & Todini, E. (2017). Data Assimilation in Water Distribution Systems. *Procedia Engineering*, 186(C), 506–513.

[49] Xing Lu, & Sela Lina. (2022). Graph Neural Networks for State Estimation in Water Distribution Systems: Application of Supervised and Semisupervised Learning. *Journal of Water Resources Planning and Management,* 148(5), 04022018.

[50] Candelieri, A., Conti, D., & Archetti, F. (2014). A Graph based Analysis of Leak Localization in Urban Water Networks. *Procedia Engineering*, *70*, 228–237. doi:10.1016/j.proeng.2014.02.026

[51] Jakob, J., Artelt, A., Hasenjäger, M., & Hammer, B. (2022). SAM-kNN Regressor for Online Learning in Water Distribution Networks. *CoRR*. doi:10.48550/arXiv.2204.01436

[52] Losing, V., Hammer, B., & Wersing, H. (2018). Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 275, 1261–1274. doi:10.1016/j.neucom.2017.06.084

[53] Malialis, K., Panayiotou, C. G., & Polycarpou, M. M. (2021). Online Learning With Adaptive Rebalancing in Nonstationary Environments. *IEEE Trans. Neural Networks Learn. Syst.*, 32(10), 4445–4459. doi:10.1109/TNNLS.2020.3017863

[54] Ganji, A., & Miles, S.B. (2018). Toward Human-Centered Simulation Modeling for Critical Infrastructure Disaster Recovery Planning. *2018 IEEE Global Humanitarian Technology Conference (GHTC)*, 1-8.

[55] Christian, B. (2020). *The Alignment Problem*. W. W. Norton & Company, ISBN-10: 0393635821

[56] Molnar C. *Interpretable Machine Learning*. Leanpub, ISBN 978-0-244-76852-2.

[57] [Park J, Lee WH, Kim KT, Park CY, Lee S, Heo TY. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Sci Total Environ*. 2022 Apr 6;832:155070. doi: 10.1016/j.scitotenv.2022.155070. Epub ahead of print. PMID: 35398119

[58] Ayati Amir Houshang, Haghighi Ali, & Ghafouri Hamid Reza. (2022). Machine Learning–Assisted Model for Leak Detection in Water Distribution Networks Using Hydraulic Transient Flows. *Journal of Water Resources Planning and Management*, 148(2), 04021104. doi:10.1061/(ASCE)WR.1943-5452.0001508

[59] Artelt A, Vrachimis S, Eliades D, Polycarpou M, Hammer B. One Explanation to Rule them All — Ensemble Consistent Explanations. Bielefeld University 2022.

[60] Artelt, A., & Hammer, B. (2022). Efficient computation of counterfactual explanations and counterfactual metrics of prototype-based classifiers. *Neurocomputing*, 470, 304–317. doi:10.1016/j.neucom.2021.04.129

[61] Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. *Advances in neural information processing systems*, 2019.

[62] Zhang, K., Keliris, C., Polycarpou, M. M., & Parisini, T. (2021). Discrimination between replay attacks and sensor faults for cyber-physical systems via event-triggered communication. *Eur. J. Control*, 62, 47–56. doi:10.1016/j.ejcon.2021.06.026

[63] Mehrabi, N., Morstatter, F., Saxena, N.,Lerman, K. Galstyan, A. . 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv*. 54, 6, Article 115 (July 2022), 35 pages.

[64] Chakraborty, A., Alam, M., Dey, V., Chattopadhyay, A., & Mukhopadhyay, D. (2018). Adversarial Attacks and Defences: A Survey. *CoRR*. http://arxiv.org/abs/1810.00069

[65] Schröder, S., Schulz, A., Kenneweg, P., Feldhans, R., Hinder, F., & Hammer, B. (2022). The SAME score: Improved cosine based bias score for word embeddings. *CoRR*. doi:10.48550/arXiv.2203.14603

[66] Koundouri, P., N. Pittis, P. Samartzis, N. Englezos, A. Papandreou. (2021). Ambiguity Aversion, Modern Bayesianism and Small Worlds, *Open Research Europe*, Vol1, Issue 13, Pages 13, F1000 Research Limited.

[67] Pugno M. (2016). On The Foundations of Happiness in Economics. London: Routledge.

[68] Koundouri, P., 2009. 'The Use of Economic Valuation in Environmental Policy: Providing Research Support for the Implementation of EU Water Policy under AquaStress'. Routledge, Taylor and Francis Group. Series: Routledge Explorations in Environmental Economics. ISBN: 978-0-415-45323-3.

[69] Rittel, H.W. and Webber, M.M., 1974. Wicked problems. *Man-made Futures*, *26*(1), pp.272-280.

2^{nd} International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doe: https://doi.org/10.4995/WDSA-CCWI2022.2022.14516

A REVIEW ON STAGED DESIGN OF WATER DISTRIBUTION NETWORKS

Lydia Tsiami[1], Christos Makropoulos[2] and Dragan Savic[3]

[1,2,3] KWR Water Research Institute, Nieuwegein, (The Netherlands)
[1,2]Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Athens, (Greece)
[3]Centre for Water Systems, University of Exeter, Exeter, (United Kingdom)
[3]Faculty of Engineering and Built Environment, Universiti Kebangsaan, Malaysia, (Malaysia)

[1] *lydia.tsiami@kwrwater.nl*, [2] *cmakro@mail.ntua.gr*, [3] *dragan.savic@kwrwater.nl*

Abstract

Water distribution networks (WDNs) evolve continuously over time. Changes in water demands and pipe deterioration require construction upgrades to be performed on the network during its entire lifecycle. However, strategically planning WDNs, especially for the long term, is a challenging task. This is because parameters that are essential for the description of WDNs in the future, such as climate, population and demand transitions, are characterized by deep uncertainty. To cope with future uncertainty, and avoid overdesign or costly unplanned and reactive interventions, research is moving away from the static design of WDNs. Dynamic design approaches, aim to make water networks adaptive to changing conditions over long planning horizons. A promising, dynamic design approach is the staged design of WDNs, in which the planning horizon is divided into construction phases. This approach allows short-term interventions to be made, while simultaneously considering the expected long-term network growth outcomes. The aim of this paper is to summarize the current state of the art in staged design of water distribution networks. To achieve that, we critically examined relevant publications and classified them according to their shared key characteristics, such as the nature of the design problem (new or existing network design, expansion, strengthening, and rehabilitation), problem formulation (objective functions, length of planning horizon), optimization method, and uncertainty considerations. In the process, we discuss the latest findings in the literature, highlight the major contributions of staged design on water distribution networks, and suggest future research directions.

Keywords
Optimization, literature review, uncertainty, flexibility, robustness, design, long-term planning.

1 INTRODUCTION

Water distribution networks (WDNs) are complex infrastructures that have been developed to supply areas with water for large planning horizons and without interruptions. At the same time, WDNs are also dynamic in nature. During their lifetime, networks age as their pipes gradually deteriorate, leakages increase and their components fail. Moreover, urban development and demographic variations make demands placed upon the network increase. Consequently, construction interventions during the lifecycle of the network are necessary in order to cope with these increasing pressures.

Due to the high capital outlay of WDN infrastructures, the construction upgrades required are made under a limited budget, which is also provided gradually during the lifecycle of the network. Furthermore, the interventions are irreversible, in the sense that once they are implemented, they cannot easily be reversed. These interventions are also interdependent because they influence each other's performance [1]. For these reasons, how interventions are prioritized is critical, because it can affect the performance of the whole network in the longer term.

In theory, if decision makers knew how the network's forces of change evolve through time, they would be able to plan strategically the required interventions in a cost-efficient way, without compromising the performance of the WDN. However, most (if not all) critical forces of change such as urban development, population variations and consumer behaviour are difficult to forecast. This is because these forces are influenced by factors such as climate, socio-economic conditions, and technology, which are characterised by the so-called 'deep uncertainty' [2].

Traditionally, decision makers cope with future uncertainty by designing networks that work for a 'best guess' of future demand. However, this approach often leads to either overdesigned or underdesigned infrastructures, which require costly reactive interventions to align with actual requirements in the future. For that reason, the research community recently started moving away from static designs of WDNs and towards more dynamic approaches. One such approach is staged design, in which the planning horizon is divided into multiple construction phases.

In this work, we review the literature on staged design and how this methodological approach can be modified from its deterministic formulation in ways that incorporate future uncertainties into the design process and allow the development of robust and flexible designs.

2 STAGED DESIGN

To more formally define staged design, we start with the definition of the static, single-objective optimization problem. An equivalent formulation also holds for the more general many-objective optimization problem [3]. A single-objective optimization problem for the optimal design of a water distribution network can be defined as:

$$minimize\ or\ maximize\ f(x), \tag{1}$$

subject to:

$$a_i(x) = 0, \qquad i \in I = \{1, \dots, m\}, \qquad m \geq 0 \tag{2}$$

$$b_j\ (x) \geq\ 0, \qquad \mathrm{j} \in \mathrm{J} = \{1, \dots, \mathrm{n}\}, \qquad n \geq 0 \tag{3}$$

$$c_k(x) \leq 0, \qquad k \in K = \{1, \dots, p\}.\ p \geq 0 \tag{4}$$

where f is the objective function (usually a cost function), equations (2)-(4) are different types of constraints and x refers to the decision variables.

Analogously, staged optimization is the problem of identifying a sequence of actions that need to be taken over a number of N_{st} consecutive stages during the planning horizon, to maximize or minimize an overall objective function, subject to specific constraints at each stage:

$$minimize\ or\ maximize\ F[f(x_1), \dots, f(x_{N_{st}})] \tag{5}$$

subject to:

$$a_{i,s}\ (x_s) = 0, \qquad i \in I = \{1, \dots, m\}, \qquad m \geq 0, \qquad \forall\, s = [1, \dots, N_{st}] \tag{6}$$

$$b_{j,s}\ (x_s) \geq\ 0, \qquad \mathrm{j} \in \mathrm{J} = \{1, \dots, \mathrm{n}\}, \qquad n \geq 0, \qquad \forall\, s = [1, \dots, N_{st}] \tag{7}$$

$$c_{k,s}\ (x_s) \leq 0, \qquad k \in K = \{1, \dots, p\}, \qquad p \geq 0, \qquad \forall\, s = [1, \dots, N_{st}] \tag{8}$$

where f is an objective function calculated at each stage s of the planning horizon, F is some aggregation (sum, average etc.) function of the values of f calculated at each stage, and expresses the overall objective function to be optimized over the whole planning horizon. x_s are the decision variables at each stage s, and equations (6)-(8) are constraints at each stage of the optimization. The solution to the staged optimization problem can be expressed as $x = (x_1, \dots, x_S)$ i.e. a sequence of decisions at each stage of the planning horizon. The above definition can be extended to describe multi-objective staged optimization problems as well.

One observation that can be derived from this definition is that actions to be taken at each stage of the planning horizon cannot conflict with actions taken at previous stages. Therefore, staged optimization is not a series of individual optimization problems. It is a series of intercorrelated problems that aim at identifying an optimal sequence of solutions which are contiguous with one another. It should also be noted that a series of individual optimal solutions at each stage of the planning horizon does not guarantee that the overall solution is also optimal. On the contrary, sub-optimal solutions at certain stages might influence the choice of actions at later stages in such a way that, in the end, the overall sequence of decisions leads to the minimization/maximization of the overall objective function.

The main advantage of staged optimisation in the optimal design of WDNs is that it allows making decisions for the present, while simultaneously considering the expected long-term network growth outcomes. In the literature, staged optimisation has also been coupled with uncertainty considerations to develop either “robust” (under a range of scenarios) or flexible designs.

3 LITERATURE OVERVIEW

This literature overview is divided into three sections. In the first section, publications that solve classic staged optimization problems without consideration of uncertainty are presented. The second section is about methodologies that aim to develop robust staged designs, i.e. designs that work well under a range of scenarios. Finally, in the last section, we review publications that use staged optimization as a tool to develop flexible designs under uncertainty.

3.1 Deterministic staged optimization

Like static optimization approaches, staged optimization has also been used for the design and upgrade (strengthening, expansion, and rehabilitation) of WDNs. One of the first publications on staged optimization is Lekane et al. (1978) [4], where the long term design of a tree water network was approached as a multi-stage linear problem. In this publication, the authors assumed that the evolution of the consumption of the network was known for the whole planning horizon. Indeed, the assumption that the changes in the network that make construction interventions necessary are known (such as future demands and pipe deterioration), is a key characteristic of staged design problems (Figure 1). More specifically, in the literature, these changes can be either prespecified at each stage [5] or modelled to follow a specific function. For example, in [6] and [7] demands are assumed to grow linearly over time. Other “known” drivers of change include pipe break rates [8] and leakage [9], network expansion [6], energy cost [5], and how customer consumption changes in response to water tariff increases [10, 11].

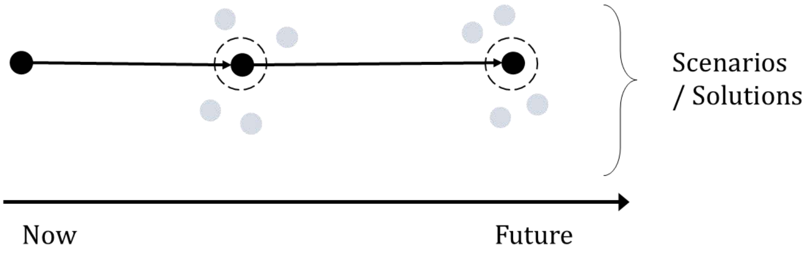


Figure 1: Deterministic staged design (adapted from Kang and Lansey [12]*). In deterministic staged optimization it is assumed that the drivers of changes such as future water demands and pipe deterioration are known.*

In staged design, the sequence of decisions that need to be made over the planning horizon is incorporated into the optimization process via the objective function. Single objective problems aim at the minimization of the accumulated cost at each stage of the planning horizon ([4, 5, 8, 10, 11, 13, 14]). For the optimization, the total cost is converted into total present cost, calculated using Present Value Analysis (PVA) concepts [15] and prespecified discount and interest rates. Multi-objective staged design problems account simultaneously for the total present cost and a second (conflicting) objective, which is also modified to account for the whole planning horizon, such as benefit [8], the volume of water lost [9], and reliability expressed as the minimum pressure surplus [6].

The optimization algorithms used in the literature to solve staged design problems include deterministic algorithms such as linear programming ([4, 10, 11, 14]), and generalized reduced gradient techniques [5]. Nature-inspired (heuristic) algorithms were also used, though mainly genetic algorithm variations ([6–9, 13]). However, given that a set of actions needs to be taken at each stage, the search space of the optimization algorithm increases exponentially, making the optimization process computationally expensive. With that in mind, Minaei et al. [7] proposed some modifications to improve computationally the algorithm proposed by Creaco et al. [6]. In their work, the modifications proposed included "efficient encoding of solutions based on practical considerations and engineering judgement, and engineering of populations". Tanyimboh and Kalungi also attempted to reduce the size of the optimization problem by using maximum entropy flows to generate optimal network designs [10, 11, 14].

The length of the planning horizon varies depending on the application area. Staged optimization problems that focus on rehabilitation ([5, 8–11, 13, 14]) considered shorter planning horizons (5-25 years). Design, expansion and strengthening problems considered longer planning horizons ranging from 20 to 100 years. Finally, the intervention time steps were usually prespecified without having necessarily equal lengths. Only Tanyimboh and Kalungi ([10, 11, 14]) explored various time step lengths using dynamic programming to further minimize the total cost.

Creaco et al. [6] and Lekane et al. [4] both compared static with staged designs and found that staged designs are more cost-effective in the long run. Creaco et al. showed that with staged design, the overdesign and underdesign of networks can be avoided, and noticed that to achieve long-term optimal solutions a higher initial investment is often required. Halhal et al. [8] demonstrated that the value of economic parameters such as inflation and interest rate influences whether higher investments will be allocated towards the beginning or end of the planning horizon. Finally, Minaei et al. [7] found that pipe roughness played an important role only in the optimization of lower-cost designs (smaller networks).

3.2 Taking account of uncertainty in staged optimization

Staged optimization aims to develop network design for the long term; hence some researchers recognize that the uncertainty that characterizes key input parameters cannot be ignored. As water consumption is a critical input parameter for the design of the network, several researchers choose to incorporate the uncertainty of water demand into the optimization process. One way to do that is to identify a set of plausible futures and attempt to deal with uncertainty by way of robustness, by finding a solution that satisfies all the generated scenarios (Figure 2).

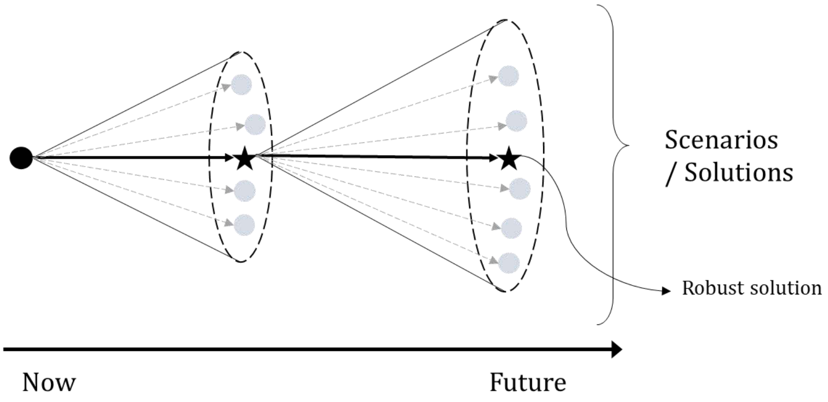


Figure 2: Robust staged optimization (adapted from Kang and Lansey [12]*). The goal of robust staged optimisation is to find a solution that works well enough for a range of scenarios.*

Just like static optimization is modified to accommodate solutions for staged design, staged design optimization is modified to account simultaneously for a set of plausible scenarios. For instance, Yi and Tiesong [16] considered that the demand follows a normal or uniform distribution with an increasing average and variance. Then 100 demand scenarios were generated, each one associated with a probability, and the goal was to minimize the expected total cost. Creaco et al. [17] modelled the growth rate of the demand as a discrete (low, medium and high growth rate) random variable with an assigned probability mass function. They solved a multi-objective problem that involved (i) the minimization of the present value cost of the solution implemented, and (ii) the minimization of the minimum pressure surplus observed over the lifecycle of the network and for each demand scenario. A similar route was followed by Dell'Aira et al. [18] who also considered the demand-growth rate as a discrete random variable, and generated multiple random scenarios to test the candidate solutions during the optimization. The solutions were evaluated by calculating the overall cost of the system and by averaging the resilience index over the whole planning horizon and over all the scenarios generated.

To avoid performing many simulations at each step of the optimization process, some researchers accounted for uncertainty after the optimization process was completed (or partially completed). Sirsant and Reddy [19] started by solving a deterministic optimization problem using life cycle costs as an objective function and a minimum value of resiliency as a constraint. Then they obtained the solution from the deterministic optimization, to run additional optimization iterations, but this time they replaced resiliency with reliability (which requires multiple simulations for its calculation). To calculate reliability, they generated scenarios by assuming that the demand is a random variable with mean the deterministic demand values at each stage and a coefficient of variation equal to 0.1. Sirsant and Reddy [20] also approached the same problem as a multi-objective problem where they minimized the cost, and maximized the resilience and then the reliability. After the optimization, the solutions of the Pareto front were ranked using an Analytical Hierarchy Process (AHA) using three criteria (i) costs, (ii) hydraulic and (iii) mechanical reliability. Marques and Cunha [21] and Cunha et al. [22, 23] also solved deterministic problems that were evaluated after the optimization under a range of scenarios. Marques and Cunha [21] generated a set of 200 equally probable scenarios, but only solved a deterministic design problem using the average demand scenario. Then they used multicriteria decision analysis to rank the performance of a number of generated alternative solutions under the 200 scenarios. Cunha et al. [22, 23] followed a similar approach by first identifying optimal solutions for a number of reference scenarios that covered a wide spectrum of possible conditions. Then they evaluated each solution under a range of scenarios using multicriteria decision analysis considering four criteria. The four criteria in [23] were cost, carbon emissions, resilience, and reliability, and the four criteria in [22] were cost, pressure deficits, velocity limits and supply deficits.

Most of the aforementioned methodologies concern the design, strengthening and expansion of the network, except for work in [18] and [21], which combined both the design and rehabilitation of the network. The optimization algorithms used include genetic algorithm variations [16–18], simulated annealing [21–23], and a hybrid differential evolution and dynamic programming [19, 20]. All networks used for the demonstration of the above methodologies have less than 30 nodes, except one real network with 914 nodes, which was used by Sirsant and Reddy [19]. Finally, the planning horizon ranges from 50 to 100 years. Only Yi and Tiesong [16] used a shorter 10-year planning horizon to demonstrate their methodology.

The literature gives useful insights regarding staged design under uncertainty. Creaco et al. [17] compared the design obtained from a staged deterministic approach with the design obtained from their methodology. They found that taking account of uncertainty in demand growth produces slightly oversized infrastructures (especially in the first phase of the construction) when future conditions are not known with certainty. Dell'Aira et al. [18] approached holistically the problem of design and rehabilitation and found that pipe ageing influences the optimal solution less than the growth of leakage. Sirsant and Reddy also arrived at a similar conclusion; they found that solutions that accounted for uncertain demands resulted in higher lifecycle costs [19] and that the break rate of pipes affected both the lifecycle costs as well as the estimated mechanical reliability of the network [20]. Finally, Cunha et al. [23] showed that the inclusion of carbon emissions as an optimal solution selection criterion favoured optimal designs that reinforced the network in the later phases of its planning horizon.

3.3 Staged optimization as a tool to obtain flexible designs

In the previous section, uncertainty is addressed by developing fixed staged solutions that perform well for as many scenarios as possible. In this section, we focus on work that attempted to provide flexible solutions allowing the water network development to become adaptive to several plausible future scenarios. To achieve that, both uncertainties and solutions are described by means of multi-stage scenario trees. Each branch in the scenario tree of uncertainties represents a plausible future that may or may not be connected to a probability of occurrence. Likewise, each branch in the solution tree represents a set of staged interventions to accommodate the corresponding (in the scenario tree) plausible future.

Assuming that all information about the network is known at the initial design stage, then the tree-like solution has a starting point, which is common for all future scenarios. The key in flexible design is to account simultaneously for 'all' different plausible futures and to identify a set of initial interventions that will allow the network to evolve to different future states in a way that requires few modifications for alternative future scenarios (Figure 3).

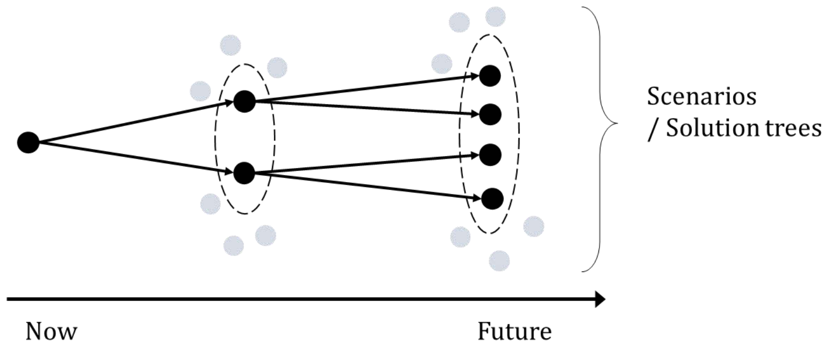


Figure 3: Flexible design optimization. The key in flexible design is to identify a set of initial interventions that will allow the network to evolve to different future states in a way that requires few modifications for alternative future scenarios.

In the literature, scenario trees are used to represent uncertainties in the demand [1, 24, 25] as well as in the spatial expansion of the network [26–29]. To account for uncertainty in the optimization process, Huang et al. [24] presented uncertainty as a scenario tree with 9 branches and optimized by minimizing the total present lifecycle cost over all the scenarios weighted by their probability of occurrence. Tsegaye et al. [26] incorporated uncertainties in a hypothetical network considering only increases in demands due to possible expansions of the network in new areas. They considered that all 4 scenarios were equally probable and optimized over their total cost.

Marques et al. published a series of papers where plausible scenarios were represented using Real Options [27–29]. In [27] the authors minimized the weighted (based on the probability of occurrence of each scenario) total lifecycle costs over all the scenarios and used in their objective a regret term to consider the differences between the cost of the flexible solution and the optimum cost for each scenario individually. Marques et al. [28] solved the problem by considering also carbon emissions in the objective function and in [29] they solve a many-objective optimization problem. The authors used four objectives that included (i) total pressure deficit (summed for each scenario, stage, and network node), (ii) total undelivered demand (summed for each scenario, time stage, and network node). (iii) total costs (iv) and total carbon emissions.

Basupi and Kapelan [1, 25] developed a decision tree solution, but instead of representing uncertainties in the demands in a tree-like form, they considered that demands follow a Gaussian probability density function with increasing mean and standard deviation over time. This means that branches in the solution tree do not represent implementations tailored to individual scenarios. Instead, a set of fine-tuned demand thresholds were used to trigger intervention decisions. Consequently, each solution pathway was robust for a range of scenarios. During optimization, a large number of samples were generated (each corresponding to one of the available intervention pathways) and the solution performance was calculated by averaging the lifecycle cost and the resilience index of the network across all samples.

To identify a flexible design that accommodates multiple scenarios, the optimisation algorithm has to consider an exponentially increased search space compared to the case of a fixed staged design problem. This is because different actions need to be taken not only at each stage of the planning horizon, but also under each individual scenario. In an attempt to tackle this problem, Kang and Lansey [12] found the optimal solution for each scenario individually and identified the common first-stage interventions across the solutions. Then, they assumed that these interventions will be implemented at the first stage of the planning horizon and excluded them from the set of decision variables to reduce the algorithm's search space.

The optimization algorithms used in almost all publications of this section were nature-inspired algorithms, such as genetic algorithm variations [12, 24–26] and simulated annealing [27–29].

Only Basupi and Kapelan in [1] generated flexible designs based on engineering judgement to demonstrate their methodology for evaluating flexibility in WDNs designs. All networks used to demonstrate methodological approaches in the reviewed literature had less than 20 nodes. The planning horizon varied between 10 and 100 years.

It was found that flexible designs have improved performance and that they are more cost-effective under uncertain conditions when compared with deterministic staged designs ([1, 12, 24–26]). Of course, for a perfectly known future, scenario-optimal solutions are less expensive than flexible ones [12, 25], but this extra cost "acts as an insurance policy" [25] when forecasts deviate significantly from reality. Marques et al. [27] found that flexible designs are more expensive in the initial phase than static designs created for the first construction period. However, when a larger horizon was considered, flexible designs cost less than solutions that ignored several future scenarios. In [28] the authors also found that the consideration of carbon emissions led to optimal designs with larger diameters. This is because larger diameters decreased the energy expenditure and therefore the cost of carbon emissions was reduced. Finally, Basupi and Kapelan [1, 25] found that their proposed flexible designs were more sensitive to the discount rate than demand uncertainty, hence they concluded that the discount rate is a parameter that needs to be carefully selected.

4 DISCUSSION

This literature overview showed that staged optimization of WDNs (in its broader sense) constitutes an improvement over traditional static approaches both in terms of lifecycle costs and overall network performance. This is because staged optimization allows to plan strategically the incremental development of WDN, therefore short-term interventions can be prioritized, without neglecting the expected long-term growth outcomes of the network. Research showed that staged designs that were slightly oversized in the beginning of the planning horizon were the ones that coped better with ("known") future changes. However, it is not possible to predict the future accurately, and as such, unexpected costs can still arise. Although additional costs at the beginning of the planning horizon are to be expected, staged design can be used as a tool to develop robust or flexible designs under uncertainty. These designs tend to be more expensive than the scenario-optimal solutions, but it has been argued that they are cost effective and perform better when a range of uncertainties is considered.

A major challenge in staged design is that the optimization process is computationally expensive. The reason is that a set of decisions needs to be identified for each stage of the planning horizon, and in the case of flexible designs, for each possible future scenario. In most publications, the proposed approaches were demonstrated using small network sizes, a limited number of intervention stages and plausible scenarios, while the decision variables focused primarily on pipe sizing alone. Although some methods to improve the computational efficiency of optimisation algorithms have shown promising results, there is still room for further research, for example in the direction of Surrogate Based Optimisation (SBO)[30].

Another challenge in staged optimization is that certain parameters require careful selection. For example, it was demonstrated that flexible design solutions were sensitive to the discount rate selection and that the discount rate also influenced how investments were allocated over the planning horizon. Moreover, it was shown that considerations of pipe roughness change were also important for small scale networks. Consequently, it is possible that the optimization process is influenced by other parameters as well, which have not been considered yet, such as water quality and short-term operations. So far, there is limited work on whether and to what extent different drivers of change influence the optimization process for the long term.

Most publications reviewed in the current work consider future uncertainties in demands and for network expansion. The different scenarios were generated stochastically, or were represented in a tree-like form, and in some cases, they were also connected to a probability of occurrence. However, demands, network expansions and intervention decisions are influenced by a range of highly uncertain parameters such as demographic changes, socioeconomic situation, technological developments, climate change, etc. This means that the robust and flexible approaches presented can still fail if reality differs from all hypothesized scenarios – as it often does. And even though robust and flexible approaches can, in principle, be revised at each phase of the planning horizon and include new information available at that phase, there is no work that provides a formal mechanism of how this can be achieved. The wider literature presents a range [31] of explicitly adaptive methods for decision making under deep uncertainty. Examples include dynamic adaptive planning [32] and dynamic adaptive policy pathways [33], and some of these methodologies have already been applied in the water sector. For instance, Beh et al. have successfully utilized a dynamic adaptive optimization approach to the problem of urban water supply augmentation [34]. With the appropriate modifications, these approaches or other similar ones might prove to be useful tools for the design of WDNs under deep uncertainty.

5 CONCLUSIONS

In this work, we have reviewed some of the most prominent publications on staged optimization of WDN. From deterministic staged designs to robust and flexible ones, this review focused on how the traditional static optimization approaches can be modified to incorporate sequential decision making and future uncertainty. We also discussed different optimization algorithms used in the literature, how different future scenarios were generated, the size of the networks used and the length of the planning horizon.

Our review demonstrates that staged optimization (in the broader sense) is a valuable tool for the generation of WDN designs that are better than traditional static designs both in terms of lifecycle costs and overall network performance under uncertainty. This is because staged optimization allows for the prioritization of short-term interventions without neglecting the expected long-term growth outcomes of the network. Further research is required to improve the computational efficiency of staged optimization algorithms, to investigate how different parameters (such as discount rate, water quality, and short-term operation) affect the long-term network performance, and to make the current flexible design methodologies more dynamic and adaptive to new emerging information.

6 AKNOWLEDGEMENTS

This paper is a result of the project Water-Futures which has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 951424).

7 REFERENCES

[1] I. Basupi and Z. Kapelan, “Evaluating Flexibility in Water Distribution System Design under Future Demand Uncertainty,” J. Infrastruct. Syst., vol. 21, no. 2, p. 04014034, 2015.

[2] W. E. Walker, R. J. Lempert, and J. H. Kwakkel, “Deep Uncertainty,” in Encyclopedia of Operations Research and Management Science, S. I. Gass and M. C. Fu, Eds. Boston, MA: Springer US, 2013, pp. 395–402.

[3] H. Mala-Jetmarova, N. Sultanova, and D. Savic, “Lost in optimisation of water distribution systems? A literature review of system design,” Water (Switzerland), vol. 10, no. 3, 2018.

[4] T. M. Lekane, D. E. Hellemans, and C. M. Whitwam, “Long-term optimization model of tree water networks,” Eur. J. Oper. Res., vol. 4, no. 1, pp. 7–15, 1980.

[5] K. E. Lansey, C. Basnet, L. W. Mays, and J. Woodburn, “Optimal maintenance scheduling for water distribution systems,” Civ. Eng. Syst., vol. 9, no. 3, pp. 211–226, 1992.

[6] E. Creaco, M. Franchini, and T. M. Walski, “Accounting for Phasing of Construction within the Design of Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 140, no. 5, pp. 598–606, May 2014.

[7] A. Minaei, A. M. Sabzkouhi, A. Haghighi, and E. Creaco, “Developments in Multi-Objective Dynamic Optimization Algorithm for Design of Water Distribution Mains,” Water Resour. Manag., vol. 34, no. 9, pp. 2699–2716, 2020.

[8] D. Halhal, G. A. Walters, D. A. Savic, and D. Ouazar, “Scheduling of water distribution system rehabilitation using structured messy genetic algorithms.,” Evol. Comput., vol. 7, no. 3, pp. 311–329, 1999.

[9] S. Alvisi and M. Franchini, “Multiobjective Optimization of Rehabilitation and Leakage Detection Scheduling in Water Distribution Systems,” J. Water Resour. Plan. Manag., vol. 135, no. 6, pp. 426–439, 2009.

[10] T. Tanyimboh and P. Kalungi, “Holistic planning methodology for long-term design and capacity expansion of water networks,” Water Sci. Technol. Water Supply, vol. 8, no. 4, pp. 481–488, 2008.

[11] T. T. Tanyimboh and P. Kalungi, “Multicriteria assessment of optimal design, rehabilitation and upgrading schemes for water distribution networks,” Civ. Eng. Environ. Syst., vol. 26, no. 2, pp. 117–140, 2009.

[12] D. Kang and K. Lansey, “Multiperiod Planning of Water Supply Infrastructure Based on Scenario Analysis,” J. Water Resour. Plan. Manag., vol. 140, no. 1, pp. 40–54, 2014.

[13] G. C. Dandy and M. Engelhardt, “Optimal Scheduling Of Water Pipe replacement,” Water Resour. Plan. Manag., vol. 127, no. 4, pp. 214–223, 2001.

[14] T. T. Tanyimboh and P. Kalungi, “Optimal long-term design, rehabilitation and upgrading of water distribution networks,” Eng. Optim., vol. 40, no. 7, pp. 637–654, 2008.

[15] R. de Neufville and S. Scholtes, Flexibility in Engineering Design. The MIT Press, 2011.

[16] Z. Yi and H. Tiesong, “Flexible design of delivery capacity in urban water distribution system,” Proc. - Int. Conf. Manag. Serv. Sci. MASS 2009, pp. 8–11, 2009.

[17] E. Creaco, M. Franchini, and T. M. Walski, “Taking Account of Uncertainty in Demand Growth When Phasing the Construction of a Water Distribution Network,” J. Water Resour. Plan. Manag., vol. 141, no. 2, p. 04014049, 2015.

[18] F. Dell’Aira, A. Cancelliere, E. Creaco, and G. Pezzinga, “Novel Comprehensive Approach for Phasing Design and Rehabilitation of Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 147, no. 3, p. 04021001, 2021.

[19] S. Sirsant and M. J. Reddy, “Optimal Design of Pipe Networks Accounting for Future Demands and Phased Expansion using Integrated Dynamic Programming and Differential Evolution Approach,” Water Resour. Manag., vol. 35, no. 4, pp. 1231–1250, 2021.

[20] S. Sirsant and M. Janga Reddy, “Performance-based multi-objective design and expansion of water distribution networks considering life cycle costs and future demands,” Water Supply, vol. 22, no. 2, pp. 1388–1408, 2022.

[21] J. Marques and M. Cunha, “Upgrading water distribution networks to work under uncertain conditions,” Water Sci. Technol. Water Supply, vol. 20, no. 3, pp. 878–888, 2020.

[22] M. Cunha, J. Marques, and D. Savić, “A Flexible Approach for the Reinforcement of Water Networks Using Multi-Criteria Decision Analysis,” Water Resour. Manag., vol. 34, no. 14, pp. 4469–4490, 2020.

[23] M. Cunha, J. Marques, E. Creaco, and D. Savić, “A Dynamic Adaptive Approach for Water Distribution Network Design,” J. Water Resour. Plan. Manag., vol. 145, no. 7, p. 04019026, 2019.

[24] D. Huang, K. Vairavamoorthy, and S. Tsegaye, “Flexible design of urban water distribution networks,” World Environ. Water Resour. Congr. 2010 Challenges Chang. - Proc. World Environ. Water Resour. Congr. 2010, pp. 4225–4236, 2010.

[25] I. Basupi and Z. Kapelan, “Flexible Water Distribution System Design under Future Demand Uncertainty,” J. Water Resour. Plan. Manag., vol. 141, no. 4, p. 04014067, 2015.

[26] S. Tsegaye, K. C. Gallagher, and T. M. Missimer, “Coping with future change: Optimal design of flexible water distribution systems,” Sustain. Cities Soc., vol. 61, no. June, 2020.

[27] J. Marques, M. Cunha, and D. Savić, “Using Real Options in the Optimal Design of Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 141, no. 2, p. 04014052, 2015.

[28] J. Marques, M. Cunha, and D. A. Savić, “Using real options for an eco-friendly design of water distribution systems,” J. Hydroinformatics, vol. 17, no. 1, pp. 20–35, 2015.

[29] J. Marques, M. Cunha, and D. Savić, “Many-objective optimization model for the flexible design of water distribution networks,” J. Environ. Manage., vol. 226, no. August, pp. 308–319, 2018.

[30] I. Tsoukalas, P. Kossieris, A. Efstratiadis, and C. Makropoulos, “Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget,” Environ. Model. Softw., vol. 77, pp. 122–142, 2016.

[31] K. Klima, “Decision Making Under Deep Uncertainty,” in The Oxford Handbook of Planning for Climate Change Hazards, W. T. Pfeffer, J. B. Smith, and K. L. Ebi, Eds. Oxford University Press, 2019.

[32] J. H. Kwakkel, W. E. Walker, and V. A. W. J. Marchau, “Adaptive Airport Strategic Planning,” Eur. J. Transp. Infrastruct. Res., vol. 10, no. 3, pp. 249–273, 2010.

[33] M. Haasnoot, J. H. Kwakkel, W. E. Walker, and J. ter Maat, “Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world,” Glob. Environ. Chang., vol. 23, no. 2, pp. 485–498, 2013.

[34] E. H. Y. Beh, H. R. Maier, and G. C. Dandy, “Adaptive, multiobjective optimal sequencing approach for urban water supply augmentation under deep uncertainty,” Water Resour. Res., vol. 51, no. 3, pp. 1529–1551, Mar. 2015.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14534

PERFORMANCE EVALUATION OF BURIED UPVC PIPES BY NUMERICAL SIMULATION OF SOIL-PIPE INTERACTION

Milad Latifi[1], Ramiz Beig Zali[2], Raziyeh Farmani[3], Akbar A. Javadi[4], Suhayl Zulfiquar[5]

[1] KTP researcher, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[2] PhD student, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[3] Professor, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[4] Professor, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[5] CEO, Datatecnics Corporation, Salford, Greater Manchester, UK.

[1] m.*latifi@exeter*.ac.uk, [2] *rb815@exeter*.ac.uk, [3] r.*farmani@exeter*.ac.uk, [4] a.a.*javadi@exeter*.ac.uk,
[5] sz@*datatecnics*.com.

Abstract

Pipelines are the most important and valuable assets in a water distribution network. Design, commissioning and maintenance of the pipes costs a huge amount of money and effort for water supply companies. Understanding the behaviour of buried pipelines is a major concern for engineers. In this paper, soil-pipe interaction has been investigated under various conditions using the numerical simulation of in-situ pipes.

A finite element model has been developed to numerically simulate the interaction of soil with pipes. More than 180 scenarios have been designed to cover the different conditions. Three segments of uPVC pipes with diameters between 160-450 mm were modelled. To cover a range of soils, three different soil types were modelled. The value of Young's modulus was decreased in various scenarios to consider the effects of age on the mechanical properties of pipe material. Four different external loads (corresponding to light to heavy vehicles) were exerted on top of the road pavement. Also, a large external load was applied to study the pipes failure under heavy loads. A variety of internal fluid pressures from 0 to 6 bars was applied inside the pipelines and the ultimate internal pressure which would cause failure in the pipe was determined.

The model domain was discretized for finite element calculations. Finer mesh sizes were selected around the pipe. An appropriate mesh size was determined by mesh sensitivity analysis. The Mohr-Coulomb failure criterion was chosen to simulate the soil behaviour. Also, an elastoplastic stress-strain relationship was considered for the uPVC material with von-Mises failure criterion.

Running the finite element model, the distribution of stresses and strains in the soil and the pipe were computed. In the absence of external load, the values of axial and hoop stresses were compared with the theoretical values. The comparison of the results showed a high correlation between the computed and theoretical values. In both absence and presence of external loads, the values of shear stresses are relatively small and negligible in comparison with axial and hoop stresses. As a result, the minimum and maximum principal stresses are close to the axial and hoop stresses, respectively. This implies that, only measuring the axial and hoop stresses can provide the principal stresses with a reasonable accuracy.

Both internal and external pressures in which the pipe fails are significantly higher than the usual loads on roads and water distribution networks. It justifies the fact that most of the pipe failures occur not due to exceeding the yield stress, but due to fatigue in elastic region. The results of numerical modelling of soil-pipe interaction help engineers to study the pipe performance under different conditions.

Keywords
Numerical simulation; Soil-pipe interaction; uPVC pipes; von-Mises stress.

1 INTRODUCTION

Nowadays, water distribution networks (WDNs) are invaluable parts of our life. Water pipelines are one of most expensive assets among urban infrastructures. Every year, failure in water pipelines costs a considerable amount for water utility companies. A study on 11 failures on main pipelines in United States with 762 – 3048 mm diameter shows a lost production between 6’800 and 10’000’000 m^3 and impact costs between 3.3 and 85.3 $ Million [1].

Better understanding of the behaviour of pipes under various loads could help the engineers to reduce the consequences of pipe failure. This knowledge may lead to more appropriate design of pipes. Analysing the performance of the pipes under loads could be carried out by physical models [2, 3, 4], analytical models [5, 6], or numerical models [7, 8]. In numerical models, by discretising the domain of the problem, the governing equations are solved for each element. A significant amount of research has been done to investigate the interaction of soil and pipe under various loading conditions using the finite element method [9, 10, 11].

Accuracy of numerical models depends on the mesh size of the model. The finer the mesh, the more accurate the model. In large models, millions of elements are needed to be generated to capture the small deformations of the pipe or soil, which makes these models very time-consuming.

Data-driven techniques have been developed to predict the behaviour of a system using observed data. Artificial Neural Network (ANN) is a widely used method for solving engineering problems. However, ANN suffers from the lack of a transparent relationship between inputs and outputs. Also, due to the black box structure of ANN, the model constants and coefficients are not easy to modify. Genetic Programming (GP) is another evolutionary modelling approach, which inspires from natural selection to fit a mathematical expression to a set of observed data. Although it is very popular in solving engineering problems, the drawbacks are weakness in finding constants and producing appropriate functions [12]. Evolutionary Polynomial Regression (EPR) is one of the machine learning methods which combines the capabilities of genetic programming and polynomial regression to find the coefficients of generated polynomial functions. Comparing the performance of different polynomial expressions, user can find the best relationship to describe the dataset.

In this research, EPR has been used to predict the behaviour of soil-pipe system. Various FE models of soil-pipe system with different conditions were generated in ABAQUS software. By simulating the developed scenarios, a dataset was created representing the interaction of soil and pipe. The dataset was used to train an EPR model in order to find simplified relationships between soil-pipe system properties and stress and strain on pipe wall. The results of this research show that data-driven models are capable of accurately predicting the mechanical behaviour of soil-pipe system at significantly reduced computational time.

2 METHODOLOGY

The present research is divided into two stages: the first is a FE analysis of the soil-pipe system, and the second is the use of EPR to capture the mechanical behaviour of PVC pipes.

2.1 Finite Element Analysis of soil-pipe system

In real world, there are various factors that affect the mechanical behaviour of PVC pipes in water distribution networks during their life time, e.g. internal hydraulic pressure, external load on the soil surface, pipe’s diameter, thickness, and Young's modulus, density, depth, friction angle, and Young's modulus of soil. Internal pressure, external load, pipe’s diameter and thickness, and Youngs modulus of soil were regarded as key variables impacting stress and strain on pipe walls in this study, while the other variables were held constant. The values of the parameters used in the model are shown in Table 1.

Table 1. Values of independent variables in generated scenarios for soil-pipe numerical modelling.

Internal Pressure (bar)	External Load (ton)	Young Modulus of soils (MPa)	Diameter of pipes (mm)	Thickness of pipes (mm)
0	Normal Sedan	15	160	6.2
2	light goods trucks	40	630	24.1
3	Six axel artic lorries	120	1200	29.4
4	Two axel heavy lorries			
5				

In this study, a range of conditions were employed in which the investigated variables were altered to create 180 scenarios for finite element simulations.

For numerical modelling, ABAQUS was used. ABAQUS is a well-known commercial software for finite element analysis of engineering problems.

Three different diameters of PVC pipe (160 mm, 630 mm, and 1200 mm) were used to create a model of a soil pipe system, which covers a wide range of PVC pipes. The pipe thickness was chosen based on the usage of PN10 PVC pipes in water distribution networks. The pipes are in a soil trench with width and depth of 5 × 5 metres, and a trench length equal to the length of the pipe. The pipe and soil components were chosen as solid homogenous, with cell partitions based on the model geometry for burying pipe at a depth of 1 metre of trench.

In ABAQUS, an eight-node continuum element (C3D8R) was used to better imitate the behaviour of plastic pipe. For modelling the homogeneous and isotropic soil around the pipe, a twenty-node quadratic brick element with reduced integration (C3D20R) was adopted. The optimal mesh size for the pipe and soil was determined through a mesh sensitivity analysis. As seen in Fig. 1, the mesh becomes finer closer to the pipe and coarser further away. In addition, depending on its thickness, the pipe wall was modelled in one, two, or three layers.

Properties of PVC pipes were taken from [13], in which 16 PVC pipes with four different diameters and four different thicknesses were investigated. They used tensile tests on dog-bone samples to produce a standard stress-strain curve for PVC pipes. The stress-strain curve employed in this study is shown in Fig. 2. The yield stress of PVC pipes is 43.67 MPa, which is substantially greater than the usual stresses in water distribution pipelines, as illustrated in this diagram.

On the soil-pipe system, two primary loads were evaluated, as previously stated. The maximum internal pressure in the urban water distribution network is limited at 5 bar. Internal pressures of 0, 2, 3, 4, and 5 bars were applied in the pipes in various conditions. Different external loads were applied on top of the soil trench, taking into account ordinary urban vehicles such as standard sedans, light goods trucks, six axel artic lorries and two axel heavy lorries.

A Mohr-Coulomb model was used to describe the mechanical behaviour of the soil. The Young's modulus of soil was altered in different FE simulation scenarios to study the effects of soil material type on pipes. The Mohr-Coulomb yield criterion is described as [14]:

$$\tau = C - \sigma \tan\varphi \tag{1}$$

where σ and τ are the effective normal stress and shear stress; C is the cohesion of soil; and φ is the effective angle of shearing resistance. The mechanical properties of considered soils are presented in Table 2.

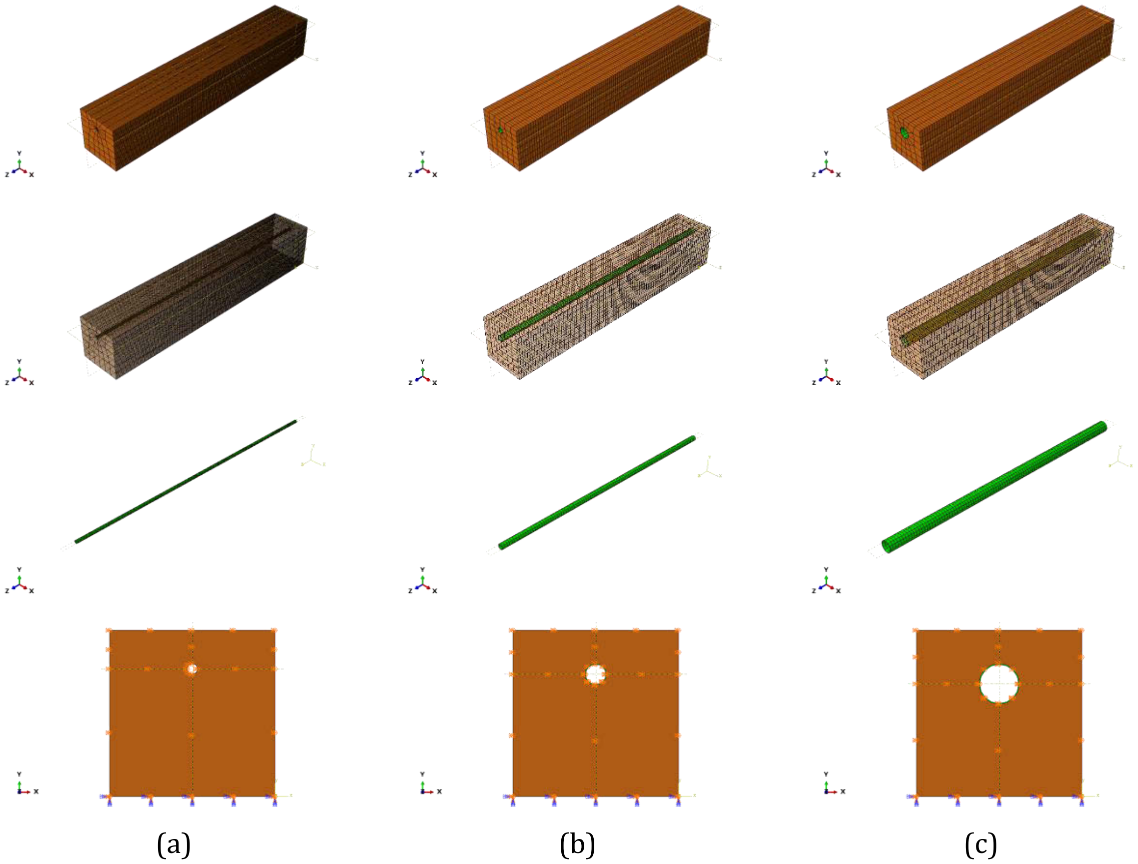

(a) (b) (c)

Fig. 1. Soil-pipe system geometry for (a) 160mm; (b) 630mm; and (c) 1200mm; pipes.

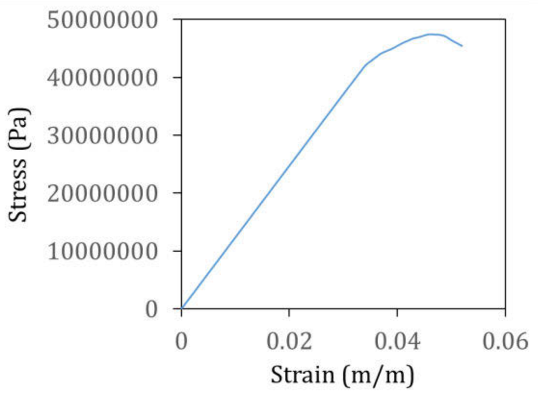


Fig. 2. Stress-strain curve used for characterizing the PVC material [13].

BY NC SA

Table 2. Mechanical properties of soil.

Elastic properties	Density (Kg/m^3): 1950 Young's modulus (MPa): 15, 40, 120 Poisson ratio: 0.3
Plastic properties	Cohesion (KPa): 15 Angle of shearing resistance (deg): 29 Dilation angle (deg): 2

Contact modelling is essential for accurately simulating the interaction between the soil and the pipe. In the 3D soil pipe interaction problems, contact between soil and underground pipe brings significant non-linearity. The "surface to surface" contact is employed in ABAQUS/Standard version 2021, with the pipe outer surface and the soil interior surface designated as "master" and "slave," respectively. The "penalty" technique approximates a "hard" contact between two surfaces in a non-linear way.

In this study, an adequately long section of underground pipe was modelled. Because the underground pipe can move with the soil, it is appropriate to treat the pipe ends as roller boundary. The displacement/rotation degrees of freedom were fixed (U1=U2=U3=UR1=UR2=UR3=0) for trench bottom boundaries; the bottom section of the soil was regarded fixed, and pipe rings and soil were selected as displacement/rotation (U1=U3=0) for side walls of trench and pipe ends.

All of the scenarios were analysed with the above assumptions and consideration. For the crown, springlines, and invert position in the middle of the pipes, axial stress, hoop stress, von-Mises stress, axial strain, and hoop strain were determined. The mechanical behaviour of the PVC pipes was captured using these data under a variety of internal and exterior loads.

2.2 Evolutionary Polynomial Regression (EPR)

EPR is a hybrid approach for fitting an equation to a set of observation data. It has been utilised to identify the link between dependent and independent variables in numerous civil engineering applications [12, 15]. To identify the optimum regression, EPR employs a two-stage technique: 1) developing equations that best match the data, and 2) applying a genetic algorithm (GA) to determine the coefficients and exponents of the equations. EPR's general expression can be shown as:

$$y = \sum_{j}^{m} F\left(X, f\left(x\right), a_j\right) + a_0 \tag{2}$$

where y is the system's estimated output; a_j is a constant value; F is a function built by the process; X is the matrix of input variables; and f is a user-defined function. The number of terms in the expression excluding bias a_0 is m. EPR is able to consider different functions, e.g., polynomial, sinusoid, logarithmic, etc., among which polynomial is the simplest form.

In most of the cases, GA uses least square (LS) of difference between observations and predictions to determine the best coefficients and exponents. Also, coefficient of determination (CoD) is used to select the equations with better correlation.

In this research, many factors affect the stresses and strains on the pipe wall:

$$y = f(W, P, D, t, E_s, E_p, H, \gamma_s, \varphi, \ldots) \tag{3}$$

where P is internal pressure of the pipe; W is external load on the soil surface; D and t are pipe diameter and thickness, respectively; E_s and E_p are soil and pipe Young's modulus, respectively; H is soil depth on the pipe crown; γ_s is density of soil; and φ is angle of shearing resistance of soil. Three main factors mainly affect the stresses and strains in pipes: internal pressure, external load, and soil weight. Nevertheless, using the superposition principle, the above equation could be written as follows:

$$y = f_1(W, D, t, E_s, H, \gamma_s, \varphi) + f_2(P, D, t, E_p) + f_3(H, \gamma_s, D, t) \tag{4}$$

Functions f_1 , f_2 and f_3 indicate the effect of external load, internal pressure and soil weight, respectively. Here, y is stress/strain on cardinal points of pipe. To simplify the problem, only W , P , D , t and E_s were considered as the variables, while other factors were treated as constant. On the other hand, assuming a polynomial regression, Eq. 3 could be reduced to:

$$y = a_1 W^{m_1} {E_s}^{m_2} D^{m_3} t^{m_4} + a_2 P^{n_1} D^{n_2} t^{n_3} + a_3 \tag{5}$$

where a_1 - a_3 are coefficients; and m_1 - m_3 and n_1 - n_3 are exponents. In this work, the EPR was used to determine the exponents between -2 and 2 with steps of 0.5. Since all the soil weight factors are constant, in has been shown as a bias (a_3).

70% of data was used to train the EPR model and the remaining 30% was used to validate the developed model. Ingesting the data from FE analysis to EPR model, it generates a large number of equations with different structures, including different combinations of independent variables and various CoDs. To select the best equation describing the soil-pipe mechanical behaviour, these criteria were considered:

1) Highest amount of CoD,
2) Model consisting of the minimum number of terms (parsimony principle),
3) Model including the maximum number of independent variables,
4) Capability of the model to respond properly in extreme conditions. For example, if P and W are being multiplied in a single term of the equation, when each of them is equal to zero, the whole term could be equal to zero, regardless of the amount of the other variable.

Considering the above criteria, the best equations were selected to calculate the axial stress, hoop stress, von-Mises stress, axial strain and hoop strain in crown, springlines and invert of the pipe. The results are presented and discussed in the next section.

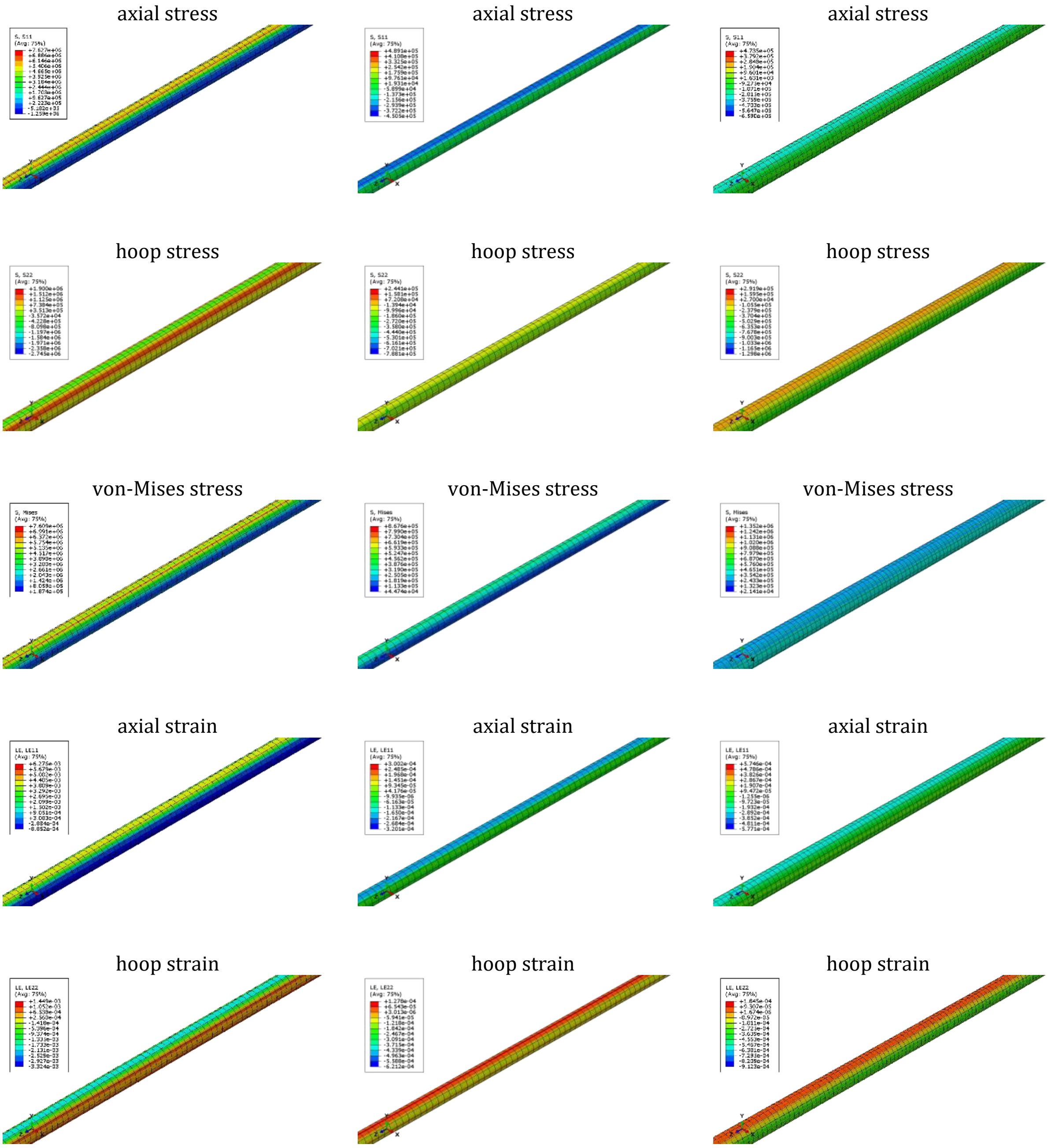
3 RESULTS

The values of axial stress, hoop stress, von-Mises stress, axial strain and hoop strain were computed by the FE models for each pipe in all scenarios. Fig. 3 presents the distribution of these variables in pipes with 160mm, 630mm and 1200mm diameter. The findings revealed that all of the pipes are long enough so that the boundary conditions do not affect their middle section.

The results demonstrate that for 1200mm pipes, von-Mises stress is high in the springlines, the maximum amount of axial strain is in the springlines, and hoop strain is high in the invert and crown sections, as shown in the Fig. 3. In 630mm pipes, the crown and invert sections have large quantities of axial stress, axial strain, hoop stress, and hoop strain, while the springlines have significant von-Mises stress. In 160mm Pipes, the pipe's shoulder and haunch sections have the

BY NC SA

highest hoop stress and hoop strain, whereas the crown and invert sections have the highest von-Mises stress as well as axial stress and axial strain.



(a) (b) (c)

Fig. 3. Distribution of axial stress, hoop stress, von-Mises stress, axial strain and hoop strain for pipes in the loading scenarios with maximum internal pressure for pipes with (a) 160mm; (b) 630mm; and (c) 1200mm diameter.

Using the EPR model, a number of equations were developed for calculating each of the dependent variables. In each case, the best equation was selected considering the criterion mentioned in Methodology section. These equations are presented in Table 3. In most of the cases, the CoD is higher than 98%, indicating a high level of correlation between prediction model and observation data.

Table 3. The best equations generated by EPR for calculating each dependent variable.

Dependent variable	Position	Equation	CoD (%)
Axial Stress	Crown	$\sigma_a = 6.0968\times10^{-12} \frac{W^{0.5} \cdot {E_s}^2}{t^2} - 1.3090\times10^3 \frac{P}{D \cdot t}$	98.64
	Springlines	$\sigma_a = 2.7545\times10^8 \frac{W^{0.5}}{D^2 \cdot t^2} - 3.1592\times10^4 \frac{P}{D^2} - 6.2909\times10^4$	98.49
	Invert	$\sigma_a = 6.2824\times10^8 \frac{W}{D^{1.5} \cdot t^2 \cdot {E_s}^{0.5}} - 2.5167\times10^4 \frac{P}{D^2} - 1.1791\times10^4$	98.16
Hoop Stress	Crown	$\sigma_h = -1.2204\times10^7 \frac{W}{D^2 \cdot t^2} + 1.6963\times10^6 \frac{P}{D^{1.5} \cdot t^2} - 2.6147\times10^5$	99.47
	Springlines	$\sigma_h = -2.8343\times10^7 \frac{W}{D^2 \cdot t^2} + 4.9327\times10^4 \frac{P}{D \cdot t^{1.5}} - 4.7687\times10^5$	98.95
	Invert	$\sigma_h = -1.9357\times10^9 \frac{W}{D^2 \cdot {E_s}^{0.5}} + 3.3832\times10^6 \frac{P}{D^2 \cdot t} - 3.1034\times10^5$	98.91
von-Mises Stress	Crown	$\sigma_{Mises} = -5.8982\times10^4 t - 0.3W - 0.35\frac{P}{D^2} + 6.9148\times10^2 \frac{P^{1.5}}{D^2}$ $+1.4096\times10^6$	97.88
	Springlines	$\sigma_{Mises} = 6.63819\times10^5 \frac{W}{D^2} - 0.00405\frac{P}{t^2} + 1.0087\frac{P^{1.5}}{t^2} + 1.46407\times10^5$	96.84
	Invert	$\sigma_{Mises} = 1.74507\times10^5 \frac{W^{0.5}}{t^2} - 7.1972\times10^2 \frac{P^{1.5}}{D^2} + 1.44697\times10^5$	98.19
Axial Strain	Crown	$\varepsilon_a = 3.6346\times10^{-18} \frac{W^{1.5} \cdot E_s}{t^{1.5}} - 0.00024375\frac{P}{D^{1.5} \cdot t^{1.5}}$	97.90
	Springlines	$\varepsilon_a = 2.1318\times10^{-18} \frac{W^{1.5} \cdot {E_s}^{1.5}}{D^2} - 1.6404\times10^{-6} \frac{P^{1.5}}{D^2 \cdot t}$	90.52
	Invert	$\varepsilon_a = 31.209\frac{W}{D^2 \cdot t^2 \cdot {E_s}^{0.5}} - 0.00312\frac{P}{D^2 \cdot t^{1.5}} + 9.6095\times10^{-5}$	99.08
Hoop Strain	Crown	$\varepsilon_h = -9.6493\times10^{-3} \frac{W}{D^2 \cdot t^2} + 1.3546\times10^{-3} \frac{P}{D^{1.5} \cdot t^2} - 1.9574\times10^{-4}$	99.47
	Springlines	$\varepsilon_h = -1.137\times10^{-17} \frac{W^2 \cdot E_s}{D^{1.5} \cdot t^{0.5}} + 5.4566\times10^{-7} \frac{P^{1.5}}{D^2}$	98.97
	Invert	$\varepsilon_h = -4.2125\times10^{-6} \frac{W^{1.5}}{t^2 \cdot {E_s}^{0.5}} - 5.5012\times10^{-4} \frac{P^{0.5}}{D} + 5.5579\frac{P}{D^2}$ -8.7481×10^{-5}	99.05

As indicated in Table 3, in most of the cases, a two-term equation (with a bias term) is achieved. High CoD values indicate the suitability of Eq. 5 for predicting the dependent variables using internal pressure, external load, pipe diameter, pipe thickness and soil Young's modulus.

Fig. 4 shows the performance of developed prediction models to forecast the axial stress, hoop stress and von-Mises stress in different positions of a pipe, (crown, springlines and invert). The

BY NC SA

figure shows the correlation for both training and testing data. It is shown that the model is able to predict the unseen testing data with the same (high) accuracy of training data. In addition, the capability of the model in prediction in crown and invert is higher than that of springlines.

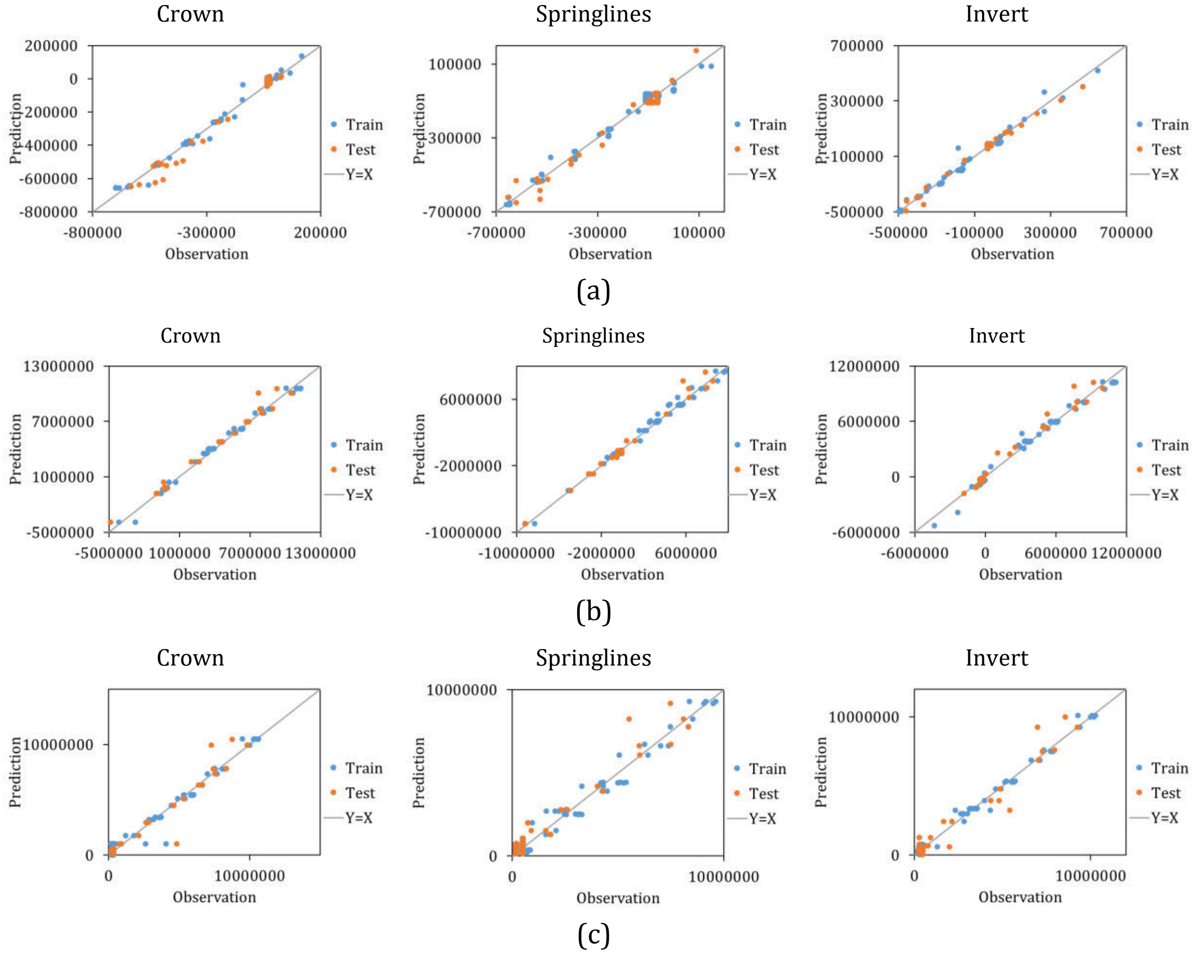
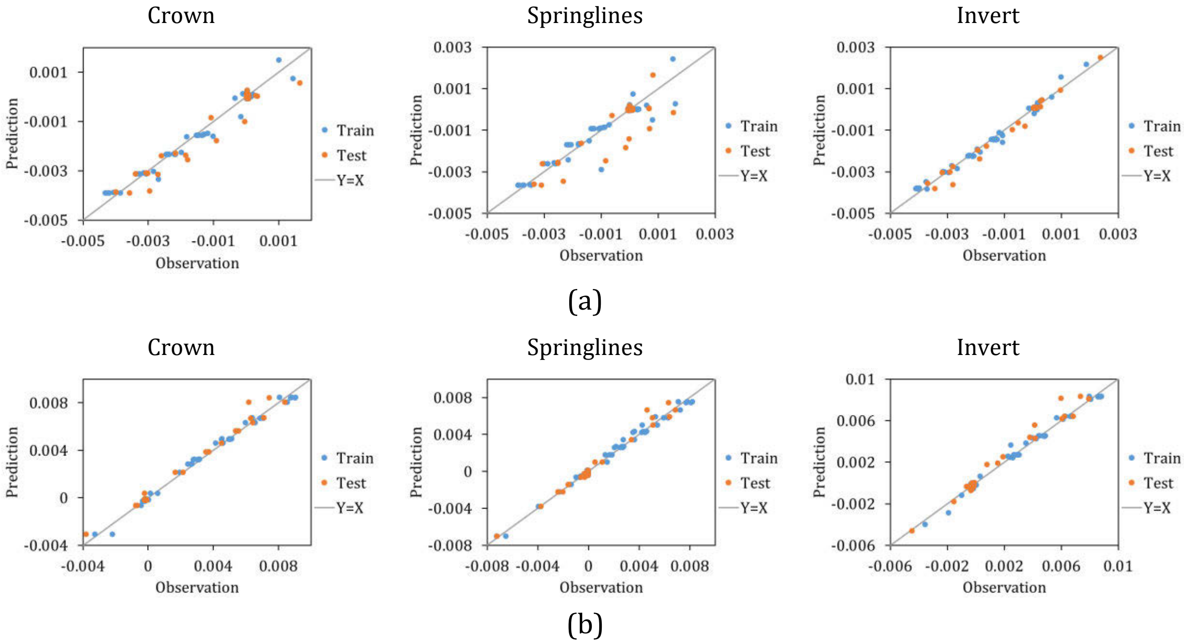


Fig. 4. Performance of prediction model to forecast observation data for (a) axial stress; (b) hoop stress; and (c) von-Mises stress.

Fig. 5 shows the capability of the proposed models to predict the axial and hoop strains in crown, springlines and invert of the pipe. Similar to the stress, the models are better in predicting the strains in crown and invert, than springlines. Also, model proposed for hoop strain has a better correlation than that of axial strain.

Fig. 6 presents the amounts of hoop stress-strain curves for crown, springlines and invert of different pipes. As mentioned earlier, the common internal pressures and external loads on soil-pipe system have been applied in the FE model, therefore, the achieved stress-strain values only cover a part of the entire curve (left) and failure has not occurred in any of the cases.

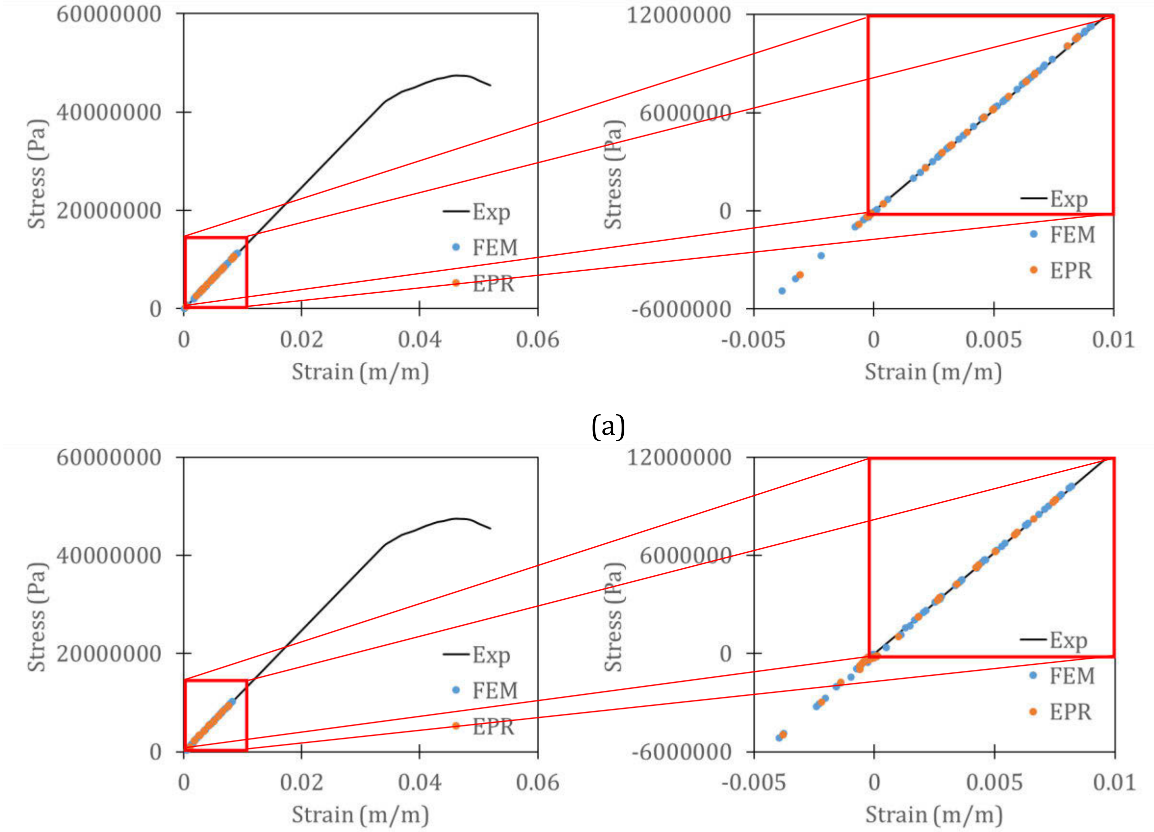
In this figure, the stress-strain curve presented by [13], which was used as a representative of PVC pipes mechanical behaviour, is shown in black lines. As it shown, the results of EPR model are perfectly fitted on the original stress-strain curve of [13], which indicates the capability of the EPR model to simulate the mechanical behaviour of PVC material using simple variables such as internal and external load and pipe's diameter and thickness.



(a)

(b)

Fig. 5. Performance of prediction model to forecast observation data for (a) axial strain; and (b) hoop strain.

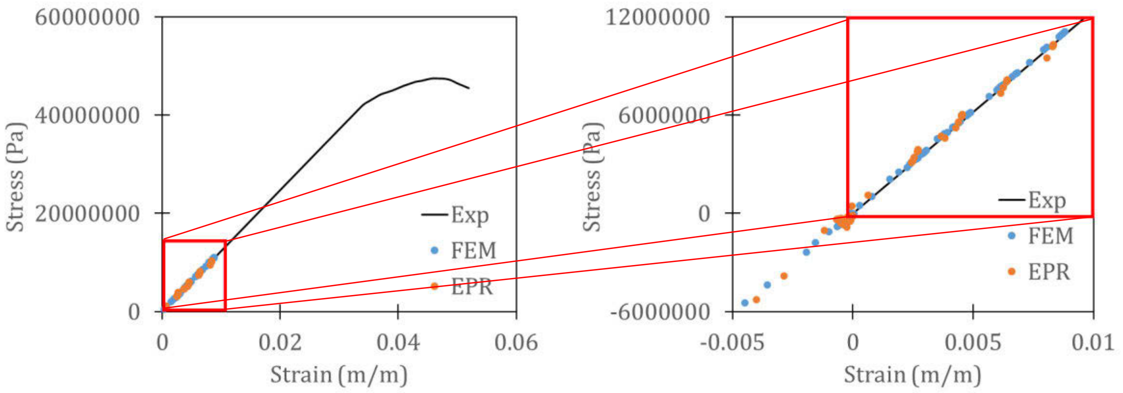


(a)

(b)

Fig. 6. Hoop stress-strain curves for (a) crown; (b) springlines; and (c) invert; of the pipes.

BY NC SA



(c)

Fig. 6 (continue). Hoop stress-strain curves for (a) crown; (b) springlines; and (c) invert; of the pipes.

4 DISCUSSION

In this research, PVC pipes with adjusted properties and various diameters and thicknesses were simulated using finite element method with Mohr-Coulomb model for soil. 180 scenarios were constructed with different soil Young's modulus and a variety of loading conditions such as internal pressure and external load. Axial stress, axial strain, von-Mises stress, hoop stress, and hoop strain were collected from three critical positions of the pipe; crown, springlines, and invert. The results of FE modelling were ingested to an EPR model to extract appropriate regression equations for predicting stresses and strains on pipe wall using independent variables.

The constructed relationships are supported by physical principles to consider the main loads on soil-pipe system. In all of the achieved equations, external load and internal pressure are showing opposite trends. This is expected, due to the fact that these two forces act in opposite directions.

The results indicate that EPR is able to find relationships between independent variables and dependent variables with high degree of accuracy. The proposed equations are actually a white-box model to predict the mechanical behaviour of pipes under different combinations of loads. Using these equations is much less time-consuming in comparison with running a finite element model. The equations are developed for pipes from 160mm up to 1200mm, which means they could be used for a wide range of pipe diameters.

It is proposed to consider other factors such as soil depth, soil density and different pipe materials, in the future investigations, in order to develop a more comprehensive model.

5 REFERENCES

[1] Piratla, K.R., Yerri, S.R., Yazdekhasti, S., Cho, J., Koo, D. and Matthews, J.C., "Empirical Analysis of Water-Main Failure Consequences", In Proc. International Conference on Sustainable Design, Engineering and Construction, Chicago, U.S., Procedia Engineering, 118, 2015, 727 – 734.

[2] Urbanowicz, K., Bergant, A., Kodura, A., Kubrak, M., Malesinska, A., Bury, P. and Stosiak, M., "Modeling Transient Pipe Flow in Plastic Pipes with Modified Discrete Bubble Cavitation Model". Energies, 2021, 14, 6756. DOI: 10.3390/en14206756.

[3] Kodura, A. and Weinerowska, K., "Some Aspects of Physical and Numerical Modeling of Water Hammer in Pipelines", In Proc. International Symposium on Water Management and Hydraulic Engineering, Ottenstein, Austria, 2005.

[4] Hariri Asli, K., Bakhman, F., Naghiyev, O. and Khodaparast Haghi, A., "Some Aspects of Physical and Numerical Modeling of Water Hammer in Pipelines", Nonlinear Dynamics, vol. 60, 2010, pp. 677–701. DOI: 10.1007/s11071-009-9624-7.

[5] Razvarz, S., Jafari, R. and Vargas-Jarillo, C., “Modelling and Analysis of Flow Rate and Pressure Head in Pipelines” In Proc. 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico. 2019.

[6] Wang, W. and Chen, G. “Analytical and Numerical Modeling for Flexible Pipes”, China Ocean Engineering, vol. 25, 2011, pp. 737–746.

[7] Dey, S., Chakraborty, S., Tesfamariam, S. “Structural performance of buried pipeline undergoing strike-slip fault rupture in 3D using a non-linear sand model”, Soil Dynamics and Earthquake Engineering, vol. 135, 2020, pp. 106180.

[8] Mosadegh, A., Nikraz, H., “Finite Element Analyses of Buried Pipeline Subjected to Live Load Using ABAQUS”, In Proc. GeoQuebec 2015, Quebec, Canada.

[9] Neya, B., Ardeshir, M., Delavar, A. and Bakhsh, M., “Three-Dimensional Analysis of Buried Steel Pipes under Moving Loads” Open Journal of Geology, vol. 7, 2017, pp. 1-11. DOI: 10.4236/ojg.2017.71001.

[10] Nair, P. and Nair, S., “Stress Analysis of Buried Pipelines”, International Research Journal of Engineering and Technology, vol. 4, no. 6, 2017, pp. 846-850.

[11] Alamatian, E., Ghadamkheir, M. and Karimpour, B., “Stress Estimation on Pipeline and Effect of Burying Depth”, International Research Journal of Applied and Basic Sciences, vol. 6, no. 2, 2013, pp. 228-235.

[12] Giustolisi, O., and Savic, D., “A symbolic data-driven technique based on evolutionary polynomial regression”, Journal of Hydroinformatics, vol. 8, no. 3, 2006, pp. 207-222.

[13] Yang, J. and Hu, S., “Estimation of Burst Pressure of PVC Pipe Using Average Shear Stress Yield Criterion: Experimental and Numerical Studies”, Applied Sciences, vol. 11, no. 21, 2021, 10477.

[14] Bridge G., Ozkaynak B. and Turhan, E., “Energy infrastructure and the fate of the nation: introduction to special issue”, Energy Res. Soc. Sci., vol. 41, 2018, pp. 1–11.

[15] Giustolisi, O. and Savic, D., “Advances in Data-driven Analysis and Modelling Using EPR-MOGA”, Journal of Hydroinformatics, vol. 11, no. 3-4, 2009, pp. 225-236.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14560

LEAK LOCALISATION METHOD USING A DETAILED HYDRAULIC MODEL COMBINED WITH HIGH RESOLUTION PRESSURE SENSORS APPLIED TO A REAL NETWORK

Grau S[1]., Pérez R.[2], Torret X.[3] and Casado D.[4]
[1] Innovació, Aigües de Manresa, S.A., Spain
[1,2,4] cs2ac, Universitat Politècnica de Catalunya, Spain
[3] BGEO, Spain

[1] *sgrau@aiguesmanresa.cat*, [2] *ramon.perez@upc.edu*, [3] *xtorret@bgeo.es*, [4] *davidcaru99@gmail.com*

Abstract

The objective has been the elaboration of a method of locating specific and background leaks in hydraulic models. We have worked with high-resolution hydraulic models with high resolution pressure sensors. The method has been applied to a real network. To execute an agile adjustment of the models, it is essential to link the hydraulic model with the data of the SCADA. The tool developed by the GISWATER ASSOCIATION, QGIS on POSGRESQL allows exporting networks to EPANET with the demanded resolution. This Work has used three levels of definition regarding the nodes introduced. As instrumentation, three pressure sensors have been used with a resolution of 1 centimetre. The situation of sensors in the network has been defined by using algorithms based on the sensitivity matrix.

Background leakage has been modelled using EPANET emitter coefficients. The value of these coefficients has been calculated based on the overall performance of the sector. They have been calculated individually for each node weighted by the lengths of the pipes linking to each node. The combination of a network with a high density of nodes, a distribution of background leakage balancing the emitting coefficients has allowed to adjust the model with errors in the pressures in the three control nodes of the order of 17 centimetres. A previous macrocalibration phase [3] was required because the sensitivity of the model became high enough to detect the existence of a tank that had not been considered due to its little effects.

To obtain the flow attributed to punctual leak, a fictional reservoir has been used [4] to subsequently calculate the most likely area where it is located from a thorough search on all nodes. At points where the leak situation is most likely, the average error of the three pressure sensors has been below 11.5 cm

Keywords
Sampling design, Macrocalibration, Leak localisation.

1 INTRODUCTION

The objective has been the elaboration of a method of locating specific and background leaks in hydraulic models. We have worked with high-resolution hydraulic models with high resolution pressure sensors. The method has been applied to a real network.

To execute an agile adjustment of the models, it is essential to link the hydraulic model with the data of the SCADA. The integration of POSGRESQL with QGIS through PYTHON executables such as WNRT or R with functions that use EPANET TOOLKIT provides this link. The tool developed by the GISWATER ASSOCIATION, QGIS on POSGRESQL [1] allows exporting networks to EPANET with the demanded resolution. This Work has used three levels of definition regarding the nodes introduced: node by change of material or diameter; nodes by subscriber; nodes according to a fixed distance. These are three of the four criteria allowed by the tool. The main water network of Manresa has 5,121 nodes the first level of definition. It increases to 10,495 nodes for the second

CC BY NC SA

and 38,788 nodes for the third, with an average distance between nodes of 2.98 meters. This will allow to observe the behaviour of the hydraulic model obtained according to the density of nodes.

As instrumentation, three pressure sensors have been used with a resolution of 1 centimetre. The situation of sensors in the network has been defined by using algorithms based on the sensitivity matrix [2].

Background leakage has been modelled using EPANET emitter coefficients. The value of these coefficients has been calculated based on the overall performance of the sector. They have been calculated individually for each node weighted by the lengths of the pipes linking to each node. The combination of a network with a high density of nodes, a distribution of background leakage balancing the emitting coefficients has allowed to adjust the model with errors in the pressures in the three control nodes of the order of 17 centimetres. A previous macrocalibration phase [3] was required because the sensitivity of the model became high enough to detect the existence of a tank that had not been considered due to its little effects.

To obtain the flow attributed to punctual leak, a fictional reservoir has been used [4] to subsequently calculate the most likely area where it is located from a thorough search on all nodes. At points where the leak situation is most likely, the average error of the three pressure sensors has been below 11.5 cm

2 MATERIALS AND METHODS

2.1 Case study

In order to choose the case study network the project stablished the following criteria:

- Urban network with at least 1.000 consumption meters.
- Both residential and industrial activity.
- With a source tank.
- Electromagnetic flowmeter.
- Chlorine analyser at the source.
- Time series of flow data at least for 10 years.

From all the networks managed by the company the hugest one District Metered Area (DMA) was chosen as it fulfilled all the criteria and its study was of high interest. It includes 39.000 consumers, its mean flow is 373 m^3/h. Its model (Figure 1) includes 5123 nodes and 5285 pipes. It is fed by gravity from a tank where an electromagnetic flowmeter provides the global consumption on-line.

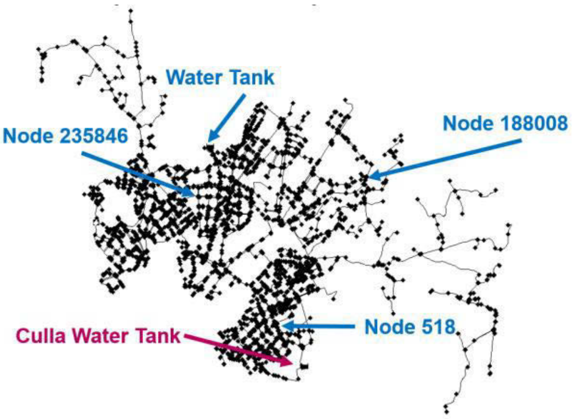


Figure 1: EPANET model for the case study network.

2.2 Sensor placement

One of the objectives of the project was to provide the network with new sensors. These sensors should be able to provide hydraulic and quality information from the network. Pressure, temperature, and chlorine sensors were installed and connected to the big data platform Zeus.

For this work only hydraulic information was used. The tank had a level sensor and an electromagnetic flow sensor at the output. Three pressure sensors were added. They are able to work between 0-10 mca and their precision is of 0.01 mca.

The sensor placement procedure methodology was based on the sensitivity matrix of pressures to demands. The objective of these pressure measurement is to detect changes in the demands that eventually can be associated to leaks.

Given the sensitivity matrix $\boldsymbol{S}$, its singular value decomposition provides the matrix $\boldsymbol{U}$. This matrix $\boldsymbol{U}$ corresponds to a base of orthonormal vectors of the measurement space (pressure). The first n (three in our case) are the maximum singular values and are related to the most relevant directions both in the pressure and demand spaces. Using the first three columns of $\boldsymbol{U}$ we obtain the Information Density Matrix that allow to find the nodes where the pressure is most sensible to changes in the demands. This methodology is explained in detail in [2].

2.3 Macrocalibration

Before using the model for supervision purposes, as leak localisation, it has to be calibrated. This calibration was carried out in a progressive improvement of the pressure estimation at the nodes where sensors had been installed. The process included 3 steps:

1. Simulation and analysis of the original model. For step one, we simulate the model considering the tank level as a boundary condition and the water flow as a global demand, but with no leaks.
2. Background leakage simulation. For step two, we introduce minor leaks to every node assuming that the water balance was due to background leakages. The assignment of the emitter coefficient is homogeneous, all the nodes have the same.

   The model obtained in step 2 highlighted a sharp change in the pressure, specially in one node, that was not seen in the measurements. The analysis of the inflow showed that there was a sharp increase of the demand that could not be imputed to a domestic demand.

3. Introduction of the secondary tank. The feedback of the company acknowledged that for simplicity they had omitted the presence of a secondary tank (10 meter below of the source) that is usually closed but when in-flow valve opens it takes around 140 l/s. The

lowest tank was included. We also updated the emitter coefficients so that they are proportional to the length of the pipes connected to each node.

2.4 Leak estimation

To estimate the magnitude of the leak, we introduce a virtual reservoir at the node where the pressure is measured. This virtual reservoir behaves as a source sink of water in order to maintain the known pressure at the node [4]. In our case we simulated the three possible scenarios given three pressure sensors. The water consumed by the reservoir is the estimation of the leak.

2.5 Leak localisation

With the calibrated model, the leak estimated will be introduced to every of the 5124. For each scenario we simulate and compared the predicted pressures with the measured ones. We use the mean absolute error (MAE) as a metric for evaluating the probability of having a leak in a node.

3 RESULTS

The results of the five methodologies applied to our case study are presented in this section. The starting point is a network illustrated by the map of Figure 1 where we installed three sensors. The evolution of the estimation of the pressure is illustrated graphically and with the average value for each step of the macrocalibration. Finally the result is another map where we provide the probability of having a leak in any of the nodes.

3.1 Sensors placement

The number of sensors to be introduced in the network was a trade-off between the budget and the information required as it often happens. In Figure 2 the first 10 singular values of the sensitivity matrix show how they decrease.

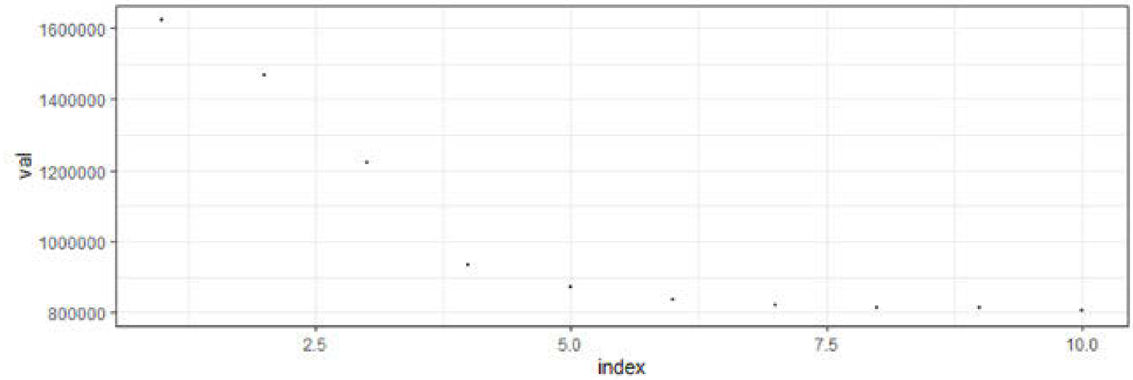


Figure 2:10 first singular values of the Sensitivity Matrix.

The Information Density Matrix was evaluated in different demand conditions. The most pressure sensible nodes were not always the same, but they appeared to be in the same three regions. One of the possible configurations was chosen. The exact location is presented in Figure 1.

3.2 Macrocalibration

The model was run introducing the real level of the input Tank as head of the reservoir in the model and adapting the demand to the measured inflow. Figure 1*Figure 3* presents the comparison between the measured pressure and its prediction in each sensor. We can observe two behaviours. For the nodes 518 and 188008 the discrepancies depend on the time instant being lower during the day while in node 235846 the discrepancies are much more constant. It is obvious that demand/leakages are responsible of part of the discrepancies due to this time dependence.

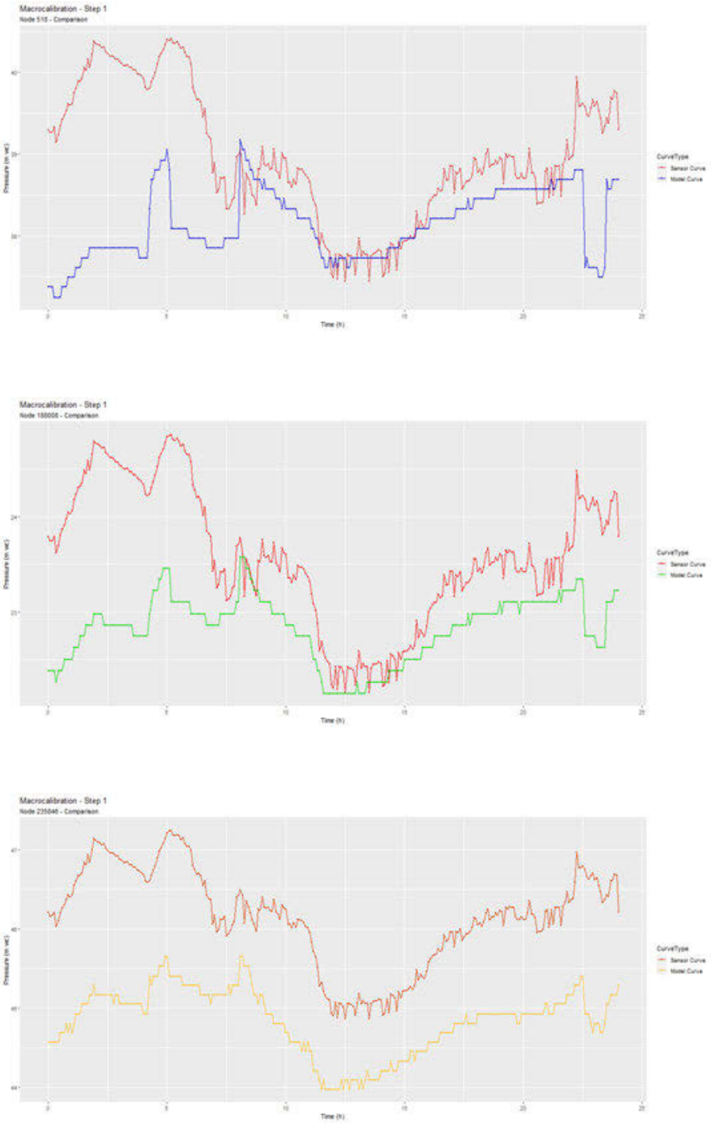


Figure 3. Comparison of measured (in red) and predicted (with the original model) pressure at the three sensors installed. From above to bellow nodes: 518, 188008 and 235846

Table 1 presents the average of the pressure errors. As was evident from *Figure 3* the maximum eror appears in node 235846.

Table 1. Average pressure absolute error (MAE) for the original model

Average Error for 24 hours	**Data (m wc)**
Node 518 Pressure	0,876533
Node 188008 Pressure	0,757516
Node 235846 Pressure	1,312204
Average from the three nodes	0,982084

The hydraulic balance of this DMA is around 80%. This has been stablished from billing data and the SCADA data for inflow. To adjust the demand behaviour we distributed 20% of the mean inflow as background leakages using a identical emitter coefficient at all the nodes. *Figure 4* presents the comparison between the measured pressure and its prediction in each sensor. *Table 2* presents the average of the pressure errors. The performance of the model has improved for each measurement. The maximum error still corresponds to node 235846. Nevertheless, we could not pass over the sharp changes in predicted pressure at node 518 and more subtle in node 188008 that was not present in measurement at all. We presented these results and received feedback from the company.

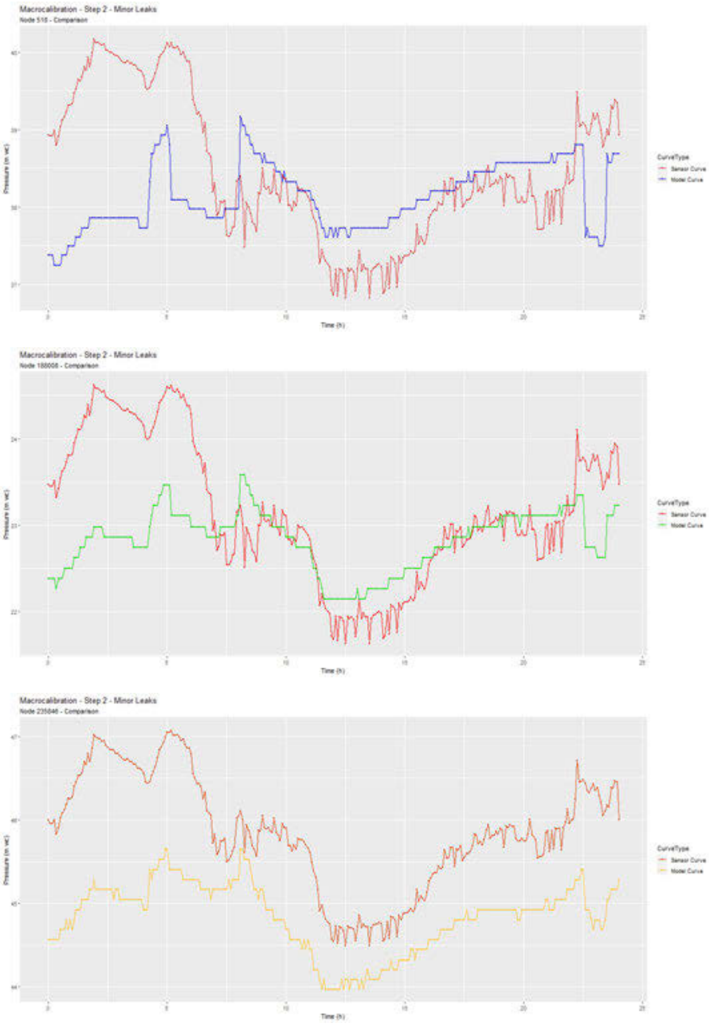


Figure 4. Comparison of measured (in red) and predicted (adding background leakage) pressure at the three sensors installed. From above to bellow nodes: 518, 188008 and 235846

Table 2. Average pressure absolute error (MAE) adding background leakage

Average Error for 24 hours	Data (m wc)
Node 518 Pressure	0,830616
Node 188008 Pressure	0,559142
Node 235846 Pressure	1,020405
Average from the three nodes	0,803388

BY NC SA

The company clarified that for simplicity the model omitted a tank that is generally closed and gives service to another DMA. The relevance of this tank is that at some times during the day, when its valve opens, it takes water from our DMA. As this tank (Culla in Figure 1) is 10 m bellow the feeding tank a great amount of water flows by gravity. Using the data of the inflow measured at Culla tank we introduced it in the model. Figure 5 presents the comparison between the measured pressure and its prediction in each sensor. Table 3 presents the average of the pressure errors. From the behaviour of sensors at nodes 188008 and 235846 we deduced that demand distribution geographically may not be the same throughout the day. Such phenomenon can be imputed to a leak that has a constant behaviour compared to domestic consumers.

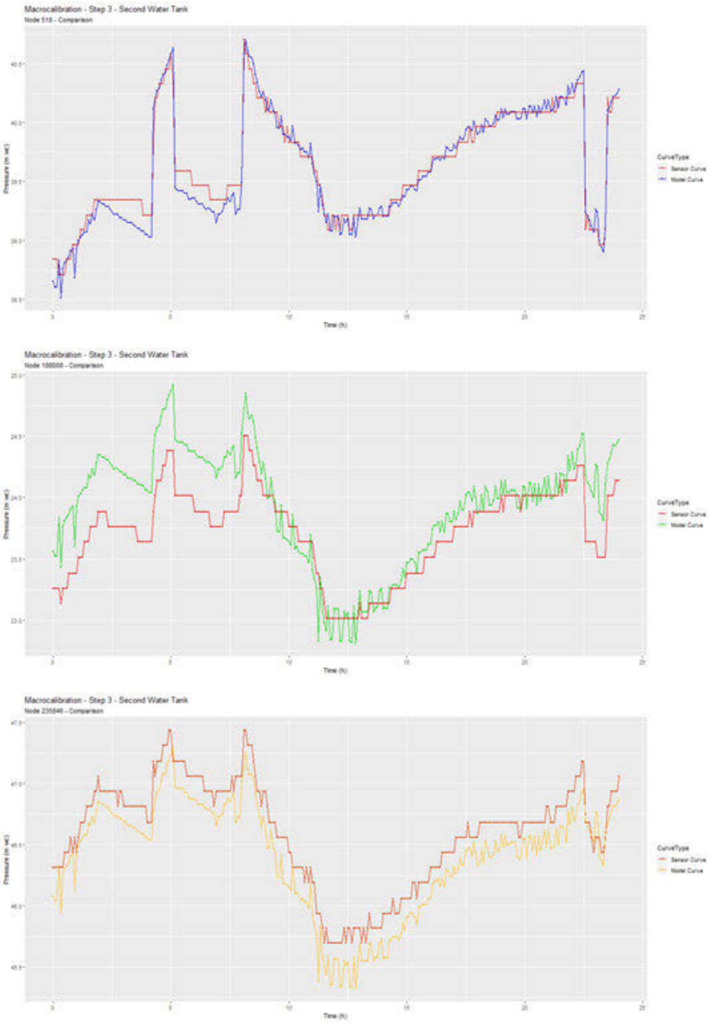


Figure 5. Comparison of measured (in red) and predicted (with the final model) pressure at the three sensors installed. From above to bellow nodes: 518, 188008 and 235846

Table 3. Average pressure absolute error (MAE) for the final model

Average Error for 24 hours	Data (m wc)
Node 518 Pressure	0,077114
Node 188008 Pressure	0,249073
Node 235846 Pressure	0,197529
Average from the three nodes	0,174572

3.3 Leak estimation

We introduced a fictional reservoir in each node provided with a pressure sensor. The head of the reservoir became the input and as output of the simulation we obtained the flow coming or leaving the reservoir. We carried out three simulations moving the fictional reservoir among the three nodes. Table 5 presents the results. Only at node 188008 the reservoir behaved as a sink (expected behaviour of a close leak). In Table 3 we had already observed that the node with worst adjustment in the prediction was this. The average inflow in this scenario is taken as an estimation of the leak.

Table 4. Average extra flow for the fictional reservoir

Node which has been added an additional water tank	Average extra flow (l/s)	Flow direction
Node 518	1,04	The water tank provide water to the network
Node 188008	4,34	The water tank absorb water from the network
Node 235846	7,01	The water tank provide water to the network

3.4 Leak localisation

Once the leak is estimated the emitter coefficients are recalculated. The proportion of leakage associated to background leakage has been reduced. In this process the emitter coefficient is not constant anymore. It has been calculated proportional to the length of the pipes connected with the node.

To locate the leak we simulate the same scenario as many times as nodes in the network (5123) introducing the leak in each node. The average absolute error (MAE) is evaluated in each scenario. Figure 6 presents these results. The colour represents the average error obtained when the leak was supposed in that node. Those nodes with lowest error are those more likely to have a leak.

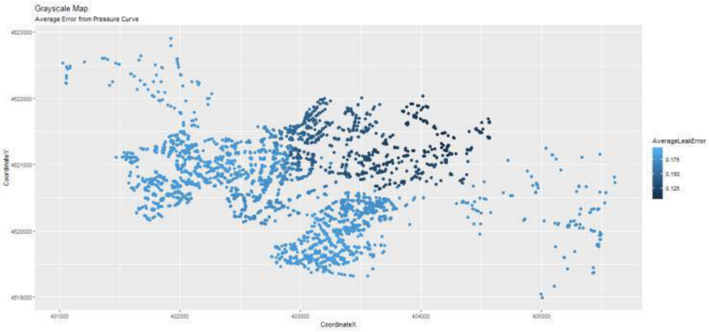


Figure 6. Average error of the measurements depending on where the leak is located.

Finally Figure 7 presents the same results but discretised. It seemed useful for the company when they wanted to plan the search campaign.

BY NC SA

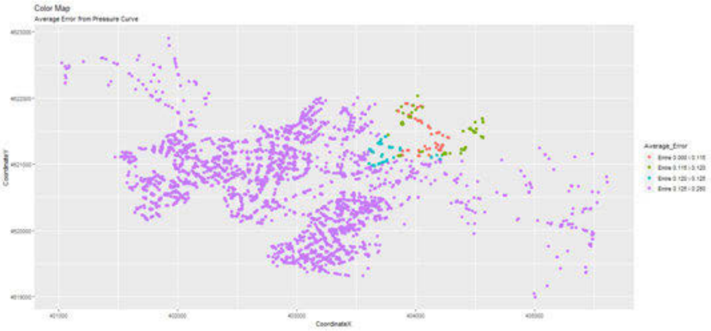


Figure 7. Discretised Average error of the measurements depending on where the leak is located.

4 CONCLUSIONS

The work presented applies in a real network methodologies of the literature for sensor placement, calibration and leak estimation and localisation. The model obtained has improved it performance arriving to errors in pressure prediction between 10 and 20 cm.

The leak localisation results encouraged the company a search campaign on those nodes with highest probability. Although no leak was finally found on the field they agree in the fact that the area signalled can have higher rate of background leakage.

Future projects will be focused in connecting the probability of leakage calculated on data available on pipes and economical assessment for the invest to do in a network.

5 ACKNOWLEDGEMENTS

The work has been possible thanks to the Hydroleaks project that has generated the model and installed the pressure sensors. This project has received a grant from the Ministrerio de Industria Comercio y Turismo within the 2020 call for support to innovative business groups (Nº Expediente AEI-010500-202-153).

6 REFERENCES

[1] https://github.com/Giswater

[2] Sanz, G., & Pérez, R. (2015). Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition. Journal of Water Resources Planning and Management, 04015020.

[3] Ormsbee, L., & Lingireddy, S. (1997). Calibrating Hydraulic Network Models. Journal of the American Water Works Association, 89(2), 42–50.

[4] Steffelbauer, D. B., Deuerlein, J., Gilbert, D., Piller, O., & Abraham, E. (2020). A Dual Model For Leak Detection and Localization. BattLeDIM 2020 - Battle of the Leakage Detection and Isolation Methods, September. https://doi.org/10.5281/zenodo.3923907

[5] Alves, D., Blesa, J., Duviella, E., & Rajaoarisoa, L. (2021). Robust data-driven leak localization in water distribution networks using pressure measurements and topological information. Sensors, 21(22), 7551. doi:http://dx.doi.org/10.3390/s21227551

[6] Ramon Pérez R., Cugueró J., Romera J., Van Eeckhout J., Cabot J., Franch O. (2021). Water Distribution Network Model Calibration and Continuous Maintenance : Terrassa , A Real Application. Modern Environmental Science and Engineering (9), 891–897.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14562

DEVELOPMENT OF A GRAPH DYNAMIC SECTORIZATION TOOL USING QGIS/POSTGRESQL TO PLAN AND OPERATE WATER DISTRIBUTION NETWORKS

Torret X.[1], Grau S[2,3]., Pérez R.[3] and Amat E.[4]
[1] BGEO, Spain
[2] Innovació, Aigües de Manresa, S.A., Spain
[3] cs2ac, Universitat Politècnica de Catalunya, Spain
[4] Cap de servei de xarxes i responsable oficina técnica, FISERSA (Figueres de Serveis, SA), Spain

[1] xtorret@bgeo.es [2] sgrau@aiguesmanresa.cat, [3] ramon.perez@upc.edu,[4] eamat@fisersa.cat

Abstract

This paper presents a definition of different types of sectorization used by water distribution companies, as well as a methodology for their identification and application in the dynamic zoning of water distribution networks.

The graph model of large water supply networks is made up of thousands or hundreds of thousands of nodes and pipelines, as well as valves, tanks, pumps and pressure groups. Its generation from the data contained in a GIS does not report difficulty with the currently existing tools, the GISWATER module of QGIS is an example of this. Problems arise when these models have to be operated from a high number of conditions and scenarios. To facilitate the operability of the system, one of the strategies is to simplify its operation through topological analysis. One of the proposals is clustering where the nodes belonging to a cluster have more and better connections with the internal nodes than with the external ones.

Our work analyses water distribution networks from five different points of view, which it represents five sector classes according to the parameter analysed. As result, one of them, the District Metering Area (DMA) delimited by headers (flowmeters) and stoppers (other flowmeters or closed valves), is extensively covered in the literature. The other four, less common in the literature, have a definition equivalent to that of the DMA but based on other class of headers and stoppers. Therefore, if we are looking to operate the network from the point of view of pressure, we can use the dynamic sectorization algorithm by configuring Tanks, Head pumps, flow pumps, PSV, PBV or PRV as headers.

Moreover, to operate and analyse the network from the point of view where water comes from, we can use again the dynamic algorithm by configuring the inlets of the system as headers (Tanks and Reservoirs).

On the other hand, if the goal is to be focused on water quality, the algorithm can be configured using chlorinators, chlorinators, water treatment equipment or water sources as headers.

Finally, special attention deserves the minimum sector which represents the minimum structural unit of the network, defined by the region covered by network where there are shutoff-valves on their borders. In order to operate water networks as optimized as possible, the status of shut-off valves is dynamic. The opening and closing of valves allow a continuous reconfiguration of the network and the classes or types of sectors defined.

As result, to analyse the network graph using this approach enables the water operator to analyse the best status as possible as well as it is amazing scenario to analyse resilience of network using stochastics methods.

Keywords
Sectorization, dynamic zoning, network, DMA, minimum structural unit.

1 INTRODUCCION

Water management has always been complex and expensive. It has always been difficult to plan new water supply networks or effectively control existing ones without making further capital investments. Nonetheless all these situations have come to an end.

Since 2014 when Giswater [1] was born, it has been the first open-source software specifically designed for water supply and water management. This software connects different IT solutions and pre-existent databases allowing you to setup a high-performance management system in combination with hydraulic software as well EPANET [2] or SWMM [3].

In order to facilitate the operability of the system, one of the strategies is to simplify its operation through topological analysis. One of the proposals consists of clustering [4] where the nodes belonging to a cluster have more and better connections with the internal nodes than with the external ones.

This work has been developed using Open-Source software from the ecosystem of Giswater environment. For this specific algorithm you can access to the full source code on Github [5].

2 NETWORK MAPZONES

INLET SECTOR (I-SECTOR)

Network subgraph defined as subgraph area limited by inlets (Tank or Reservoir). It is the minimum unit for hydraulic calculation since it contains the water supply. Flow meters or pressure meters are not considered. pressure reducing valves and pressure groups are not considered.

On the INLET-SECTOR (I-SECTOR) there are no customers. Only exists water transportation from some point to another point.

SUPPLY SECTOR (S-SECTOR)

Network subgraph defined as subgraph area limited by inlets (Tank or Reservoir) and closed valves with customers which they consume water. As well as [I-SECTOR] it is also the minimum unit for hydraulic calculation by exporting some EPANET file since it contains the water supply.

Flow meters or pressure meters are not considered. pressure reducing valves and head pumps are not considered,

DISTRICT METER AREA(DMA)

Network subgraph defined as subgraph area limited by meters and closed valves. It is the subgraph unit to make measurements of water balances for network. It enables the possibility to calculate water losses and Not Revenue Water (NRW).

PRESSURE SECTOR (PRESSZONE)

Network subgraph defined as subgraph area limited by pressure regulation elements (Tank, Reservoir, PSV, PRV, PBV, Pump) and closed valves. It is the subgraph used to calculate the static pressure as maximum pressure for the whole system.

DISTRICT QUALITY AREA (DQA)

Network subgraph defined as subgraph area limited by quality regulation and quality measurement elements (Tanks, Reservoirs or chlorinators) and closed valves. It is the subgraph used to define quality water areas. It enables to calculate the residence time and the evolution of the water quality parameters.

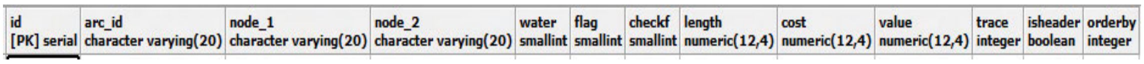
Moreover, network subgraph defined as subgraph area limited by shut-off valves. It is the minimum network affected in case of breakdown. It shows the minimum number of clients affected by the supply interruption due to the breakdown. Also enables the breakdown probability calculation. This subgraph is called **minimum sector** and can provide also amazing analysis for each one of those five functional mapzones defined above.

3 NETWORK AND GRAF DELIMITERS

Drinking water network has several elements which some of them establish hierarchy and also work between each other. That can give many opportunities to the organization from a functional and system analysis point of view.

In order to analyse water distribution networks from different functional points of view, we need to built-up the graph and define which network elements acts as graph objects and which acts are delimiters.

Regarding network graph will be built-up by using topologic data from arc and node table (*arc_id*, *node_1* and *node_2*). This information is stored on *temp_anlgraf* table which is the core table for this algorithm in combination with the *fct_gw_grafanalytics_mapzones* function.



id	arc_id	node_1	node_2	water	flag	checkf	length	cost	value	trace	isheader	orderby
[PK] serial	character varying(20)	character varying(20)	character varying(20)	smallint	smallint	smallint	numeric(12,4)	numeric(12,4)	numeric(12,4)	integer	boolean	integer

Figure 1. Columns for temp_anlgraf table

Moreover, headers and stoppers are defined below:

Storage tanks. Modifies the residence time of the water in the system. They act as graph delimiter for [I-SECTOR] [S-SECTOR].

Supply and storage tanks. Establishes a pressure or primary supply level for one or more customer sectors and modifies its residence time of water in the system. They act as graph delimiter for [I-SECTOR] [S-SECTOR].

Queue tanks. Establishes a secondary supply level for a sector of customers and modifies the residence time of the network water. In fact, they are not graph delimiter because they are by themselves end-point of system.

Supply and queue tanks. Queue tanks, establishes primary and secondary supply level at the same time for two or more customer sectors and modifies the residence time of water in the network. They may act as graph delimiter for [S-SECTOR].

Flow meters. Determines the differential flow in the first case and integral flow in the second case, by a specific synchrony. They act as graph delimiter for [DMA].

Pressure reduce, sustain and break valves. Modifies the pressure and supply level downwards. It is generally associated with customer sectors but not exclusively. They act as graph delimiter for [PRESSZONE].

Head pump. Modifies upwards the pressure and supply level to a sector of clients. It has a small storage (pilot whale) that usually should not be considered, since the storage time is usually lower than the measurement frequency of the flow meters. They act as graph delimiter for [PRESSZONE].

Pumping station. Water transfer between two storage points with increase of elevation. They act as graph delimiter for [PRESSZONE].

Quality measurement elements. Quality measurement elements like chlormeter. They act as graph delimiter for [DQA].

Quality regulator. Quality regulator like chlorinator. They act as graph delimiter for [DQA].

Shut-off valves. Is an actuated valve designed to stop the flow upon the detection of a dangerous event or some other operation. This provides protection against possible harm to people, equipment or the environment. They act as graph delimiters for all the different mapzones in function of they are open or closed.

Unique customers. Customers with a very high or unique consumption in their behaviour. They should be treated as specific nodes.

CC BY NC SA

4 DYNAMIC ZONIFICATION ALGORITHM

One of the main characteristics of this algorithm is that it is executed on server side. It means that the whole code is stored into PostgreSQL database and QGIS client only acts as user interface.

As a result, with this algorithm, the five different types of zonifications described before are ready-to-execute. To work with, this mapzones algorithm can be executed after completing the configuration process which means that:

- Populate the *graf_delimiter* column on *cat_feature_node* table with values defined above on network graph delimiters chapter.
- Populate the *grafconfig* column on mapzones table (*sector, dma, dqa, presszone*). Syntax is a bit special but allows every possible cardinality (multiple nodeParent with multiple toArc for a single mapzone).

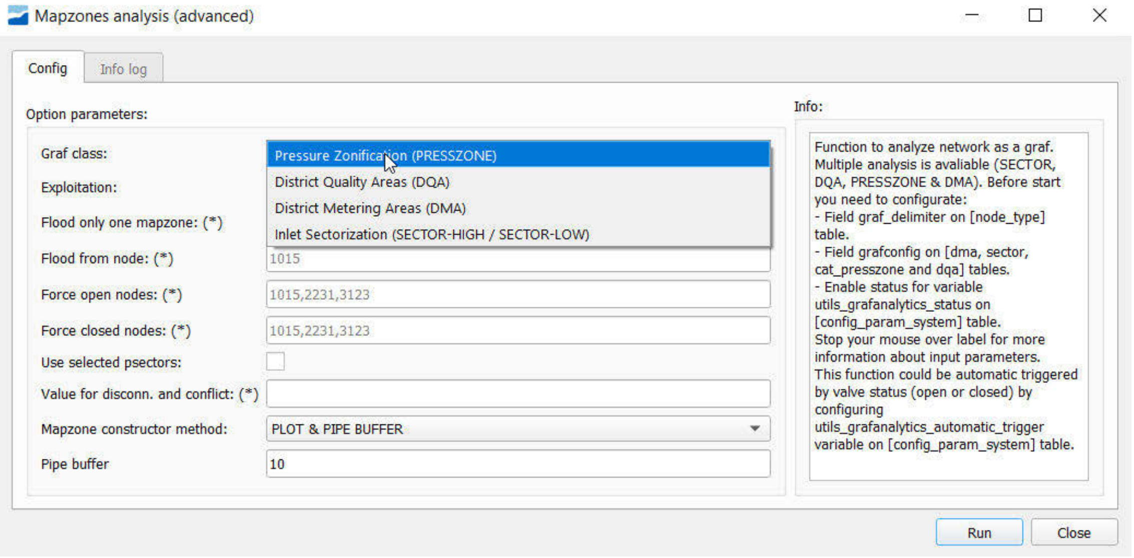
In case that nodes whose *node_type* is a graph delimiter and has not been named as the head of any mapzone, the system will allow the calculation but will give a warning. Since this is a possible option (transport tanks or queue tanks without a specifically defined node_type) it is advisable to use the "ignore" key of the grafconfig field so that at least the system does not give warnings of elements that we have clear and checked.

Once configured it is moment to trigger the algorithm. The workflow of the algorithm it is simple.

It starts an inundation process from headers by flooding the network, tracing element by element stopping in front of other headers or in front of closed shut-off valves.

This inundation process is done by *gw_fct_grafanalytics_mapzones* function by using a simple loop. In order to control the process on each loop, algorithm updates columns *water* and *trace* on *temp_anlgraf* table. Column *water* to know when one node has been processed or not and column *trace* to know which is the node_id of the header. Besides these two objects (*gw_fct_grafanalytics_mapzones* and *temp_anlgraf)* one relation more on PostgreSQL database is used on this process. This is the *view v_anl_grafanalytics_mapzones.*

Moreover, code side, algorithm has a well-structured user side. By using dialog showed below user can choose the graph class analysis as well as other parameters for their analysis.


Mapzones analysis (advanced)
Config
Info log
Pressure Zonification (PRESSZONE)
District Quality Areas (DQA)
District Metering Areas (DMA)
Inlet Sectorization (SECTOR-HIGH / SECTOR-LOW)
Run
Close

Option parameters:

Graf class: Pressure Zonification (PRESSZONE)
Exploitation:
Flood only one mapzone: (*)
Flood from node: (*) 1015
Force open nodes: (*) 1015,2231,3123
Force closed nodes: (*) 1015,2231,3123
Use selected psectors: ☐
Value for disconn. and conflict: (*)
Mapzone constructor method: PLOT & PIPE BUFFER
Pipe buffer 10

Info:

Function to analyze network as a graf. Multiple analysis is avaliable (SECTOR, DQA, PRESSZONE & DMA). Before start you need to configurate:
- Field graf_delimiter on [node_type] table.
- Field grafconfig on [dma, sector, cat_presszone and dqa] tables.
- Enable status for variable utils_grafanalytics_status on [config_param_system] table.
Stop your mouse over label for more information about input parameters.
This function could be automatic triggered by valve status (open or closed) by configuring utils_grafanalytics_automatic_trigger variable on [config_param_system] table.

Figure 2. Dialog on QGIS user interface for input parameters before trigger algorithm

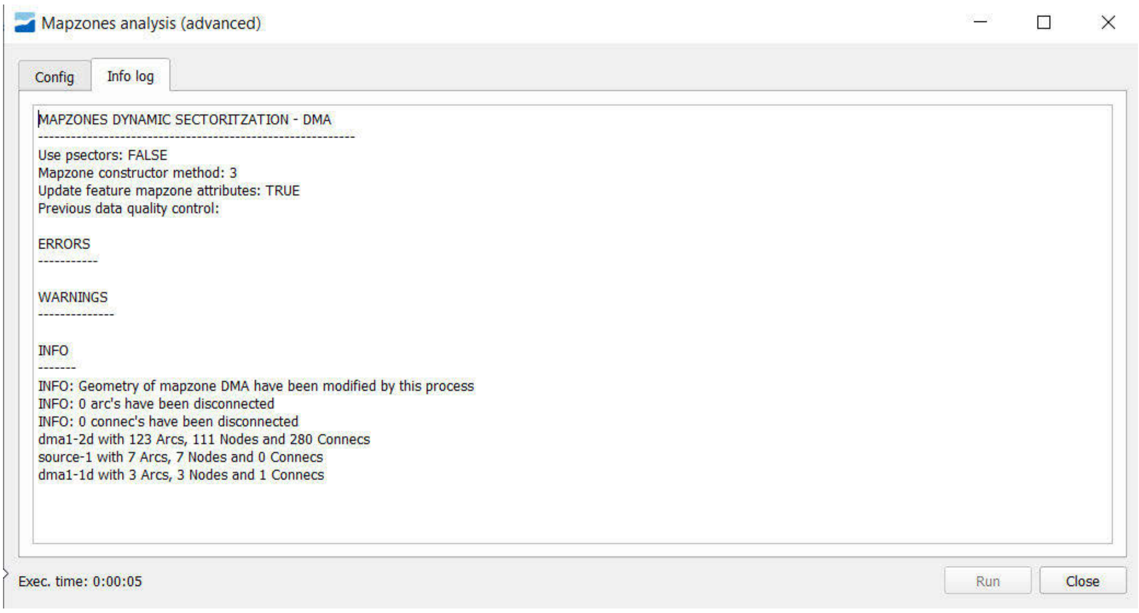
In fact, several parameters can be passed when function is triggered. Table on below shows all parameters and additional information about description, example data and range of values:

cc BY NC SA

Table 1. Detailed information about the parameters accepted for algorithm

PARAMETER	DESCRIPT	EXAMPLE	RANGE OF VALUES
grafClass	Graph class to perform	DMA	DMA, SECTOR, DQA, PRESSZONE
exploitation	Exploitations to participate in the algorithm	[1,2]	All the available exploitations
updateFeature	If true, update the values of dma_id, presszone_id, sector_id & dma_id of all NODES, ARCS, CONNEC that are flooded by the algorithm	true	false, true
updateMapZone	0: does not update the geometry field (the_geom) of the mapzone 1: Updates by making an enveloping polygon with all the elements 2: Updates by buffering the pipes with the value of geomParamUpdate 3: Updates by buffering the pipes with the value of geomParamUpdate and incorporating the plot geometry (if it exists) 4: Updates by buffering the pipes with the value of geomParamUpdate and incorporating the link geometry (if it exists) 5: Updates by buffering the pipes with the value of geomParamUpdate and incorporating some customized geometry (if it exists)	2	0,1,2,3,4,5
geomParamUpdate	Value related to options 2, 3 of the previous key. It is the value of the buffer (expressed in meters)	10	Any float 0.1 – 100
macroExploitation	Not mandatory value. Useful to define macro-exploitations to participate in the algorithm. In case of not null value this will be priorized against exploitation key	[1,2]	All available macro-exploitations
checkData	Not mandatory value. If true, check if the system data is correct (topology, state_type, etc), and in case of errors, this will abort the process.	false	false, true
usePlanPsector	Not mandatory value. If true, the selected sectors from user will be used by the analysis of the algorithm.	false	false, true
floodOnlyMapzone	Not mandatory value. If not null, algorithm only will flood defined mapzones.	[1,2]	All defined mapzones
valueForDisconnected	Not mandatory value. If not null, disconnected elements will take another value different from 0. This is useful when you are building your sectorization model.	1	All mapzone values
forceOpen	Not mandatory value. Valves that can be forced to open (e.g. for closed valves that we want to open for whatever reason)	[1,2,3]	All closed valves
forceClosed	Not mandatory value. Nodes in general that can be forced to close (e.g., in the debug phase in case the trace gets out of control and does not converge as expected)	[1,2,3]	All nodes (except closed valves)

Once function is finished returns a detailed log about executed process. This log provides information about possibly conflicts against mapzones, disconnected elements as well as the number of elements affected, in terms of pipes, nodes and connecs.



```
MAPZONES DYNAMIC SECTORITZATION - DMA
-----------------------------------------------------------
Use psectors: FALSE
Mapzone constructor method: 3
Update feature mapzone attributes: TRUE
Previous data quality control:

ERRORS
-----------

WARNINGS
--------------

INFO
-------
INFO: Geometry of mapzone DMA have been modified by this process
INFO: 0 arc's have been disconnected
INFO: 0 connec's have been disconnected
dma1-2d with 123 Arcs, 111 Nodes and 280 Connecs
source-1 with 7 Arcs, 7 Nodes and 0 Connecs
dma1-1d with 3 Arcs, 3 Nodes and 1 Connecs
```

Exec. time: 0:00:05

Figure 3. Dialog on QGIS user interface for output log after trigger algorithm

BY NC SA

One of the most interesting input parameters is '**updateMapZone**' which allows user to represent the mapzone on map by four different ways.

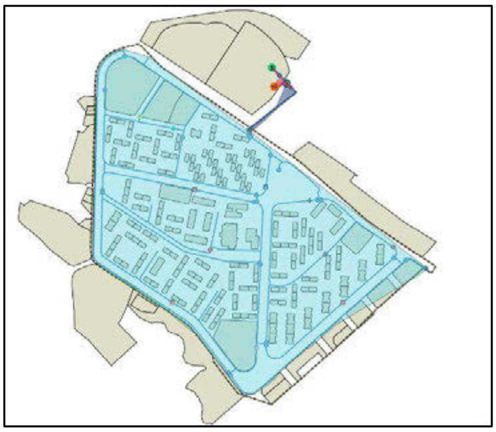


Figure 4. Mapzone geometry by polygon envelope

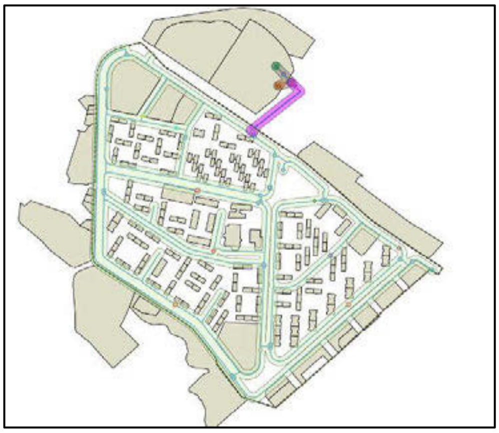


Figure 5: Mapzone geometry by pipe buffer

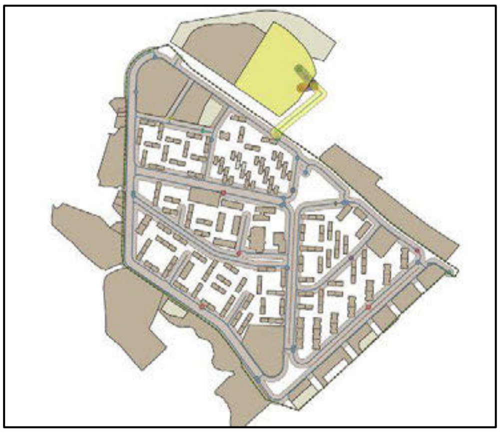


Figure 6. Mapzone geometry by pipe buffer and plot.

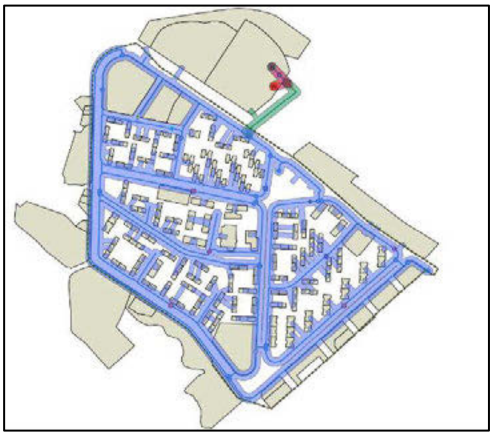


Figure 7: Mapzone by pipe and link buffer

Once algorithm finish a detailed log is showed for user in order to provide as much information as possible for the obtained result

5 ENHANCEMENT OF WATER NETWORK ANALYSIS USING EPANET

One of the most interesting issues applying this methodology is the enhancement of network analysis by using EPANET.

As you know this tool have been developed into the Giswater ecosystem. Giswater project is natively oriented to integrate GIS inventory data with US EPA hydraulic programs EPANET & SWMM. From version 1.0 launched on 2014, database was already designed to full-integration of inventory and hydraulic modelling capabilities. As a result, there are some special issues that makes Giswater as powerful tool to integrate hydraulic capacities with asset inventory.

The impact of this dynamic graph analytics on EPANET analytics is huge because it enhances several issues.

In terms of the mapzone SECTOR, this is the minimum unit to generate EPANET files. It is possible to generate as many sectors as you like but at least on. In fact, the definition of sector comes natively from EPANET. Sector is what EPANET needs, at least with one INLET, one pipe and one junction.

In terms of PRESSZONE, it could be useful to calibrate and to check EPANET water analytics because this mapzone provides us information about the static pressure which is the maximum pressure could be supplied

Regarding of DMA, it is very useful because brings us the possibility to make water balances by using this tool as well as it is possible to use composed patterns for in-side junctions on each DMA by composing patterns using timestep water balances from inlet meters less outlet meters.

Finally, DQA is less used than other mapzones, but the logic is also the same.

6 CONCLUSIONS

The work presented applies in a real network, methodologies described on the literature for different types of functional analysis for water distribution networks.

The key point of this work is how we have managed to harmonize and structure the functional analysis of a water network based on the objective that we are pursuing, [INLET, WATER BALANCE, OPERATION ZONE, QUALITY AREAS] always oriented to the analysis needed.

Future works can go on giving more performance to the algorithm with the use of the minimum sector described in section 2 of this paper as a unitary element of network graph in spite of work with node and arcs objects.

7 ACKNOWLEDGEMENTS

The work has been possible thanks to the financing of Giswater project which is receiving recurrent money from Giswater Association [1] and its associates.

8 REFERENCES

[1] Web page of Giswater project: https://www.giswater.org/

[2] EPANET web page: https://www.epa.gov/water-research/epanet

[3] SWMM web page: https://www.epa.gov/water-research/storm-water-management-model-swmm

[4] Source code for mapzone analytics algorithm developed with PL-SQL on the Github repository of Giswater: https://github.com/Giswater/giswater_dbmodel/blob/master-3.5/ws/fct/ws_gw_fct_grafanalytics_mapzones.sql

[5] Lina Perelman, Avi Ostfeld, Topological clustering for water distribution systems analysis, Environmental Modelling & Software, Volume 26, Issue 7, 2011, Pages 969-972, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2011.01.006.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14574

EFFECT OF SWABBING CLEANING METHOD ON BIOFILM COMMUNITIES OF A DRINKING WATER DISTRIBUTION SYSTEM IN MADRID (SPAIN)

Carolina Calero Preciado[1], Emelina Rodríguez Ruiz[2] and Manuel Arias Guedón[3]

[1,2,3] Canal de Isabel II S.A./Área de Construcción de Redes de Abastecimiento/Madrid/España

[1]ccalero@canal.madrid, [2] errodriguez@canal.madrid, [3] mjarias@canal.madrid

Abstract

Drinking Water Distribution Systems (DWDS) are complex engineering networks where it has been shown that the microbial presence plays a key role for the drinking water quality and safety. Most of microbial biomass in these systems is found attached to the inner-pipe surface forming biofilms, which have been associated with several water quality problems including discolouration events, taste and odour problems, degradation and corrosion of pipes, and the presence of opportunistic pathogens. The implementation of biofilm control strategies is thus vital to guarantee the supply of good-quality drinking water. Water flushing is the most common and longest applied management strategy for cleaning the pipes worldwide. However, the amount of water resource required for the cleaning work is favouring the emergence of alternative methods such as swabbing. Nevertheless, there is a lack in the knowledge about the effectiveness or influence on microbial dynamics of this method. This study aimed to investigate the effect of swabbing in the biofilm communities of a real chloraminated DWDS in Madrid (Spain). Biofilm samples from real pipes were taken before and after swabbing cleaning. Then, the biofilm re-growth over a two-years after the cleaning was evaluated thanks to a coupon device that allowed in situ biofilm sampling without interrupting the supply. Coupons samples were taken every 6 months to study the microbial re-growth and the biofilm succession over a two-year period after the swabbing cleaning. Inorganic fraction of biofilm samples was determined by means of Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Flow cytometry (FCM) analysis were carried to estimate the total cells counts (TCC) and the 16s-rRNA bacterial gene sequencing was performed for the taxonomic characterization of biofilm samples. ICP-MS results showed that swabbing did not affect the inorganic content of the biofilm and suggests that it is more influenced by the pipe material over other factors. Regarding the biofilm quantity, FCM results suggested that although swabbing can reduce the number of cells in the biofilm, it is not completely effective removing the biofilm deposits from the pipe walls, as it happens with other cleaning methods such as flushing. Taxonomic analysis showed that swabbing affected the structure and composition of biofilm bacterial community. Important bacterial genera for water quality such as Nitrosomonas highly reduced their relative abundance after the swabbing. Moreover, during the biofilm regrowth over 2 years, taxonomical changes were also observed in the bacterial composition. From this study it can be concluded that swabbing modifies the biofilm communities of DWDS systems. Although this method cannot completely remove the biofilm deposits, the results showed that swabbing can be an effective tool aimed to minimise the presence and thus the risk associated with the presence of certain bacterial genera. This is a first approach to improve the management of DWDS, but more research is needed that includes other cleaning methods to design effective biofilm control strategies that ensure biological safety of the water that reaches consumers' taps.

Keywords
Water quality, bacteria, metagenomics, nitrification.

1 INTRODUCCTION

Drinking Water Distribution Systems (DWDS) are large and complex engineering networks in which it has been shown that microorganisms play an important role for the water quality and safety that is distributed to consumers [1]. Despite of drinking water is treated in treatments plants using diverse processes to remove contaminants and microbes to make it safe for consumption, microorganisms are able to survive this disinfection processes and/or enter these systems by intrusion or contamination events [2], including bacteria, fungi, archaea, viruses and even protist. They can inhabit DWDS in a planktonic way of life, but most of the microbial biomass in these systems is found attached to the inner-pipe surface forming complex biofilms (REF). Biofilm is defined as an assemblage of microorganisms attached to a surface and/or to each other and enclosed in a self-produced matrix of hydrated extracellular polymer substances (EPS) [3]. EPS matrix is the most extensive part of the biofilm, and it is responsible for adhesion and fixation of microbial cells to the pipe surface and between them. In addition to the structure and mechanical stability, biofilm EPS matrix provide a wide range of advantages to microorganisms, including protection against chemicals such as disinfectants, desiccation or oxidizing [4]. EPS are also a rich source of nutrients and exoenzymes, that enhancing metabolic capacities helping the acquisition of these nutrients. Several exoenzymes can also help for complete degradation of complex compounds potentially toxic or bactericidal, for example antibiotics [5]–[7]. This makes that more than 95% of the microbial biomass in DWDS can be found as biofilms [8]. Nevertheless, biofilm development in DWDS have been associated with several problems in DWDS, which can alter the quality and safety of drinking water [9]. For example, if the flow shear stress at the pipe wall exceeds the normal value, biofilms attached at the pipe wall can be mobilised into the bulk water. The mobilisation and presence of biofilms is related to discolouration events and thus the increase of metals and contaminants concentrations in water [10]. Biofilm activity can also generate problems and complains about water taste and odour, induce degradation and corrosion of pipes or metabolic processes, for example nitrification, that trigger the formation of toxic or harmful compound for health [11].

In order to minimise the risks associated with the presence of biofilms, water utilities use different operational and management strategies [12]. Mechanical methods are the most widely used to control biofilms in DWDS, including flushing, air scouring, ice pigging, and swabbing [13]. Water flushing is the most common and longest applied method for cleaning the pipes worldwide [12]. It consists of increasing the water flow to produce an increase in the water velocity that will lead to further shear stress on the biofilms attached to the pipe surfaces [14], [15]. However, flushing can be difficult to apply in pipes with large diameters such as trunks mains because it is impossible to reach the necessary water velocity [16]. In addition, for a flushing event a very large volume of drinking water is wasted [17], and this is promoting that the rest of the techniques are emerging as an alternative cleaning method.

Air scouring consists of injecting filtered and compressed air onto the water mains through a hydrant at the beginning of the section to rinse. Air and water form “slug flows”, which are driven along the pipes by the compressed air at high velocity removing biofilm, fouling and material sedimentation [18], [19]. This method has been observed to be more aggressive and thus more efficient in particles, biofilm, tubercles, and incrustations removal due to increased liquid velocity in the slugs and increased wall shear stress. Air scouring does not depend on the network pressure thus it can be used in pipes of large diameter [19] . Moreover, it is estimated that it uses 40% less water reducing the amount of washing water compared to flushing [20] . Among its disadvantages is that it is more expensive since it requires specialized equipment, such as a compressor, air cooler and filters [18], [21], [22].

Pig-based method consist of the introduction inside the pipes of a solid device or material (the pig) that travels through the system. The pig can be ice, a foam sponge or even smart devices that can also inspect and monitor the pipe conditions [22], [23].

Ice pigging process consists of inserting a slurry of ice into the pipe through an element of the network, taking advantage of the existing pressure in the supply network to move the ice through the pipe, which generates great friction on the walls, and eliminates with a high effectiveness the adhered material. As the pig passes through the pipe, the natural glacial effect of the ice causes the pig to scrub the pipe and capture sediments and biofilm. This makes the ice pigging uses approximately 50% less water than standard water flushing and less time [24]. Other advantages of this method include the reduction of execution times and, therefore, the reduction of the supply interruption time, as well as the minimum environmental impact that it entails. However, as it happens with air scouring, it requires specialized equipment thus is more expensive comparing to flushing [20], [24].

Swabbing is other alternative method with high efficiency in removing attached materials in DWDS pipes. Swabbing process involves driving a cylindrical foam sponge also known as a swab through pipes using water pressure [13], [25]. The swab, which has to be approximately 25% grater tan the pipe it is being forced through, is inserted via an element of the network and is driven through the pipe to be cleaned using the network pressure. Its extraction is carried out by another element of the network, without the need for additional works [13]. Swabbing has been shown to be more effective than other methods to eliminate a high number of deposits using a smaller amount of water volume. This technique also offers the possibility of cleaning pipes with diameters of up to 1000 mm. However, the swabs can break or get stuck in the pipe, especially in pipes with a high degree of internal corrosion or scale and requires the installation of adequate points for the insertion and extraction of the swabs [13], [21], [25].

To date, of all these methods, only the effect of flushing on biofilm communities of DWDS has been studied. It has been widely demonstrated that pipe flushing influences the mobilization and the structure and composition of microbial communities, and thus the water quality of the systems [10]. However, there is a research lack on the effect that other cleaning methods have on microbial communities and its implication for water quality.

This study was aimed to evaluate the effect of an alternative method to flushing on the biofilm communities of DWDS. For this study, swabbing method was selected, which consists of driving a cylindrical foam sponge (known as a swab) through pipes using water pressure [13]. It has been observed to be potentially more effective than flushing or air scouring to remove a high number of deposits using less water [13]. Hereby, the aim of this research was aimed at understanding how swabbing method affects the structure and composition of biofilm communities and subsequent microbial regrowth in DWDS pipes. This new knowledge will provide new information to adapt and/or change biofilm control strategies, thus helping to ensure the supply of safe drinking water to consumers.

2 MATERIALS AND METHODS

2.1 Study site and biofilm sampling

The study was carried out in a real drinking water distribution system in a municipality located northwest of the Community of Madrid (Spain) (**Figure 1**). The network selected for the study is composed of 200 mm diameter ductile iron pipe disinfected by chloramination and managed by Canal de Isabel II S.A.

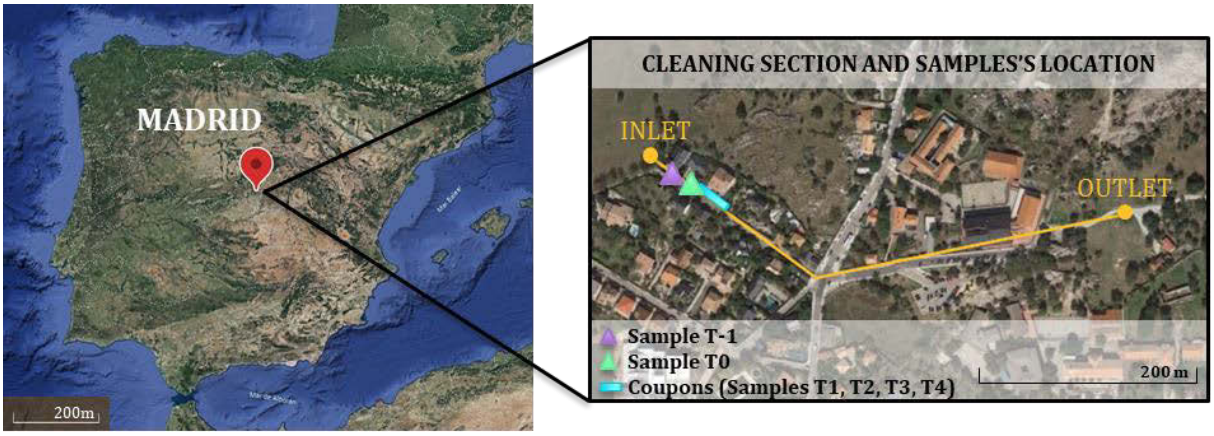

MADRID
200m
CLEANING SECTION AND SAMPLES'S LOCATION
INLET
OUTLET
Sample T-1
Sample T0
Coupons (Samples T1, T2, T3, T4)
200 m

Figure 1. Study site and sampling point locations.

Cleaning process was carried out using the swabbing method in a total length of pipe of 460 m in July 2019. As it is shown in Figure 2, two special pieces at both ends of the section were installed for the insertion and extraction of the pigs (inlet and. For this study, 4 poly-pigs of 230 mm with different grades of swab were used, starting from the softest to the most scouring. Each pig took approximately 20 minutes to traverse the entire section of pipe.

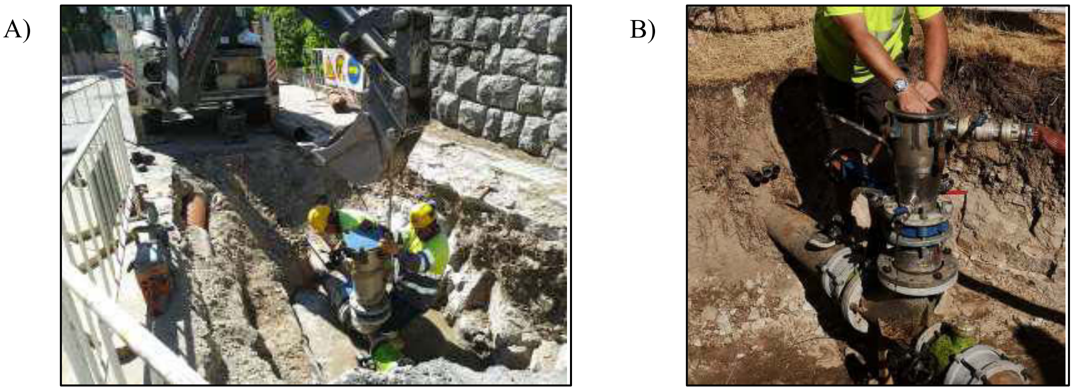


Figure 2. A) Input and B) output devices for the pig insertion and extraction.

For biofilm characterisation, a pipe section was removed to get biofilm samples before (t-1) the cleaning in July 2019. Once the cleaning was performed, an adjacent pipe section was removed 12 day later to get the biofilm sample after the cleaning (t0). At that time, the removed section of pipe was replaced by a coupon device (**Figure 2**), that allowed in situ biofilm sampling without interrupting the supply. Coupons samples were taken approximately every 6 months to study the microbial re-growth and the biofilm succession over a two-year period after the swabbing cleaning: sample t1 in December 2019, sample t2 in June 2020, sample t3 in October 2020 and sample t4 in March 2021 (**Figures 1 and 3**).

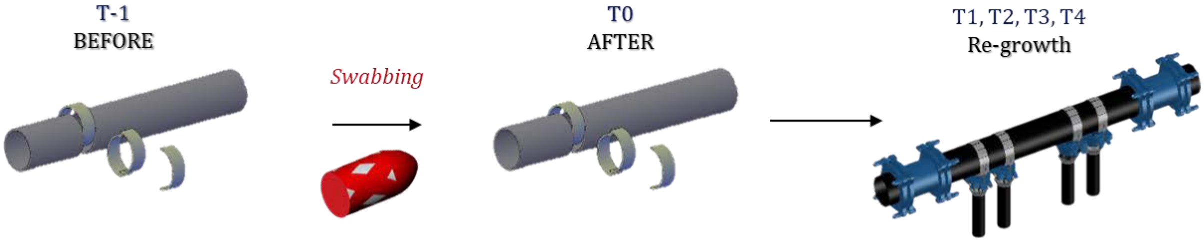


Figure 3. Scheme of the experiment and sampling over time.

BY NC SA

2.2 Biofilm analysis

To study biofilm communities, Labaqua S.A. (Alicante, Spain), a Spain-accredited drinking water laboratory, was in charge of taking and analysing all samples. Inorganics compounds were determined by means of Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Flow cytometry (FCM) analysis were carried to estimate the total cells counts (TCC) in the biofilm samples using SYBR® Green I for staining nuclear double-stranded DNA and the BD Accuri™ C6 software (BD Biosciences, UK) following the manufacture's protocol. To study the structure and composition of the biofilm bacterial communities, DNA was extracted from each biofilm sample and was sent to ADM Lifesequencing S.A. (Valencia, Spain) for Next Generation Sequencing. The bacterial 16s rRNA gen was selected for amplification using the primers spanning the V3 to V4 hypervariable regions [26]. Bioinformatic analysis of 16s rRNA sequences was performed using an optimised laboratory pipeline. Briefly, pair-end sequences were joined, adapters were removed, and a quality trimming was performed. Then, chimeric sequences were identified and filtered and de-novo clustering by 97% similarity to obtain the Operational Taxonomic Units (OTUs). OTU tables were aligned against NCBI database for taxonomical classification to get an overview of the bacterial community structure and composition.

3 RESULTS

Metal screening results showed that main inorganic compound before and after the cleaning and during the biofilm regrowth was iron (Fe), although during the regrowth others like silica or zinc slightly increased their presence (**Figure 4**).

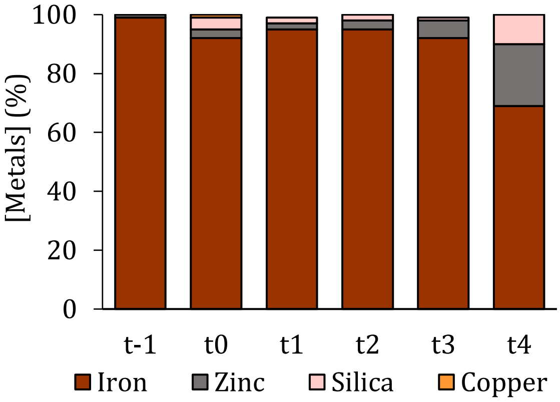


Figure 4. Metal screaning results before swabbing (t-1), after swabbing (t0) and during the biofilm regrowth every 6 months (t1 to t4).

Results from flowcytometry showed an average value for TCC of 1 x 10^7 cells/cm^2 before the cleaning. After the cleaning, 1.8 x 10^6 cells/cm^2 were observed in biofilm samples. Then during biofilm regrowth, t1 sample showed 1 x 10^6 cells/cm^2, t2 sample 4.2 x 10^6 cells/cm^2, t3 sample 3.6 x 10^6 cells/cm^2 and t4 sample 4.0 x 10^6 cells/cm^2.

Taxonomic analysis showed changes in the bacterial community after the swabbing cleaning (**Figure 5**). Before the cleaning, biofilm bacterial community was dominated by *Phreatobacter* (35.34%), *Nitrosomonas* (21.44%), *Methylobacterium* (17.25%) and *Sphingomonas* (14.83%). After swabbing, only three bacteria genera we observed in the community: *Methylobacterium* (69.69%), *Sphingomonas* (11.08%) and *Pseudomonas* (6.74%). During the biofilm regrowth over 2 years, a large number of bacterial genera were observed in the samples. In sample T1, *Blastomonas* (25.10%), *Hydrogenophaga* (9.96%) and *Dechloromonas* (8.34%) dominated the community. Then, in sample T2, only two bacterial genera were identified, *Acinetobacter*

CC BY NC SA

(75.11%) and *Pseudomonas* (23.79%). Sample T3 was dominated by *Aquabacterium* (20.63%), *Methylibium* (12.05%), *Hyphomicrobium* (10.54%) and *Mesorhizobium* (9.38%). In sample T4 the most abundant genera were *Propionivibrio* (20.76%), *Lactobacillus* (16.69%) and *Sphingomonas* (10.33%).

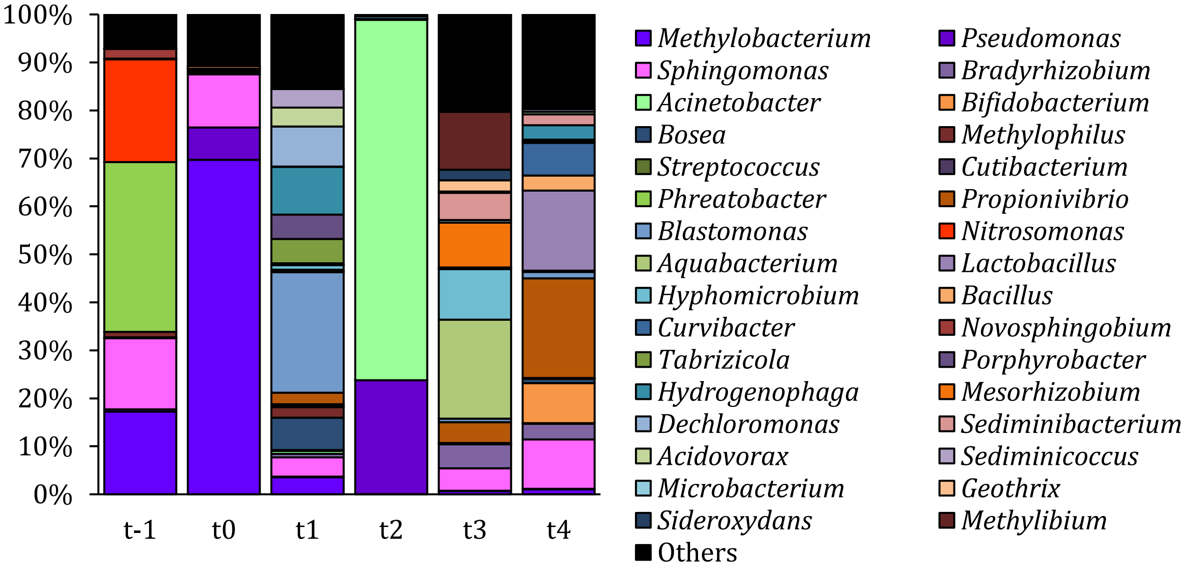


Figure 5. Relative abundance of bacteria at genus level (>1 % of the total sequences) of biofilm samples before swabbing (t-1), after swabbing (t0) and during the biofilm regrowth every 6 months (t1 to t4). Remaining genera were combined in category "Others".

4 DISCUSSION

Results of the biofilm inorganic composition showed that Fe was the most abundant inorganic compound in all samples. This result was expected since it is well known that biofilm facilitate the adsorption and entrapment of materials and particles, especially iron and manganese[10], [27]. Metal release in DWDS has been demonstrated to be affected mainly by pipe material, and especially in pipes of metallic materials [28] such as the ductile iron study pipe. Then, results obtained here show that swabbing cleaning does not affect the inorganic profile of the biofilm and suggests that it is more influenced by the pipe material over other factors. Although Fe does not have a direct impact on human health, the presence of this metal in the biofilm can affect the flavour and colour of the water, and high levels of these contaminants can result in discoloured water. This is the main cause of the loss of aesthetic acceptability, being one of the problems that produces the most complaints from drinking water consumers [28], [29].

Flow cytometry results suggested that although swabbing can reduce the number of cells in the biofilm an order of magnitude, it is not completely effective removing the biofilm deposits from the pipe walls, as it happens with other cleaning methods such as flushing [30].

Focusing on the most abundant taxa (average >1% relative abundance), differences in the taxonomical composition of bacteria were detected in biofilm samples at sampling times (Figure 5). Mature biofilm developed in the studied system (sample t-1 before swabbing) showed a bacterial composition similar to that observed in previous studies of other monochlorinated system in Pinellas (USA), as described by Kelly *et al.* (2014) [31].

Bacterial community was dominated by *Phreatobacter*, *Nitrosomonas*, *Methylobacterium* and *Sphingomonas*. Swabbing affected this composition, and important bacterial genera such as *Nitrosomonas* highly reduced their relative abundance with the cleaning. *Nitrosomonas* is responsible for the oxidation of ammonium to nitrite [32], that can suppose a direct risk to

BY NC SA

consumers, especially in babies, because it can affect how blood carries oxygen and cause methemoglobinemia [33], [34]. In addition, nitrite can react with chloramine, reinforcing its decay, and/or be oxidised to nitrate by nitrite oxidising bacteria, enhancing the biofilm development thus its related water quality problems [32]. Therefore, the reduction of these genera thanks to the swabbing cleaning can suppose positive aspects for the water quality, especially in chloraminated systems such as the one in this study.

Regarding *Phreatobacter,* it has not been previously reported by many studies because it was described few years ago by Toth *et al.* (2014) [35]. This genus was isolated from ultrapure water of a water purification system in Hungary. It has been observed that it is one of the most abundant bacterial genera in drinking water samples from different countries such as Paris, Belgium, Paris or Japan [36]–[38]. Chlorine concentration, residual biodegradable organic matter and temperature have been observed to be abiotic factors influencing the relative abundance of *Phreatobacter* in other DWDSs [36], [38]–[41]. From this study, it can be suggested that swabbing process is able to remove this genus from the biofilm community, since this genus disappears after the cleaning, and does not reappear as the most abundant during biofilm regrowth. Although it could be a clean water bioindicator due to its presence in ultrapure water system, the implications of this genus for water quality have not yet been observed, so more research is needed around this to understand its role in DWDS.

Methylobacterium was the only genus that increased its relative abundance with cleaning. The increased relative abundance of this genus with the swabbing cleaning suggested that *Methylobacterium* grows in the deepest layers of the biofilm and/or because it is a highly resistant genus. These results were agreement with Chaves and Simoes (2010) [42] that found that *Methylobacterium sp*. formed the most resistant biofilms in a model laboratory DWDS in Portugal [43]. This increase does not imply a water quality problem since it has been observed that *Methylobacterium*, even at low concentrations, is able to decrease the concentration of trihalomethanes in drinking water [44].

Sphingomonas is commonly found on DWDs under different conditions [45]. Its relative abundance did not decrease with cleaning, dominating at T0 together with *Methylobacterium* and *Pseudomonas*. *Sphingomonas* and *Pseudomonas* play a key role in DWDS because they are considered as pioneers organisms involved in the initial stages of biofilm development. These two bacteria secrete a great amount of EPS that enhance the growth of biofilms [42], [46]–[48]. Therefore, their presence in the pipe early biofilm after cleaning was not surprising. Although these genera are widely distributed and can be commonly found in water with variable relative abundance, it has been observed that some *Sphingomonas* and *Pseudomonas* species can include species considered opportunistic pathogens and can cause infections in healthy or in immunocompromised people [49]–[52].

During the biofilm regrowth over 2 years, taxonomical changes were also observed in the bacterial composition. Overall, a greater number of genera were observed during the regrowth, including some common bacteria of the community core of DWDS such as *Pseudomonas*, *Sphingomonas*, *Aquabacterium* or *Blastomonas* [53]. These changes suggest a colonization process and competition between microorganisms during biofilm formation after cleaning to shape a more stable community. In addition, of the temporality and natural diversification, changes in the community could also be dependent on seasonality. Samples were taken every 6 months, thus covering summer and winter months. This is especially important in a Mediterranean climate, like the one at the study site, characterized by large temperature changes between seasons. Temperature has been observed to be a determinant abiotic factor that can affect and produce changes in the DWDS microbiome, in terms of microbial growth rate and bacterial community structure and composition over other abiotic factors [8], [33], [40], [54], [55]. In fact, *Acinetobacter* genus dominated the community in sample t2 that was taken in the warmest sampling month, while in the rest of the samples it had a much lower relative abundance.

Acinetobacter is one of the most common groups of bacteria isolated from drinking water [56], and a number of *Acinetobacter* species have been shown to produce biofilms [57], [58]. According to our results, Kelly *et al.* (2014) [31] observed that the high free ammonia concentration combined with the high temperature increased nitrite concentration and the presence of Acinetobacter, accounting more than 70% of the sequences in biofilm summer samples in Pinellas, Florida (USA), as it happened in the sample t2 of the present study.

Therefore, results from this research show that the diversification observed after cleaning in the biofilm of the studied system is due to a process of biofilm colonization and growth, which seems to be influenced by time and temperature and seasonal conditions.

5 CONCLUSION

From this study it can be concluded that swabbing cleaning in DWDS pipes modify the biofilm communities of these systems. Although this method is not capable of completely removing the biofilm deposits, the results showed that swabbing can be an effective tool aimed to minimise the presence and thus the risk associated with the presence of certain bacterial genera. This is a first approach to improve the management of DWDS, but more research is needed that includes other cleaning methods to design effective biofilm control strategies that ensure biological safety of the water that reaches consumers' taps.

6 REFERENCES

[1] I. Douterelo, J. B. Boxall, P. Deines, R. Sekar, K. E. Fish, and C. A. Biggs, “Methodological approaches for studying the microbial ecology of drinking water distribution systems,” Water Res., vol. 65, pp. 134–156, 2014.

[2] A. Gheisi, M. Forsyth, and G. Naser, “Water distribution systems reliability: A review of research literature,” J. Water Resour. Plan. Manag., vol. 142, no. 11, p. 04016047, Nov. 2016.

[3] H.-C. Flemming and J. Wingender, “The biofilm matrix,” Nat. Rev. Microbiol., vol. 8, no. 9, pp. 623–633, 2010.

[4] H.-C. Flemming, “Biofouling in water systems -- cases, causes and countermeasures,” Appl. Microbiol. Biotechnol., vol. 59, no. 6, pp. 629–640, Sep. 2002.

[5] H.-C. Flemming, “EPS-Then and Now,” Microorganisms, vol. 4, no. 4, p. 41, Nov. 2016.

[6] A. L. S. Dos Santos et al., “What are the advantages of living in a community? A microbial biofilm perspective!,” Mem. Inst. Oswaldo Cruz, vol. 113, no. 9, pp. e180212–e180212, Jul. 2018.

[7] P. Sivadon, C. Barnier, L. Urios, and R. Grimaud, “Biofilm formation as a microbial strategy to assimilate particulate substrates,” Environ. Microbiol. Rep., vol. 11, no. 6, pp. 749–764, Dec. 2019.

[8] S. Liu, C. Gunawan, N. Barraud, S. A. Rice, E. J. Harry, and R. Amal, “Understanding, monitoring, and controlling biofilm growth in drinking water distribution systems,” Environ. Sci. Technol., vol. 50, no. 17, pp. 8954–8976, 2016.

[9] L. Chaves Simões and M. Simões, “Biofilms in drinking water: problems and solutions,” RSC Adv., vol. 3, no. 8, pp. 2520–2533, 2013.

[10] P. S. Husband and J. B. Boxall, “Asset deterioration and discolouration in water distribution systems,” Water Res., vol. 45, no. 1, pp. 113–124, 2011.

[11] E. I. Prest, F. Hammes, M. C. M. van Loosdrecht, and J. S. Vrouwenvelder, “Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges ,” Frontiers in Microbiology , vol. 7. p. 45, 2016.

[12] I. J. H. G. Vreeburg and D. J. B. Boxall, “Discolouration in potable water distribution systems: A review,” Water Res., vol. 41, no. 3, pp. 519–529, 2007.

[13] R. G. Ainsworth, Safe piped water. Iwa Publishing, 2004.

[14] M. Friedman et al., Establishing site-specific flushing velocities. AWWA Research Foundation Denver, 2003.

[15] J. Vreeburg, “Discolouration in drinking water systems: a particular approach,” Jan. 2007.

[16] G. Quarini et al., “Investigation and development of an innovative pigging technique for the water-supply industry,” Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng., vol. 224, no. 2, pp. 79–89, Mar. 2010.

[17] P. Jurek Vidlářová and S. Heviánková, “Comparison of Modern Drinking Water Network Maintenance Methods: Evaluation of Removed Deposits in the Form of Total Suspended Solids (TSS),” Int. J. Environ. Res. Public Health, vol. 18, no. 8, p. 4311, 2021.

[18] P. Kitney, R. Woulfe, and S. Codd, “AIR SCOURING OF WATER MAINS-AN ASSET MANAGEMENT APPROACH,” in 64th Annual Water Industry Engineers and Operators Conference, Bendingo, Australia, 2001.

[19] F. Pourcel, I. E. Smith, and S. Duchesne, “Slug Flow Simulation, A Way to Improve Air Scouring of Water Mains?,” Procedia Eng., vol. 186, pp. 601–608, 2017.

[20] C. Bae, D. Lee, D.-Y. Choi, H. Jun, S. Park, and T. Choi, “An assessment of the effect of air scouring and swabbing pig cleaning technique on water distribution pipes,” J. Korean Soc. Water Wastewater, vol. 29, no. 4, pp. 459–468, Aug. 2015.

[21] J. Rubulis, J. Verberk, J. Vreeburg, K. Gruškevica, and T. Juhna, “Chemical and microbial composition of loose deposits in drinking water distribution systems,” in Proceedings of the, 7th International Conference on Environmental Engineering, May, 2008, pp. 22–23.

[22] I. Garcia de Carellan, P. Catton, C. Selcuk, and T.-H. Gan, “Methods for detection and cleaning of fouling in pipelines,” Emerg. Technol. Non-Destructive Test. V, p. 231, 2012.

[23] J. N. H. Tiratsoo, Pipeline pigging technology. Gulf Professional Publishing, 1992.

[24] R. Miller, E. Duderstadt, and P. Treloar, “Ice Pigging Creates Clean Mains and Water Quality Insights,” Opflow, vol. 45, no. 7, pp. 10–14, Jul. 2019.

[25] D. Ellison, Investigation of pipe cleaning methods. American Water Works Association, 2003.

[26] A. Klindworth et al., “Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies,” Nucleic Acids Res., vol. 41, no. 1, pp. e1–e1, Jan. 2013.

[27] M. P. Ginige, J. Wylie, and J. Plumb, “Influence of biofilms on iron and manganese deposition in drinking water distribution systems,” Biofouling, vol. 27, no. 2, pp. 151–163, Jan. 2011.

[28] J. Fawell and M. J. Nieuwenhuijsen, “Contaminants in drinking water: Environmental pollution and health,” Br. Med. Bull., vol. 68, no. 1, pp. 199–208, Dec. 2003.

[29] P. S. Husband and J. B. Boxall, “Asset deterioration and discolouration in water distribution systems,” Water Res., vol. 45, no. 1, pp. 113–124, Jan. 2011.

[30] I. Douterelo, R. L. Sharpe, and J. B. Boxall, “Influence of hydraulic regimes on bacterial community structure and composition in an experimental drinking water distribution system,” Water Res., vol. 47, no. 2, pp. 503–516, 2013.

[31] J. J. Kelly, N. Minalt, A. Culotti, M. Pryor, and A. Packman, “Temporal variations in the abundance and composition of biofilm communities colonizing drinking water distribution pipes,” PLoS One, vol. 9, no. 5, 2014.

[32] S. C. Potgieter, Z. Dai, S. N. Venter, M. Sigudu, and A. J. Pinto, “Microbial Nitrogen Metabolism in Chloraminated Drinking Water Reservoirs,” bioRxiv, p. 655316, Jan. 2019.

[33] K. D. M. Pintar and R. M. Slawson, “Effect of temperature and disinfection strategies on ammonia-oxidizing bacteria in a bench-scale drinking water distribution system,” Water Res., vol. 37, no. 8, pp. 1805–1817, 2003.

[34] WHO, “Nitrite in drinking-water background document for development of WHO guidelines for drinking-water quality,” Geneva World Heal. Organ., 2011.

[35] E. M. Toth et al., “Phreatobacter oligotrophus gen. nov., sp. nov., an alphaproteobacterium isolated from ultrapure water of the water purification system of a power plant,” Int. J. Syst. Evol. Microbiol., vol. 64, no. Pt 3, pp. 839–845, Mar. 2014.

[36] Y. Perrin, D. Bouchon, V. Delafont, L. Moulin, and Y. Héchard, “Microbiome of drinking water: A full-scale spatio-temporal study to monitor water quality in the Paris distribution system,” Water Res., vol. 149, pp. 375–385, 2019.

[37] A. Van Assche, S. Crauwels, J. De Brabanter, K. A. Willems, and B. Lievens, “Characterization of the bacterial community composition in water of drinking water production and distribution systems in Flanders, Belgium,” Microbiologyopen, vol. 8, no. 5, pp. e00726–e00726, May 2019.

[38] I. Rahmatika, I. Kasuga, F. Kurisu, and H. Furumai, “Impacts of Organic Matter Migrating from Pipe Materials on Microbial Regrowth in Drinking Water,” J. Water Environ. Technol., vol. 18, no. 1, pp. 45–53, 2020.

[39] L. F. Stanish et al., “Factors Influencing Bacterial Diversity and Community Composition in Municipal Drinking Waters in the Ohio River Basin, USA,” PLoS One, vol. 11, no. 6, p. e0157966, 2016.

[40] C. Calero Preciado, J. Boxall, V. Soria-Carrasco, S. Martínez, and I. Douterelo, “Implications of Climate Change: How Does Increased Water Temperature Influence Biofilm and Water Quality of Chlorinated Drinking Water Distribution Systems? ,” Frontiers in Microbiology , vol. 12. p. 1361, 2021.

[41] X. Ma et al., “Spatial variation of loose deposit characteristics in a 40 km long operational drinking water distribution system,” Environ. Sci. Water Res. Technol., vol. 5, no. 10, pp. 1689–1698, 2019.

[42] L. C. Simões, M. Simões, and M. J. Vieira, “Influence of the diversity of bacterial isolates from drinking water on resistance of biofilms to disinfection,” Appl. Environ. Microbiol., vol. 76, no. 19, pp. 6673–6679, Oct. 2010.

[43] L. C. Simões, M. Simões, R. Oliveira, and M. J. Vieira, “Potential of the adhesion of bacteria isolated from drinking water to materials,” J. Basic Microbiol., vol. 47, no. 2, pp. 174–183, 2007.

[44] E. Tsagkari and W. T. Sloan, “Impact of Methylobacterium in the drinking water microbiome on removal of trihalomethanes,” Int. Biodeterior. Biodegrad., vol. 141, no. August 2018, pp. 10–16, 2019.

[45] M. W. Cowle, G. Webster, A. O. Babatunde, B. N. Bockelmann-Evans, and A. J. Weightman, “Impact of flow hydrodynamics and pipe material properties on biofilm development within drinking water systems,” Environ. Technol., 2019.

[46] A. R. Johnsen, M. Hausner, A. Schnell, and S. Wuertz, “Evaluation of fluorescently labeled lectins for noninvasive localization of extracellular polymeric substances in Sphingomonas biofilms,” Appl. Environ. Microbiol., vol. 66, no. 8, pp. 3487–3491, Aug. 2000.

[47] Y. Irie et al., “Self-produced exopolysaccharide is a signal that stimulates biofilm formation in Pseudomonas aeruginosa,” Proc. Natl. Acad. Sci. U. S. A., vol. 109, no. 50, pp. 20632–20636, Dec. 2012.

[48] Y. E. Navarro-Noya et al., “Pyrosequencing Analysis of the Bacterial Community in Drinking Water Wells,” Microb. Ecol., vol. 66, no. 1, pp. 19–29, 2013.

[49] D. J. Jurgens, S. A. Sattar, and T. F. Mah, “Chloraminated drinking water does not generate bacterial resistance to antibiotics in Pseudomonas aeruginosa biofilms,” Lett. Appl. Microbiol., vol. 46, no. 5, pp. 562–567, May 2008.

[50] I. Vaz-Moreira, O. C. Nunes, and C. M. Manaia, “Diversity and antibiotic resistance in Pseudomonas spp. from drinking water,” Sci. Total Environ., vol. 426, pp. 366–374, 2012.

[51] K. Luhring, Bacterial communities in drinking water biofilms, no. May. 2016.

[52] J. P. Steinberg and E. M. Burd, “238 - Other Gram-Negative and Gram-Variable Bacilli,” J. E. Bennett, R. Dolin, and M. J. B. T.-M. Blaser Douglas, and Bennett’s Principles and Practice of Infectious Diseases (Eighth Edition), Eds. Philadelphia: Content Repository Only!, 2015, pp. 2667-2683.e4.

[53] WHO, WHO drinking water guideline. 2017.

[54] N. . Hallam, J. . West, C. . Forster, and J. Simms, “The potential for biofilm growth in water distribution systems,” Water Res., vol. 35, no. 17, pp. 4063–4071, Dec. 2001.

[55] I. Delpla, A. V. Jung, E. Baures, M. Clement, and O. Thomas, “Impacts of climate change on surface water quality in relation to drinking water production,” Environ. Int., vol. 35, no. 8, pp. 1225–1233, 2009.

[56] U. Szewzyk, R. Szewzyk, W. Manz, and K.-H. Schleifer, “Microbiological Safety of Drinking Water,” Annu. Rev. Microbiol., vol. 54, no. 1, pp. 81–127, Oct. 2000.

[57] S. K. Hansen, P. B. Rainey, J. A. J. Haagensen, and S. Molin, “Evolution of species interactions in a biofilm community,” Nature, vol. 445, no. 7127, pp. 533–536, 2007.

[58] A. P. Tomaras, C. W. Dorsey, R. E. Edelmann, and L. A. Actis, “Attachment to and biofilm formation on abiotic surfaces by Acinetobacter baumannii: involvement of a novel chaperone-usher pili assembly system,” Microbiology, vol. 149, no. 12, pp. 3473–3484, 2003.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14627

DYNAMIC EDGE BETWEENNESS CENTRALITY AND OPTIMAL DESIGN OF WATER DISTRIBUTION NETWORKS

Mohsen Hajibabaei[1], Sina Hesarkazzazi[2], Amin Minaei[3], Dragan Savić[4] and Robert Sitzenfrei[5]

[1,2,3,5]Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Innsbruck, Austria
[4]KWR Water Research Institute, Nieuwegein, The Netherlands
[4]Centre for Water Systems, University of Exeter, Exeter, United Kingdom

[1] *Mohsen.Hajibabaei@uibk.ac.at,* [2] *Sina.Hesarkazzazi@uibk.ac.at,* [3] *Amin.Minaei@uibk.ac.at,*
[4] *, Dragan.Savic@kwrwater.nl,* [5] *Robert.Sitzenfrei@uibk.ac.at.*

Abstract

The multi-objective design of water distribution networks (WDNs) as a nonlinear optimization problem is a challenging task. With two contradicting objectives (e.g., minimizing costs and maximizing resilience), Pareto fronts of optimal solutions can be obtained with, e.g., evolutionary algorithms. However, the main drawback of these algorithms is the high computational effort required to optimize large WDNs. Recently, a highly efficient method based on complex network theory (CNT) was developed, where within seconds, a wide range of Pareto near-optimal solutions can be obtained for the design of WDNs (i.e., determining optimal diameters). The developed method is based on a customized graph measure called demand edge betweenness centrality (EBC^Q). This measure is based on the frequency of occurrence of an edge in the shortest path from a source node to a demand node. In addition, EBC^Q sums up the demands routed through that edge, giving a valid flow estimation for an optimal design. In the graph of a WDN, the edges can have different weights. The weighting function used for EBC^Q calculations can be 'static' or 'dynamic'. A constant value is utilized for edge weights in the static weighting approach, while a dynamic weighting function implies that edge weights are modified when iterating through all demand nodes. In this context, using dynamic weighting functions for EBC^Q (i.e., dynamic EBC^Q) avoids concentrating EBC^Q values in just a few edges (shortest-path trees) by considering redundancy in flow paths and better approximation of the hydraulic behavior. However, it is not clear how the parameters of dynamic weighting functions should be defined to achieve the best approximation of the Pareto-optimal front. This work performs a systematic investigation of dynamic weighting functions and gives guidance for optimal parameter selection. The comparative study between the CNT approach (with static and dynamic weights) and evolutionary optimizations on four WDN design problems confirms the capability of the proposed dynamic functions in providing optimal/near-optimal solutions.

Keywords
Graph theory, Multi-objective evolutionary algorithms, Pareto fronts, Complex network analysis.

1 INTRODUCTION

Water distribution networks (WDNs) as pivotal urban infrastructures are comprised of various components interacting in a complex way [1]. Apart from the inherent complexity, major financial investments are required for the construction, operation, and renovation of WDNs. Therefore, various aspects of these infrastructures have come into focus of optimization methods to identify the best solutions ensuring maximum benefits (e.g., resilience, performance) with minimum costs [2]. Evolutionary algorithms (EAs) are the most widely used approaches to solve multi-objective optimization problems in WDNs [3]. However, using EAs for optimal design problems (i.e., selecting optimal pipe diameters), is a major challenge, particularly when considering real-world large-scale networks [4]. The reason derives from the fact that the design problem is a nonlinear and NP-hard problem [5], which cannot be tackled in a reasonable time frame for large WDNs, as the size of search space grows exponentially [6]. Several attempts have been made to improve the efficiency of EAs by reducing the search space during the optimization procedure [3], [6], [7], [8]. Nevertheless, finding optimal/near-optimal solutions for complex and large WDNs remains a computationally expensive task since multiple runs and numerous function evaluations are still required [9]. Hence, a fast and efficient approach is needed to overcome the limitations of EAs and achieve the best approximation of optimal design solutions. To this end, complex network theory (CNT) approaches are of interest. In CNT, WDNs are converted to mathematical graphs, including sets of nodes (junctions) and edges (pipes), where the relationships between components of complex networks can be described more efficiently. In addition, CNT is applicable to large-scale WDNs (with thousands to millions of pipes) in a relatively short execution time [4].

CNT has been employed in various aspects of analyses of WDNs, such as resilience assessment [10], temporal evolution [11], district metered areas creation [12], water quality analysis [13], or vulnerability assessment [14]. CNT has also gained attention in WDN optimizations. Sitzenfrei et al [15] explored patterns and network characteristics of Pareto-optimal solutions based on different graph measures. They showed that using a centrality metric called edge betweenness centrality (EBC) could drive the layout of low resilient optimal solutions. Giustolisi et al. [16] customized certain centrality metrics according to WDNs characterization to predict hydraulic behavior. One of those customized metrics (i.e., tailored edge betweenness) was later used as a topological measure to narrow down the search space of EAs for pipe re-sizing optimization [3]. Sitzenfrei et al. [17] developed a fast and efficient CNT-based design approach for WDNs, where a range of near-optimal solutions are obtained without conducting hydraulic simulation. The proposed approach is a based on a modified EBC measure denoted as demand EBC ($\mathrm{EBC}^{\mathrm{Q}}$). This measure determines how often a pipe (edge) is part of the shortest path from a reservoir/tank (source node) to a demand node, and it sums up the demands routed through each path. For identifying shortest paths, the edges of a graph are weighted, and often the Euclidean length is used as the edge weighting function. However, weighting functions can be constant (i.e., static) or changed iteratively (i.e., dynamic), which could result in obtaining different design solutions.

It was shown that choosing proper dynamic weighting functions could noticeably improve the quality of design solutions compared to the results obtained with static weighting functions [18]. However, no systematic attempts have been made to discover the potential of using different dynamic weighting functions. To fill this gap, two dynamic weighting functions for WDN design are proposed and their parameters are tailored to reproduce hydraulic behavior of WDNs. The proposed functions are tested in four case studies, including three known benchmark problems and one real-world large network, and the results are compared with those obtained with multi-objective optimization using EAs. The outcome of this paper gives a better insight into applying proper dynamic weights for the optimal design of WDNs, leading to a move toward higher quality design solutions based on CNT.

2 MATERIAL AND METHODS

In the first section of methods (2.1), graph measures based on CNT used for designing WDNs are described, and the procedure is explained with a simple illustrative example. In section 2.2 the tool and literature used to find multi-objective optimization results based on EAs are introduced. Also, detailed descriptions of considered case studies are provided in section 2.3.

2.1 WDN design using CNT

Graph measures used for CNT

BY NC SA

In CNT, urban water infrastructures are presented as a mathematical graph G consisting of sets of vertices/nodes ($\#N$) connected via a set of edges/pipes ($\#E$). In a WDN, demand nodes (sinks) and source nodes (reservoirs/tanks) are represented with D and S, respectively, which are subsets of N. Depending on the analysis objective, various weights can be assigned to edges/nodes. For instance, pipe length (Euclidean distance) can be used as edge weight for WDN design. In this work, the following two graph measures are utilized for CNT-based design [17]:

The first measure is the shortest path length indicated with SP, which is utilized to describe the shortest distance between two nodes. Distance in this context refers to the sum of positive edge weights in a connecting path between two nodes [19]. As an example, calculating SP with the weight of pipe length from a source to a demand node results in finding a path between them where the pipe length is minimal.

The second measure is the single-source EBC, which describes the frequency of occurrence of an edge in the SP from S (source node) to every demand node ($i \in N$) [15]. This measure was modified for WDN design denoted as demand EBC (EBC^Q) [17]. EBC^Q finds SP connecting S (source node) and every demand node $i \in D$, and adds the nodal demands Qi ($Q_i > 0$) to the EBC^Q values of all edges along the $SP_{S,i}$. For instance, EBC^Q of an edge e is formulated regarding Equation (1):

$$EBC^Q(e) = \sum_{S,i \in D} SP_{S,i}(e) \cdot Q_i \tag{1}$$

One can realize that EBC^Q value of each edge (pipe) can be used for estimating water flow in that pipe. On the other hand, the EBC^Q value itself depends on the weights chosen for SP. As mentioned, pipe length is often employed for weighting edges in the design procedure. However, the value of edge weights can be constant or modified with a function in the course of the EBC^Q calculation, denoted as static and dynamic weights, respectively [17]. Dynamic weights use alternative paths for routing the demand, which are not necessarily the shortest path but could be for example the second shortest path. Therefore, dynamic weights avoid concentrating EBC^Q values in just a few pipes (shortest-path tree) by considering redundant flow paths. This idea is based on the concept of energy balance in WDNs. One method to define dynamic weights could be increasing pipe length in a path (shortest path) to a maximum value (i.e., a threshold) per iteration with a function approximating increased friction losses in that path due to an increased flow. This paper gives guidance for selecting the proper parameters of such functions and thresholds used for defining dynamic weights. Detailed information regarding suggested dynamic weighting functions is described in the following section. Note that hereinafter, 'dynamic $\mathrm{EBC}^{\mathrm{Q}}$' refers to EBC^Q calculation with dynamic weighting functions.

Design procedure using static and dynamic $\mathrm{EBC}^{\mathrm{Q}}$

The EBC^Q value of an edge gives an estimation of the volumetric flow rate in that edge (pipe), which can be employed for pipe sizing of WDNs. Thus, according to the continuity equation, pipe diameter DN_e required for the edge e with $EBC^Q(e)$ is estimated as follows [17]:

$$DN_e = \left\lceil \sqrt{\frac{4}{\pi} \cdot \frac{EBC^Q(e)}{V_{design}}} \right\rceil \in DN_{available} \tag{2}$$

V_{design} in Equation (2) is the assumed velocity used for pipe sizing. In order to design a pipe based on the CNT approach, the corresponding EBC^Q of the pipe is first calculated, and then the required diameter is estimated by assigning a value between 0.5 to 2.5 m/s to V_{design}. Therefore, various ranges of solutions can be found by varying the design velocity. For instance, with 0.01 m/s steps, 201 different design options can be created. Note that no hydraulic stimulation is involved during the course of pipe sizing. It is only after the design that a hydraulic solver (Epanet2) is used to examine solutions to ensure required constraints (e.g., required minimum pressure) are met. An illustrative simple WDN is shown in Figure 1 to explain the CNT design method with static and dynamic EBC^Q.

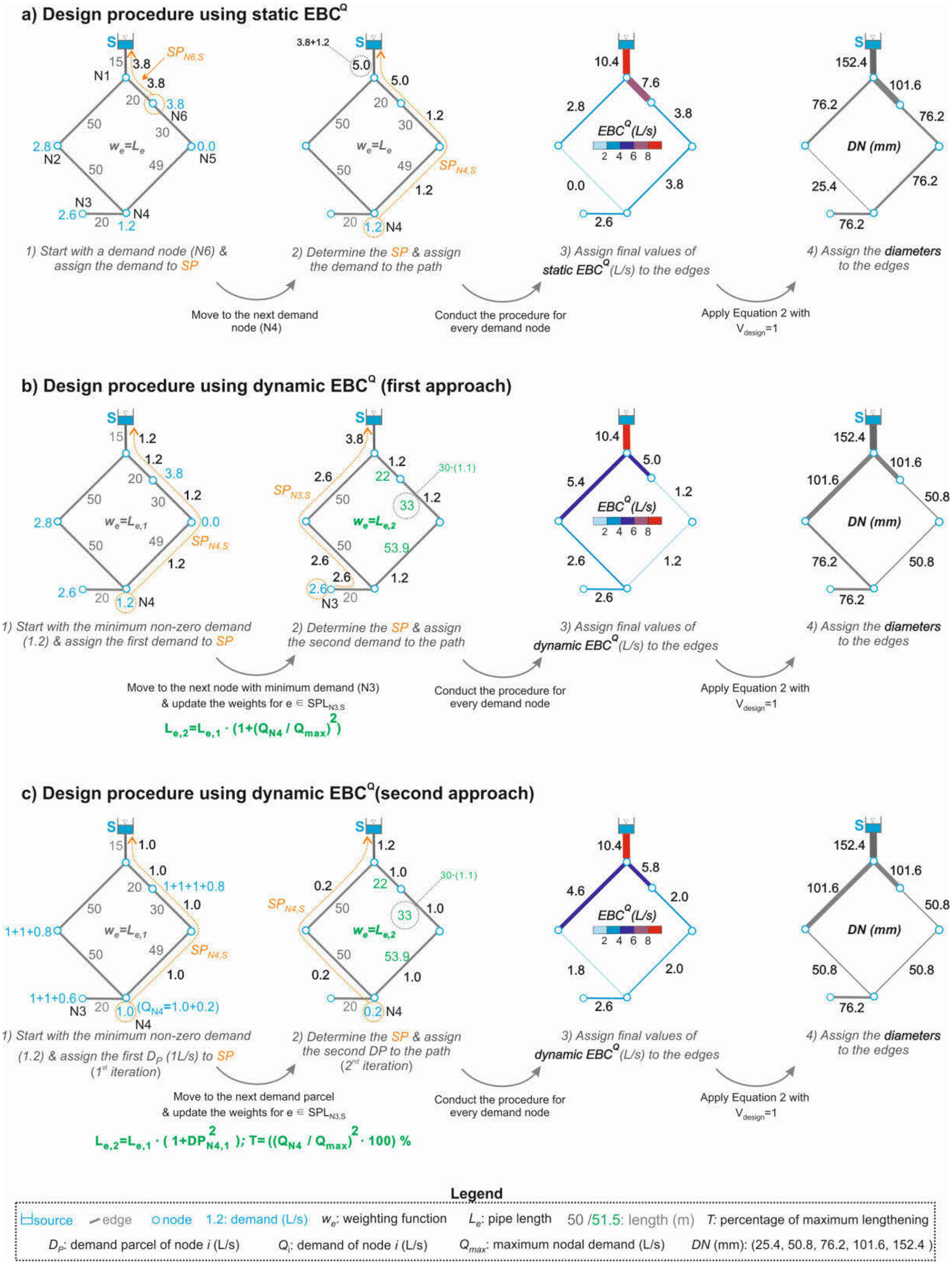

a) Design procedure using static EBC^Q
1) Start with a demand node (N6) & assign the demand to SP
2) Determine the SP & assign the demand to the path
3) Assign final values of static EBC^Q (L/s) to the edges
4) Assign the diameters to the edges
Move to the next demand node (N4)
Conduct the procedure for every demand node
Apply Equation 2 with V_design=1
S
N1
N2
N3
N4
N5
N6
SP_N6,S
SP_N4,S
w_e=L_e
EBC^Q(L/s)
DN (mm)
3.8+1.2
b) Design procedure using dynamic EBC^Q (first approach)
1) Start with the minimum non-zero demand (1.2) & assign the first demand to SP
2) Determine the SP & assign the second demand to the path
3) Assign final values of dynamic EBC^Q (L/s) to the edges
4) Assign the diameters to the edges
Move to the next node with minimum demand (N3) & update the weights for e ∈ SPL_N3,S
L_e,2=L_e,1 · (1+(Q_N4 / Q_max)^2)
Conduct the procedure for every demand node
Apply Equation 2 with V_design=1
w_e=L_e,1
w_e=L_e,2
SP_N3,S
30·(1.1)
c) Design procedure using dynamic EBC^Q(second approach)
1) Start with the minimum non-zero demand (1.2) & assign the first D_P (1L/s) to SP (1st iteration)
2) Determine the SP & assign the second DP to the path (2nd iteration)
3) Assign final values of dynamic EBC^Q (L/s) to the edges
4) Assign the diameters to the edges
Move to the next demand parcel & update the weights for e ∈ SPL_N3,S
L_e,2=L_e,1 · (1+DP^2_N4,1); T= ((Q_N4 / Q_max)^2 · 100) %
Conduct the procedure for every demand node
Apply Equation 2 with V_design=1
1+1+1+0.8
1+1+0.8
1+1+0.6
(Q_N4=1.0+0.2)
Legend
source
edge
node
1.2: demand (L/s)
w_e: weighting function
L_e: pipe length
50 /51.5: length (m)
T: percentage of maximum lengthening
D_P: demand parcel of node i (L/s)
Q_i: demand of node i (L/s)
Q_max: maximum nodal demand (L/s)
DN (mm): (25.4, 50.8, 76.2, 101.6, 152.4)

Figure 1: CNT design procedure using static and dynamic EBC^Q

Before initiating the procedure, the WDN Epanet2 file needs to be converted to a graph object, including the network's information (e.g., nodal demands, pipe length). In the next step, pipe lengths are assigned to the graph edges as weights. In order to design pipe diameters, we need to distribute the nodal demands through the edges (pipes). This process is conducted with the EBC^Q measure, which can be either based on using static (static $\mathrm{EBC^Q}$) or dynamic weights (dynamic $\mathrm{EBC^Q}$). Required steps for the static $\mathrm{EBC^Q}$ design are described in Figure 1a. In static $\mathrm{EBC^Q}$ the procedure can be initiated with any node. In Figure 1a, the design process starts for example with N6, and then its corresponding nodal demand (3.8L/s) is routed to the path between N6 and S with the shortest path length. We conduct this process for all the demand nodes resulting in EBC^Q value of each edge using static weights. As shown in the third step of Figure 1a, EBC^Q values are concentrated in the shortest path (i.e., right branch), implying that the existing redundant capacity in the left branch is overlooked since it is a slightly longer route. However, based on the energy balance concept in the looped networks, alternative paths' capacity should also be considered to balance friction losses along different pipes. This issue can be addressed in CNT design by utilizing dynamic $\mathrm{EBC^Q}$. Dynamic $\mathrm{EBC^Q}$ design is conducted based on two approaches shown in Figures 1b and 1c. In the first dynamic $\mathrm{EBC^Q}$ approach (see Figure 1b), we initiate the design process with the node that has the minimum non-zero demand (N4) and route its corresponding nodal demand (1.2L/s) through the edges located in the right-hand branch (the orange-colored path where SP is minimal). Moving to the next larger demand node (N3), the edge weights of the orange-colored path of the first iteration are increased (updated) using a dynamic function, resulting in routing the demand of N3 to the left-hand branch (i.e., the second shortest path before weights updating). The proposed dynamic function updates the weights (pipe length) in $j+1^{th}$ iteration using the following equation:

$$L_{e,j+1} = L_{e,j}.\left(1 + \left(Q_{i,j}/Q_{max}\right)^2\right) \tag{3}$$

Where, $L_{e,j}$ is the length of each edge e in the shortest path between demand node i and source node S in j^{th} iteration (m), $Q_{i,j}$ is the demand of node i in j^{th} iteration j (L/s), and Q_{max} is the maximum nodal demand (L/s). Note that j in this equation is between 1 and the total number of demand nodes.

It is worth mentioning that the idea of using this dynamic weighting function is derived from the quadratic relationship between water flow and friction losses in the Darcy-Weisbach equation. By conducting this procedure for all the demand nodes and updating weights in each iteration, EBC^Q values of edges are calculated. Comparing the obtained values for static and dynamic $\mathrm{EBC^Q}$ in Figures 1a and 1b indicates that the potential capacity of the left-hand branch is further utilized using the dynamic approach, better approximating the energy balance in loops.

The second method for determining dynamic $\mathrm{EBC^Q}$ divides the demand of each node into smaller parcels (D_P) to reproduce flow division in a looped network (see Figure 1c). For instance, two demand parcels ($D_P = D_{P,1} + D_{P,2}$) with the values of 1 L/s and 0.2 L/s are considered for N4 with a total nodal demand of 1.2 L/s. After routing D_P of each nodal demand to its corresponding SP in the j^{th} iteration, the related edge weights are increased by the following equation in $j+1^{th}$ iteration as follows:

$$L_{e,j+1} = L_{e,j}.\left(1 + D_{P,i,j}{}^2\right) \tag{4}$$

Where, $D_{P,i,j}$ is the demand parcel of node i in j^{th} iteration (L/s). Note that j in this equation is between 1 and the total number of demand parcels.

The proposed function in Equation (4) could excessively increase the weight of a path, especially when a network has a few large demands. In order to avoid this issue, we limit the maximum lengthening percentage in each iteration. Therefore, the corresponding threshold (T) is calculated as $T = ((Q_i/Q_{max})^2 \cdot 100)\%$, where (Q_i) is nodal demand and (Q_{max}) is maximum demand.

The capacity of neglected paths in the static approach is activated with the proposed dynamic EBC^{Q}, resulting in creating loops in the course of design procedures. This issue leads to assigning more capacity to the left-hand branch of the WDN in Figures 1b and c, compared to those obtained in Figure 1a. In this work, CNT design with static and dynamic EBC^{Q} approaches are systematically tested on four design problems to select the appropriate method and give guidance for future applications. The selected approaches are then compared with the results found by multi-objective optimization using EAs.

2.2 Multi-objective optimization based EAs

The best outcome of the CNT approach is compared to those derived from multi-objective optimization based on EAs to validate the results and compare the computational efficiency of procedures. For benchmark case studies, the best-known solutions obtained based on five state-of-the-art EAs in literature are used, referred to as best-known Pareto fronts (BPF) [20]. For the real case, we solve the optimization problem using the state-of-the-art tool called GALAXY [9]. GALAXY was proposed based on the framework of multi-objective EAs to deal with the multi-objective and combinatorial design of WDNs [9]. Detailed information regarding the methodology and algorithms used for this tool was described in [9].

In both BPF and GALAXY, the multi-objective design is conducted using two conflicting objectives, i.e., maximizing resilience and minimizing total costs. The resilience is formulated according to a measure proposed by Prasad and Park [21], taking the surplus nodal head, as well as pipes' uniformity, into account. Besides, total costs are calculated based on unit pipe expenses (as a function of diameter) and the corresponding pipe length.

Parameters of optimization problems (e.g., diameter classes, pressure constraints) are selected according to the characteristics of case studies, which are described in the next section.

2.3 Case studies

Four WDNs, including three benchmark problems and one real network, are considered the case studies. The first case (Figure 2a) is known as the two-loop network (TLN) [22], which is a hypothetical WDN consisting of six demand nodes, a single reservoir, and eight pipes with an equal length of 1000 m. The second case (Figure 2b) is a subdivision of the WDN in Blacksburg town located in Virginia [23]. This network comprises 33 demand nodes, a reservoir, and 32 pipes, 12 of which have fixed diameters. The third case (Figure 2c) is BakRyun (BAK) network in South Korea which has 35 demand nodes, a single reservoir, and 58 pipes [24]. The fourth case is a real-world large network located in Austria, comprising 3,558 nodes, one reservoir, and 4,021 pipes with a total length of 211 km (the layout of this case study is anonymized, however, the real hydraulics are preserved).

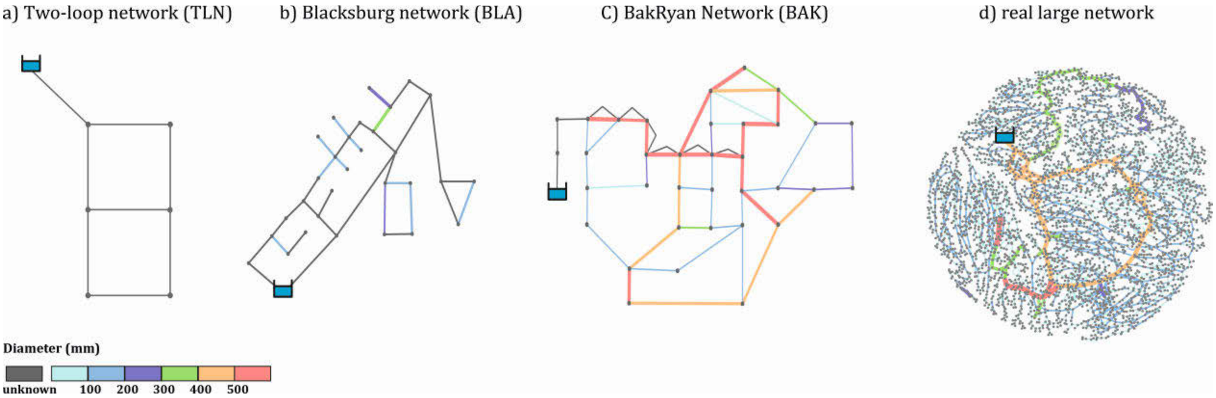

a) Two-loop network (TLN)
b) Blacksburg network (BLA)
C) BakRyan Network (BAK)
d) real large network

Figure 2: Layouts of considered case studies

BY NC SA

BAK has been presented as an 'extended design (rehabilitation) problem' with the objective of finding the best possible diameters for the parallel and new pipes shown in Figure 2c. While for TLN, BLA, and real networks, the aim is to determine the possible diameters of all pipes, which is considered a 'design problem'. The minimum pressure constraint is considered 15 m for BAK and 30 m for other WDNs. In addition to minimum pressure, the maximum pressure of each demand node in the BLA network is limited to a particular value presented in [23]. For benchmark problems, diameter options and their costs are defined according to the values suggested in the literature [20]. For the real network, 15 diameter classes from 76.2 to 914.4 mm with the unit costs ranging from 8 to 1,200 $/m are considered. The optimization of this network is solved using GALAXY, with 500,000 and 100,000 generations and population size of 100. Regarding the size of the search space, the problems can be categorized into three groups: small (TLN and BAK), Intermediate (BLA), and very large (real WDN). As an example, a complete enumeration of $15^{4,021}$ ($DN^{\#E}$) different design solutions are needed to optimize the real network using EAs, making it a very large design problem.

3 RESULTS AND DISCUSSION

3.1 Comparing the results of static and dynamic EBC^Q approaches

The obtained CNT results using static and dynamic EBC^Q are indicated in Figure 3. Dynamic EBC^Q (1) and EBC^Q (2) in this figure refer to the results found based on the first and second dynamic methods, respectively.

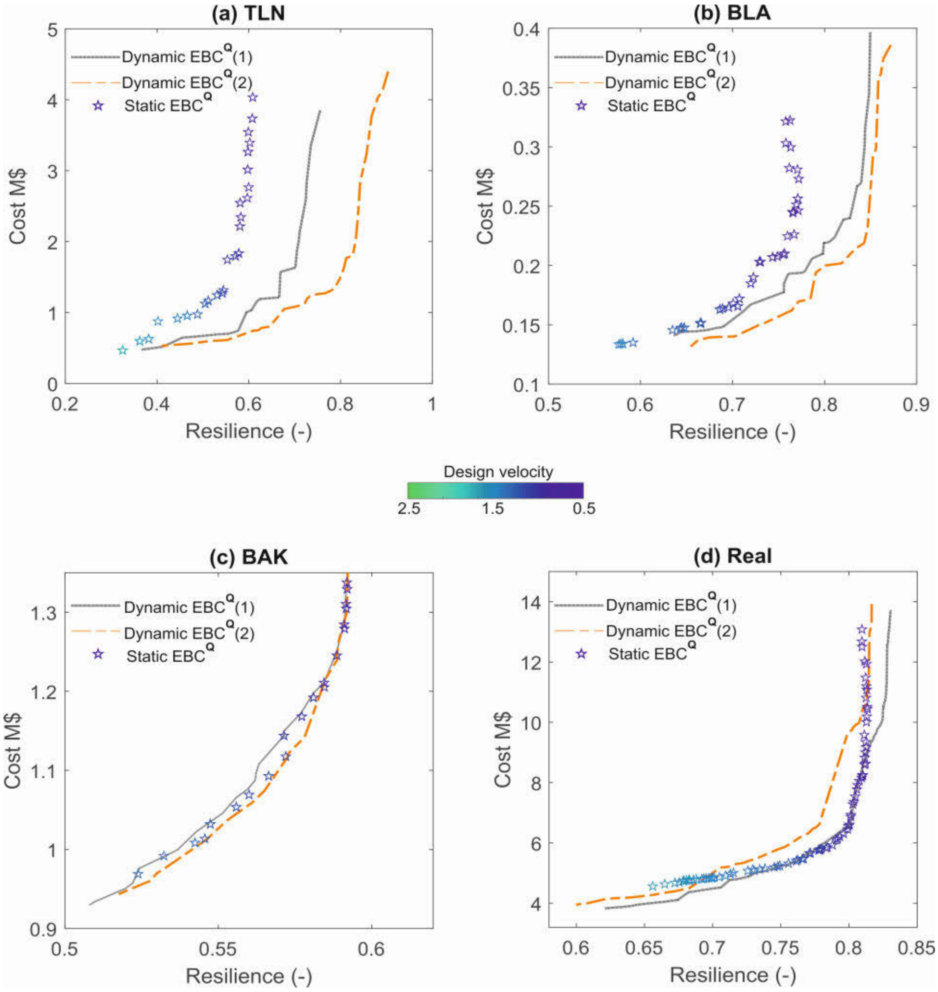


Figure 3: Comparing the results of static and dynamic EBC^Q approaches for different case studies

At the end of the CNT procedure, we examined all the obtained solutions using Epanet2 to exclude those which do not meet the pressure constraints. For instance, the pressure constraint for BLA cannot be fulfilled for the obtained solutions based on static EBC^Q with the design velocity larger than 1.7 m/s.

A lower number of unique design solutions (i.e., design options with different values of resilience and cost) is obtained for benchmark cases based on static EBC^Q compared to those found for the large WDN. For instance, varying the design velocity in the range of 0.5-2.5 m/s with 0.01 m/s steps using the static approach results in 20 and 113 unique design options for BAK and real WDNs, respectively (see Figures 3c and 3d). This issue can be clarified by simple structures and less redundancy associated with benchmark networks in comparison with the real WDN with more than 4,000 pipes.

Figures 3a and 3b indicate that applying proposed dynamic approaches to TLN and BLA networks yields a better trade-off between the resilience and costs compared to the static approach. TLN is a looped network consisting of equal pipe lengths. Therefore, applying the static approach to such networks results in concentrating water flow (EBC^Q) in particular pipes and assigning zero value to the EBC^Q of others. Similar behavior can be seen in the BLA network, where the reservoir is connected to two pipes and EBC^Q values are mostly assigned to the shorter pipe. In contrast, dynamic EBC$^{\mathrm{Q}}$ can activate the overlooked paths and distribute nodal demands uniformly, deriving cheaper design options with higher resilience values.

For the BAK network as an extended design problem, only the diameters of six parallel and three new pipes need to be determined (see Figure 2c). In this case, the resilience and costs of the design solutions are mainly influenced by those three new pipes that are connected to the reservoir. Using either the static or dynamic approach for BAK results in the same value of EBC^Q for each of those pipes. Therefore, dynamic EBC$^{\mathrm{Q}}$(2) can slightly improve the quality of obtained solutions compared with static EBC^Q, which is mainly due to the less capacity assigned to the parallel pipes.

Figure 3d shows that in contrast with the benchmark problems, the outcome of using dynamic EBC$^{\mathrm{Q}}$(2) for the real WDN cannot outperform those found by EBC$^{\mathrm{Q}}$(1). The real WDN is composed of thousands of demand nodes and edges, implying that a great number of alternative paths exist for routing a demand to the source in every iteration. Using dynamic EBC$^{\mathrm{Q}}$(2) for designing this network increases edge weights excessively as demand division is conducted during the procedure (see Figure 1d). Consequently, the excessive increase of edge weights leads to incorrect identification of the shortest path, hence a sub-optimal route.

In this study, dynamic EBC$^{\mathrm{Q}}$(2) and EBC$^{\mathrm{Q}}$(1) are cherry-picked for the small-to-intermediate and large design problems, respectively.

3.2 Comparing dynamic EBC$^{\mathrm{Q}}$ results with optimal design solutions

Figure 4 illustrates a comparison between the best dynamic EBC$^{\mathrm{Q}}$ results for each case study with their optimal solutions obtained based on multi-objective optimization (OPT) using EAs. As shown in Figure 4a, near-optimal solutions can be found by applying dynamic EBC$^{\mathrm{Q}}$(2) to TLN. For the BLA network, EBC$^{\mathrm{Q}}$(2) results cover part of the optimal Pareto front with the resilience values ranging from 0.63 to 0.73. However, EBC$^{\mathrm{Q}}$(2) solutions with a Resilience > 0.75 cannot compete with OPT solutions. This is because redundant capacity plays a more crucial role in high resilient design solutions and the EBC$^{\mathrm{Q}}$(2) results with high resilience value can only partly benefit from that potential capacity. The reason lies in the fact that no hydraulic information is involved in the CNT design, and the procedure is solely performed based on the spatial distribution of water flow using EBC^Q.

Figure 4c indicates that a couple of optimal/near-optimal solutions are obtained using dynamic EBC$^{\mathrm{Q}}$ (2) for BAK WDN, highlighting the potential of the CNT design approach for extended design/rehabilitation problems. For the real network, optimal solutions are obtained with GALAXY, using two different generation numbers. Figure 4d shows that promising solutions can be found by applying dynamic EBC^{Q}(1) to the real network. According to the figure, EBC^{Q}(1) outcomes outperform those found with GALAXY with a generation number of 100,000 in the resilience range of 0.62 to 0.79. In addition, dynamic EBC^{Q}(1) covers the knee bend area in the Pareto front of these optimal solutions, which is of interest to operators and decision-makers. Dynamic EBC^{Q}(1) results can also outperform OPT results with 50 million function evaluations (g = 500,000) in a specific resilience range from 0.62 to 0.72.

The advantages of using dynamic EBC^{Q} over EAs approaches for large-scale networks become more noticeable by comparing their computational efficiency. Multi-objective design optimization of the real network with 100,000 and 500,000 generations requires 8 and 35 weeks of execution time, respectively, while the design procedure based on dynamic EBC^{Q} is performed only in 47 seconds.

The high efficiency of CNT design approach with dynamic EBC^{Q} in terms of computational time makes it possible to find a range of optimal/near-optimal solutions for large networks within seconds/minutes, enabling a broad range of applications. In addition, CNT can be used in combination with EAs to reduce the search space and improve the computational efficiency of optimizations for large networks.

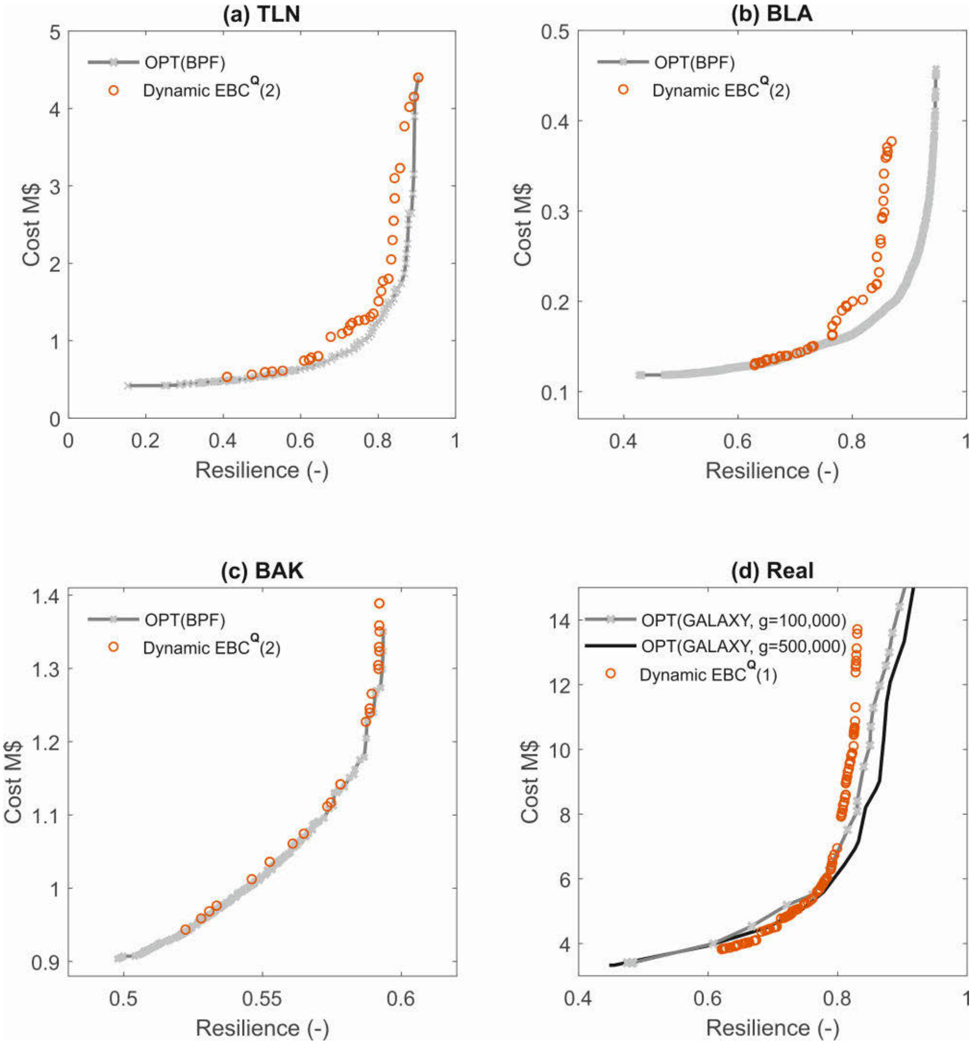


Figure 4: Comparing dynamic EBC^{Q} results with optimal design solutions (OPT) according to the best-known Pareto front (BPF) and GALAXY with g generation number.

4 SUMMARY AND CONCLUSIONS

This paper explores the potential of using dynamic weighting functions for a complex network theory-based (CNT) design approach. For this purpose, two dynamic weighting functions are proposed and integrated into a graph measure known as EBC^Q which is used for WDNs design. This measure provides a valid estimation of water flow by determining the frequency of occurrence of an edge in the shortest path from a source node (reservoir) to a demand node and summing up the demands routed through an edge (pipe). In order to determine the shortest path, pipe length is used as a weighting function. Using a constant value as a weighting function in the course of EBC^Q calculation (static $\mathrm{EBC^Q}$) results in centralizing EBC^Q values in a few pipes (shortest-path tree). In comparison, changing (modifying) edge weights when calculating EBC^Q (dynamic $\mathrm{EBC^Q}$) employ the potential redundancy of alternative paths neglected with static EBC^Q, which could result in higher resilient and cheaper cost design solutions.

The first dynamic $\mathrm{EBC^Q}$ proposed in this work increases edge weights according to the corresponding demand node i (Q_i) and maximum demand (Q_{max}) in a WDN with a quadratic function. The second dynamic $\mathrm{EBC^Q}$ splits Q_i into smaller parcels (D_{Pi}) and increases the weights during the design process based on $(D_{Pi})^2$. The results of applying dynamic $\mathrm{EBC^Q}$ approaches to three designs and one extended design problem confirm the ability of proposed weights in enhancing the quality of solutions, especially for complex design problems. For instance, the results of applying dynamic $\mathrm{EBC^Q}$ to a real large WDN with 4,021 pipes can partly outperform those obtained with evolutionary algorithms with 50 million function evaluations. Further, in terms of execution time, dynamic $\mathrm{EBC^Q}$ needs 47 seconds to find optimal/near-optimal design options for this network, whereas evolutionary algorithms require 35 weeks.

Regarding the obtained results in this work, it is recommended to utilize the second dynamic method for small-to-intermediate size networks and the first one for large/very large-scale WDNs. In future work, the focus will be on the design of multi-source WDNs with CNT approach using dynamic $\mathrm{EBC^Q}$.

5 ACKNOWLEDGEMENTS

This study was funded by the Austrian Science Fund (FWF): P 31104-N29.

6 REFERENCES

[1] M. Hajibabaei, S. Nazif, and R. Sitzenfrei, “Improving the Performance of Water Distribution Networks Based on the Value Index in the System Dynamics Framework,” Water, vol. 11, no. 12, p. 2445, 2019.

[2] H. Mala-Jetmarova, N. Sultanova, and D. Savic, “Lost in optimisation of water distribution systems? A literature review of system operation,” Environ. Model. Softw., vol. 93, pp. 209–254, 2017.

[3] K. Diao, L. Berardi, D. B. Laucelli, B. Ulanicki, and O. Giustolisi, “Topological and hydraulic metrics-based search space reduction for optimal re-sizing of water distribution networks,” J. Hydroinformatics, 2022.

[4] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savić, “A Complex Network Approach for Pareto-Optimal Design of Water Distribution Networks,” in World Environmental and Water Resources Congress, 2021, pp. 901–913.

[5] C. H. Papadimitriou and K. Steiglitz, Combinatorial optimization: algorithms and complexity. Courier Corporation, 1998.

[6] M. S. Kadu, R. Gupta, and P. R. Bhave, “Optimal design of water networks using a modified genetic algorithm with reduction in search space,” J. Water Resour. Plan. Manag., vol. 134, no. 2, pp. 147–160, 2008.

[7] K. Diao, G. Fu, R. Farmani, M. Guidolin, and D. Butler, “Twin-hierarchy decomposition for optimal design of water distribution systems,” J. Water Resour. Plan. Manag., vol. 142, no. 5, p. C4015008, 2016.

[8] A. Minaei, A. M. Sabzkouhi, A. Haghighi, and E. Creaco, "Developments in multi-objective dynamic optimization algorithm for design of water distribution mains," Water Resour. Manag., vol. 34, no. 9, pp. 2699–2716, 2020.

[9] Q. Wang, D. A. Savić, and Z. Kapelan, "G ALAXY: A new hybrid MOEA for the optimal design of Water Distribution Systems," Water Resour. Res., vol. 53, no. 3, pp. 1997–2015, 2017.

[10] A.-J. Ulusoy, I. Stoianov, and A. Chazerain, "Hydraulically informed graph theoretic measure of link criticality for the resilience analysis of water distribution networks," Appl. Netw. Sci., vol. 3, no. 1, 2018.

[11] J. Zischg, C. Klinkhamer, X. Zhan, P. S. C. Rao, and R. Sitzenfrei, "A Century of Topological Coevolution of Complex Infrastructure Networks in an Alpine City," Complexity, vol. 2019, 2019.

[12] K. Diao, Y. Zhou, and W. Rauch, "Automated creation of district metered area boundaries in water distribution systems," J. Water Resour. Plan. Manag., vol. 139, no. 2, pp. 184–190, 2013.

[13] R. Sitzenfrei, "Using complex network analysis for water quality assessment in large water distribution systems," Water Res., vol. 201, no. June, p. 117359, 2021.

[14] O. Giustolisi, A. Simone, and L. Ridolfi, "Network structure classification and features of water distribution systems," Water Resour. Res., vol. 53, no. 4, pp. 3407–3423, 2017.

[15] R. Sitzenfrei, M. Oberascher, and J. Zischg, "Identification of Network Patterns in Optimal Water Distribution Systems Based on Complex Network Analysis," in World Environmental and Water Resources Congress 2019: Hydraulics, Waterways, and Water Distribution Systems Analysis, 2019, pp. 473–483.

[16] O. Giustolisi, L. Ridolfi, and A. Simone, "Tailoring centrality metrics for water distribution networks," Water Resour. Res., vol. 55, no. 3, pp. 2348–2369, 2019.

[17] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savić, "Using complex network analysis for optimization of water distribution networks," Water Resour. Res., vol. 56, no. 8, p. e2020WR027929, 2020.

[18] M. Hajibabaei, S. Hesarkazzazi, and R. Sitzenfrei, "Optimization of Water Distribution Networks with Dynamic Edge Betweenness Centrality – A sensitivity analysis," in World Environmental and Water Resources Congress 2021, 2021.

[19] E. W. Dijkstra, "A note on two problems in connexion with graphs," Numer. Math., vol. 1, no. 1, pp. 269–271, 1959.

[20] Q. Wang, M. Guidolin, D. Savic, M. Asce, and Z. Kapelan, "Two-Objective Design of Benchmark Problems of a Water Distribution System via MOEAs : Towards the Best-Known Approximation of the True Pareto Front," vol. 141, no. 2009, pp. 1–14, 2015.

[21] T. D. Prasad and N.-S. Park, "Multiobjective genetic algorithms for design of water distribution networks," J. Water Resour. Plan. Manag., vol. 130, no. 1, pp. 73–82, 2004.

[22] E. Alperovits and U. Shamir, "Design of optimal water distribution systems," Water Resour. Res., vol. 13, no. 6, pp. 885–900, 1977.

[23] H. D. Sherali, S. Subramanian, and G. V. Loganathan, "Effective Relaxations and Partitioning Schemes for Solving Water Distribution Network Design Problems to Global Optimality," J. Glob. Optim., vol. 19, no. 1, pp. 1–26, 2001.

[24] Z. W. Geem, "Optimal cost design of water distribution networks using harmony search," Eng. Optim., vol. 38, no. 3, pp. 259–277, 2006.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14636

HYDRAULIC AND CO-LOCATED PIPE CRITICALITIES IN THE REHABILITATION OF WATER DISTRIBUTION MAINS

Amin Minaei[1], Mohsen Hajibabaei[2], Enrico Creaco[3] and Robert Sitzenfrei[4]

[1]PhD Fellow, Austrian Academy of Sciences (ÖAW) and Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria
[2]PhD Student Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria
[3]Associate Professor, Dipartimento di Ingegneria Civile e Architettura, Via Ferrata 3, 27100 Pavia (Italy)
[4]Professor, Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria

[1] *amin.minaei@uibk.ac.at,* [2] *Mohsen.Hajibabaei@uibk.ac.at,* [3] *creaco@¢unipv.it,*
[4] *robert.sitzenfrei@uibk.ac.at*

Abstract

Infrastructures in urban areas can have spatial and also functional correlation. Water distribution networks (WDNs) along with other infrastructures therefore constitute a complex and interlinked multi-utility system in cities. This brings up the risk of cascading failures to the different networks' elements; for example, a pipe failure could interrupt traffic in a main street, eventually leading to a road network failure. On the other hand, WDNs should be hydraulically robust so that the potable water is supplied to the customers with high reliability. Traditionally, there is a single perspective design approach for WDN rehabilitation and upgrade activities such as pipe replacement, duplicating, and lining, which does not consider the interlinked system in a city. This study aims to assess this issue in terms of an integrated asset management perspective with a multi-utility approach. For this purpose, beyond minimizing costs, two reliability indices will be defined to represent the reliability of a WDN against the hydraulic and multi-utility cascading failures. The hydraulic reliability represents the robustness of the network against the water pressure deficit, and cascading reliability represents the extent to which WDN elements are decoupled from other assets elements. Then, the rehabilitation problem is solved with the contribution of a nature-inspired optimization algorithm, the Non-Dominated Sorting Genetic Algorithm (NSGA-II), through a dynamic approach. In every decision of pipe rehabilitation action, the priority could be given to either the first or second reliability index. These two cases will be assessed and compared to examine the hydraulic and co-location pipe criticality roles in the decision-making for the upgrade of aged water distribution mains on a simplified real network located in the southwest Iran.

Keywords
Water distribution networks, Complex network, Cascading failures, Asset management perspective, Rehabilitation, Multi-objective optimization.

1 INTRODUCTION

Water distribution networks (WDNs) are considered complex infrastructures and their design, construction, and rehabilitation are very complicated and multi-criteria problems [1]. Together with the advancements in developing computer-based models, solving such complex problems through optimization algorithms has been the target of many researches over the past decades [2-7].

While pipe diameters have been the common decision variables in WDN optimization problems, there are many other practical considerations which could significantly impact on the construction cost of the output plans [8]. For example, WDNs change dynamically as time goes by and this deviates the optimum solutions due to the uncertainties in layout expansion, consumers' demand and budget allocation. Some authors have proposed phasing design and construction approaches to deal with this challenge [9-13].

Conflicts between WDN elements (pipes, valves, pumps and etc) and adjacent urban infrastructure systems (road, sewer, etc) increase the risk of cascading failures in the case of failure events [14, 15]. Hence, the renewal plans for WDNs should be organized in such a way that not only demanded water with desirable pressure is delivered to the consumers, but also the layout of WDNs should be dynamically re-designed to achieve decoupled WDNs from the elements of neighbour networks. On the other hand, there is not always enough budget to reach all the goals. In this regard, there are important questions as follows:

1- What is the trade-off between the hydraulic and decoupled reliability measures when an aged WDN must be rehabilitated?
2- How can an aged WDN be improved when there is a budget constraint?

This study attempts to respond to the abovementioned questions through a novel approach.

2 METHODOLOGY

To dynamically solve the rehabilitation problem of a WDN through an optimization algorithm, the design period with the planning horizon T is divided into p phases where every phase is Δt years long, as is shown in Figure 1.

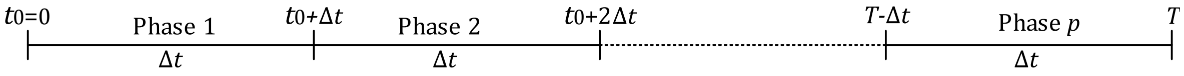


Figure 1. Design period in dynamic optimization of an aged WDN rehabilitation problem

Renewing the WDN starts from phase one. The WDN is simulated with the average of peak demand of consumers over the phases. The multi-objective optimization problem is solved in phase one generating some renewal plans. Based on the available budget or desired reliabilities, a renewal plan is selected and constructed in phase one. The network is updated at the beginning of phase two to repeat the optimization as done in the previous phase. This dynamic phase-by-phase rehabilitation continues till the phase p and gradually upgrades an aged WDN saving money and ending up a well status network [13]. This is a nonlinear, constrained, and integer-real optimization problem stochastically solved by the common nature-inspired approach, Non-Dominated Sorting Genetic Algorithm (NSGA-II) [16]. In the following sections, the decision variables, constraints and objectives of the optimization problem are explained.

2.1 Optimization Decision Variables for Rehabilitation of WDN Problem

Every deteriorated pipe in a WDN should be reclaimed and rehabilitated over its life span to supply potable water continuously to the customers. In this regard, there are some pipe rehabilitation and upgrade techniques such as replacement, duplicating, repairing and etc. The suitable strategy depends on the desired targets requested by water utility managers. In the current study, the client requests a low-cost rehabilitation program making the network decoupled from adjacent infrastructures (in line with the objective of decreasing the risk of cascading failures under hazard-based circumstances) and hydraulically robust overcoming pressure deficit due to the increase in demands, pipe bursts, aging and leakages. Hence the decision variables are defined in the optimization algorithm as follows:

$$r_i \in \{0,1,2,3\}, i = 1{:}\, n_{ps} \tag{1}$$

$$d_i \in \{D_{c,min}, \ldots, D_{c,max}\}, i = 1{:}\, n_{ps} \tag{2}$$

where, r_i is rehabilitation indicator actions for pipe i which could get integer numbers between 0 to 3 explained in Table 1, n_{ps} is the number of pipe sites which are already occupied by the pipes giving services to the customers, d_i is the pipe i diameter belonging to the set of commercial diameters which are real values changing from $D_{c,min}$ (the minimum available commercial diameter in the market) to $D_{c,max}$ (the maximum available commercial diameter in the market).

Table 1. Decisions for the upgrade of an aged water distribution network

Decision ID	Action Indicator	Explanation
1	$r_i = 0$	The old pipe is removed from site i
2	$r_i = 1$	The old pipe is kept in site i to continue its service
3	$r_i = 2$	The old pipe is replaced with the new pipe which has d_i diameter
4	$r_i = 3$	The old pipe gets a parallel pipe with d_i diameter

The first decision contributes to make the network robust against interconnectivity with adjacent networks. The second decision contributes to save rehabilitation costs. Here, the main assumption is neglecting the pipe maintenance and operating cost. The third and fourth decisions are for improving the hydraulic aspect of the network.

2.2 Optimization Constraints for Rehabilitation of WDN Problem

The hydraulic simulation of WDN is carried out by EPANET 2.2 and therefore the physical constraints of the pipe network hydraulics, i.e., the conservation of mass and energy, are automatically satisfied in the simulation model. Moreover, every solution is feasible as long as it keeps the piezometric pressure head of network's nodes above the minimum pressure service, Equation (3), below which the demand of the node is not satisfied [17]. In Equation (3), H_j is the pressure head at node j, H_{min-s} is minimum service pressure head and n_n is the number of demand nodes in WDN layout.

$$H_j \geq H_{min-s}, j = 1{:}\, n_n \tag{3}$$

There are some practical and decision constraints for rehabilitation actions. The first decision cannot be applied to the pipes playing the main role in supplying the water demands of associated node. For this, first, the graph analysis of WDN layout is carried out to recognize the shortest path (where the edge weights are the Euclidean length) from a source to a demand node of the network's graph. Those pipes which belong to the shortest path set are marked as critical sites.

Once the decisions two or three is constructed for the pipe i in phase $k, k = 1{:}\, p$, repeating the mentioned decisions is avoided in the following phases, phases $k + 1, k + 2, \ldots, p$. These decisions constraints are due to the fact that a multiple pipe replacement is not economical and laying numerous parallel pipes beside each other is practically infeasible.

As being clear, there is a trade-off between choosing the decision variables in terms of reaching the main objectives of the current rehabilitation problem. For example, while removing a pipe could make a decoupled WDN for an element in a site, it could worsen the hydraulic reliability of the network; or laying parallel pipes increase the hydraulic capacity of the network, but it makes the rehabilitation program expensive and increases the risk of cascading failures. In this regard, two scenarios are considered for laying parallel pipes in this study (Figure 2).

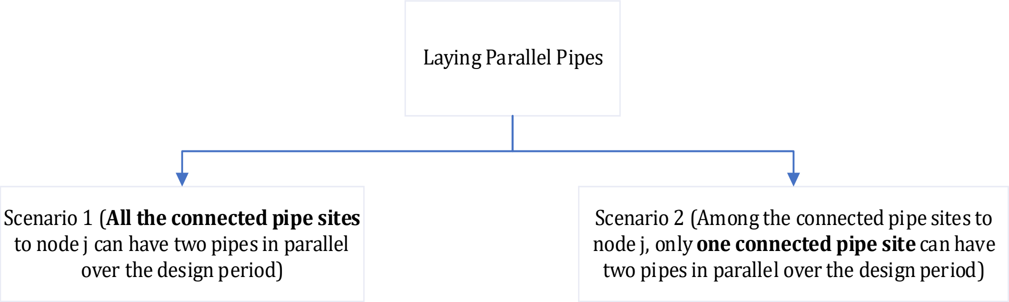

Laying Parallel Pipes
Scenario 1 (**All the connected pipe sites** to node j can have two pipes in parallel over the design period)
Scenario 2 (Among the connected pipe sites to node j, only **one connected pipe site** can have two pipes in parallel over the design period)

Figure 2. Laying parallel pipe scenarios for rehabilitation of WDN

While scenario one is more in favor of making a hydraulically robust network, scenario two is more in favor of making a decoupled network. Scenario two refers to constraints for laying parallel pipes in the decision-making process of optimization algorithm; Equation (4) represents the mentioned constraint, where $pp_{i,j}$ refers to the existence of a parallel pipe to associated node j in site i (the parallel pipe is defined as the new pipe laid beside the old pipe in site i). $pp_{i,j}$ gets binary values such that if there is a parallel pipe, $pp_{i,j}=1$; otherwise, $pp_{i,j}=0$.

$$\sum_{i=1}^{n_{ps}} pp_{i,j} \leq 1 \tag{4}$$

There are some controls for laying parallel pipes in Scenario two. To explain the controls, first some graph and hydraulic-based indices should be introduced as follows:

- **Node j degree (D_{n_j}):** the number of pipes connected to node j represents the degree of the node j (Equation (5)), where $p_{i,j}$ refers to the existence of single pipe in site i and it gets binary values as done for $pp_{i,j}$. The nodes with the highest node degree could refer to a hub where crowded sites in urban areas including hospitals and administration offices are located.

$$D_{n_j} = \sum_{i=1}^{n_{ps}} pp_{i,j} + \sum_{i=1}^{n_{ps}} p_{i,j} \tag{5}$$

- **Pipe i co-located degree (D_{Co-p_i}):** every pipe i in WDN can have a correlation with the adjacent networks' elements; for example, pipe i under a street and beside a sewer conduit has co-located degree of two. Equation (6) mathematically explains how this degree is calculated where e_{z_i} and n_z represent the site i adjacent element in network z and the total number of neighbour infrastructure systems, respectively.

$$D_{Co-p_i} = \sum_{z=1}^{n_z} e_{z_i} \tag{6}$$

- **Demand Edge betweenness centrality ($\mathbf{EBC_Q}$):** the number of times edge i (pipe i) is a part of the shortest paths between all node pairs j and the source node is known as the source edge betweenness centrality (EBC). This metric was modified by [18], referred to as demand EBC (EBC_Q). The EBC_Q of a pipe i finds the shortest path connecting the reservoir (source node S) and all demand node *j,* and adds the demand Q_j to the EBC_Q of all pipes located in that shortest path. $\mathrm{EBC}_Q(i)$ is formulated as follows(to know more about this index, please refer to [18]).

$$\mathrm{EBC}_Q(i) = \sum_{j=1}^{n_n} shortest\ path_{S,j}(i) \cdot Q_j \tag{7}$$

Considering the introduced indices, EBC_Q and D_{Co-p_i} represent the hydraulic and co-located criticality of a pipe. The algorithm of scenario two is shown in Figure 3. In short, if node j is a hub, the associated pipes cannot be strengthened by parallel pipes. If there is a decision for laying parallel pipes for the nodes which have a degree smaller than hub nodes, the priority of laying a parallel pipe is for the pipe site with the lowest co-located degree; if the co-located degrees are the same for all connected pipes to node j, the pipe with the highest EBC_Q gets a parallel pipe.

After upgrading the network in phase k, some updates for the status of the sites are necessary. If the pipe i has been removed, its site cannot be occupied with new pipe till the last phase unless the layout of adjacent networks has been changed in favor of decreasing the site i co-located degree. If the pipe site i has gotten a parallel pipe, this site is not among the critical sites anymore and there could be the option of removing the pipe. Moreover, it should be checked if there has become a hub (the nodes with the highest degree) to not receive a parallel pipe over the next subsequent phases.

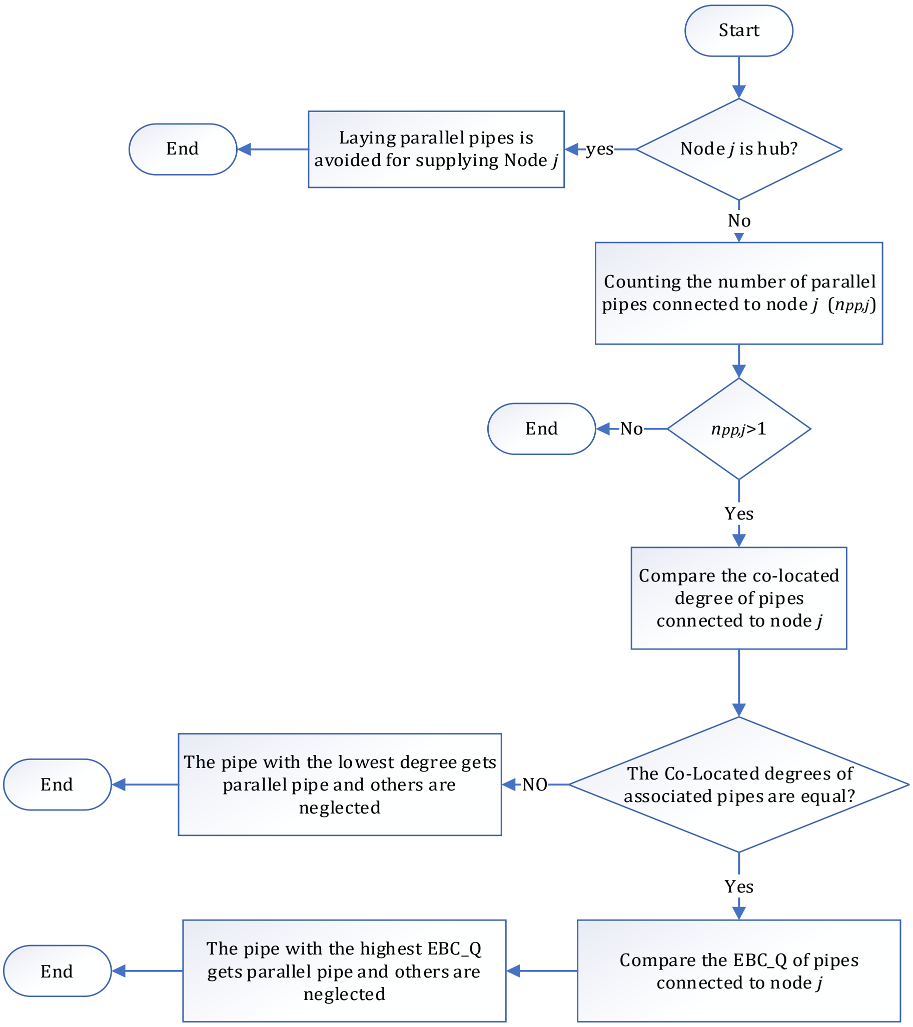

Start
Node *j* is hub?
yes
Laying parallel pipes is avoided for supplying Node *j*
End
No
Counting the number of parallel pipes connected to node *j* (*npp,j*)
npp,j>1
No
End
Yes
Compare the co-located degree of pipes connected to node *j*
The Co-Located degrees of associated pipes are equal?
NO
The pipe with the lowest degree gets parallel pipe and others are neglected
End
Yes
Compare the EBC_Q of pipes connected to node *j*
The pipe with the highest EBC_Q gets parallel pipe and others are neglected
End

Figure 3. The decision-making algorithm of laying parallel pipes in scenario 2 for rehabilitation of WDN

2.3 Optimization Objectives for Rehabilitation of WDN Problem

The current multi-objective optimization problem is formulated with three objectives as follows:

$$\text{Minimize } (Cost_k(\Gamma_k(r,d)), -Rel_{\text{hyr},k}, -Rel_{\text{dec},k}) \tag{8}$$

where, $Cost_k$ is the rehabilitation cost of upgrade program Γ, indicating the budget of upgrading the network at the beginning of phase k, and evaluated by Equation (9):

$$Cost_k(\Gamma(r,d)) = \sum_{i=1}^{n_{ps}} c_i L_i \tag{9}$$

Where c_i is the unit cost of the commercial diameter size assigned to the pipe with a length L_i. Hydraulic reliability is the second objective of optimization calculated by a hybrid index which is the combination of two indices in hydraulically weak and robust conditions [13]. The first one is the frequency index of counting the number of demand nodes with water pressure above desirable pressure calculated by Equation (10):

$$Rel1_{hyd,k} = 1 - \frac{\sum_{i=k}^{p}\sum_{j=1}^{n_n} max\,(0, -sign(H_{ij} - H_{\mathrm{des}}))}{n_n(p-k+1)} \tag{10}$$

where $Rel1_{hyd,k}$ represents the first frequency hydraulic reliability of the network in phase k, and H_{ij} actual head at node j in phase i, where i changes from k to p, and H_{des} is desirable pressure. The second part of the hybrid reliably is the resilience index introduced by [19] and then improved by [20]:

$$Rel2_{hyd,k} = \frac{\sum_{i=k}^{n}\left(\frac{\sum_{j=1}^{n_n} C_{ij} \times q_{ij}\,(H_{ij} - H_{des})}{\left[\sum_{l=1}^{nr_i} Qr_{il} \times Hr_{il} + \sum_{m=1}^{pn_i}(\frac{P_{im}}{\gamma})\right] - \sum_{j=1}^{n_n} q_{ij}H_{des}}\right)}{(p-k+1)} \tag{11}$$

where q_{ij} is the design demand at node j at the end of phase i, nr_i is the number of reservoirs in phase i, Qr_{il} and Hr_{il} are respectively the discharge from and head at reservoir l in phase i, pn_i is the number of pumps in phase i and P_{im} is the power of pump m at the end of phase i. Also, C_{ij} is a weighting coefficient associated with the uniformity of the diameter of pipes connected to node j in phase i as follows:

$$C_{ij} = \frac{\sum_{r=1}^{np_{ij}} D_r}{np_{ij} \times max\{D_r\}} \tag{12}$$

where np_{ij} and $max\{D_r\}$ are respectively the number of pipes and the maximum pipe diameter size connected to node j in phase i. According to the above equation, $C_{ij} = 1$ if only one pipe is connected to node j or all pipes connected to that node have the same diameters and, $C_{ij} < 1$ if pipes connected to node j have different diameters.

In each phase k the two mentioned indexes are calculated and final hydraulic reliability would be as follows:

$$Rel_{hyr,k} = \begin{cases} Rel1_{hyr,k} & Rel1_{hyr,k} < 1 \\ 1 + Rel2_{hyr,k} & Rel1_{hyr,k} = 1 \end{cases} \tag{13}$$

Using the hybrid reliability index (Equation (13)) the upgrade program becomes more flexible and manageable so that the system can be gradually upgraded from a weak state ($Rel1 < 1$) to a normal state ($Rel1 \approx 1$) and then to a robust state ($Rel2 > 1$) depending on the money invested for the project and the reliability expected. Hence the value of $Rel_{hyr,k}$ changes between 0 and 2.

The third objective refers to the decoupled reliability (Equation (14)) of WDN showing how much the rehabilitation plan can make a decoupled network in phase k.

$$Rel_{\mathrm{dec},k} = 1 - mean(Avg(D_{Co-p_{i,k}})) \tag{14}$$

where $Avg(D_{Co-p_{i,k}})$ calculates the average correlation for the pipe i in phase k meaning that for a network correlated with two adjacent networks (for example road and sewer), if a pipe element is correlated with only a road element, the average correlation for the pipe is 0.5. After calculation of the average correlation in every site i, the mean value of $Avg(D_{Co-p_{i,k}})$ gives an overall view about the decoupled status of WDN where its values change between 1 and 0.

3 APPLICATION

3.1 Case Study

To investigate the proposed approach, an aged water WDN, Baghmalek network (Figure 4), located in the southwest of Iran is considered for upgrading. The network has 90 pipe sites and 72 consumption nodes and is fed by one reservoir (node 1) located at the highest elevation of the region. The network is more than 30 years old and its current hydraulic performance turns out to be in an urgent need for rehabilitation and upgrading. It is assumed that the network only has correlations with sewer and road networks. The multiplex system is conceptualized in Figure 4 representing the correlations and the network configuration. The full information about the network can be found in [13] and is available upon request.

The network is first analysed for the existing conditions in year zero. Currently, the hydraulic performance of the WDN is weak and only 36% of the consumption nodes meet the desirable pressure of 18 m required for the network according to the national regulations (Rel1=36%). Also, the correlation of WDN with the adjacent infrastructures is about 63% and therefore the decoupled reliability is about 37%. Hence, the hydraulic and decoupled reliabilities of the network should be improved through upgrade actions. There are some main assumptions for upgrading the current network as follows:

- A 25-year design period is considered for upgrading the network (T=25)
- During the design period, the network has no extension in plan. The network layout is fixed with time.
- A list of polyethylene pipes containing 14 diameter sizes as shown in Table 2 is used for upgrading the network.
- For all new pipes, the Hazen–Williams coefficient in year zero is 130. Also, the Hazen–Williams coefficients are supposed to change linearly with time with a reduction rate of -0.6% yearly.
- The installation of a parallel pipe is more difficult and expensive than replacing a new pipe. To take this issue into account the unit cost of parallel pipes is increased by 20%.
- The network consumption is supposed to change with time linearly. The annual rate of consumption increase is estimated with 0.0332 l/s/year.
- The design period is divided into 5 construction phases, 5-year periods

- The correlation of a water pipe with a street (road element) is assumed if the water pipe is located under or in the 3-meter distance from the margin of the street.
- The correlation of a water pipe with a sewer network element is assumed if they are located in 3-meter radius distance from the axis line of each other (this assumption comes from the possible cascade failure in earthquake circumstances)

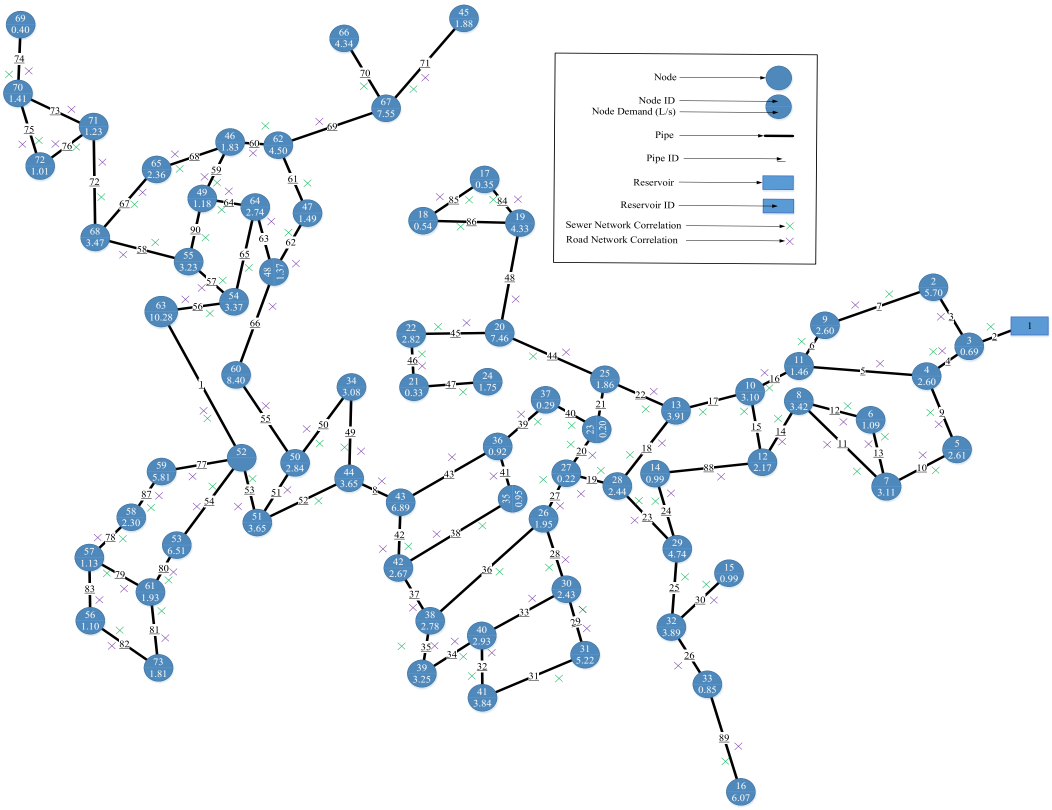

Node
Node ID
Node Demand (L/s)
Pipe
Pipe ID
Reservoir
Reservoir ID
Sewer Network Correlation
Road Network Correlation
1
69 0.40
70 1.41
71 1.23
72 1.01
66 4.34
45 1.88
67 7.55
46 1.83
62 4.50
65 2.36
49 1.18
64 2.74
47 1.49
68 3.47
55 3.23
48 1.37
54 3.37
63 10.28
17 0.35
18 0.54
19 4.33
60 8.40
22 2.82
20 7.46
21 0.33
24 1.75
34 3.08
37 0.29
25 1.86
23 0.20
36 0.92
52
59 5.81
50 2.84
44 3.65
43 6.89
27 0.22
35 0.95
28 2.44
14 0.99
13 3.91
10 3.10
11 1.46
9 2.60
2 5.70
3 0.69
4 2.60
8 3.42
6 1.09
12 2.17
7 3.11
5 2.61
58 2.30
51 3.65
53 6.51
57 1.13
61 1.93
56 1.10
73 1.81
42 2.67
26 1.95
29 4.74
15 0.99
30 2.43
38 2.78
40 2.93
32 3.89
31 5.22
39 3.25
41 3.84
33 0.85
16 6.07

Figure 4. Baghmalek WDN and correlation with adjacent networks' elements in year 0

Table 2. Commercial polyethylene pipes with their unit construction cost in year zero, D_C: Commercial diameter, D: Internal diameter

D_C (mm)	D (mm)	*Unit cost (Rials/m)*
63	53.60	163473.90
75	63.80	192204.30
90	76.60	232484.70
110	93.80	300636.10
125	106.60	374078.20
160	136.40	543685.60
200	170.60	769448.30
250	213.20	1147234.00
315	268.20	1752000.00

400	341.20	2734489.00
450	383.89	3455312.00
630	537.50	6466257.00
710	605.77	8102474.00
800	682.58	10187126.00

3.2 Optimization Results

The Pareto fronts obtained from the multi objective optimization are shown in Figure 5. The left side Pareto fronts (Figure 5-(a)) refer to the rehabilitation programs of Scenario 1, while the right-side Pareto fronts (Figure 5-(b)) indicate the rehabilitation programs of Scenario 2.

As being clear, the Pareto fronts show a serious trade-off between objectives. The higher budget investment in the rehabilitation of WDN, the higher improvement in the hydraulic aspect of the network, and the higher values of hydraulic reliability, the lower values of decoupled reliability.

It is assumed that the policy in Phase 1 is improving both hydraulic and decoupled reliabilities of WDN to values around 0.93 and 0.47. The solutions with the mentioned desired reliabilities are selected from Pareto fronts. As observed, the rehabilitation program based on scenario two provides a cheaper price than the one from Scenario 1. Hence, the low co-located degree priority-based strategy for laying parallel pipes not only makes the cost saving rehabilitation plans, but also improves the optimization performance in terms of efficiency.

It is assumed in Phase 2 that there is a limitation in budget allocation and the WDN client can only invest 1.50 ×10^{10} Rials for rehabilitating of Baghmalek network. As being clear, while Scenario 1 provides solutions that make the WDN hydraulically resilient ($Rel1_{hyr,2}$≥1), Scenario 2 rehabilitations keep the hydraulic status of the network weak ($Rel1_{hyr,2}$<1). On the other hand, decoupled reliability resulting from Scenario 2 is higher than the one related to Scenario 1 (0.46 vs 0.34).

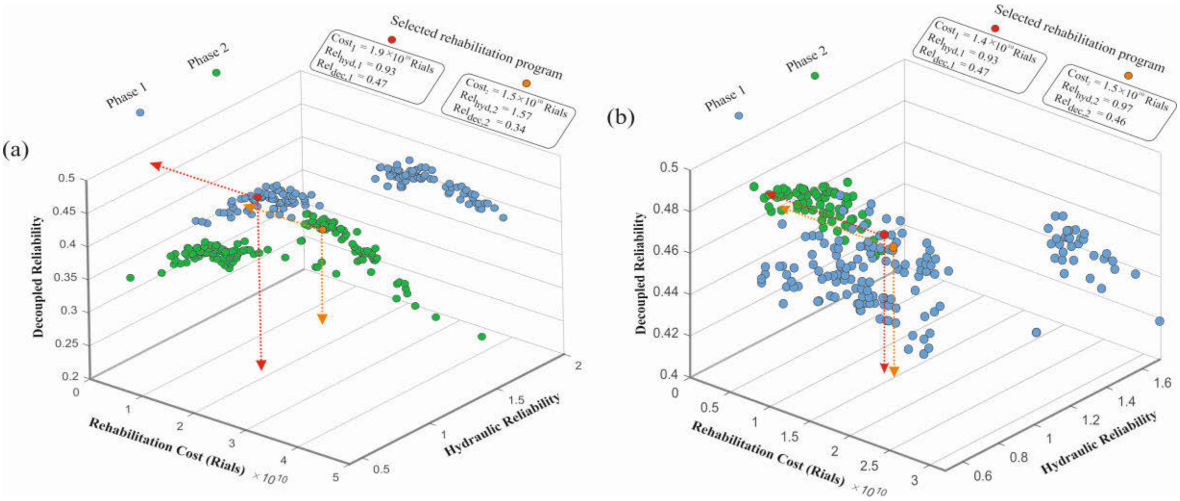


Figure 5. The three objectives Pareto fronts of rehabilitation programs for the WDN obtained by Scenarios 1 and 2 for two phases

Figure 6 shows the constructed rehabilitation plan in two phases obtained by Scenario 1 and 2 approaches which are Figures 6-a and 6-b respectively. As it can be seen, removing pipes can

CC BY NC SA

remove co-located degree and therefore significantly decreases the risk of cascade failures under events at the sites. For example, site 7 in Figure 6-b is removed in Scenario 2, and co-located degree becomes 0, while in Scenario 1, there is the replacement technique and no change in co-located degree in site 7. As seen, in the constructed plan of Scenario 2, there are no nodes with two connected parallel pipes whereas numerous are present in the Scenario 1 plan.

Another important consideration is recognition of hubs in every phase of rehabilitation. The hub for the Baghmalek network is node 52 with the degree four which is the highest degree. Hence, while node 52 gets higher degree in Scenario 1, it is avoided in Scenario 2. Moreover, in Scenario 2, some nodes reach degree four in phase 2 (for example node 43) and they cannot receive higher degrees by parallel pipes over the next phases unless the coming rehabilitation actions reduce the degree of the node (for example the pipe 43 could be removed).

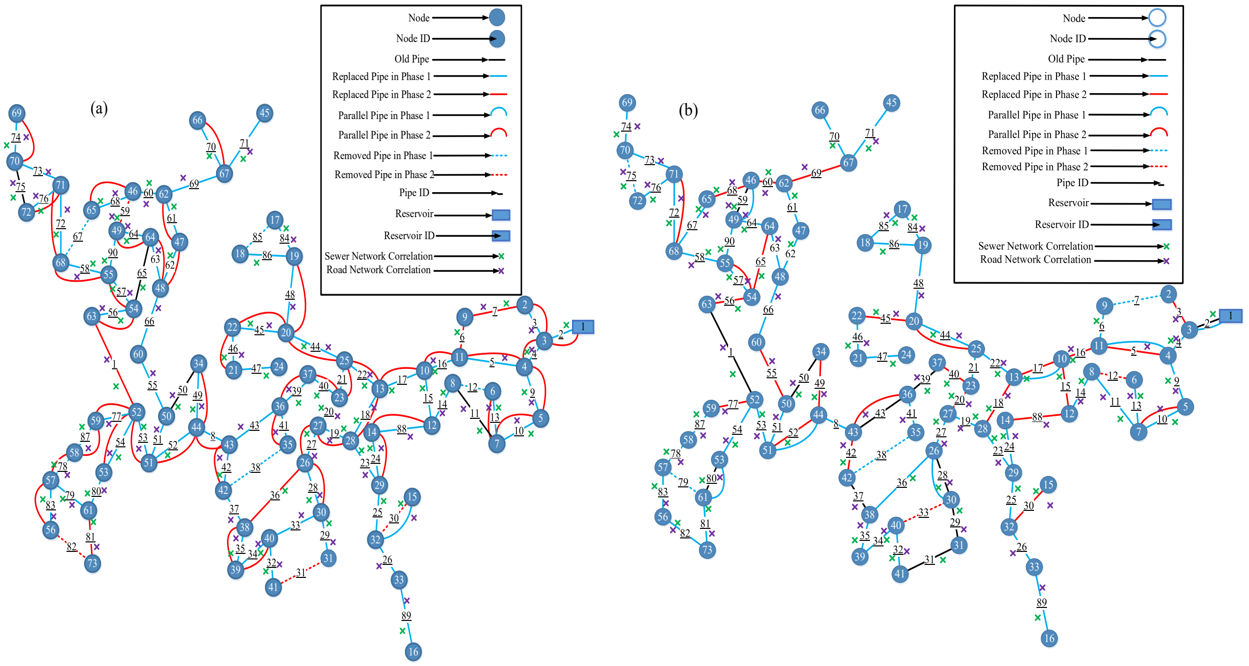


Figure 6. Constructed rehabilitation plan for Baghmalek WDN obtained by Scenario 1 and 2 approaches

4 SUMMARY AND CONCLUSIONS

Current study proposed a method for rehabilitation of aged water distribution mains when there are three conflicting objectives, minimizing rehabilitation cost, maximizing hydraulic reliability, and decoupled reliability showing the interconnectivity of WDN with adjacent networks (like road and sewer networks). In this method, the design period is divided into some intervals, and construction, as well as design of rehabilitation plans, are done phase by phase dynamically.

In every phase, multi-objective optimization problem with the mentioned objectives is solved where the decision variables are leaving the old pipe to continue its service, pipe replacement, laying parallel pipes and removing pipes.

Two Scenarios were defined for rehabilitation of the network. In Scenario 1, the node degree can be enhanced in every phase of rehabilitation by laying parallel pipes, while this can happen only one time over the design period in Scenario 2 where the hubs (nodes with the highest degree) are not allowed to receive a higher degree. In this regard, other locations are assessed in terms of co-located degree and demand edge betweenness centrality as co-located and hydraulic pipe criticalities. In every decision of enhancing the degree of nodes by parallel pipes, only the associated pipe with the lowest co-located criticality can get a pipe in parallel and if all the associated pipes have the same co-located criticality, the pipe with the highest demand edge betweenness centrality gets a parallel pipe.

The results showed that designing the rehabilitation programs in favour of asset managers could have negative impacts on the hydraulic performance of the WDN (Scenario 2). On the other hand, making hydraulically a resilient network by laying many parallel pipes (Scenario 1) results in a vulnerable network against cascading failure and is not desirable when there is the matter of budget constraints. This implies that the multi-utility rehabilitation planning is a highly complex task and should be synchronized with other asset rehabilitation and upgrade programs to save cost and obtain a better balance between decoupled and hydraulic reliability of WDN.

5 ACKNOWLEDGMENTS

This study was funded by Austrian Academy of Sciences (ÖAW) fund: DOC Fellowship and partly funded by the Austrian Science Fund (FWF): P 31104-N29.

6 REFERENCES

[1]. Hajibabaei, M., S. Nazif, and R. Sitzenfrei, Improving the Performance of Water Distribution Networks Based on the Value Index in the System Dynamics Framework. Water, 2019. 11(12): p. 2445.

[2]. Rahmani, F., K. Behzadian, and A. Ardeshir, Rehabilitation of a water distribution system using sequential multiobjective optimization models. Journal of Water Resources Planning and Management, 2016. 142(5): p. C4015003.

[3]. Tanyimboh, T.T. and P. Kalungi, Optimal long-term design, rehabilitation and upgrading of water distribution networks. Engineering Optimization, 2008. 40(7): p. 637-654.

[4]. Farmani, R., G. Walters, and D. Savic, Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. reliability vs. water quality. Journal of Hydroinformatics, 2006. 8(3): p. 165-179.

[5]. Tanyimboh, T.T. and P. Kalungi, Multicriteria assessment of optimal design, rehabilitation and upgrading schemes for water distribution networks. Civil Engineering and Environmental Systems, 2009. 26(2): p. 117-140.

[6]. Wang, Q., et al., Impact of problem formulations, pipe selection methods, and optimization algorithms on the rehabilitation of water distribution systems. Journal of Water Supply: Research and Technology-Aqua, 2020. 69(8): p. 769-784.

[7]. Halhal, D., et al., Scheduling of water distribution system rehabilitation using structured messy genetic algorithms. Evolutionary computation, 1999. 7(3): p. 311-329.

[8]. Walski, T. How does water distribution design really work? in World Environmental and Water Resources Congress 2014. 2014.

[9]. Creaco, E., M. Franchini, and T.M. Walski, Accounting for Phasing of Construction within the Design of Water Distribution Networks. Journal of Water Resources Planning and Management, 2014a. 140(5): p. 598-606.

[10]. Creaco, E., M. Franchini, and T. Walski, Taking account of uncertainty in demand growth when phasing the construction of a water distribution network. Journal of Water Resources Planning and Management, 2014b. 141(2): p. 04014049.

[11]. Kang, D. and K. Lansey, Multi-Period Planning of Water Supply Infrastructure Based on Scenario Analysis. Journal of Water Resources Planning and Management, 2014. 140(1).

[12]. Zischg, J., W. Rauch, and R. Sitzenfrei, Morphogenesis of Urban Water Distribution Networks: A Spatiotemporal Planning Approach for Cost-Efficient and Reliable Supply. Entropy, 2018. 20(9): p. 708.

[13]. Minaei, A., A. Haghighi, and H.R. Ghafouri, Computer-Aided Decision-Making Model for Multiphase Upgrading of Aged Water Distribution Mains. Journal of Water Resources Planning and Management, 2019. 145(5): p. 04019008.

[14]. Sun, S., et al., A fast approach for multiobjective design of water distribution networks under demand uncertainty. Journal of hydroinformatics, 2011. 13(2): p. 143-152.

[15]. Sitzenfrei, R., et al., Cascade vulnerability for risk analysis of water infrastructure. Water Science and Technology, 2011. 64(9): p. 1885-1891.

[16]. Deb, K., et al., A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 2002. 6(2): p. 182-197.

[17]. Wagner, J.M., U. Shamir, and D.H. Marks, Water distribution reliability: simulation methods. Journal of water resources planning and management, 1988. 114(3): p. 276-294.

[18]. Sitzenfrei, R., et al., Using complex network analysis for optimization of water distribution networks. Water resources research, 2020. 56(8): p. e2020WR027929.

[19]. Todini, E., Looped water distribution networks design using a resilience index based heuristic approach. Urban water, 2000. 2(2): p. 115-122.

[20]. Prasad, T.D., S.-H. Hong, and N. Park, Reliability based design of water distribution networks using multi-objective genetic algorithms. KSCE Journal of Civil Engineering, 2003. 7(3): p. 351-361.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14699

A BAYESIAN GENERATIVE ADVERSARIAL NETWORK (GAN) TO GENERATE SYNTHETIC TIME-SERIES DATA, APPLICATION IN COMBINED SEWER FLOW PREDICTION

Amin E. Bakhshipour[1], Alireza koochali[2], Ulrich Dittmer[3], Ali Haghighi[4], S. Ahmed[5] and A. Dengel[6]

[1, 2, 3, 5, 6] TU Kaiserslautern, Kaiserslautern, Germany
[2] Ingenieurgesellschaft Auto und Verkehr(IAV) GmbH, Berlin,Germany
[2, 5, 6] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[4] Shahid Chamran University of Ahvaz, Ahvaz, Iran

[1] *amin.bakhshipour@bauing-uni.kl.de*, [2] *alireza.koochali@iav.de*, [3] *ulrich.dittmer@bauing.uni-kl.de*, [4] *a.haghighi@scu.ac.ir*, [5] *sheraz.ahmed@dfki.de*, [6] *andreas.dengel@dfki.de*

Abstract

Despite various breakthroughs of machine learning and data analysis techniques for improving smart operation and management of urban water infrastructures, some key limitations obstruct this progress. Among these shortcomings, the absence of freely available data due to data privacy or high costs of data gathering and the nonexistence of adequate rare or extreme events in the available data plays a crucial role. Here, the Generative Adversarial Networks (GANs) can help overcome these challenges. In machine learning, generative models are a class of methods capable of learning data distribution to generate artificial data. In this study, we developed a GAN model to generate synthetic time series to balance our limited recorded time series data and improve the accuracy of a data-driven model for combined sewer flow prediction. We considered the sewer system of a small town in Germany as the test case. Precipitation and inflow to the storage tanks are used for the Data-Driven model development. The aim is to predict the flow using precipitation data and examine the impact of data augmentation using synthetic data in model performance. Results show that GAN can successfully generate synthetic time series from real data distribution, which helps more accurate peak flow prediction. However, the model without data augmentation works better for dry weather prediction. Therefore, an ensemble model is suggested to combine the advantages of both models.

Keywords
Machine Learning, Urban Water Infrastructures, Generative Adversarial Networks, Time Series Prediction, synthetic time series generation, Combined Sewer Flow Prediction

1 INTRODUCTION

Nowadays, many industry sectors such as health care, automation, financial markets, aerospace, water resources management, and weather forecasts deal with time-series data in their operations and development. Improving the data availability can increase the efficiency of existing infrastructures and foster research for environmental sustainability. Data analysis and mining would facilitate a shift from pure infrastructure development to smart operation and management of our highly interconnected systems with environmental challenges. In recent years, we have witnessed numerous breakthroughs in machine learning techniques in various domains, such as computer vision, language processing, reinforcement learning, and many more [1].

However, still, some fundamental shortcomings hinder progress in environmental management. They include (1) lack of freely available data (e.g., due to data privacy or high expenses of data gathering) for an extended period as well as lack of rare or extreme events in the training data, (2) lack of robust methods for anomaly detection, particularly for drift detection, (3) absence of probabilistic time series forecasting data-driven methods to consider different sources of uncertainty for optimal and robust operation of critical urban water infrastructures, (4) lack of benchmark cases and (5) various sources of uncertainty affecting the data and the resulting models.

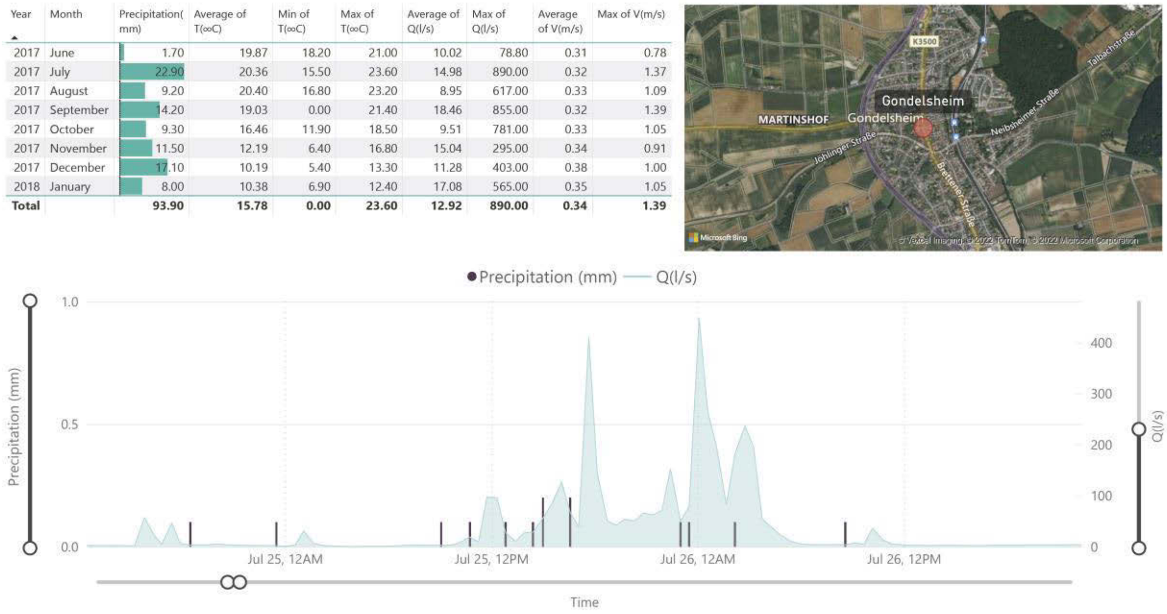
Generative algorithms are powerful approaches in data science that can help us overcome the challenges mentioned above in various domains of water and environmental management. GANs are a deep learning architecture for training powerful generator models. The main goal of GANs is to learn from a set of training data and generate new data with the same characteristics as the training data. Initially, they were applied to domains where their results are intuitively assessable, e.g., images. However, GANs have been successfully used to generate time series sequences in the health care, finance, and energy industry and outperform state-of-the-art benchmarks. Nevertheless, GANs' potential in Urban Water Management (UWM) problems are not yet discovered to our knowledge.

In this study, as a proof of concept, we test one primary application of GANs, i.e., data augmentation to urban drainage systems. We use GANs to generate synthetic time series to balance our data set and improve the accuracy of a data-driven combined sewer flow prediction model. As relevant events are relatively rare in the historical data, pure data-driven rainfall-runoff models often underestimate the runoff when predicting these events. The optimal operation of the sewer system during these events, which result in the most critical states for the urban area and environment, depends highly on accurate flow predictions. We used GANs to generate synthetic time series from approximately the same statistical distribution of our data set to overcome these challenges. The generator model enables us to balance our training data with extreme synthetic events. To evaluate the performance of the proposed approach, we train a specific deep neural network model with and without synthetic data and test their performance using a similar test data set. The remainder of this manuscript is structured as follows:

2 MATERIALS AND METHODS

2.1 Case Study

In this study, the sewer system of a small town in Baden-Württemberg state in Germany is considered to evaluate the performance of the proposed model. Precipitation (mm), temperature (°C), and inflow to the storage tanks are the data set used for the model development. All data are measured from the beginning of June 2017 to the end of January 2018 at a 5-min time resolution [2]. Figure 1 summarizes our data set. It also depicts an example of our measured time series. The aim is to build a black-box simulator to predict combined sewer flow in the system using historical and synthetic data. The model then can be employed for optimal operation, e.g., to minimize the volume and duration of combined sewer overflows (CSOs) of the system using RTC.



Year	Month	Precipitation(mm)	Average of T(∞C)	Min of T(∞C)	Max of T(∞C)	Average of Q(l/s)	Max of Q(l/s)	Average of V(m/s)	Max of V(m/s)
2017	June	1.70	19.87	18.20	21.00	10.02	78.80	0.31	0.78
2017	July	22.90	20.36	15.50	23.60	14.98	890.00	0.32	1.37
2017	August	9.20	20.40	16.80	23.20	8.95	617.00	0.33	1.09
2017	September	14.20	19.03	0.00	21.40	18.46	855.00	0.32	1.39
2017	October	9.30	16.46	11.90	18.50	9.51	781.00	0.33	1.05
2017	November	11.50	12.19	6.40	16.80	15.04	295.00	0.34	0.91
2017	December	17.10	10.19	5.40	13.30	11.28	403.00	0.38	1.00
2018	January	8.00	10.38	6.90	12.40	17.08	565.00	0.35	1.05
Total		**93.90**	**15.78**	**0.00**	**23.60**	**12.92**	**890.00**	**0.34**	**1.39**

Figure 1. Case study

2.2 Problem Description

The release of untreated sewerage into receiving water bodies during rain events can be reduced by dynamically controlling sewer flow and retention volume with sensor networks and automated valves [1]. MPC is an advanced RTC technique in which the optimization is based not only on the knowledge of the system's current state but also on its forecast state. Thus, MPC allows us to improve the monitoring process and optimally utilize the storage capacities of rainwater reservoirs, detention ponds, and in-sewer storage volumes by considering anticipated rainfall, thus, e.g., reducing CSOs [1]. As relevant events are relatively rare in the historical data, pure data-driven rainfall-runoff models often underestimate the runoff when predicting these events. The optimal operation of the sewer system during extreme events, which result in the most critical states for the urban area and environment, depends highly on accurate flow predictions.

2.3 Generative Adversarial Networks (GANs)

Generative models are a class of algorithms that aim to generate realistic artificial data. These models approximate the probability distribution of a given data set as closely as possible. Then, we can fabricate new data samples from the approximated distribution. To put it in more concrete terms, assume data set $\mathcal{D}$, which is consists of N i.i.d samples (x) from a probability distribution i.e., $\mathcal{D} = \{x_1, \ \dots \ , x_n\}$. We denote these samples as real data and their probability distribution as P_{real}. The goal of a generative model is to accurately approximate P_{real} given samples from this distribution i.e., real data. Once we have a good approximation of data distribution, we can sample from this distribution and generate artificial data that follow the real data distribution.

The GAN [3] approach to achieving this goal is to define a mapping function f that transforms samples from a latent space (P_z) to the samples in the data space (P_{real}). In other words, our goal is to define f as:

$$f(z) = x,$$

where $z \sim P_{latent}$ and $x \sim P_{real}$. Conventionally, GAN utilizes a Gaussian distribution as the latent space. This latent space is also referred to as Noise space in the GAN domain. GAN employs a

CC BY NC SA

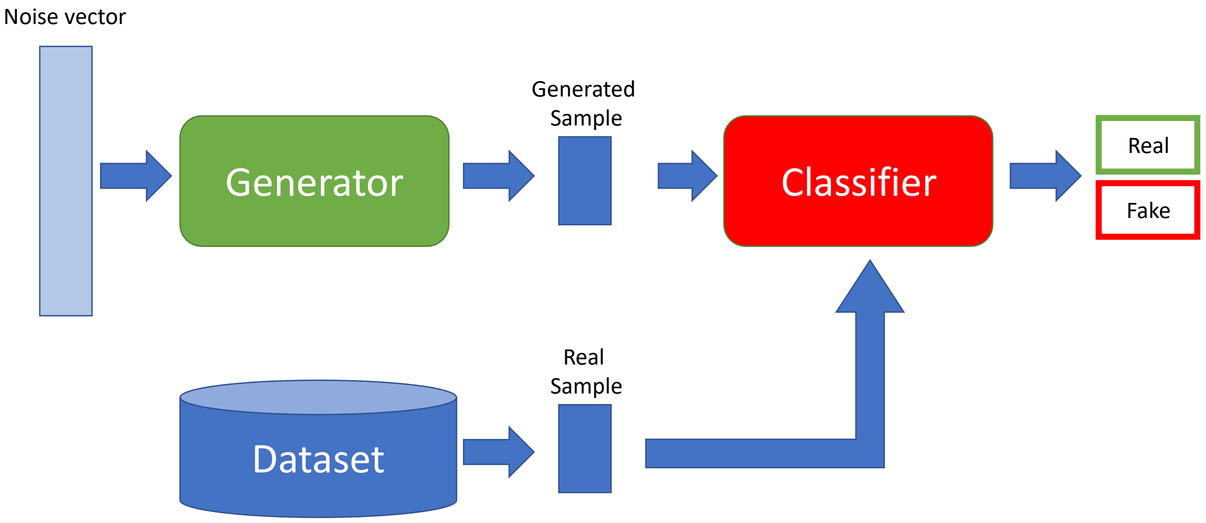


Figure 2. The abstract representation of GAN architecture

neural network called Generator (G) to approximate f. To train G, GAN uses another neural network called Discriminator (D). The Discriminator is a classifier trained to differentiate between data from the data set (real data) and generated from the generator (fake data). By learning to discriminate between real and fake data, the Discriminator can provide generator feedback regarding how realistic the fake data is. **¡Error! No se encuentra el origen de la referencia.** illustrates GAN's structure. First, we sample vector z from latent space P_z. Then, generator takes the vector z and maps it to a data sample. Finally, the Discriminator assesses the authenticity of generated with having access to both fake data and generated.

During the training of the GAN, the Generator and Discriminator are trained interchangeably in an adversarial fashion. We first train Discriminator to classify real and fake data as accurately as possible in one training step. Then, we train the Generator to fabricate the data sample to be identified as real by the Discriminator. In other words, the Generator tries to fool the Discriminator by generating realistic data points while the Discriminator tries not to be fooled by learning the classification between real and fake data. This two-player minimax game is set in motion by optimizing the following value function:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim P_{real}}[\log(D(x))] + \mathbb{E}_{z\sim P_z}[\log(1 - D(G(z)))].$$

If both models have enough capacity, the Generator will learn the probability distribution data and $D(x) = \frac{1}{2}$ for all x.

The original GAN proposal is quite challenging to train since the divergences which GANs typically minimize are potentially not continuous with respect to the Generator's parameters [4]. Therefore, various methods have been suggested to improve GAN's training stability. In this work, we adopt the Wasserstein GAN gradient penalty (WGAN-gp) to improve the training process. In WGAN, the training process minimizes the Wasserstein distance between a real distribution and fake distribution, i.e., $W(P_{real}, P_{fake})$ which continues under mild assumptions. Using Kantorovich-Rubinstein duality [5], the value function of WGAN is defined as:

$$\min_{G}\max_{D\in\mathcal{L}} = \mathbb{E}_{x\sim P_{real}}[D(x)] - \mathbb{E}_{\overline{x}\sim P_{fake}}[D(\overline{x})],$$

where, $\mathcal{L}$ is the set of 1-Lipschitz functions and P_{fake} is the distribution learned by the Generator implicitly. In this case, the discriminator network is replaced with the critic network. The critic aims to assess the authenticity of its input and assigns a numerical score based on the similarity of the input to real data. The WGAN-gp [6] imposes a penalty on gradient norm to enforce the Lipschitz constraint. Hence, the final objective function is:

$$L = E_{\overline{x}\sim P_{fake}}[D(\overline{x})] - E_{x\sim P_{real}}[D(x)] + \lambda\, \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(|\nabla_{\hat{x}} D(\hat{x})|_{\{2\}} - 1)^2]$$

where λ is the penalty coefficient and $P_{\hat{x}}$ is implicitly defined by sampling uniformly on lines between patis of points sampled from P_{real} and P_{fake} .

2.4 GAN Architecture

¡Error! No se encuentra el origen de la referencia. and 4 illustrate the Generator and Discriminator structures, respectively. The Generator receives the start token and the noise vector from latent space. The start token is the mean value of each data channel and serves as a common starting point for our generation. Note that the start token is not part of the final generation. The Generator uses the start token as input of a GRU layer and employs the noise vector as an initial hidden state of the GRU layer. Then, the output of GRU passes through a fully connected layer to produce the first step of our fabricated time series. Next, we feed the generated time-step back to the generator to fabricate the next time step. This auto-regressive process continues until we generate the desired number of time-step (24 time steps in this scenario). The Discriminator receives a time frame from the generator or dataset and passes it through a GRU layer. Then, the output of the GRU layer is fed to a fully connected layer to obtain the final output of Discriminator.

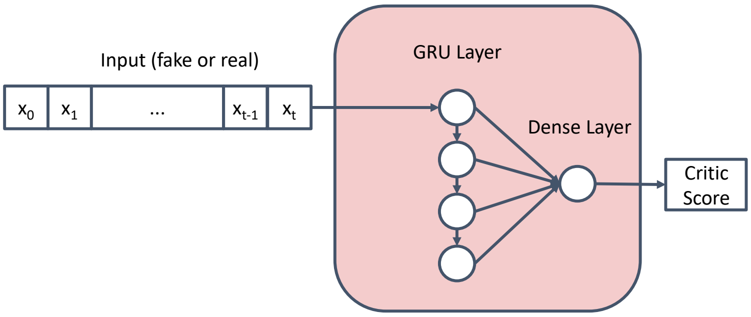


Figure 3. The architecture of critic

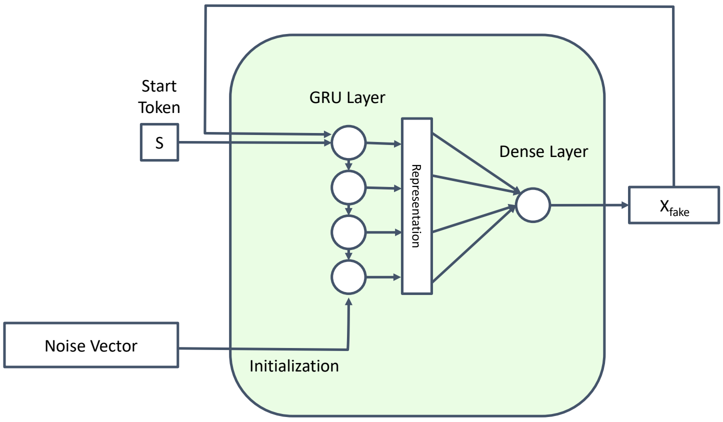


Figure 4. The architecture of generator

3 RESULTS AND DISCUSSION

3.1 Pre-processing pipeline

In this paper, we aim to generate artificial data to enhance the size of the dataset at hand, especially for rare situations where data sample scarcity makes it difficult for any model to learn the data pattern. **¡Error! No se encuentra el origen de la referencia.** illustrates our data preprocessing pipeline. Since the flow channel had a skewed distribution toward zero value (Figure), first, we applied log transform on this channel to obtain a less skewed distribution. Then, we standardized data distribution by a linear transformation to have mean=0 and standard deviation = 1. The next step of preprocessing the data turned into a series of time-frames using the rolling window technique with windows size = 24.

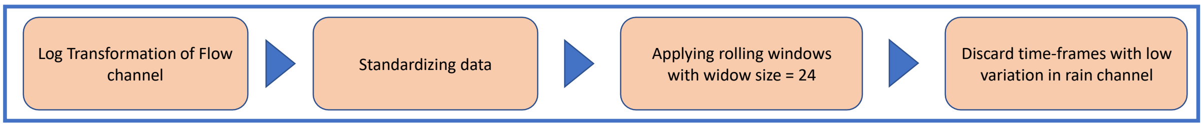
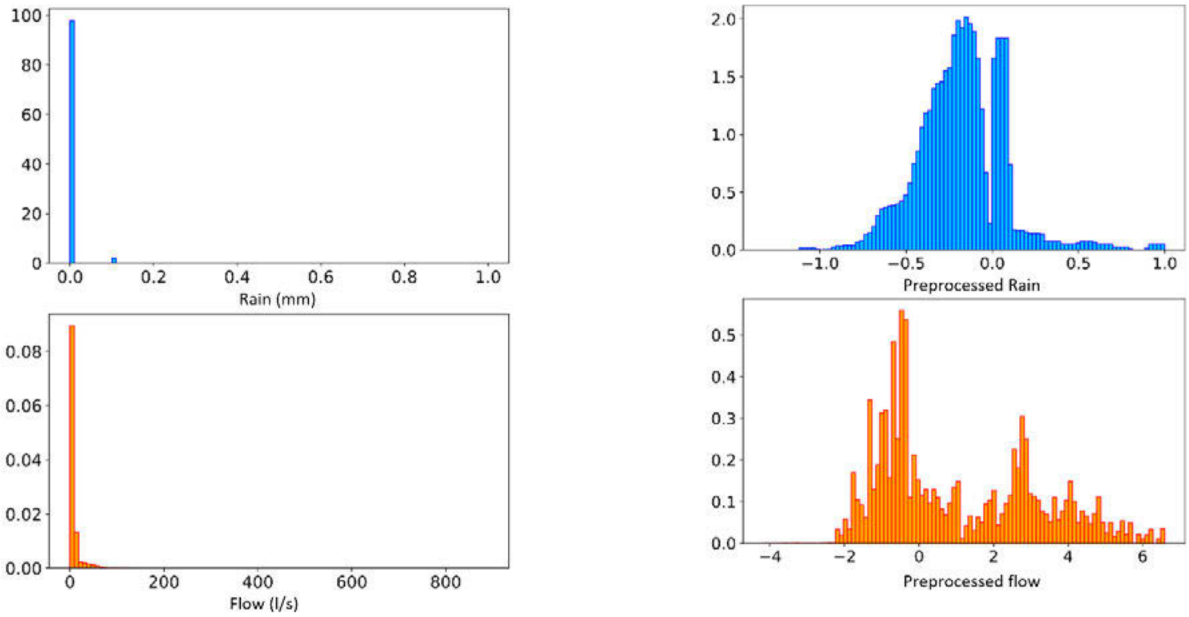

Log Transformation of Flow channel
Standardizing data
Applying rolling windows with widow size = 24
Discard time-frames with low variation in rain channel

Figure 5: The preprocessing pipeline

The rain channel data consists of long periods of dry days (low variation section) and a few short periods of rainy days (high variations). Our goal is to generate more rare cases (rainy days) to augment our data set. Therefore, we discard those data frames which did not have any variation in their time frame.



a) Original data distribution

b) Preprocessed data distribution

Figure 6: The data distribution before and after preprocessing

3.2 Experiment Set-up

In order to train a GAN successfully, we need to select the Discriminator and the Generator hyperparameters carefully. If one of the networks improves substantially faster than the other network during training, it is likely that training stops before reaching equilibrium. Hence, we perform a hyperparameter tuning for the Generator and the Discriminator. Table x presents the list of hyperparameters we tuned alongside the search space for each hyperparameter and the final value we select for the corresponding hyperparameter.

During training, the value function of GAN informs us about the performance of the Generator against the current Discriminator for a given batch. It does not provide any information regarding the performance of the Generator in general. Hence, we need a secondary assessment method to determine the best-performing model during the hyperparameter tuning process. The task of assessing a generative model in the time-series domain is a problem, and there is not any suitable measurement method for this task currently. For this work, we employed Jensen-Shannon Divergence (JSD) between the batch of real points from the dataset and the batch of fake points from the Generator.

Table 1: This table presents the list of GAN's hyperparameters alongside the search of space of each hyperparameter and the best hyperparameter we found during hyperparameter tuning

Hyperparameter	Search space	Best hyperparameter
GRU Layers in G	[1,4]	3
GRU layers in D	[1,4]	3
GRU hidden size in G	[32, 512]	450
GRU hidden size in D	[32, 512]	120

$$JSD(P_{real}||P_{fake}) = \frac{1}{2}KLD(P_{real}||M) + \frac{1}{2}KLD(P_{fake}||M),$$

where $M = \frac{1}{2}(P_{real} + P_{fake})$ and KLD is Kullback–Leibler divergence:

$$KLD(P||Q) = \sum_{x\in\chi} P(x)\log\left(\frac{P(x)}{Q(x)}\right),$$

where χ is the probability space. JSD does not consider the auto-correlation between the consecutive point in time series or the channels' correlation. However, it provides us with an easy to compute assessment method that can weakly measure the performance of GAN during the training and allows us to prune poor-performing configurations in the early stages. Then, we can select the best model based on the improvement of downstream tasks (in this case, forecasting) from the set of good-performing candidates.

For hyperparameter tuning, we utilized Bayesian Optimization Hyper Band Algorithm (BOHB) [7]. For implementing the networks, we used Pytorch [8] and for performing hyperparameter tuning, we employed RayTune package [9]. All experiments were executed on two Nvidia RTX 1080Ti graphic cards.

3.3 Data Augmentation with GAN

The best model that is obtained during hyperparameter tuning has JSD equal to 0.33. Figure 7 shows the generated time series.

BY NC SA

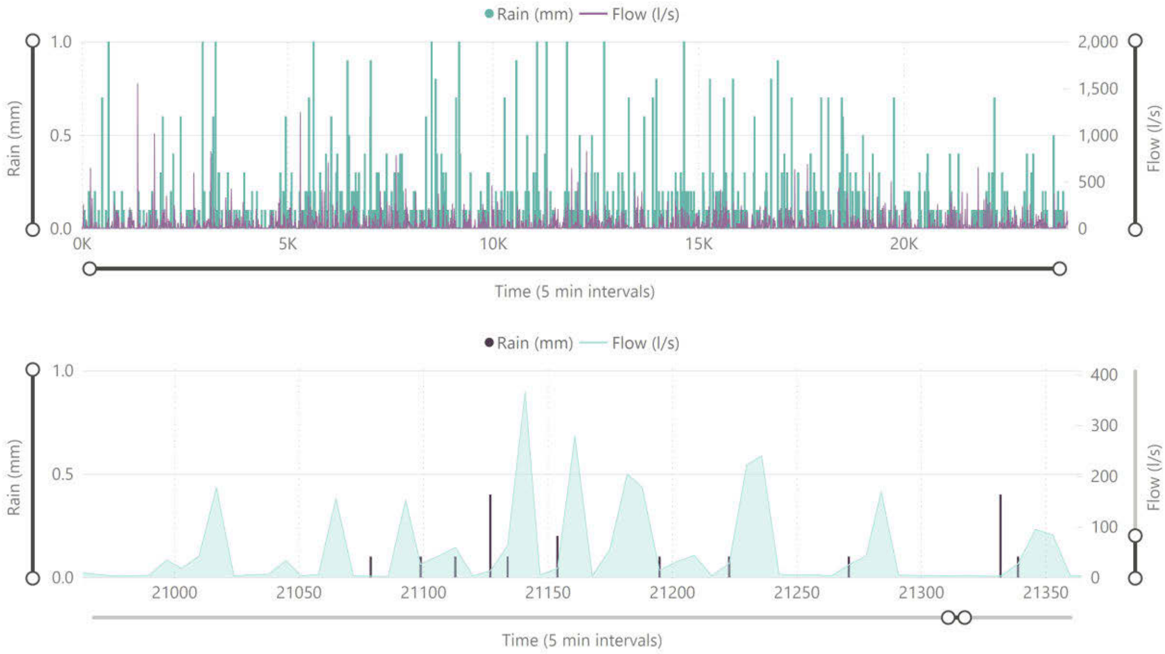


Figure 7: Synthetic time series represented in different resolutions

3.4 Data-Driven Forecasting Model

For training the data-driven forecasting model, we transformed the time series prediction problem into a supervised learning problem. We split the rainfall time series into 24-time steps (120 minutes) windows. The aim is to predict flow using these windows, as depicted in figure 8.

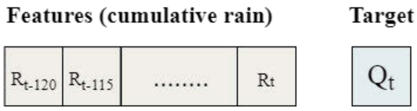
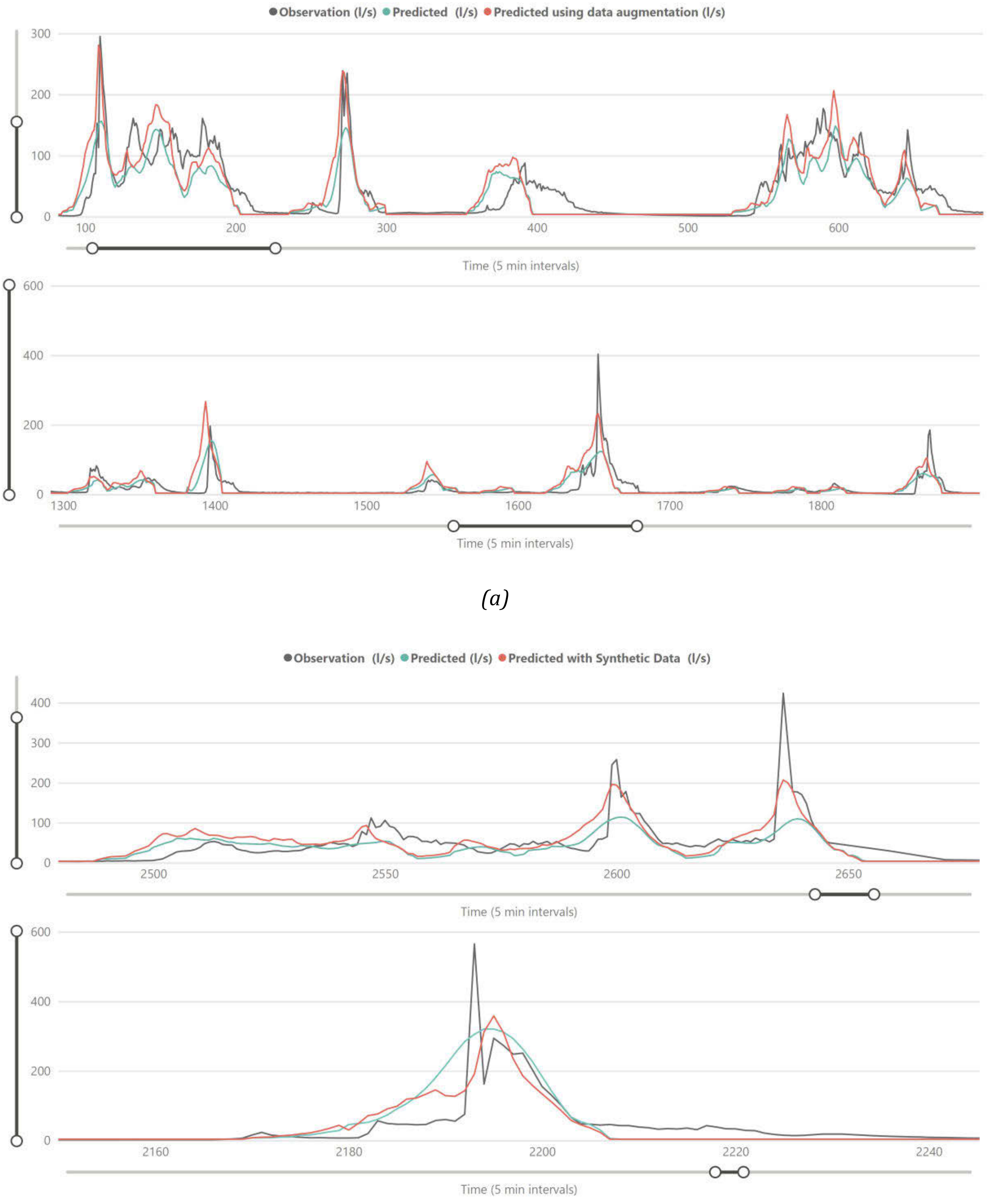


Figure 8: Splitting rainfall time series into 24-time steps windows for supervised learning

We trained two different models to solve this problem, one with synthetic data and one without. 8000 windows generated from the synthetic data set are used to balance the training data for the first model. We also balance the second model using the oversampling method for better comparison. The train data set for both models consist of 38000 windows from the real dataset. The test data set has 13000 windows only from the real data set. We built two sequential models in TensorFlow with some LSTM layers connecting to some Dense layers. ADAM method is used as an optimizer and means average error (MAE) as the cost function. A simple grid search is used to tune the hyperparameters, including the number of LSTM and Dense layers and their neurons and batch size. The results show that the model without synthetic data performs slightly better than the model with synthetic data. The MAE for the model without synthetic data is 5.01, while this value is 5.40 for the model with synthetic data. However, a closer look at the results reveals that the model with synthetic data has higher accuracy in predicting peak values which are the main concerns of this study. Figure 9 presents the results of both models for the test dataset. As it can be seen, the model with synthetic data predicts the peak values with higher accuracy for all extreme events. Still, the other model has better accuracy for dry weather conditions.



(a)

(b)

Figure 9: comparing flow predictions with observations for models with and without synthetic data

4 CONCLUSION

As moderate or extreme rain events are relatively rare in the historical data, pure data-driven rainfall-runoff models often underestimate the runoff when predicting these events. The optimal operation of the sewer system during extreme events, which result in the most critical states for the urban area and environment, depends highly on accurate flow predictions. To overcome these challenges, we used Generative Adversarial Networks (GANs). GANs are a class of methods capable of learning data distribution to generate artificial data. We developed a GAN model to generate synthetic time series to balance our limited and imbalanced recorded time series data to improve the accuracy of a data-driven model for combined sewer flow prediction using historical rainfall and measured flow data.

Results show that balancing the training dataset using the synthetic data generated by the developed GAN improves the accuracy of the data-driven model in peak flow prediction. However, this data augmentation method reduces the model performance in dry weather conditions. Some potential research themes to complement, expand, and build upon the presented study are given in the following.

- Clustering time series into dry and wet weather series and using different data-driven models for each series
- Using more complicated models like physic-informed ML
- Ensemble learning
- In the future, GANs can be applied for probabilistic anomaly detection and missing data imputation in urban water management [10]
- GANs can be used also for Probabilistic rainfall-runoff modeling for model predictive combines sewer overflow control [11]

5 REFERENCES

[1] S. Eggimann et al., “The Potential of Knowing More: A Review of Data-Driven Urban Water Management,” Environmental science & technology, vol. 51, no. 5, pp. 2538–2553, 2017, doi: 10.1021/acs.est.6b04267.

[2] Z. Ayazpour, A. E. Bakhshipour, and U. Dittmer, “Combined Sewer Flow Prediction Using Hybrid Wavelet Artificial Neural Network Model,” in Green Energy and Technology, New trends in urban drainage modelling, G. Mannina, Ed., New York NY: Springer Berlin Heidelberg, 2018, pp. 693–698.

[3] I. Goodfellow et al., “Generative adversarial nets,” Advances in neural information processing systems, vol. 27, 2014.

[4] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in International conference on machine learning, 2017, pp. 214–223.

[5] C. Villani, Optimal transport: old and new: Springer, 2009.

[6] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” Advances in neural information processing systems, vol. 30, 2017.

[7] A. Klein, S. Falkner, S. Bartels, P. Hennig, and F. Hutter, “Fast bayesian optimization of machine learning hyperparameters on large datasets,” in Artificial intelligence and statistics, 2017, pp. 528–536.

[8] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” Advances in neural information processing systems, vol. 32, 2019.

[9] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, “Tune: A Research Platform for Distributed Model Selection and Training,” arXiv preprint arXiv:1807.05118, 2018.

[10] A. Koochali, P. Schichtel, A. Dengel, and S. Ahmed, “Probabilistic Forecasting of Sensory Data With Generative Adversarial Networks – ForGAN,” IEEE Access, vol. 7, pp. 63868–63880, 2019, doi: 10.1109/ACCESS.2019.2915544.

[11] A. Koochali, A. Dengel, and S. Ahmed, “If You Like It, GAN It—Probabilistic Multivariate Times Series Forecast with GAN,” in The 7th International conference on Time Series and Forecasting, p. 40.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14702

SUSTAINABLE WASTEWATER TREATMENT SOLUTIONS FOR WATER-SMART CIRCULAR ECONOMY

Daniel Aguado[1], Henri Haimi[2], Anna Mikola[3], Ana Soares[4] and Ulf Jeppsson[5]

[1] CALAGUA – Unidad Mixta UV-UPV, Institut Universitari d'Investigació d'Enginyeria de l'Aigua i Medi Ambient – IIAMA, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain.
[2] 2 FCG Finnish Consulting Group Ltd, Osmontie 34, P.O. Box 950, FI-00601 Helsinki, Finland
[3] Department of Built Environment, Aalto University, PO Box 15200, FI-00076, AALTO, Finland
[4] Cranfield Water Science Institute, Vincent Building, Cranfield University, Cranfield, Beds, MK43 0AL, UK
[5] Division of Industrial Electrical Engineering and Automation (IEA), Department of Biomedical Engineering, Lund University, P.O. Box 118, SE-22100 Lund, Sweden.

[1] *daaggar@hma.upv.es*, [2] *Henri.Haimi@fcg.fi*, [3] *anna.mikola@aalto.fi*, [4] *a.soares@cranfield.ac.uk*,
[5] *ulf.jeppsson@iea.lth.se*

Abstract

The Protection of aquatic water bodies and human health is a paramount objective accomplished by wastewater treatment systems. Traditionally, pollutants are managed and removed in wastewater treatment plants (WWTPs), following a paradigm in which wastewater is considered a waste. Wastewater treatment requires significant amounts of resources, such as energy and chemicals, while sludge is produced, requiring further treatment.

A decade ago, a new paradigm emerged, suggesting that municipal wastewater is a source of resources, particularly reclaimed water, materials (e.g., nutrients) and energy. Many processes applicable for this new paradigm already existed, and others have been further developed (struvite-crystallization, membrane contactors, air-stripping, ionic exchange, electrodialysis, direct osmosis, etc.). Recently, resource recovery processes have been extensively developed and investigated to optimize their operation.

Reclaimed water can be used for recharging aquifers, irrigation in agriculture and cooling applications. Potential risks posed by the use of reclaimed water – and of other recovered wastewater resources – must be assessed and managed during the lifecycle of the application. For example, membrane separation processes are recognised as suitable for this application to remove pathogens and particles to ensure water quality.

Traditional WWTP design is based on effluent quality requirements and investment costs, with energy efficiency being only rarely considered. Larger facilities exhibit lower normalized electric consumption than smaller WWTPs, and older ones normally consume more than modern facilities (although is process dependent). For instance, in Spain it is possible to find facilities with specific electric consumptions 5-10 times higher than in modern and optimized facilities. This clearly reflects the great margin for potential energy savings.

Electricity consumption at WWTPs can be reduced by improving the processes and their operation, as well as through mechanical equipment improvement. The aeration of the biological process is the major electricity consumer; thus, control strategies have been deployed to its optimization. Also, less oxygen-demanding process alternatives have been explored, like the simultaneous nitrification-denitrification operated at very low dissolved oxygen concentration. Partial nitritation and deammonification processes with low oxygen consumption per nitrogen load removed, are especially suited for treating supernatant from sludge dewatering units. However, these low energy solutions might have a downside with direct greenhouse gas GHG emissions, especially N2O.

Anaerobic digestion of sludge, usually applied in large WWTPs, produces biogas that can generate both electricity and heat for local use or external use, through combined heat and power production, or liquefied biogas for external use. It is also possible to increase biogas

CC BY NC SA

production through co-digestion of external substrates, advanced control or sludge pre-treatment. Thermolysis processes, piloted for sewage sludge treatment, enable also waste-to-chemicals applications.

There are also other possibilities for energy recovery at WWTPs, such as thermal energy via heat exchangers and heat pumps, hydropower generation using turbines, and heat from sludge incineration. Energy can be also recovered by anaerobic digestion of microalgae grown in nutrient-rich wastewater.

In this paper, the transition towards sustainability and water-smart circular economy is illustrated showing how current WWTPs can be turned into Water Resource Recovery Facilities (WWRFs). The incorporation of sustainable pathways and technologies, make energy-positive facilities achievable, thus, reducing their climate impact.

Keywords
Circular economy, resource recovery, sustainability, wastewater treatment plant, water resource recovery facility.

1 INTRODUCTION

The main target of wastewater treatment plants (WWTPs) is to remove pollutants from wastewater in order to protect human health and the receiving natural aquatic ecosystems (Metcalf & Eddy, 2013). To achieve this purpose, these facilities feature a sequence of treatment processes that require significant input of energy and chemicals, producing a large amount of sludge which requires further treatment (Veera, 2021). In this wastewater management paradigm, wastewater is considered a waste.

A decade ago, a new paradigm emerged, suggesting that municipal wastewater is a source of resources, particularly reclaimed water, materials (e.g., nutrients) and energy (Neczaj and Grosser, 2018). Moving towards this new paradigm, the wastewater sector is contributing to the Circular Economy (CE). A CE system aims at improving the productivity of resources by keeping products, materials and infrastructure in use for longer than in the traditional linear 'take-make-consume-waste' economic model. CE has been promoted by policymakers in (e.g., European Commission 2020) and adopted by many industries (Mhatre et al., 2021) during past decade. The concept of water-smart Circular Economy (CE) brings further attention to water, resources in water and water-related ecosystems in context of CE (Salminen et al., 2022).

Resource recovery is not something new to the wastewater treatment sector, in fact, at some facilities it has been practiced for many years. The use of anaerobic digestion process to produce biogas as energy recovery strategy together with the application of the digestate (stabilized biosolids) as organic fertilizer has been practiced with success for over 100 years in different parts of the world (Pikaar et al., 2020). Potable drinking water has been produced from wastewater for more than 50 years in the wastewater treatment plant (WWTP) of the capital of Namibia (Windhoek). Namibia is one of the most arid countries in Africa, and most of the little rain that does fall is lost to evaporation, thus, reused water was sought as alternative water source to secure water supply in shortages. To guarantee healthy and high-quality drinking water, the WWTP treatment train was designed as a sequence of state-of-the-art barriers: ozone treatment, ultrafiltration membranes and residual chlorination (Veolia, 2014). This sequence of treatments has been coupled with rigorous monitoring programs to achieve and assure safe drinking water to population. These facts evidence that resource recovery technologies exist and have been practiced successfully at full-scale facilities for many years.

However, in global terms worldwide, only a small fraction of the resources that are conveyed within wastewater are recovered (Pikaar et al., 2020). It is important to highlight that reducing the energy consumption at full-scale facilities and recovering water, nutrients and energy, the wastewater treatment sector is reducing its contribution to greenhouse gas emissions (GHG) and thus its impact on climate change. GHG emissions from WWTPs are mainly associated with the high energy consumption of the facilities as well as the direct emissions of methane (CH4) and nitrous oxide (N2O) (Ye et al., 2022). Taking action to minimize the climate change is urgent, as its negative effects are so evident that can be experienced every year in the urban water cycle itself, with higher frequency of extreme weather events (floodings, droughts), combined sewers overflows, sea level rise, water shortages, and water quality deterioration issues.

Innovative cost-effective processes and strategies that allow resources to be recovered while meeting effluent quality requirements should be deployed at pilot, demonstration and full-scale, to demonstrate that they really work under real-life conditions as to date most research has been limited to the laboratory (Krueger et al., 2020).

Life cycle assessment (LCA) has been widely used to assess the environmental dimension of sustainability of resource recovery from wastewater alternatives, resulting in many papers published in recent years. However, these studies are usually case-specific and as the review of Lam et al. (2020) highlight, there is a need to improve the methodological consistency in these studies as well as to ensure the transparency of inventory and methods among other issues.

In this paper, the transition towards sustainability and water-smart circular economy is illustrated showing how two current WWTPs, of completely different scales (small ≤ 2.000 and large ≥ 100.000 P.E.), can be turned into Water Resource Recovery Facilities (WWRFs). The incorporation of sustainable pathways and technologies, make energy-positive facilities achievable, thus, reducing their climate impact.

2 WWTPS DESCRIPTION

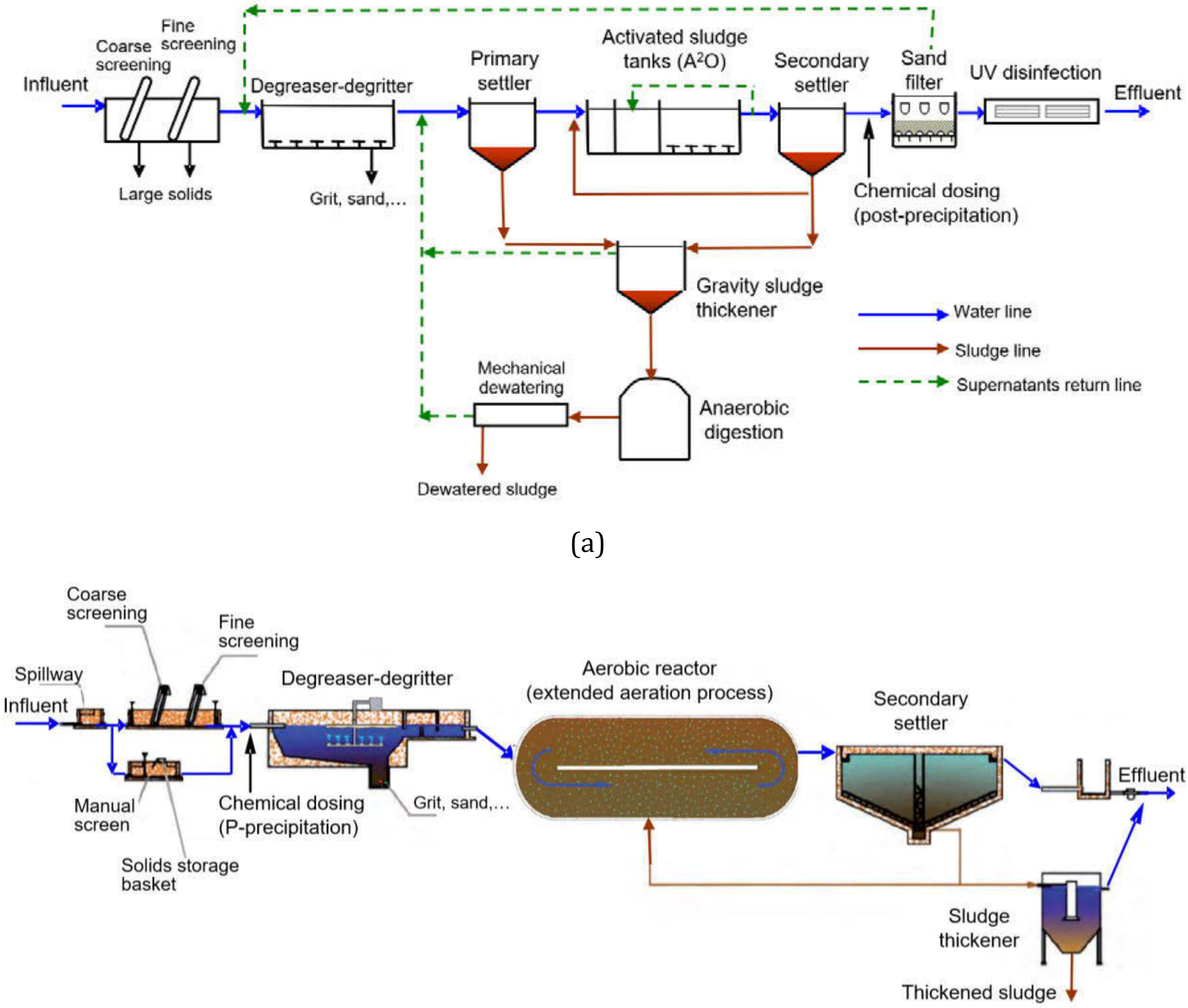
Figure 1 shows the layout of two completely different size WWTPs that will be used as case-base to propose modifications to transform them into WRRFs levering the existing infrastructure. As can be seen both WWTPs are based on activated sludge process.

Figure 1a shows the typical configuration for a large-size facility (~ 100.000 P.E.). Its water treatment line includes standard pre-treatment, primary settler to remove particulate organic matter from the water (to allow its valorisation in the anaerobic digester), organic matter and nutrient removal via biological processes featuring an A^2O process for enhanced biological phosphorus (P) removal, which will be complemented by chemical precipitation after the secondary settler to precipitate the soluble P that has not been previously removed, a sand filter that will allow the removal of particles that have not been retained previously enabling to reach the strictest P limits by eliminating the organic P contained in those particles, and finally UV disinfection prior to discharge the treated wastewater into the natural aquatic environment. The sludge line is mainly composed by a gravity sludge thickener, followed by a anaerobic digester that will generate biogas from the organic matter that receive from both settlers, the stabilised sludge will be dewatered mechanically, resulting in a sludge that can be evacuated by truck from the facility. Note that the gravity settler and the mechanical dewatering produce a reject water rich in nutrients that is returned to the water line.

Effluent discharge limits for the large-size WWTP include DBO_5<25 mg/L, DQO<125 mg/L, SS<35 mg/L, N_T<10mg/L and P_T<0.2mg/L.

Figure 1b shows a small-size WWTP (~ 2.000 P.E.). These types of facilities are much simpler, and usually the sludge line is simple composed by a sludge thickener. The water line features an activated sludge process, operated at high cellular retention times to allow the sludge to be somewhat stabilized, P-removal is accomplished via chemical precipitation and nitrogen (N) is not removed (i.e., the process and its operation is not designed for (biological) N removal, and only the N removed in that contained in the excess sludge).

Effluent discharge limits for the small-size WWTP include $DBO_5<25$ mg/L, DQO<125 mg/L, SS<35 mg/L, and $P_T<2$mg/L


Fine screening
Coarse screening
Influent
Degreaser-degritter
Large solids
Grit, sand,...
Primary settler
Activated sludge tanks (A^2O)
Secondary settler
Sand filter
UV disinfection
Effluent
Chemical dosing (post-precipitation)
Gravity sludge thickener
Water line
Sludge line
Supernatants return line
Mechanical dewatering
Anaerobic digestion
Dewatered sludge
Coarse screening
Fine screening
Spillway
Influent
Degreaser-degritter
Aerobic reactor (extended aeration process)
Secondary settler
Effluent
Manual screen
Chemical dosing (P-precipitation)
Grit, sand,...
Solids storage basket
Sludge thickener
Thickened sludge

(a)

(b)

Figure 1. Layout of two typical WWTPs configurations: (a) Large-size WWTP ~100.000 P.E. (b) Small-size WWTP ~2.000 PE

3 RESOURCE RECOVERY PROCESSES FOR SUSTAINABLE WASTEWATER TREATMENT

3.1 Large-size WWTP

3.1.1 From raw data to quality checked data fitted for purpose

The specific electrical energy consumption among different WWTPs is quite variable. In a survey of more than 140 WWTPs in Spain specific energy consumptions ranging from 0.1 to higher than 3.5 kWh/m^3 were reported (Ferrer *et al.*, 2016). Factors like the biotechnological process implemented in the WWTP, its configuration, the effluent quality requirements or its treatment capacity significantly impact the specific energy consumption. Conventional treatment technologies usually range from 0.3 to 0.6 kWh/m^3 (Sarpong and Gnaneswar, 2021). Therefore, WWTPs exhibiting noticeably higher specific consumptions have (in principle) a significant margin of energy savings.

The distribution of energy consumption among the different parts of a WWTP is approximately as follows: 44% aeration tank, 15% sludge thickening, 12% anaerobic digestion, 12% pumping, 12% deodorization. Electricity consumption at WWTPs can be reduced by optimizing the processes and their operation, as well as through mechanical equipment improvement (eg., using more efficient pumps, blowers and diffusers).

Since aeration is the most energy-intensive process (accounting in some cases more than 50% of the total energy consumption of the WWTP), efforts should focus on this process to achieve significant energy savings. Although less oxygen-demanding process alternatives exist (e.g., simultaneous nitrification-denitrification, ANaerobic AMMonium OXidation – ANAMMOX) and these biological processes have been explored even implemented in full-scale facilities, they have an important downside related to the emissions of direct greenhouse gas (GHG): N_2O. This potent GHG has a global warming potential 265 times greater than CO_2 (Myhre *et al.*, 2013), it is very persistent in the atmosphere (mean residence time > 100 years) and is considered the greatest ozone-depleting gas in the atmosphere (Ravishankara *et al.*, 2009). Since GHG emissions are a major cause of global warming, these low oxygen demand biological processes are not proposed to save energy in the WWTP upgrade. Therefore, the reduction in the energy consumption of the aeration process will be achieved via:

- Reducing the amount of organic matter that reach the aerobic tank which will lead to lower aeration needs in the biological reactor. This can be accomplished by enhancing the primary settling tank performance (as it will be discussed later).
- Using more efficient diffusers and blowers/compressors to supply the required oxygen to the aerobic tank and sizing the equipment in such a way that it is possible not only to meet the peak oxygen needs of the biological process (that traditionally was the only issue taken into account in its design), but also during night-time wastewater flow satisfy just the minimum oxygen needs without exceeding the amount required (thus, avoiding energy waste during nights as well as in other periods of low organic load).
- Adjusting the amount of oxygen supplied into the aerated tank according to the biological (and mixing) needs, via process control.

Energy from the wastewater can be recovered from the influent organic matter that it contains, from the kinetic energy of the flow in motion and also from the heat content of the wastewater.

Anaerobic digestion is the best option for recovering energy from the organic content of the sludge gathered in the primary and secondary settling thanks of medium and large-size WWTPs. During the anaerobic digestion of the organic compounds a biogas is produced (mainly composed by 60-70% CH_4 and 30-40% CO_2), which has a heating value around 37,3 KJ/m^3. Thus, using the biogas as a fuel source, electricity can be generated. The recovered energy is usually used directly in the WWTP itself. The biogas production can be increased by:

- Increasing the removal of biodegradable organic compounds in the primary settler (which typically is around 40% of the particulate material), with the added benefit of reducing the energy consumption in the aerobic reactor as less organic matter will reach the aerobic reactor via the water line, thus, less oxygen will be required by the microorganisms, and less energy will be consumed. This can be achieved by the implementation of micro-sieving followed by cloth media filter (reaching 80-90% of particulate material retention), or ultrafiltration membranes (100% of particulate material retention and part of the soluble organics are also retained since the pore size of the membranes is around 0.03 μm) after the primary settler. Although the first option (micro-sieving followed by cloth media) has a lower particulate retention capacity, it has a significative lower capital investment than the membranes, thus the payback would be noticeably shorter.

- Incorporate sludge thermal pre-treatment to increase its degradability. Thermal hydrolysis process has been successfully applied in full-scale facilities, reporting an increment in biogas production ranging 20 to 50% (Linh *et al.*, 2021).
- Adding external biodegradable organic compounds (e.g., food waste or other energy-rich organic wastes) which can also supply missing nutrients to the microbiota of the digester, thus favouring the anaerobic degradation process. This will provide an extra revenue generation while reducing odour problems in the streets associated with food waste in the garbage containers as well as the arrival of organic compounds to landfill.

Effluent water heat recovery with heat pumps. The implementation of heat pumps after the treatment train will allow to recover heat that can be used to provide hot water within the facility, to dry the sludge and to heat/cool the administrative, maintenance and utility buildings within the facility even residential buildings that are close to the facility.

Installing turbines to take advantage of kinetic energy of the effluent of the WWTP. Recovering energy from the flow of wastewater facility using micro-hydro-power (MHP) turbines has been shown to be a viable strategy to achieve energy savings at facilities with large flows rates and/or large available heads, the latter being relatively uncommon (Power *et al.*, 2017). According to these authors, traditional MHP turbines (eg., Kaplan and Francis) are viable for applications with large flow rates and power-production capacities exceeding 10 kW. Pump-as-turbines (PATs) are a low-cost alternative for MHP energy recovery allow a shorter payback but lower efficiencies (close to 60%) in contrast with traditional turbines (with efficiencies close to 80%). Splitting the flow through several turbines in parallel allows to increase the average efficiency of the MHP recovery system by allowing them operate close to their maximum efficiency point despite daily wastewater flow variations. The power recovery can be estimated as $P\ (\mathrm{kW}) = \rho\ (\mathrm{Tn/m^3})\ g\ (\mathrm{m/s^2})\ Q\ (\mathrm{m^3/s})\ H\ (\mathrm{m})\ e\ (\%)$, being ρ the fluid density, g the gravity acceleration (9.81 $\mathrm{m/s^2}$), Q the flow rate, H the available head at the turbine and e the overall efficiency of the MHP system. Therefore, for a WWTP serving 100.000 PE if the available head is lower than 5 m, PATs would the preferred option for recovering energy as long as low-cost low-head turbine solutions do not exist on the market.

Harvesting on-site wind and solar power. Although this is not energy from the wastewater, this option would allow to take advantage of the large space that these facilities normally have. Evidently, the techno-economic viability of their installation depends on local weather conditions.

3.1.2 Phosphorus recovery

P- recovery from wastewater has attracted scientific research interest for several reasons:

- It is a nutrient essential to life: humans, animals and plants need P. It is an important component of fertilizers used in agriculture.
- It is mainly obtained from phosphate rocks. This natural reserve has been predicted to be exhausted within the next century (Li *et al.*, 2019).
- If not removed from wastewater, its discharge into aquatic environments promotes their eutrophication.
- It can cause uncontrolled precipitation within the WWTP, increasing rugosity and reducing the effective cross-section of pipelines, increasing the energetic costs associated to pumping, even blocking pipes (Neethling and Benisch, 2004).

Traditionally P-removal is accomplished in WWTPs biologically (enhanced biological phosphorus removal, EBPR), chemically or a combination of both methods (Cornel and Schaum, 2009). Many different processes to recover P from different parts of a WWTP (effluent, digested sludge, sludge ash,…) have been sufficiently studied to be implemented and tested at full-scale. A description of a wide variety of technologies to recover nutrients from wastewater (discussing their potentials and limitations) can be found in Robles *et al.* (2020). The success of any P-recovery technique is highly dependent on the amount and the quality of the recovered product (usually a marketable fertilizer) which influence its final price. Taking into account investment and operational costs, the production costs of the ecological P-rich fertilizer of most recovery technologies developed have been calculated to double the production costs of the industrial fertilizer (Schaum *et al.*, 2019).

Thus, to recover P from wastewater in the large-size WWTP, a modification in the sludge line configuration is proposed, which enables a noticeably reduction in P precipitation in anaerobic digesters (> 40%, Bouzas *et al.*, 2019) with minimal structural modifications in the facility. The new configuration is shown in Figure 2 and consists in recirculating part of the thickened sludge to elutriate it in the gravity thickener. In this way, polyphosphate accumulating bacteria will release the P stored intracellularly towards the liquid phase, thus reducing the P-load to the anaerobic digester.

P will be recovered as struvite in a crystallizer should be constructed and that would receive two P-rich flow streams: the supernatant from the primary thickener and the centrifuge filtrate. Struvite is composed by phosphate, ammonium and magnesium ions in equal molar concentrations ($MgNH_4PO_4 \cdot 6H_2O$) and it is precisely a common scale that forms within the digester and pipe walls. The crystallization of struvite (magnesium ammonium phosphate - MAP) is nowadays considered one of the best alternatives to recover both ammonia nitrogen (N) and phosphate phosphorus (P) from wastewater (Robles *et al.*, 2020). It is a cost-effective process that yields a high-quality product that can be used as ecological and valuable slow-release fertilizer. The higher struvite production the lower demand for conventional fossil-fuel based fertilizers.

Another potential option for recovering P precipitate is from the back-wash water of the sand filters. In this case, its suitability depends on P concentrations in the back-wash, thus is very case-specific and not as general as the centrifuge filtrate obtained from the anaerobic digester effluent. The lower the effluent P limit the higher P concentration in backwash waters. To obtain the precipitate, the solids should be settled, thickened and dewatered. That precipitate should then be further processed.

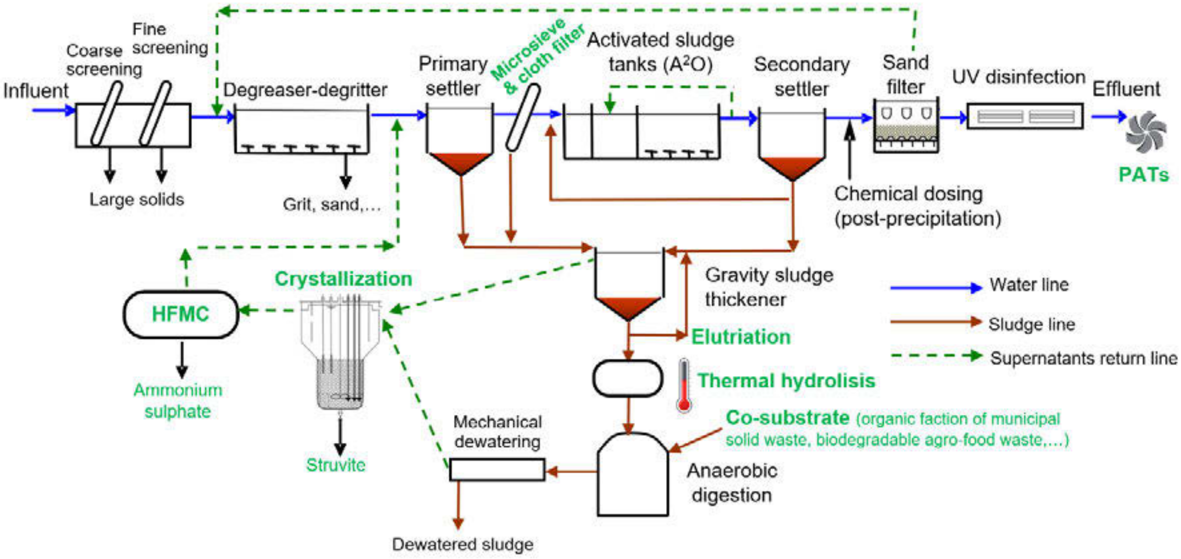

Fine screening
Coarse screening
Influent
Large solids
Degreaser-degritter
Grit, sand,...
Primary settler
Microsieve & cloth filter
Activated sludge tanks (A²O)
Secondary settler
Sand filter
UV disinfection
Effluent
PATs
Chemical dosing (post-precipitation)
Crystallization
HFMC
Ammonium sulphate
Struvite
Gravity sludge thickener
Elutriation
Thermal hydrolisis
Co-substrate (organic faction of municipal solid waste, biodegradable agro-food waste,...)
Anaerobic digestion
Mechanical dewatering
Dewatered sludge
Water line
Sludge line
Supernatants return line

Figure 2. Large-size WWTP layout showing the modifications to turn it into a WRRF.

3.1.3 Nitrogen recovery

N is a nutrient present in many commercial fertilizers. Despite being the most abundant gas in the atmosphere, its conversion into ammonium through the Haber-Bosch process is very energy-demanding. Its removal from wastewater is usually accomplished via biological nitrification-denitrification processes requiring high-energy consumption due to the oxygen that must be supplied the nitrifying bacteria.

The struvite crystallization, previously proposed to recover P from the reject water from anaerobic digestion also allows recovering part of the nitrogen present in this return-stream. As reported in Robles *et al.* (2020) P-recovery efficiencies ranging from 80 to 90% of the P-contained in the reject water from the anaerobic digester but only from 20 to 30% of the ammonia nitrogen present in this return-stream.

Therefore, other N-recovery process should be implemented in the large-size WWTP to recover the ammonia nitrogen and avoid its return to the waterline, where it would be removed via nitrification-denitrification increasing the energy consumption. Among the wide variety of technological processes existing, the hollow fibre membrane contactor (HFMC) has emerged as a promising technology for nitrogen recovery (as ammonium sulphate, which is a marketable fertilizer) from high nitrogen streams. Noriega *et al.* (2021) have shown that HFMC application to the reject water from the anaerobic digestion of a full-scale WWTP it is economically viable (as the expenses are offset by the benefits), with the great advantage of significant reduction in the global warming potential.

3.1.4 Water recovery

Water is the most obvious resource that can be recovered from wastewater. It has been practiced for many years for agriculture irrigation purposes, thus, contributing to alleviate the local water stress. In this way, higher quality water river resources are released for domestic drinking water purposes.

Although the reuse of treated wastewater for domestic direct potable use is technically possible, is not usually practiced as it is forbidden in most parts of the World and there is also the “yuck factor” of customers. Just the notion of drinking treated wastewater triggers this emotional discomfort on people. There are very few cases, like in Windhoek (Namibia), where potable use has been put into practice, and in these cases to guarantee the maximum the customer’s health a sequence of barriers (stages) has been deployed in treatment chain leading to a multi-barrier approach.

Non-potable uses are generally more acceptable by people than potable uses because the underlying perception of health risk associated with wastewater is significantly lower: agriculture, artificial wetlands, groundwater replenishment, land irrigation, industrial purposes, toilet flushing, …

Water can be recovered from wastewater together with nutrients and a reduction in global energy consumption of the facility, via fertigation. This is a reliable and controlled source of water and nutrients that are supplied simultaneously to agriculture crops and plants, and significantly reduces GHG emissions as it avoids energy-intensive wastewater treatment process to remove nutrients, that would be later supplied to crops in the form of fossil-fuel industrial fertilizers (Mainardis *et al.*, 2022). These authors recommend assessing the correspondence between facility treated effluent in terms of quality and flowrate and the agriculture irrigation demand before the adoption of fertigation.

3.2 Small-size WWTP

Small-size facilities are considered those that serve a population of less than or equal to 2.000 PE. Despite treating a relatively small wastewater flow, in global terms they are important as they are many facilities in this range size, for instance, only in Spain there are about 6000 municipalities with a population size in this range.

This type of facilities is usually designed with a simplified configuration to allow a robust operation with very little staff. For this reason, high maintenance requirements in terms of expertise (and cost) are no feasible for these facilities. Therefore, no advanced resource-efficiency technology would really be feasible in practice. Not even the use of PATs for recovering energy from the flow would be profitable unless the available head would be close to (or higher than) 5 m. According to Llácer-Iglesias *et al.* (2021), recent developments in small scale hydropower indicate that 100 W can be considered a reasonable cut-off value to suggest the technical feasibility of the installation of PATs in a WWTP.

Therefore, the best solution that could be done in these cases is increasing resource-efficiency by means of automatic control and outsourcing the further processing of sludge in a centralized plant that makes use of energy and nutrient content of the sludge. Additionally, if the climatic conditions allow it, harvest on-site wind and solar power.

Another option for this type of facilities, but not leveraging the existing infrastructure, would be to substitute the mainstream biological aerobic process by an anaerobic bio-membrane reactor (AnMBR) in combination with fertigation (as described in Jiménez-Benítez *et al.*, 2020). This configuration (Figure 3) will fit within the available space (the AnMBR is a compact solution) and will significatively reduce the CO_2 emissions due to both the organic matter valorisation and the minimization of the mineral fertilizer requirements. Moreover, the biogas production in the AnMBR could be easily increased by incorporating the OFMSW as co-substrate for the anaerobic process.

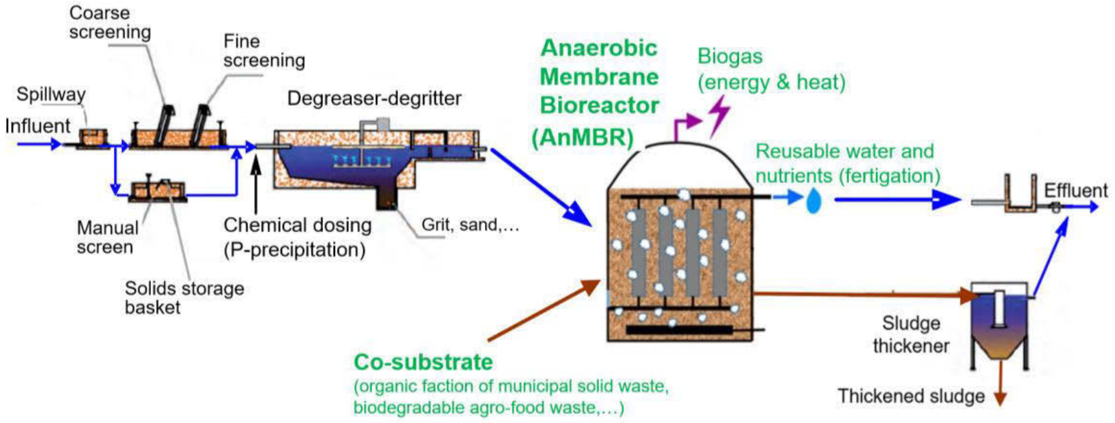

Coarse screening
Fine screening
Spillway
Influent
Degreaser-degritter
Anaerobic Membrane Bioreactor (AnMBR)
Biogas (energy & heat)
Reusable water and nutrients (fertigation)
Effluent
Manual screen
Solids storage basket
Chemical dosing (P-precipitation)
Grit, sand,...
Sludge thickener
Thickened sludge
Co-substrate
(organic faction of municipal solid waste, biodegradable agro-food waste,...)

Figure 3. Small-size WWTP layout showing the modifications to improve its resource recovery capabilities.

4 CONCLUSIONS

In this paper, the transition towards sustainability and water-smart circular economy has been illustrated showing how two extremely different sizes WWTPs can be turned into Water Resource Recovery Facilities (WWRFs). The main conclusions that can be drawn are as follows:

- The best way to save energy is not needing to use it:

  - therefore, diverting organic matter and nutrients from the water line to the sludge line for its anaerobic treatment should be promoted, when possible,
  - check all the pumping systems in the facility, as well as all the intensive-energy processes to find inefficiencies and solve them.
- A nice choice for increasing resource-efficiency is to implement advanced process control, for instance by reducing the aeration supplied and the chemicals dosed or by increasing the production of biogas.
- The high nutrient content in the reject water from the anaerobic digester makes this flow-stream ideal for implementing nutrient recovery techniques.
- Every WWTP, regardless of its size and configuration, can improve its resource recovery capability.
- Resource recovery in small-size WWTP is more tricky than larger ones, since their simplified configuration (to enable its operation with very little staff) and their lower influent flowrate, limit the techno-economically sensible alternatives. In most cases, further sludge processing would be preferable to do it in large-scale centralized plants off-site.
- Many different resource recovery alternatives already exist, but the optimal conversion is case-specific and should leverage the existing infrastructure.

5 REFERENCES

[1] Bouzas A., Martí N., Grau S., Barat R., Mangin D., Pastor L. (2019) Implementation of a global P-recovery system in urban wastewater treatment plants, Journal of Cleaner Production, 227, 130-140. 10.1016/j.jclepro.2019.04.126

[2] Cornel P., Schaum C., 2009. Phosphorus recovery from wastewater: needs, technologies and costs. Water Sci. Technol. 59, 1069e1077. https://doi.org/10.2166/ wst.2009.045.

[3] European Commission (2020) A new circular economy action plan – for a cleaner and more competitive Europe. Brussels, Belgium.

[4] Ferrer J., Aguado D., Barat R.; Serralta J.; Lapuente E.J. (2017). Huella Energética en el Ciclo Integral del Agua en la Comunidad de Madrid (Energy footprint in the Integral Water Cycle in the Community of Madrid (Spain)). Fundación Canal. Canal Isabel II. ISBN 978-84-945176-8-6

[5] Jiménez-Benítez A., Ferrer F.J., Greses S., Ruiz-Martínez A., Fatone F., Eusebi A.L., Mondéjar N., Ferrer J., Seco A. (2020). AnMBR, reclaimed water and fertigation: Two case studies in Italy and Spain to assess economic and technological feasibility and CO2 emissions within the EU Innovation Deal initiative, Journal of Cleaner Production,270,https://doi.org/10.1016/j.jclepro.2020.122398

[6] Krueger B. C., Fowler G. D., Templeton M.R., Moya B. (2020). Resource recovery and biochar characteristics from full-scale faecal sludge treatment and co-treatment with agricultural waste, Water Research, 169, https://doi.org/10.1016/j.watres.2019.115253

[7] Lam K. L., Zlatanović L., van der Hoek J.P. (2020). Life cycle assessment of nutrient recycling from wastewater: A critical review, Water Research,173, https://doi.org/10.1016/j.watres.2020.115519

[8] Llácer-Iglesias R.M.; López-Jiménez P.A.; Pérez-Sánchez M. (2021) Hydropower Technology for Sustainable Energy Generation in Wastewater Systems: Learning from the Experience. Water, 13, 3259. https://doi.org/10.3390/w13223259

[9] Li B., Boiarkina I., Yu W. , Huang H. M., Munir T., Wang G. Q. , Young B. R., (2019). Phosphorous recovery through struvite crystallization: Challenges for future design, Science of The Total Environment,2019, 1244-1256, https://doi.org/10.1016/j.scitotenv.2018.07.166

[10] Linh Ngo P., Udugama I. A., Gernaey K.V., Young B.R., Baroutian S. (2021). Mechanisms, status, and challenges of thermal hydrolysis and advanced thermal hydrolysis processes in sewage sludge treatment. Chemosphere, 281,13089.

[11] Mainardis M., Cecconet D., Moretti A., Callegari A., Goi D., Freguia S., Capodaglio A.G.(2022). Wastewater fertigation in agriculture: Issues and opportunities for improved water management and circular economy, 296, 118755. doi: 10.1016/j.envpol.2021.118755

[12] Mhatre, P., Panchal, R., Singh, A. and Bibyan, S. (2021): A systematic literature review on the circular economy initiatives in the European Union. Sustainable Production and Consumption, 26, pp. 187-202.

[13] Metcalf & Eddy (2013. Wastewater Engineering: Treatment and Resource Recovery, 5ed. ISBN: 978-0073401188

[14] Neczaj E., Grosser A. (2018). Circular Economy in Wastewater Treatment Plant–Challenges and Barriers. Proceedings.; 2(11):614. https://doi.org/10.3390/proceedings2110614

[15] Neethling J.B., Benisch M. (2004) Struvite control through process and facility design as well as operation strategy. Water Sci Technol; 49 (2): 191–199. doi: https://doi.org/10.2166/wst.2004.0122

[16] Noriega-Hevia, G; Serralta, J;Seco, A; Ferrer, J. (2021). Economic analysis of the scale-up and implantation of a hollow fibre membrane contactor plant for nitrogen recovery in a full-scale wastewater treatment plant, Separation and Purification Technology, 275,119128. https://doi.org/10.1016/j.seppur.2021.119128

[17] Pikaar I., Huang X., Fatone F., Guest S. (2020). Resource recovery from water: From concept to standard practice, Water Research, 178, 115856, https://doi.org/10.1016/j.watres.2020.115856

[18] Ravishankara A.R., Daniel J.S., Portmann R.W. (2009). Nitrous Oxide (N2O): The Dominant Ozone-Depleting Substance Emitted in the 21st Century. Science, 326 pp. 123-125. doi:10.1126/science.1176985

[19] Robles A., Aguado D., Barat R., Borrás L., Bouzas A., Giménez J.B., Martí N., Ribes J., Ruano M.V., Serralta J., Ferrer J., Seco A. (2020) New frontiers from removal to recycling of nitrogen and phosphorus from wastewater in the Circular Economy, Bioresource Technology, 300. https://doi.org/10.1016/j.biortech.2019.122673

[20] Salminen, J., Määttä, K., Haimi, H., Maidell, M., Karjalainen, A., Noro, K., Koskiaho, J., Tikkanen, S. and Pohjola, J. (2022) Water-smart circular economy - Conceptualisation, transitional policy instruments and stakeholder perception. J. Clean. Prod., 334, 130065.

[21] Sarpong G. and Veera G.G. (2022). Energy consumption and recovery in wastewater treatment systems. In Resource Recovery from wastewater: Toward Sustainability. CRC Press. ISBN 978-1771889285

[22] Schaum C., Hubert C., Krause S., Steiniger B. (2022). Phosphorus removal and recovery in water resource recovery facilities. In Resource Recovery from wastewater: Toward Sustainability. CRC Press. ISBN 978-1771889285

[23] Veera G.G. (2022). Resource Recovery from wastewater: Toward Sustainability. CRC Press. ISBN 978-1771889285

[24] Veolia (2014). Veolia installs wastewater plant in Windhoek, Namibia. Membrane Technology, 10, https://doi.org/10.1016/S0958-2118(14)70208-3.

[25] Ye L., Porro J., Nopens I. (2022). Modelling of Fugitive Greenhouse Gas Emissions from Urban Water Systems. IWA Publishing. ISBN: 9781789060454

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry

Valencia (Spain), 18-22 July 2022

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14703

MACHINE LEARNING METHODOLOGIES TO PREDICT POSSIBLE WATER QUALITY ANOMALIES AS A SUPPORT TOOL FOR ONLINE MONITORING OF ORGANIC PARAMETERS

Leonid Kadinski[1], Jonas Schuster[2], Gopinathan R. Abhijith[3], Cao Hao [4], Anissa Grieb[5], Thomas Meier[6], Pu Li[7], Mathias Ernst[8] and Avi Ostfeld F. ASCE [9]

[1] PhD Student, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[2] PhD Student, Institute for Water Resources and Water Supply Hamburg University of Technology Am Schwarzenberg-Campus 3E, 21073 Hamburg, Germany
[3] Post-doctoral researcher, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
Schwarzenberg-Campus 3E, 21073 Hamburg, Germany
[4] PhD Student, Process Optimization Group, Institute of Automation and Systems Engineering, Technical University Ilmenau, 98693 Ilmenau, Germany
[5] Post-doctoral researcher, Institute for Water Resources and Water Supply Hamburg University of Technology Am Schwarzenberg-Campus 3E, 21073 Hamburg, Germany
[6] Head of Microbiology, Drinking Water Laboratory Hamburg Wasser, Billhorner Deich 2, 20539 Hamburg, Germany
[7] Professor, Process Optimization Group, Institute of Automation and Systems Engineering, Technical University Ilmenau, 98693 Ilmenau, Germany
[8] Professor, Institute for Water Resources and Water Supply Hamburg University of Technology Am Schwarzenberg-Campus 3E ,21073 Hamburg, Germany
[9] Professor (corresponding author), Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

[1] kleonid@campus.technion.ac.il, [2]jonas.schuster@tuhh.de, [3]gnrabhijith@gmail.com,
[4]hao.cao@tu-ilmenau.de, [5]anissa.grieb@tuhh.de, [6]thomas.meier@hamburgwasser.de,
[7]pu.li@tu-ilmenau.de, [8]mathias.ernst@tuhh.de, [9] ostfeld@technion.ac.il

Abstract

Water Distribution Systems (WDSs) function to deliver high-quality water in major quantities. While standard water quality parameters are monitored at waterworks, it is still a challenge to monitor water quality in the WDS network itself. While mostly hydraulic parameters are frequently monitored and modelled in drinking water networks in Germany, the measurements of specific organic and bacteriological water quality parameter are still done offline which can take hours or even days, which might be too late to react to possible water events. This study utilizes water quality data of a Utility in Hamburg, Germany to train machine learning algorithms to predict possible anomalies in specific water quality parameters which can indicate the necessity for more thorough investigations. While a large amount of water parameters is utilized and checked for deviations from the normal distribution, the input features to train the machine learning algorithms are parameters which can be measured online like pH, temperature, total cell count of bacteria and the organic content of the water sample. A parallel study uses innovative online testing methods like fluorescence spectroscopy and flow cytometry in batch and flow experiments with the overarching goal of validating the trained algorithm to develop a wholesome online monitoring and warning system for drinking water anomalies. Various algorithms like Random Forest and Artificial Neural Networks are trained to predict whether the water samples indicate possible water quality anomalies. First results of this study show promising possibilities for a data driven online water quality prediction methodology which can help to digitalize the water sector immensely.

BY NC SA

Keywords
Water distribution systems, contamination response, water quality monitoring, machine learning.

1 INTRODUCTION

To operate, manage and monitor water distribution systems (WDSs) is a highly complex and multidisciplinary task for water utilities. Especially monitoring water quality parameters in a continuous manner needs consideration of various boundary conditions. Specific parameters are frequently monitored at the water works because conducting measurements inside water networks is a complicated endeavour. Fixed water quality sensors and the optimized placement of them at strategically important places has shown to be highly efficient [1–4]. In Germany and in some other European countries, the drinking water in a WDS do not contain any residual disinfectants (chlorine) [5]. In these specific cases it is very important to monitor organic and bacteriological water quality parameters. Monitoring organic parameter usually requires time-intensive laboratory tests and the results are not available at once. Schuster et al (2022) [6] describe an innovative, sensitive and low-cost method how to determine organic compounds in drinking water with fluorescence sensors and flow cytometers. Recent studies have used machine learning methodologies to detect water quality anomalies and their sources by training these algorithms with specific water quality parameters to predict possible water events. Artificial and convolutional neural networks (ANN/CNN) and support vector machines (SVM) have been used to detect whether a contamination event has occurred in a water network [7,8]. Hamburg Wasser, the water utility of the city of Hamburg in Germany, provided this study with a major amount of water quality data of a time period of five years. The data from 22 sampling point in 19 water works was pre-processed and analysed with various data evaluation packages in Python. The objective is to create a support tool for water utilities as a method for exploring water quality online with e.g. flow-cytometry in real time and understanding whether additional laboratory check-ups of the water quality need to be conducted.

2 METHODOLOGY

The goal of the presented method is to develop an efficient method to detect water quality anomalies in real time by evaluating parameters that are measured online and to predict whether a drinking water sample needs follow up checks or not. After the first pre-processing and evaluation of the water quality data, the correlation of various parameters was conducted to understand for which parameter it is reasonable to train the machine learning algorithms. Although a major amount of data was provided, there was still a lack of various organic water quality parameter in the dataset. The sparse dataset was imputed with the k-Nearest Neighbor (kNN) Imputation method where the kNN algorithm is used to replace missing values. With this method, the features of the missing neighbour values are uniformly averaged or weighted according to the distance to each other [9]. While a big part of the data had to be imputed, the authors considered the imputation as sufficient for a proof of concept for the presented method. After preparing and pre-processing the data, an ANN and random forest algorithm is trained to predict water quality anomalies. The algorithms are trained with the open source Python packages Scikit-Learn [9] and TensorFlow [10]. A random forest is a classification method which belongs to the decision tree family and has various advantages regarding accuracy and efficiency compared to other models [11,12]. An artificial neural network is a machine learning model which is composed of an input layer, various hidden layers and an output layer which represents the prediction output. All of these layers consist of interconnected neurons [13–15]. A neural network is a mathematical model which predicts a specific value based on the features and data it is trained with and consist of smaller and simpler mathematical functions. Figure 1 shows a conceptual layout of an ANN where the respective data for the input and output layer of the neural network is shown. The training dataset has around 3% values which were considered as anomalies. These are not necessarily hazardous contaminations but can also be values which vary from the normal distribution of the specific water quality parameter in the dataset.

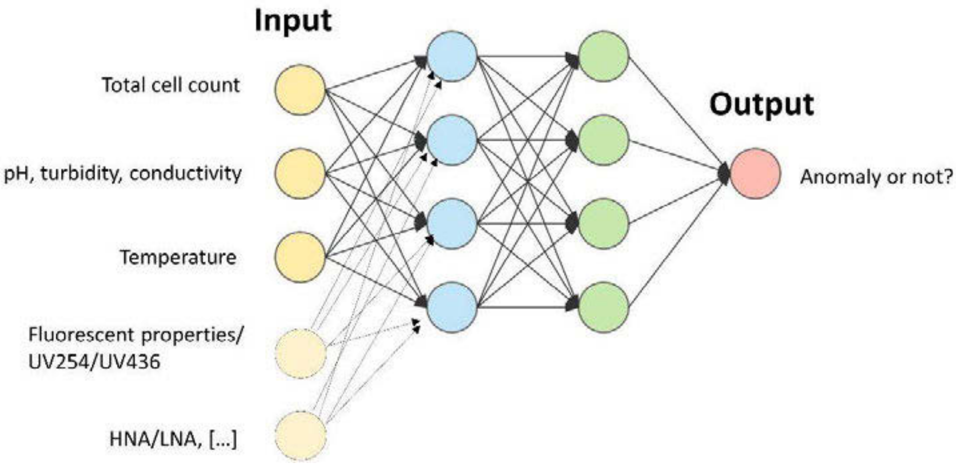


Figure 1: Conceptual layout of ANN where input features are water quality parameters that can be measured online and the output is a prediction of a possible anomaly in the water network system.

The process workflow of developing the water quality event prediction support tool follows this sequence:

- Data preparing and pre-processing
- Imputing missing data
- Training machine learning algorithm with training data set (ANN/random forest algorithm) and tuning of hyperparameter
- Testing algorithms and determining accuracy of water quality predictions with as specifically determined test data set.

3 RESULTS

The presented study introduces and presents the results of a simple case study which was conducted with the obtained water quality data training a random forest algorithm. The results which are presented were produced training a random forest algorithm to simplify the proof of concept which is meant to be shown.

Figure 2 shows the correlation of specific chosen water quality parameters in the provided dataset). The parameters chosen are all parameters which can be measured online and consequently evaluated in “real time”. They include the conductivity, total organic carbon TOC, the total cell count (TCC), all intact cells (TCCi), Turbidity, pH, UV436, temperature, O2 and high or low nucleic acid content (HNA/LNA) of the bacteria in the water. The correlation heatmap was created with the Python open source package seaborn [16].

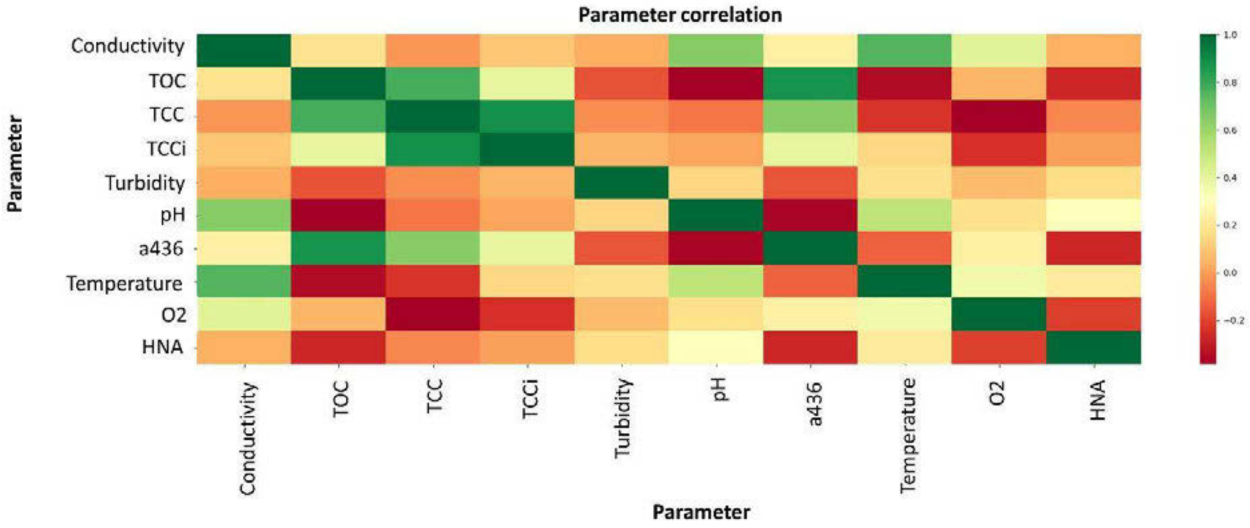


Figure 2. Correlations drawn between chosen water quality parameters

As illustrated in Figure 2, specifically the organic parameters strongly like TCC and TOC correlate with each other. However, there are correlations which might not be as clear as that, e.g., temperature with the pH or conductivity parameters of a drinking water sample. This correlation heatmap can be used to determine parameters which can be used in a sensible way for training a machine learning algorithm to give accurate predictions about the possibility of a water quality event in a WDS.

Figure 3 shows the heatmap of the confusion matrix of the water quality anomaly predictions. It was created with the Python package seaborn [16]. While the accuracy given by Scikit-Learn presents a very high value of 97%, the actual predictions need to be analysed more thoroughly. A major part of the predictions are true negatives while there have been no true positives. Which gives a false impression of the accuracy value which was put out by the software. Although it needs to be pointed out that the actual occurrences of anomalies in the dataset were comparably rare. The accuracy was tested on a test set of around three thousand samples.

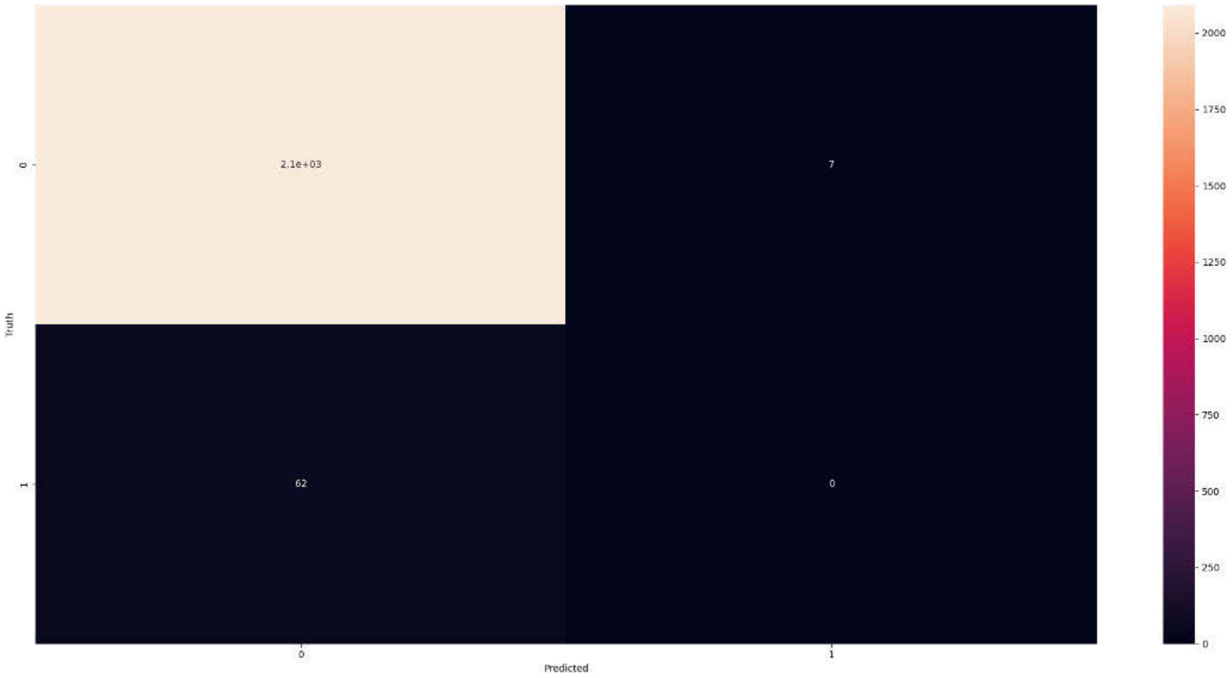


Figure 3. Confusion matrix heatmap of the water quality event predictions of the trained random forest algorithms

4 CONCLUSIONS

This study presented the proof of concept for a method to determine water quality anomalies with the online measurements of specific water quality parameters paired with a machine learning algorithm to predict a possible water event. Water quality data of the Hamburg Wasser water utility was utilized to train an ANN and a random forest algorithm, of which the latter has been presented in the results of this study. The accuracy of the random forest algorithm for predicting a water event came out to 97% and is considerably very high but it needs to be acknowledged that it came from mostly true negatives and not from true positive water event predictions. Utilizing a data driven model as a support tool for water utilities is a very promising technology and should continuously be explored with support from water utilities which have the possibility to consistently collect relevant water quality data. Future work will require more data samples so the algorithms have more bandwidth of water quality parameter that they can learn from and a more comprehensive sensitivity analysis of various machine learning algorithms to predict water quality anomalies.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support provided by the by the Israeli Ministry of Science and Technology (MOST) and the German Federal Ministry of Education and Research (BMBF) under the project numbers 3-17011 (MOST) and 02WIL1553A (BMBF).

6 REFERENCES

[1] Janke, R.; Murray, R.; Uber, J.; Taxon, T. Comparison of Physical Sampling and Real-Time Monitoring Strategies for Designing a Contamination Warning System in a Drinking Water Distribution System. 2006, 310–314.

[2] Hall, J.; Zaffiro, A.D.; Marx, R.B.; Kefauver, P.C.; Radha Krishnan, E.; Haught, R.C.; Herrmann, J.G. On-line water quality parameters as indicators of distribution system contamination. J. / Am. Water Work. Assoc. 2007, 99, doi:10.1002/j.1551-8833.2007.tb07847.x.

[3] Ostfeld, A.; Asce, M.; Salomons, E. Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security. 2004, 130, 377–385.

[4] Ostfeld, A.; Über, J.G.; Salomons, E.; Berry, J.W.; Hart, W.E.; Phillips, C.A.; Watson, J.P.; Dorini, G.; Jonkergouw, P.; Kapelan, Z.; et al. The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. J. Water Resour. Plan. Manag. 2008, 134, 556–568, doi:10.1061/(ASCE)0733-9496(2008)134:6(556).

[5] Rosario-Ortiz, F.; Rose, J.; Speight, V.; Gunten, U. v.; Schnoor, J. How do you like your tap water? Science (80-.). 2016, 351, 912–914, doi:10.1126/science.aaf0953.

[6] Schuster, J.; Kadinski, L.; Cao, H.; Abhijith, G.R.; Grieb, A.; Li, P.; Ostfeld, A.; Asce, F.; Ernst, M. Real-time monitoring and controlling of water quality in water distribution networks based on flow cytometry and fluorescence spectroscopy. In Proceedings of the World Environmental & Water Resources Congress; 2022.

[7] Asheri Arnon, T.; Ezra, S.; Fishbain, B. Water characterization and early contamination detection in highly varying stochastic background water, based on Machine Learning methodology for processing real-time UV-Spectrophotometry. Water Res. 2019, 155, 333–342, doi:10.1016/j.watres.2019.02.027.

[8] Ashwini, C.; Singh, U.P.; Pawar, E.; Shristi Water quality monitoring using machine learning and iot. Int. J. Sci. Technol. Res. 2019, 8, 1046–1048.

[9] Pedregosa, F.; Weiss, R.; Brucher, M. Scikit-learn : Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.

[10] Martín Abadi, Ashish Agarwal, Paul Barham, E.B.; Zhifeng Chen, Craig Citro, Greg S. Corrado, A.D.; Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, I.G.; Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Y.J.; Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, M.S.; Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, J.S.; Benoit Steiner, Ilya Sutskever, Kunal Talwar, P.T.; Vincent Vanhoucke, Vijay Vasudevan, F.V.; Oriol Vinyals, Pete Warden, Martin Wattenberg, M.W.; Yuan Yu, and X.Z. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 2015.

[11] Lee, Y.J.; Park, C.; Lee, M.L. Identification of a Contaminant Source Location in a River System Using Random Forest Models. 2018, 1–16, doi:10.3390/w10040391.

[12] Grbčić, L.; Kranjčević, L.; Družeta, S. Machine Learning and Simulation-Optimization Coupling for Water Distribution Network Contamination Source Detection. Sensors 2021, 21, 1157, doi:10.3390/s21041157.

[13] Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press, 2016;

[14] Fan, X.; Zhang, X.; Yu, X.B. Machine learning model and strategy for fast and accurate detection of leaks in water supply network. 2021, 6.

[15] Grbčić, L.; Lučin, I.; Kranjčević, L.; Družeta, S. Water supply network pollution source identification by random forest algorithm. J. Hydroinformatics 2020, 22, 1521–1535, doi:10.2166/HYDRO.2020.042.

[16] Waskom, M. seaborn: statistical data visualization. J. Open Source Softw. 2021, 6, 3021, doi:10.21105/joss.03021.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry

Valencia (Spain), 18-22 July 2022

doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14709

MULTI-OBJECTIVE OPTIMIZATION OF WATER DISTRIBUTION SYSTEM UNDER UNCERTAINTY USING ROBUST OPTIMIZATION

Sriman Pankaj Boindala[1] and Avi Ostfeld F. ASCE[2]

[1] PhD Student, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

[2] Professor (Corresponding Author), Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

[1] *srimanpankaj@gmail.com*, [2] *ostfeld@technion.ac.il*

Abstract

Uncertainty is inevitable while trying to tackle any real-world problem, Water distribution system design problem is not an exception for that. Among the many uncertainties involved in the design problem, demand uncertainty is the most important. The current study aims to provide efficient designs that can handle the predicted variations in demand without compromising the resilience of the network system. Most previous studies explored stochastic solutions to handle the uncertainty that consumed immense computational power. Innovations in developing non-probabilistic techniques like robust optimization paved the way to reduce computational time and handle uncertainty efficiently. A new methodology is proposed in this study to obtain efficient designs for the multi-objective design problem of WDS under uncertainty. The proposed methodology uses a combination of robust optimization approaches to handle the uncertainty and a multi-objective cuckoo search algorithm. The proposed methodology is applied to a common benchmark water distribution system problem, and the designs are compared with the nominal designs obtained when demand is assumed as certain. Furthermore, the effect of considering different uncertainty sets is discussed.

Keywords

Water Distribution System, Optimization Under Uncertainty, Multi-objective Optimization, Robust Optimization, Self Adaptive Multi-Objective Cuckoo Search Algorithm (SAMOCSA).

Optimal design and management of water distribution systems (WDS) is an extensively explored research area in the water resources field. The main objective of the optimal design of WDS is to minimize the design cost (i.e. Pump capacity, Tank size, Pipe diameter, etc.) such that the system satisfies the hydraulic and water quality constraints. Over the past four decades, multiple variants of this problem have been explored using many optimization approaches. The research explored linear programming, non-linear programming, as well as dynamic programming approaches(Alperovits and Shamir 1977; Avi Ostfeld and Shamir 1996.; Kessler and Shamir 1989). Later with the development of evolutionary meta-heuristic algorithms, these techniques have also been explored to solve the problem (Ostfeld et al. 2008.; Savic and Walters 1997; Vasan and Simonovic, 2010.; Wu et al. 2005). Even though these approaches could provide the least cost design alternative that could satisfy the required constranits, the obtained designs could not meet the expected reliability. Walski (2001) emphasized incorporating additional objectives to the least-cost design problem like reliability, capacity, and resistance to uncertainty. With the development of multi and many-objective optimization algorithms and Pareto-front representation of designs, the research moved towards incorporating multiple conflicting objectives into the problem definition. Over the last two decades, optimal multi-objective WDS design has been extensively explored with studies considering two objectives: minimization of cost and resilience/robustness/ reliability (Pankaj et al. 2020; Perelman et al. 2008; Prasad and Park 2004; Wang et al. 2014), three objectives(Farmani et al. 2006; Wu et al. 2013) and also six

CC BY NC SA

objectives (Fu et al. 2013). Most of these studies considered the design problem to be deterministic and ignored the uncertainties associated with various design variables. In most real cases, this assumption is not true. Babayan et al. 2005 stated that almost all the design parameters associated with the WDS design problem have some uncertainty. Even though the multi-objective optimization approach considering maximization of resilience/ reliability provided some protection against uncertianiy, they lacked quantitative realization of the protection level achieved. This motivated the resercheres to solve stochastic formulation of the WDS design problem.

Lansey and his team were the pioneers in developing a methodology to solve the stochastic least-cost design problem. They used chance constraints to address the stochastic nature of the problem. They used GRG-2 methodology to obtain the stochastic optimal designs with various levels of protection for the two-loop network (Lansey et al. 1989). Later, Sumer and Lansey (2005) proposed a stochastic optimization model using First Order Second Moment (FOSM) uncertainty analysis. Other stochastic approaches have been explored using various probabilistic analysis methods to solve the stochastic least-cost design of the WDS problem (Jung et al. 2012, 2014; Seifollahi-Aghmiuni et al. 2013). Table -1 summarizes few of the studies that used probabilistic approaches to solve the WDS design problem.

Although the probabilistic approach successfully handled the uncertainty, the considerable computational time and the uncertainty in the assumption of probability density function (PDF) hindered its practical application. The researchers moved toward applying non-probabilistic approaches to handle the uncertainty to overcome these disadvantages. The robust optimization (RO) is a non-probabilistic approach that has been getting attention in the recent past to handle the uncertainty. This technique was successfully applied to WDS least-cost design by (Perelman et al. 2013a, b) under single loading conditions and (Schwartz et al. 2016) under multiple loading conditions. This approach has also been used to handle water quality constraints (Pankaj et al. 2022). Non-probabilistic approach application in the design of WDS is limited to the least cost design problem. There have not been any studies that used these methods to solve a multi-objective problem considering cost minimisation and resilience maximisation as objectives.

Table 1 List of few works in the area so WDS design and analysis under uncertainty

Uncertain parameters	**PDF assumed**	**Uncertainty handling techniques**	**Optimization techniques**	**References**	**Type**
q,H, RC	Normal	FORM	GRG2	(Xu and Goulter 1999)	Hydraulic analysis
q,H, RC	Normal	MCS	GRG2	(Lansey et al. 1989)	SO Design
q,RC	Normal	MCS	SFLA	(Seifollahi-Aghmiuni et al. 2013)	Hydraulic analysis
q,RC	Normal	FOSM & MCS	--	(Hwang et al. 2018)	Hydraulic analysis
q	Gaussian	LHS	GA	(Babayan et al. 2004)	SO Design
q,RC	Normal	FORM	GA	(Tolson et al. 2004)	SO Design

q	Gaussian	LHS	RNSGA-II	(Kapelan et al. 2005a)	MO Design
q	--	Robust optimization	cross entropy	(Perelman et al. 2013a, b)	SO Design
q- Demand; RC - roughness coefficient; H- Pressure head; LHS-Latin Hypercube Sampling; MCS - Monte Carlo Simulations; GA- Genetic Algorithm; FORM -First order reliability method; SO -single objective; MO- multiobjective design; SFLA -Shuffled frog leap algorithm					

The current study proposes a new robust multi-objective design optimization formulation that simultaneously minimizes construction cost and maximizes network resilience considering consumer demands as uncertain. In this study, the uncertain problem is reformulated using a robust counterpart for both objectives. A ‘self-adaptive multi-objective cuckoo search algorithm’ (Pankaj et al. 2020) and ‘fmincon’ optimization algorithms have been used to solve the problem. The proposed methodology is demonstrated on the Hanoi water distribution system.

Robust optimization is a recent non-probabilistic optimization under uncertainty (OUU) technique. The main advantage of this technique is that the uncertain problem is reformulated into a tractable form, and the solution obtained is feasible for all possible realizations of the uncertain parameter within the specified uncertainty set. This method is gaining popularity in engineering applications due to its fast computational ability and ability to handle uncertain parameters that do not follow any standard probability distribution.

The main problem with optimal water distribution system design under uncertainty is computational time, and consumer demands do not follow any probability distribution. This robust optimization technique can solve both problems. The robust approach to solving the multi-objective design optimization of WDS involves converting the uncertain problem into a deterministic form using a robust counterpart approach and a self-adaptive multi-objective cuckoo search algorithm. This methodology is applied to a Hanoi WDS.**Robust optimization**

The aim of robust optimization is to obtain a unique solution for the optimization problem whose feasibility is independent of the uncertainty in the data. The general robust optimization formulation is as follows,

$$\begin{gathered} min\, f(x) \\ s.t\; g_i(x,\alpha_i) \geq 0\; \forall \alpha_i \in U_i, i \in [1,m] \end{gathered} \tag{1}$$

Here x is decision vectors, $f(.)$ is the objective function and $g_i(.)$ is the ith constraint function. α_i is the uncertain parameter and U_i is the uncertainty set corresponding to the ith constraint function.

In general, the problem in equation (1) is intractable, leading to infinite constraints (all possible realizations within the uncertainty sets), we can write this in-tractable constaint as shown in equation (2), if we can obtain the minimum value of g(x,a) for all values of a in the uncertainty set U and check its value to the inequality, it implies that all the values of g(x,a) within the set of U satisfy the inequality. By using this we can covert the infine constraints into one single tractable cosntriant.

$$s.t \min_{\alpha_i \in U_i, i \in [1,m]} g_i(x,\alpha_i) \geq 0 \tag{2}$$

This reformulation is called the robust counterpart of the problem. A detailed description of robust optimization and robust counterpart formulation and its applications is provided by Bertisimas et al (2011). The reformulation approach varies with the type of uncertainty set. For the current study, we use ellipsoidal uncertainty sets as they are less conservative compared to box uncertainty sets and can also incorporate the correlation between the uncertain variables (Baron et al. 2011).

2.2 Ellipsoidal uncertainty set [(Ben-Tal and Nemirovski 1998)]:

Let us assume that for every ith constraint, α_i, can vary within the interval $[\hat{\alpha}_i - \delta, \hat{\alpha}_{ij} + \delta]$, where $\hat{\alpha}$ is the nominal value of α and δ is the maximum deviation from the nominal value.

For any uncertain coefficient α with the nominal value $\hat{\alpha}$ and covariance matrix Σ, the ellipsoidal uncertainty can be defined using Mahalanobis distance in the form:

$$U(\Omega) = \{\alpha | (\alpha - \hat{\alpha})^T \Sigma^{-1} (\alpha - \hat{\alpha}) \leq \Omega^2\} \tag{3}$$

'Ω' is a value controlling the size of the ellipsoidal uncertainty set, which is also referred to as the protection level.

Let us consider a simple function as the constraint (2), $g(x, \alpha) = \alpha^T x$ and the uncertainty set be ellipsoidal uncertainty set described in equation (3), then the constraint in the equation (2) can be written as

$$\min_{\alpha_i \in U(\Omega)_i, i \in [1,m]} \alpha_i^T x \geq 0 \tag{4}$$

Then the solution to this minimization problem can be easily attained by Karush Kuhn tucker conditions, the minimum value of $\alpha^T x$ can be attained when $\alpha = \hat{\alpha} - \frac{\Omega}{\sqrt{(x^T \Sigma x)}} \Sigma x$

By substituting this value, we can re-write the constraint as

$$\min_{\alpha_i \in U(\Omega)_i, i \in [1,m]} \alpha_i^T x \geq 0 \Rightarrow \hat{\alpha}_i x - \Omega \sqrt{x^T \Sigma x} \geq 0 \tag{5}$$

2.3 Cost vs resilience optimal design of WDS problem formulation

The optimal design of WDS is an np-hard problem containing complex non-linear equations in energy constraint and discrete search space. Initial efforts were made considering this as single objective as explained in the introduction section, and then the research moved towards multi-objective optimation, considering maximization of reliability or resilance as the second objective. Although there is no exact maximization way to realize resilience, many authors suggest a few surrogate measures to indicate the system's resilience. The most popular surrogate measure is the resilience index. In the current study we incorporate this resilience index as the second objective. The mathematical representation of the problem is expressed as follows:

$$Minimize \sum_{i=1}^{np} U_C(D(i)) * L(i) \tag{6}$$

$$Maximize\ RI = \frac{\sum_{i=1}^{nn} q_i (h_i - h_i^{min})}{(\sum_{s=1}^{nr} Q_s H_s + \sum_{b=1}^{npu} \frac{P_b}{\gamma}) - \sum_{i=1}^{nn} q_i h_i^{min}} \tag{7}$$

$$subject\ to\text{:}\ \ A_{21} Q - q = 0 \tag{8}$$

$$A_{11}Q + A_{12}\text{h}=0 \tag{9}$$

$$h \geq h^{min} \tag{10}$$

$$D_i \in \{D_C\} \tag{11}$$

Here U_c- Unit cost per length of pipe corresponding to the diameter, D – set of design diameters, $\{D_C\}$ is the set of commercially available diameters, L – length of pipe, np – number of pipes, q_i- demand of node 'I', h- pressure head at node, $h^{\min}$ – minimum pressure required, nr – number of reservoirs, Q_s- flow from reservoir 's', H_s – pressure head of reservoir 's', npu – number of pumping units, P_b – energy of pump 'b', γ – efficiency of the pump, nn – number of nodes in the network, A_{21}=A_{12}^T is the connectivity matrix of the network based on topology, A_{11}- nonlinear elements representing the frictional resistance of the pipe, Q- Flow values in each pipe.

The equation (9) represents the energy constraint where the $A_{11}Q$ is the non-linear head loss term that can be expressed as :

$$\Delta h = \Delta h(Q) = R_c Q^{a1}\} \tag{12}$$

Where $R_c(resistance\ coefficent) = {a_3 L}/{f_c^{a1} \times D^{a2}}$, f_c- pipe friction coefficient, $a1$ = 1.852, a2 = 4.87 and $a3$ is the Hazen-Willams coefficient

2.4 Robust Counterpart formulation considering demand(q) as uncertain

Among all the uncertain parameters affecting WDS design, demand is the most important parameter (Babayan et al. 2005). For the current study, demand is assumed to be uncertain. To explicitly formulate the constraint with demand uncertainty, we use the linearization method proposed in (Perelman et al. 2013) to replace the head loss function. Among the two linearization methods they proposed, for the current study we incorporate the linearisation within a range [Q1, Q2] that under estimates the head within the range. For the case study we used Q1 = mean demand -2*standard deviation and Q2 = mean demand +2*standard deviation.

$$\Delta h = \left(\frac{\Delta h(Q_2) - \Delta h(Q_1)}{Q_2 - Q_1}\right) Q + \frac{\Delta h(Q_1)Q_2 - \Delta h(Q_2)Q_1}{Q_2 - Q_1};\ L_1 Q + L_0 \tag{13}$$

$$\begin{gathered} \begin{bmatrix} A_{12} & L_1 \\ 0 & A_{21} \end{bmatrix} \begin{bmatrix} h \\ Q \end{bmatrix} = G \begin{bmatrix} h \\ Q \end{bmatrix} = \begin{bmatrix} -L_0 + h_o \\ q \end{bmatrix} \\ \Rightarrow \begin{bmatrix} h \\ Q \end{bmatrix} = K \begin{bmatrix} L_0^* \\ q \end{bmatrix} = \begin{bmatrix} K_{11} & K_{21} \\ K_{12} & K_{22} \end{bmatrix} \begin{bmatrix} L_o^* \\ q \end{bmatrix} \end{gathered} \tag{14}$$

Where $G^{-1} = K = \begin{bmatrix} K_{11} & K_{21} \\ K_{12} & K_{22} \end{bmatrix}$ is the inverse of the matrix $\begin{bmatrix} A_{12} & L_1 \\ 0 & A_{21} \end{bmatrix}$, K_{11} is of the size $[nn \times np]$, K_{12} is of size $[nn \times nn]$ and $L_o^* = -L_o + h_o$, where h_o is a given vector of fixed known heads.

From the equation (15), the nodal heads' h' can be computed as

$$h = K_{11} L_0^* + K_{12} q \tag{15}$$

Using this formulation, we can rewrite the optimization problem as

$$Minimize \sum_{i=1}^{np} U_C(D(i)) * L(i) \tag{16}$$

$$Maximize\ RI = \frac{\sum_{i=1}^{nn} \tilde{q}_i((K_{11}L_0^* + K_{12}\tilde{q})_i - h_i^{min})}{(\sum_{s=1}^{nr} Q_s H_s + \sum_{b=1}^{npu} \frac{P_b}{\gamma}) - \sum_{i=1}^{nn} \tilde{q}_i h_i^{min}}; \tag{47}$$

$$K_{11}L_0^* + K_{12}\tilde{q} \geq h^{min} \tag{18}$$

$$\tilde{q} \in U \tag{19}$$

Now consider the equation (18,19), the equation contains demand (q) as an uncertian parameter, the robust optimization formulation for this is assuming the demand varies in the uncertainty set $U(\Gamma) = \{q_{ij} | (\tilde{q}_{ij} - \hat{q}_{ij})^T \Sigma^{-1} (\tilde{q}_{ij} - \hat{q}_{ij}) \leq \Gamma^2\}$ and $P: \Sigma = P.P^T$. Then as explained in the ellipsoidal robust optimization formulation (equation 3-6), we can write the formulation as follows.

$$\min_{\tilde{q} \in U} K_{11}L_0^* + K_{12}\tilde{q} \geq h^{min} \Rightarrow K_{11,i}L_0^* + \tilde{q}^T K_{12,i}^T - \Gamma \| P^T K_{12,i}^T \| \geq h^{\min} \tag{20}$$

Robust optimization formulation for resilience index, for simplifying the problem, lets assume that the network consists of only one source and no pumps. Then the simplified resilience index equation can be written as:

$$Maximize\ RI = \frac{\tilde{q}^T K_{12}\tilde{q} + \tilde{q}^T K_{11} L_o^* - \tilde{q}^T h^{min}}{(\Sigma q)H_s - \tilde{q}^T h^{min}}; K_{11}L_0^* + K_{12}\tilde{q} \geq h^{min}; \tilde{q} \in U \tag{21}$$

The robust optimization formulation for the problem in equation 21 is,

$$\max \tau$$

$$subject\ to: \min_{\tilde{q} \in U(\Gamma)} \frac{\tilde{q}^T K_{12}\tilde{q} + \tilde{q}^T K_{11} L_o^* - \tilde{q}^T h^{min}}{(\Sigma q)H_s - \tilde{q}^T h^{min}} \geq \tau;\ K_{11}L_0^* + K_{12}\tilde{q} \geq h^{min} \tag{5}$$

The resilience index formulation is still nonlinear with a form similar to quadratic over linear, but all the elements in the matrix K_{12} are negative (Perelman et al, 2013). The denominator is always positive as energy at the source ($(\Sigma q)H_s$) is always greater than energy reached at the nodes ($\tilde{q}^T h^{min}$), this problem will never be of the form quadratic over linear with the positive definite quadratic matrix.

In order to solve the optimization problem in equation 22, an inbuilt nonlinear optimization algorithm in MATLAB named "fmincon" is used.

The overall robust multiobjective formulation used in this study is as follows

Objective function (1)

$$\mathit{Minimize} \sum_{i=1}^{np} U_C(D(i)) * L(i) \tag{63}$$

$$\text{Subject to: } K_{11,i}L_0^* + \tilde{q}^T K_{12,i}^T - \Gamma\|P^T K_{12,i}^T\| \geq h^{\min}$$

Objective function (2)

$$\max \tau$$

$$\mathit{subject\ to}: \min_{\tilde{q}\in U(\Gamma)} \frac{\tilde{q}^T K_{12}\tilde{q} + \tilde{q}^T K_{11} L_o^* - \tilde{q}^T h^{min}}{(\Sigma q)H_s - \tilde{q}^T h^{min}} \geq \tau; \tag{24}$$

$$K_{11}L_0^* + K_{12}\tilde{q} \geq h^{min}$$

The proposed method is applied to a standard benchmark problems Hanoi WDS proposed by (Fujiwara and Khang 1990)

3.1 Multiobjective optimization method:

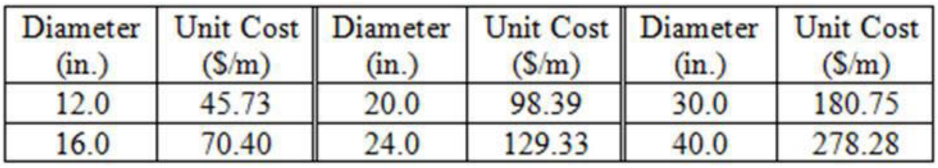
Self-adaptive multiobjective cuckoo search algorithm (SAMOCSA) combined with fmincon nonlinear optimization model is used to solve the robust multiobjective WDS design optimization problem. SAMOCSA is an improved version of the multiobjective cuckoo search that adapts the algorithm's exploration and exploitation governing parameters at every iteration. This algorithm has been tested on the two-loop network, Hanoi network and Pamapur network (Indian network) for deterministic multiobjective design problem of WDS. The complete details of the algorithm and its efficiency can be obtained from (Pankaj et al. 2020).

3.2 Hanoi WDS case study

Hanoi WDS is a medium gravity-based WDS proposed by Fujiwara and Khang 1990. The network consists of 32 demand nodes and 34 pipes connected to a single source with a head of 100m. The minimum pressure head required at every node is 30m. The network needs to be designed with 6 different sized pipes. The unit cost corresponding to the available diameter are shown in Table -1.

The full data for this example can be found https://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/design-resiliance-pareto-fronts/medium-problems/

Table 2 Diameter options and associated unit costs for Hanoi WDS



Diameter (in.)	Unit Cost ($/m)	Diameter (in.)	Unit Cost ($/m)	Diameter (in.)	Unit Cost ($/m)
12.0	45.73	20.0	98.39	30.0	180.75
16.0	70.40	24.0	129.33	40.0	278.28

To model uncertainty in demands, the WDS nodes were partitioned into three demand regions: region 1—nodes 1:15, region 2—16:24, and region 3—25:32 (Figure 1). Demands in region 2 were assumed to be certain and in regions 1 and 3 as uncertain with a standard deviation of 12% from the mean demand of each region [i.e., 80 and 50 (m3/h)], respectively. Two different protection levels are studied $\boldsymbol{\Omega} = [\mathbf{1}, \mathbf{2}]$. Furthermore, the correlation between the nodes within the region and the correlation between the regions are also altered. The intraregional correlation values are set to be $\boldsymbol{\rho} = \mathbf{0.8}$, and the interregional correlation varies between positive, no-correlation and negative correlation $\boldsymbol{\rho} = [\mathbf{0.6}, \mathbf{0}, -\mathbf{0.6}]$. SAMOCSA algorithm is used to solve the outer design problem, and the "fmincon" algorithm is used to solve the nonlinear inner optimization problem for minimization of the resilience index within the demand search space.

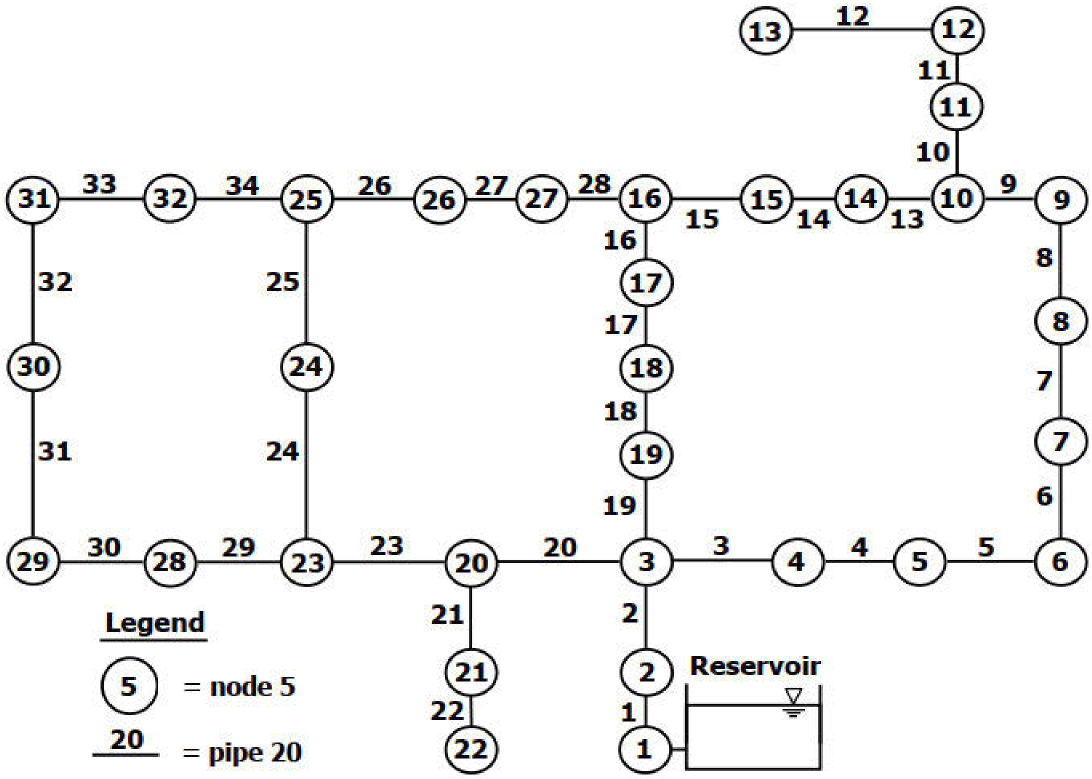

13
12
12
11
11
10
31
33
32
34
25
26
26
27
27
28
16
15
15
14
14
13
10
9
9
16
32
25
17
8
17
8
30
24
18
7
18
7
31
24
19
19
6
29
30
28
29
23
23
20
20
3
3
4
4
5
5
6
Legend
21
2
5
= node 5
21
2
Reservoir
22
1
20
= pipe 20
22
1

Figure 1 Graph of the topology of Hanoi WDS (Fujiwara & Khang 1990)

The graphs are shown in fig(2) compare the results obtained for different uncertainty sets (Ellipsoidal-1, Ellipsoidal -2) for positive, negative and no correlation. We can infer that uncertainty changes the cost to resilience trade-off from the graphs. As the ellipsoidal set size increases ($\Omega = 1\ to\ 2$), the cost vs resilience trade-off gets worst. The worst case is obtained when the demands are positively correlated among the three cases. We need to provide a higher cost design, even for a small resilient design.

BY NC SA

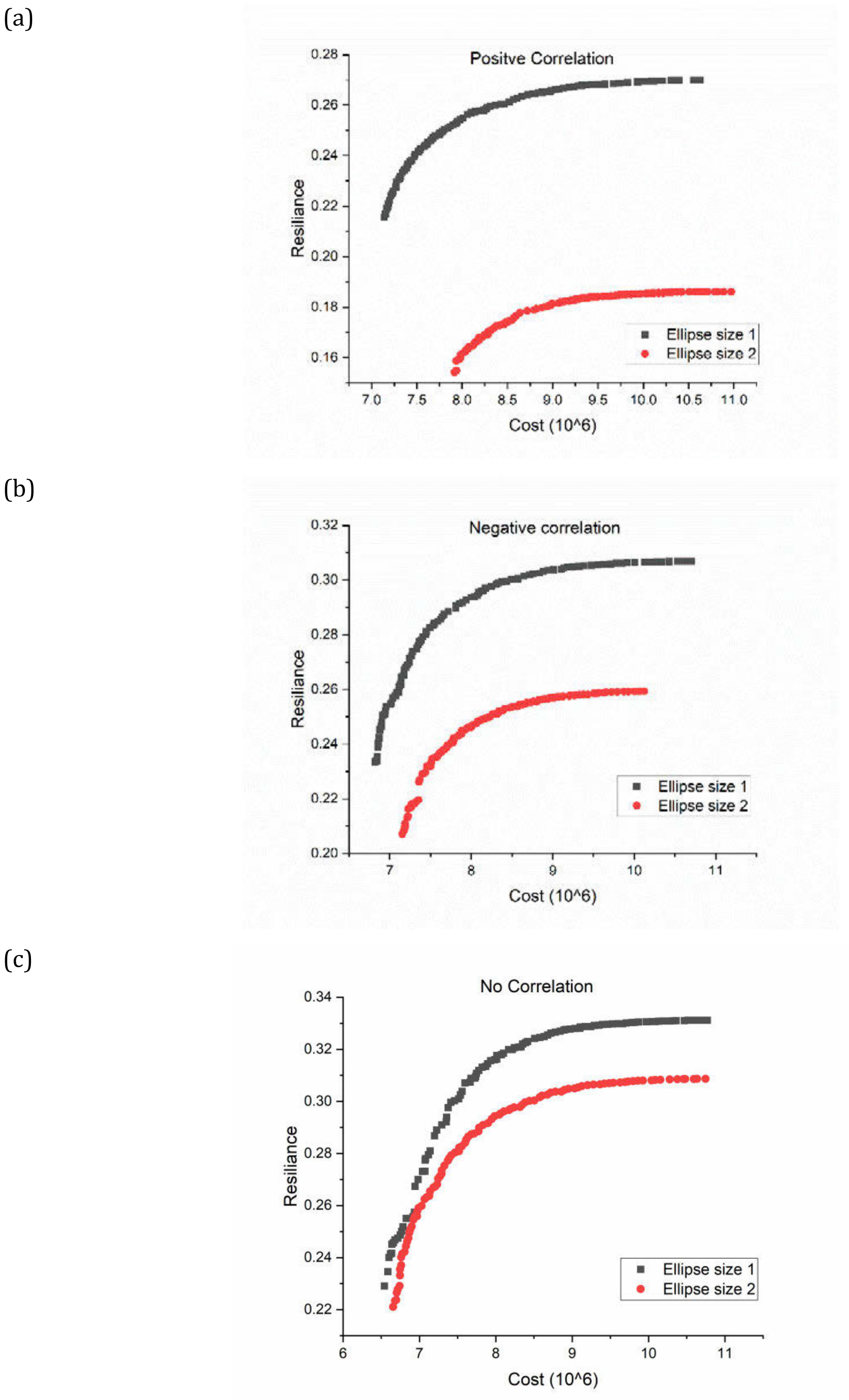


Figure 2 Cost vs Resilience Pareto Fronts for two different sizes of uncertainty sets [Ω = 1,2] when the demands are considered (a) Positively Correlated (b) Negatively Correlated (c) no correlation

This work proposes using a robust counterpart approach to handle the demand uncertainty in WDS. The results show significant promise of this new approach in terms of ease of computation and model formulation. Once the problem is converted to tractable deterministic robust counterparts, the methodology for solving the problem is similar to solving a conventional multi-objective design problem. Even though the formulation is similar to max-min approach, the ellipsoidal uncertainty set makes it less conservative. As the size of the uncertainty set to increase, the trade-off also increases; even for little resilience, we need to incorporate high-cost designs. The worst-case scenario is attained when the demands are assumed to be positively correlated. The results of the ellipsoidal set with $\Omega = 1$ is a less conservative and robust solution. The further scope of this work is trying to incorporate other objectives like (minimizing the leakages or carbon emissions)

5 ACKNOWLEDGEMENT

This research was supported by the Israel Science Foundation (Grant No. 555/18).

6 REFERENCES

[1] Alperovits, E., and Shamir, A. U. (1977). Design of Optimal Water Distribution Systems.

[2] Avi, and Shamir, U. (1996). “Design of optimal reliable multiquality water-supply systems.” Journal of Water Resources Planning and …, 122(5)(October), 322–333.

[3] Babayan, A., Kapelan, Z., Savic, D., and Walters, G. (2005). “Least-Cost Design of Water Distribution Networks.” (October), 375–382.

[4] Baron, O., Milner, J., and Naseraldin, H. (2011). “Facility location: A robust optimization approach.” Production and Operations Management, 20(5), 772–785.

[5] Farmani, R., Walters, G., and Savic, D. (2006). “Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. reliability vs. water quality.” Journal of Hydroinformatics, IWA Publishing, 8(3), 165–179.

[6] Fu, G., Kapelan, Z., Kasprzyk, J. R., and Reed, P. (2013). “Optimal Design of Water Distribution Systems Using Many-Objective Visual Analytics.” Journal of Water Resources Planning and Management, 139(6), 624–633.

[7] Jung, D., Chung, G., and Kim, J. H. (2012). “Optimal design of water distribution systems considering uncertainties in demands and roughness coefficients.” Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference, WDSA 2010, (1989), 1390–1399.

[8] Jung, D., Kang, D., Kim, J. H., and Lansey, K. (2014). “Robustness-Based Design of Water Distribution Systems.” Journal of Water Resources Planning and Management, 140(11), 04014033.

[9] Kessler, A., and Shamir, U. (1989). “Analysis of the linear programming gradient method for optimal design of water supply networks.” Water Resources Research, Wiley Online Library, 25(7), 1469–1480.

[10] Lansey, K. E., Duan, N., Mays, L. W., and Tung, Y. (1989). “Water Distribution System Design Under Uncertainties.” Journal of Water Resources Planning and Management, 115(5), 630–645.

[11] Ostfeld, A., Asce, M., and Tubaltzev, A. (n.d.). “Ant Colony Optimization for Least-Cost Design and Operation of Pumping Water Distribution Systems.”

[12] Pankaj, B. S., Jaykrishnan, G., and Ostfeld, A. (2022). “Optimizing Water Quality Treatment Levels for Water Distribution Systems under Mixing Uncertainty at Junctions.” Journal of Water Resources Planning and Management, American Society of Civil Engineers, 148(5), 04022013.

[13] Pankaj, B. S., Naidu, M. N., Vasan, A., and Varma, M. R. (2020). “Self-Adaptive Cuckoo Search Algorithm for Optimal Design of Water Distribution Systems.” Water Resources Management, Water Resources Management, 34(10), 3129–3146.

[14] Perelman, L., Housh, M., and Ostfeld, A. (2013). “Robust optimization for water distribution systems least cost design.” Water Resources Research, 49(10), 6795–6809.

[15] Savic, D. A., and Walters, G. A. (1997). “Genetic Algorithms for Least-Cost Design of Water Distribution Networks.” Journal of Water Resources Planning and Management, American Society of Civil Engineers, 123(2), 67–77.

[16] Seifollahi-Aghmiuni, S., Bozorg Haddad, O., and Mariño, M. A. (2013). “Water Distribution Network Risk Analysis Under Simultaneous Consumption and Roughness Uncertainties.” Water Resources Management, 27(7), 2595–2610.

[17] Sumer, D., and Lansey, K. (2005). “Effect of uncertainty on water distribution system model decisions.” Proceedings of the 8th International Conference on Computing and Control for the Water Industry, CCWI 2005: Water Management for the 21st Century, 1(February), 38–47.

[18] Vasan, A., and Simonovic, S. P. (n.d.). “Optimization of Water Distribution Network Design Using Differential Evolution.”

[19] Wu, W., Maier, H. R., and Simpson, A. R. (2013). “Multiobjective optimization of water distribution systems accounting for economic cost, hydraulic reliability, and greenhouse gas emissions.” Water Resources Research, 49(3), 1211–1225.

[20] Wu, Z. Y., Asce, M., Walski, T., and Asce, F. (2005). “Self-Adaptive Penalty Approach Compared with Other Constraint-Handling Techniques for Pipeline Optimization.” Journal of Water Resources Planning and Management, American Society of Civil Engineers, 131(3), 181–192.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14730

EVALUATION OF PHYSICAL AND WATER QUALITY INDICATORS IN ADUCTORS DURING HYDRAULIC TRANSIENT EVENTS

Mariele de Souza Parra Agostinho[1], Danieli Mara Ferreira[2] and Cristovão Vicente Scapulatempo Fernandes[3]

[1]Federal University of Parana and Sanitation Company of Parana. Curitiba, Parana, Brazil
[2]Vale Institute of Technology, Belem, Para, Brazil
[3]Federal University of Parana. Curitiba, Parana, Brazil

[1] marielespa@gmail.com, [2] danielimaraferreira@gmail.com, [3] cvs.fernandes@gmail.com

Abstract

An indicator that correlates physical quantities with water quality has great relevance for the definition of projects and understanding of real hydraulic transient phenomena. The work proposing a method to estimate water quality in real system focusing on turbidity measurements during a transient, as surrogate to understand potential dynamics in real systems. In order to get an empirical formula to turbidity during hydraulic transients, was observed 5 years' data from a real system. This system is one of the mains important pipelines of the Water Supply System of Curitiba and Metropolitan Region, in South Brazil.

Keywords
Hydraulic Transient, Water quality, Real system, Turbidity.

1 INTRODUCTION

Hydraulic Transient is a phenomenon characterized by the occurrence of pressure waves propagating along the pipe, caused by interference or maneuver in the flow of the fluid, by varying operating pressures, which can generate or not a water hammer, depending upon how unstable fluid flows propagates through high-pressure forces and rapid acceleration of the fluid [1].

Typical events that require transient considerations include: pump start-up or shutdown; valve opening or closing; changes in boundary pressures; rapid changes in demand conditions; changes in transmission conditions; pipe filling or draining-air release from pipes; and, check valve or regulator valve action [2].

The consequence of strong influence of inertia and compressibility effects can: a) rupture a pipe directly through excessive pressure or they can exploit an existing weakness, like corrosion, earth pressures, construction faults, to damage the pipe indirectly [1], [3], [4]; b) cause water quality degradation [2], [5], [6], [7], [8], [9], [10].

As water utilities are under increasing pressure to improve the quality of service (less bursts and discoloration events) while reducing leakage, it is becoming essential to better understand and control the main factors affecting failures. Despite the importance, unfortunately, the occurrence of unsteady and quasi-unsteady state hydraulic conditions in operational networks remains poorly understood due to the limitations of existing telemetry systems and the capabilities of the data processing tools [11], especially in developing countries.

Furthermore, there is a lack of knowledge of the impact which the variety of transient events might have on specific assets and the complete system both in the short and long term [11].

1.1 Objectives

The main goal is to assess physical and water quality indicators in pipelines during hydraulic transient events. The specific objectives are: a) Proposing a method to estimate water quality in real system focusing on turbidity measurements during a transient, as a surrogate to understand potential dynamics in real systems; b) Modeling and simulation of hydraulic transient in water pipeline system, all in the Metropolitan Region of Curitiba, in Parana, Brazil; c) Measurement and evaluation of water quality parameters during transitory situations.

2 THEORICAL BACKGROUND

2.1 Hydraulic transient causes and consequences

In general, any disturbance in the water generated during a change in mean flow conditions will initiate a sequence of transient pressures in the WDS. Disturbances will normally originate from changes or actions that affect hydraulic devices or boundary conditions. Typical events that require transient considerations include: pump start-up or shutdown; valve opening or closing (variation in cross-sectional flow area); changes in boundary pressures (e.g., losing overhead storage tank, adjustments in the water level at reservoirs, pressure changes in tanks, etc.); rapid changes in demand conditions (e.g., hydrant flushing); changes in transmission conditions (e.g., main break or line freezing); pipe filling or draining-air release from pipes; and, check valve or regulator valve action [2].

Clearly, there is a direct relation between the consequences of transient effects and safety that impose equipment damage or operational difficulties. Some of the common consequences are [5]: a) maximum pressures in hydraulic systems. This is the most common consequence of hydraulic transients; b) occurrence of local vacuum conditions at specific locations that may result in cavitation either within specific devices such as pumps or within a pipe; c) hydraulic vibration of a pipe, its supports, or in specific devices; d) occurrence of contaminant intrusion at joints and cross-connections.

If sub-atmospheric pressure conditions evolve, the risk of pipeline collapse increases for some pipeline materials, diameters, and wall thicknesses [12]. Even if a pipeline does not collapse, column separation (sudden vaporous cavitation) caused by differential flow into and out of a section could occur if the pressure in the pipeline is reduced to the vapor pressure of the liquid. Besides that, excessive negative pressures groundwater can draw into the system. Studies assessed the intrusion risk by evaluating the possible volume of intruding groundwater given transient duration and severity (minimum water hammer pressure) [10].

2.2 Hydraulic transient affecting water quality

2.2.1 Potential for pathogen intrusion during pressure transients

Studies introduced three requirements to be met for stating risk conditions to human health due to contaminant intrusion: adverse pressure conditions (the driving force), a pathway (leakage points, badly fitted joints, air valves, cross-connections), and the contaminant source [13].

Drinking water distribution systems are vulnerable to external contaminant entry if there is a loss of physical integrity. The main driver for an intrusion event to occur is the failure to maintain an adequate pressure in the distribution system. Low and negative pressure events have the potential to result in intrusion of pollutants: negative pressures create a suction effect inside the pipe and the contaminant intrusion through pipe leaks [14].

Failures in the integrity of distribution pipes resulting in a system leak are therefore classed as high-risk routes of entry. Examples of typical leak types in water distribution system pipes are pin holes, cracks, corrosion clusters, and joint/ connection failures [15].

Transient negative pressure can draw leaked water back into the pipe at any point where water is leaking out of the system. Once these leaks or breaks occur, any microbial contamination in the vicinity of the break or leak can potentially enter the distribution system given the pressure changes that occur during breaks or leaks. A major fecal source is usually near sewer lines, which are notorious for leaking. Main breaks can also introduce high concentrations of injured coliform bacteria (undetectable by standard coliform techniques) into the distribution system [8].

Studies collected and tested soil and water samples in the immediate vicinity of water mains at eight locations in six US states and found that often these soils contain potentially harmful bacteria and pathogens such as coliforms (detected in 58% and 70%, of water and soil samples, respectively) and fecal coliform bacteria (detected in 43% of the water and 50% of the soil sample) [7].

Experimental procedure was conducted that analyzed the risk of contamination in two cases: intermittent water and transient. It wasshown that both can allow a large number of contaminants inside the pipes. In the intermittent system, the contaminants get into the pipe by means of infiltration when the pipe is partially empty and the service gets discontinued. The pressurization of pipes ejected part of them from the pipe but a large amount remained, and it was supplied to the users. To transient involving negative pressure the physical process was similar, but the temporal scale of the process was much smaller and the number of contaminants flowing into the pipe was smaller. In this case it was strictly related with the extension and the magnitude of the negative pressure transient but the contamination was still present and it could produce risks for the users [16].

Weston et al. improved the understanding of intrusion volumes and supported previous research, providing physical evidence that contaminant intrusion can occur due to dynamic pressure transient events within water distribution systems. The work shows how intrusion volumes change as a function of the hydraulic conditions within in a full-scale laboratory pipeline that replicates operational environments. Expected underlying trends are clearly evidenced: as the change in velocity increases (increasing initial pressure surge), the volumes of intrusion increase, and as the change in pressure increases (increasing the initial pressure offset), the volumes of intrusion decrease [17].

Studies investigated the volume of contamination intruded into the pipeline using mathematical and numerical modeling of the phenomenon. The study estimated the intrusion amount for 72 different scenarios including: two lengths of pipeline (i.e. short and long), three different leak locations, three different fluid velocities in the pipe, two leak diameters and two pipeline materials (elastic and viscoelastic). The results showed that the: a) amount of intrusion in viscoelastic pipes was clearly less than that in elastic pipes, especially in long pipelines; b) It may seem that higher Joukowsky pressure or leak size produces more intrusion volume, but this is not always true and the numerical results demonstrate several counter-examples. Large leak sizes are prone to significant leak induced wave reflections, thus making the leak position zone of high intrusions quite unpredictable (case dependent); c) The critical zone of high intrusion risk is identified close to the downstream valve for small leak sizes, nevertheless, it is difficult to estimate this zone in the case of large leaks due to significant interactions between nodal components (valve, leak, reservoir) [18].

The numbers of pathogens introduced from a sewage contaminated groundwater during a transient intrusion can theoretically result in unacceptably large numbers of microorganisms being transported to consumers, even when adequate chlorine residuals are present [19]. Despite studies have provided a comprehensive description of the mathematical solution of hydraulic transients with intrusion and advection [6], very few results from real system performance under water quality and transient conditions have been reported.

2.2.2 Relation between discoloration and hydraulic transient

Discoloration is mainly attributed to the re-suspension of accumulated particles and not true change of color [20]. Discoloration pathways can be classified as follows: a) Gravitational sedimentation: re-suspension of particles settled in the bottom of pipes; b) Non-gravitational accumulation: detachment of particles accumulated over the entire circumference of pipes; c) Corrosion-related: release of iron particles either directly from corrosion, or accumulation of iron produced during corrosion in the form of scales on the pipe wall which is then released by dissolution or shear stress; and, d) Biofilm-related: entrapment of particles in biofilms and subsequent release due to biofilm sloughing.

Biofilms play a major part in the microbial characterization of drinking water quality in distribution. These films or slimes, as presented in Figure 1, become established in sediments, corrosion tubercles, static areas of slow water flow, dead-ends, standpipes, and storage tanks [20].

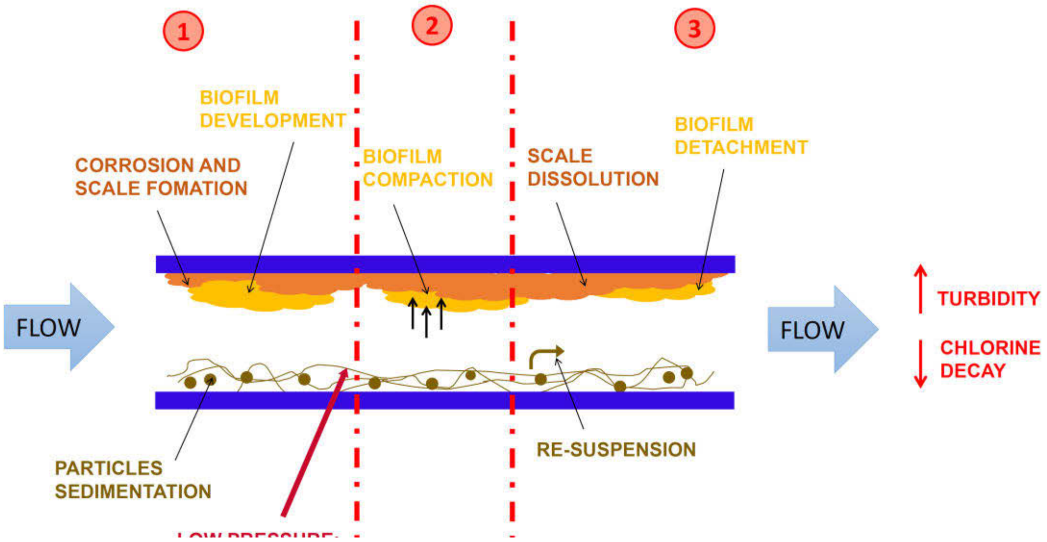


Figure 1 – Main Interactions Affecting Water Quality Decay Inside Distribution Systems

FONTE: The author, adapted from: Mays [20]

Weston et al. proposes that there are three key components of hydraulic transients that could theoretically lead to mobilization of biofilms growing or other material accumulated on the pipe wall, releasing such material into the bulk water and possibly causing water quality/health concerns, as presented in Figure 2. These mechanisms are consequence of induced shear stress due to velocity changes, induced stress due to a pressure differential across the wave front and structural changes from cavitation expansion and collapse. Once mobilized, biofilms and previously entrapped material have the potential to cause discoloration and water quality failures [21].

CC BY NC SA

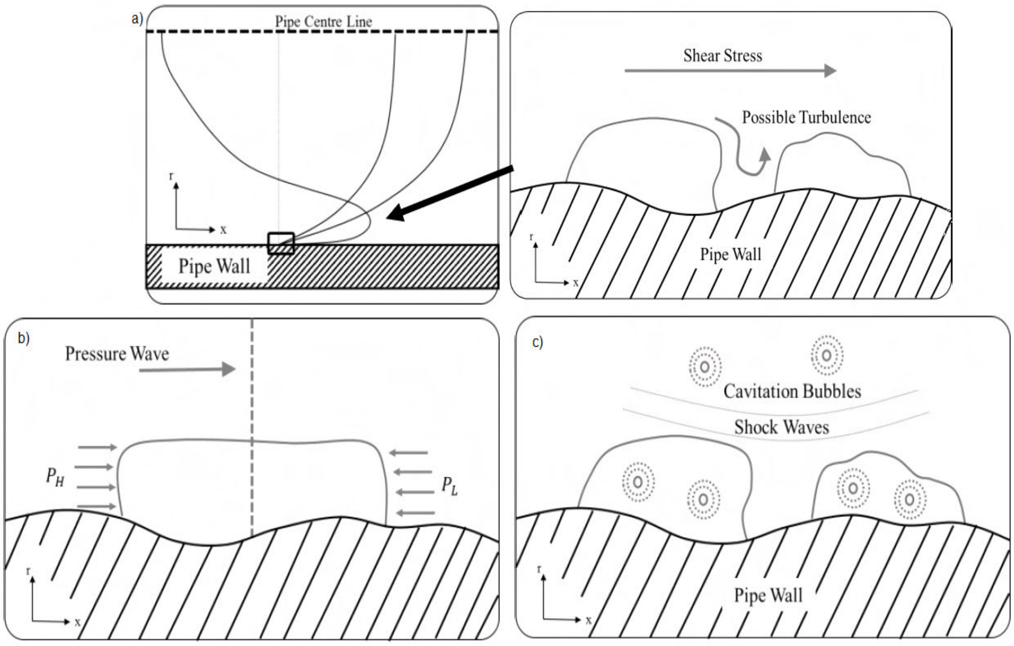

a)
Pipe Centre Line
r
x
Pipe Wall
Shear Stress
Possible Turbulence
Pipe Wall
b)
Pressure Wave
P_H
P_L
c)
Cavitation Bubbles
Shock Waves
Pipe Wall

Figure 2 - Mobilization of biofilms growing or other material accumulated on the pipe wall [22]

Experimental findings conclusively prove mobilization of pipe-wall adhered material due to hydraulic transients, which was only suggested from indirect observations in other works. Both valve closing and valve opening transients can mobilize this material where steady-state hydraulic conditions cannot. This study presents a schematic of possible transient and steady state mobilization force relationships, as illustrated in Figure 3 as very in-depth explanation for potential physical changes in pipes, under dynamics conditions. Linear relationships are drawn as the simplest association possible, which may or may not pass through the origin. The area above each relationship indicates that the magnetic forces are too high for mobilization to occur due to the hydraulic force. The area below each relationship indicates the magnetic forces are too low, thus mobilization would occur due to the hydraulic force. If transient relationships sat above the steady state relationship an area would exist where the transients can cause mobilization, but steady state cannot (green highlight). A direct implication of this research is that by reducing transient events a water utility provider could reduce the risk of discoloration contacts, the major cause of water quality related customer dissatisfaction [22].

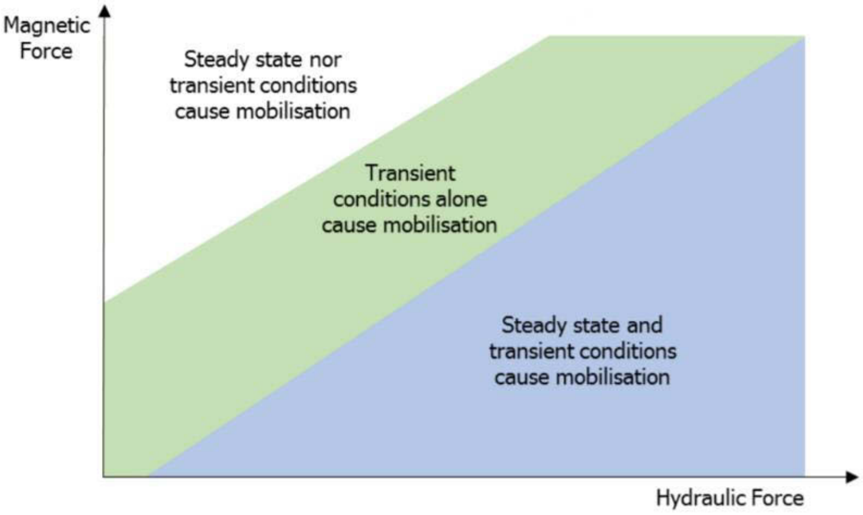

Magnetic Force
Steady state nor transient conditions cause mobilisation
Transient conditions alone cause mobilisation
Steady state and transient conditions cause mobilisation
Hydraulic Force

Figure 3 – Schematic of possible transient and steady state mobilization force relationships [22]

As highlighted by Speight "Understanding the processes and interactions occurring within complex, ageing drinking water distribution systems is vital to ensuring the supply of safe drinking water. While many water quality samples are taken for regulatory compliance, the resulting data are often simply archived rather than being interrogated for deeper understanding due to their sparse nature across time and space and the difficulties of integrating with other data sources" [23].

An experimental study demonstrates that pre-transient conditions of both flow and pressure combine in a complex manner and play key roles in the mobilization effects of both accelerating and decelerating flow transients. Further analysis shows that in transients with decelerating flow conditions, post-transient flow rate is also important for mobilization. All accelerating flow transients caused mobilization, those starting at zero flow and those starting with non-zero flow rates. Conversely only decelerating flow transients that ended at zero flow conditions caused mobilization, independent of the magnitude of the change in flow. The timing of mobilization was shown to be slower than the propagation of pressure wave front, suggesting the pressure wave front is not responsible for the mobilization. Conversely, links are made to the boundary shear stress of dynamic changes in velocity profiles observed in other research [24].

2.2.3 Water Quality Monitoring

Samples can be collected and analyzed in two ways: grab samples and continuous samples. Grab samples are collected manually in the field and analyzed in the field or in the laboratory. Continuous monitoring is conducted by means of sensors and remote recording stations. Grab samples are more labor intensive and provide data only at the time of collection. Continuous monitoring requires a capital expense for the monitoring equipment and for equipment maintenance time, but it provides a continuous time-series profile of changes in water quality. A comprehensive monitoring plan can make use of both grab and continuous methods [20].

A routine monitoring program should be established to meet regulatory requirements and to collet additional water- quality information that is useful in the operation of a water system. Table 1 lists parameters that are frequently measured in a routine monitoring.

Table 1 – Water Quality Parameters for Finished Water Storage Facilities [9]

PARAMETER	PURPOSE	SAMPLING PROCEDURE USED
Alkalinity	Indicates the potential buffering capacity	On-line ion-selective electrode or grab sample and laboratory analysis
Aluminum	Indicates the potential coagulant overfeeding	
Ammonia, total or free	Indicates potential for nitrification	
Chorine residual, total or free	Indicates protection from bacterial growth and provides early warning sign of water-quality deterioration: control re-chlorination when practiced	
Conductivity, specific	Can quickly indicate relative changes in total dissolved solids, e.g. alkalinity	
Iron	Indicates potential corrosion reactions	
Nitrate	Indicates possibility of nitrification	
pH	Indicates changes from the water source. Indication of corrosion of concrete or an unlined new facility.	
Coliform, total and/or fecal	Indicates presence of indicator bacteria	Grab sample and laboratory analysis
Disinfection by products	Represents potential for ongoing chemical reactions and DBP formation	
Heterotrophic bacteria	Indicates conformance to MCL; provides early warning sign of water-quality deterioration	
Taste and odor	Evidence of water-quality problem in progress	
Temperature	Differences within storage facility indicate possible stratification and stagnant zones. Early warning sign of potential	On-line sensor
Turbidity	Provides early warning sign of water-quality deterioration	On-line turbidimeter sensor and analyzers

CC BY NC SA

2.3 State of Art

The last six years of literature review was selected and the main papers about water quality in distribution system was summed up in Table 2. The first highlight is the incipient number of researches that associate water quality with the hydraulic transient state in real networks. Research involving the subject is essentially in networks installed in laboratory with small pipe diameter and flow, with high control of parameters through sensors. Another finding is that the vast majority of cases involving high frequency sensors are in laboratory networks.

Table 2 – Art State Overview linking water quality and water distribution system

	Theme	Real or Lab	Hydraulic state	Observed or simulated?	Pipeline or network?	High-resolution sensor	Math model	Data statistical generation	Index?	Reference
1	Discoloration	Lab.	Intermittent	Observed	Pipeline	No	No	No	Yes	[25]
2	Discoloration	Lab.	Transient	Observed	Pipeline	Yes	NO	No	Yes	[24]
3	Discoloration	Lab.	Steady state	Observed	Network	No	No	No	No	[26]
4	Discoloration	Lab.	Transient	Observed	Pipeline	Yes	No	No	No	[27]
5	Discoloration	Real	Steady state	Observed	Network	No	No	Machine learning	No	[23]
6	Discoloration	Lab.	Steady state	Observed	Pipeline	No	No	No	No	[28]
7	Chlorine decay	Real	Steady state	Both	Network	No	No	No	No	[29]
9	Discoloration	Review	-	-	-	No	no	no	No	[30]
10	Discoloration	Real	Steady state	Observed	Network	no	no	no	no	[31]
11	Discoloration	Real	Steady state	Observed	Network	No	No	No	Yes	[32]
12	Discoloration	Real	Steady state	Observed	Network	No	No	No	Yes	[33]
13	Discoloration	Real	Steady state	Observed	Pipeline	No	yes	no	Yes	[34]
14	Discoloration	Real	Steady state	Observed	Network	No	No	No	No	[35]
15	Discoloration	None	Transient	Review	None	No	No	No	No	[21]
16	Discoloration	Real	-	Observed	Pipelines	No	Yes	No	Yes	[36]
17	Discoloration	Lab.	Transient	Observed	Pipeline	Yes	No	No	No	[37]
18	Discoloration	Lab.	Transient	Simulated	Pipeline	no	yes	no	No	[38]
19	Intrusion	-	Transient	Simulated	pipeline	No	Yes	No	No	[18]
20	Intrusion	Lab.	Transient	Observed	Pipeline	Yes, video	No	No	No	[17]
21	Chlorine decay	Lab.	Transient	Observed	Pipeline	Yes, pressure - 100 Hz Chlorine - 1 Hz	Yes	No	No	[39]
22	Chlorine decay	Lab.	Steady state	Observed	Pipeline	No	No	No	No	[40]
23	Chorine decay	-	Steady state	Observed					No	[41]
24	Chorine decay	Lab.	Steady state	Observed		No	No	No		[42]
25	Turbidity	Lab.	Steady state	Observed	Network	No	No	No	No	[12]
26	Disinfection byproducts (DBP)	Lab.	Steady state	Observed	-	No	No	No	No	[43]
27	Disinfection byproducts (DBP)	Lab.	Steady state	Observed	Pipeline	No	No	No	no	[44]
28	Corrosion	Lab.	Steady state	Observed	Pipeline	No	No	No	No	[45]
29	Bacterial regrowth	Review	-	-	-	No	No	No	No	[46]
30	Bacterial regrowth	Review	-	-	-	No	No	No	No	[47]
31	Bacterial regrowth	Review	-	-	-	No	No	No	No	[48]
32	Bacterial regrowth	Lab.	Steady state	Observed	Pipeline	No	No	No	No	[49]
33	Bacterial regrowth	Real	Steady state	Observed	-	No	No	No	No	[50]

3 MATERIAL E METHODS

3.1 The system

The area under study belongs to the Water Supply System of Curitiba and Metropolitan Region, named SAIC. The current production and treatment system is around 9.9 m^3/s. After the treatment, the water is sent to around 50 existing reservation centers, being in all 377650 m3 reserved. There are currently 160 pressure zones serving 3.5 million people. In this study was selected one of the mains important pipelines of the SAIC system, that are responsible for treated water transportation by Iguacu Treatment Plant to Corte Branco Tank, as represented in Figure 4, and composed by: three pipelines being two of 800 mm ductile iron and one of 1,100 mm iron and six pumps, being five used and one back up. The average flow is 2,600 L/s.

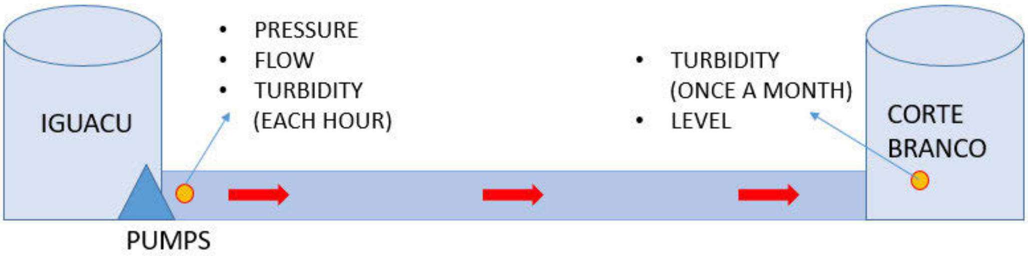


Figure 4 – Schematic Sketch of Pipeline

The Figure 5 presents the real data of Flow, Pressure after pumps and Level' Corte Branco, collect in 2020.

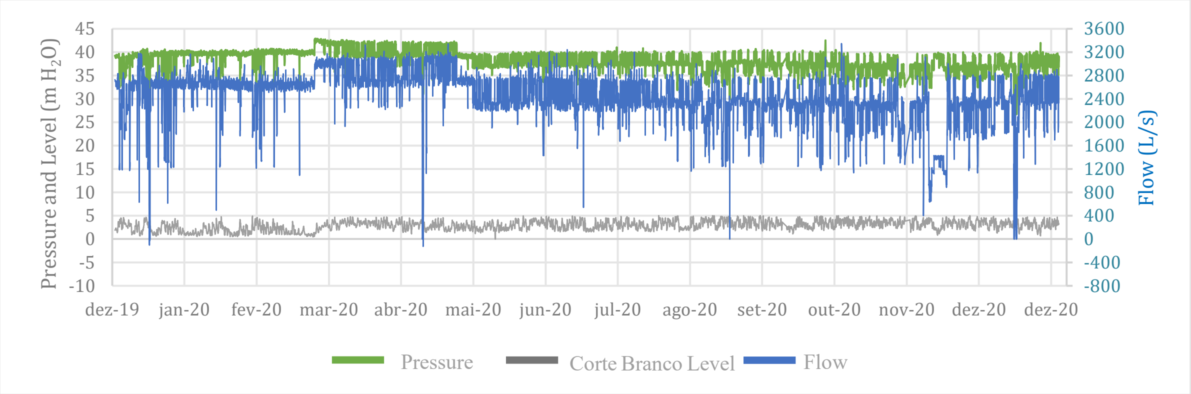


Figure 5 – Real Pipeline Data

3.2 Fundamental Database

In order to get an empirical formula to turbidity during hydraulic transients, was observed 5 years' data from the real system. The data used, as present in Figure 6, are the ones required by Brazilian regulation.

From that data was create simple correlations between the Corte Branco turbidity and: a) Flow; b) Pressure after pumps; and c) Level; and d) between the difference of turbidity (Corte Branco and Iguaçu) and Pressure pumps. These correlations are presents in Figure 7.

CC BY NC SA

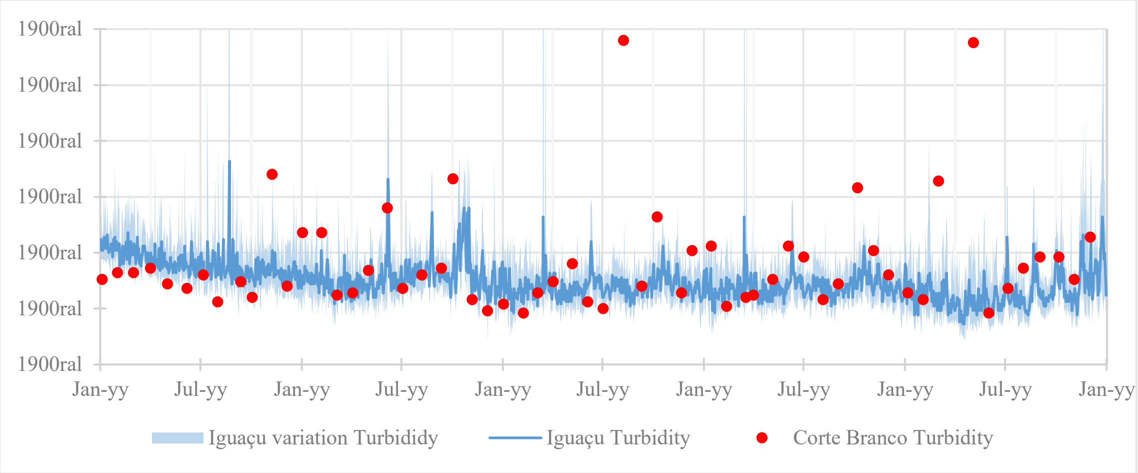


Figure 6 – Turbidity Data

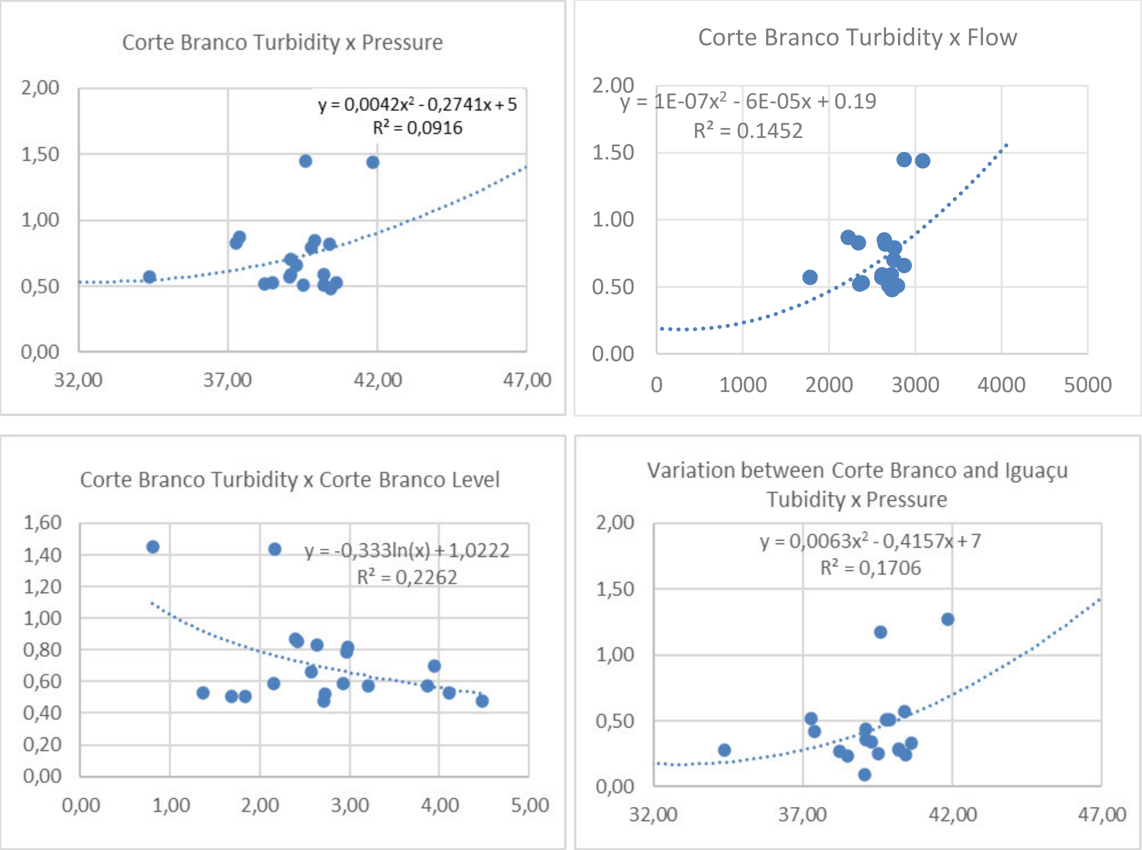


Figure 7 – Water Quality x Physical Correlations [51]

The boxplot chart was used to assess the empirical distribution of the data. The boxplot consists of the first and third quartiles and the median. The lower and upper stems extend, respectively, from the lower quartile to the lowest value not lower than the lower limit and from the upper quartile to the highest value not higher than the upper limit, as present in Figure 8. Applying the obtained equations in the 2020' data, as illustrated in Figure 9, the results propose that the maximum turbidity was around 1.2 uT.

CC BY NC SA

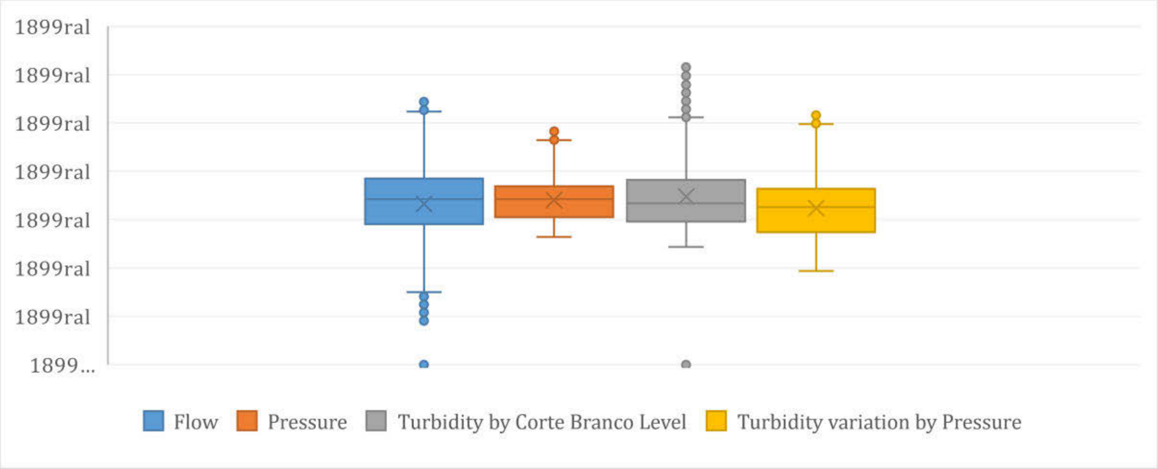


Figure 8 – Correlation's Comparison

Analyzing the data, the graph that presented the most confidential equation with an of 22% was based in Corte Branco's Level. This make sense because the sample is taken directly of the Tank, but cannot be extrapolated to transients, considering that tank does not vary in time during a transient. The second graph that presented more confidential the correlation is the variation of turbidity between the Corte Branco's Level and the Iguaçu, and the pressure.

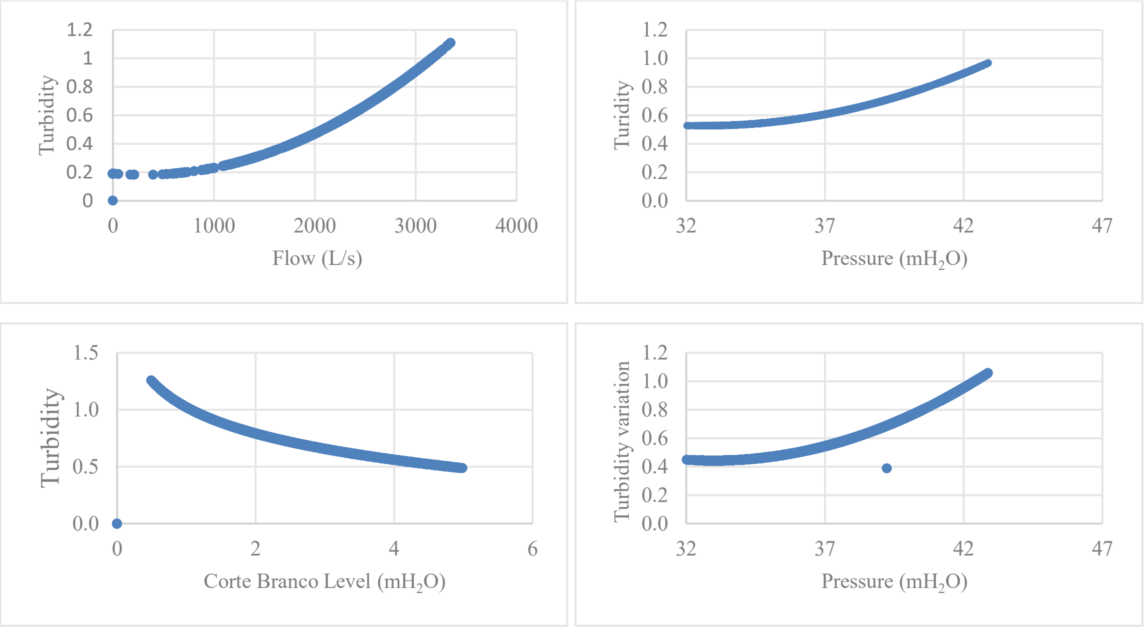


Figure 9 – Turbidity Applying the Generated Correlations

3.3 Field Campaigns

Two water quality measurement campaigns were realized during hydraulic transient events in order to obtain the relationship between the physicals and quality results. The samples were used to measure the turbidity and chlorine.

In the first campaign the pipelines were with no flow for about 3 hours. The test consisted in turn on the pumps following the operational routine, turning on one by one. Then, a time was wait and after that, all of them were turned off together. In the second one the pipelines were operating. The test consisted in turn off all pumps. Then turn on one by one and them turn off all of them.

The samples were collected in the end of the pipeline each five minutes in the first test and two minutes in the second one.

1.2 Hydraulic Transient Calculation

The simulation of the hydraulic transient in the pipelines was carried out through the software Hammer, Bentley, which uses as numerical method the Method of Characteristic (MOC).

4 PRELIMINARY RESULTS

Simulating the hydraulic transient caused by the shutdown of the pumps during 600 seconds, the results expressed in Figure 10 were obtained.

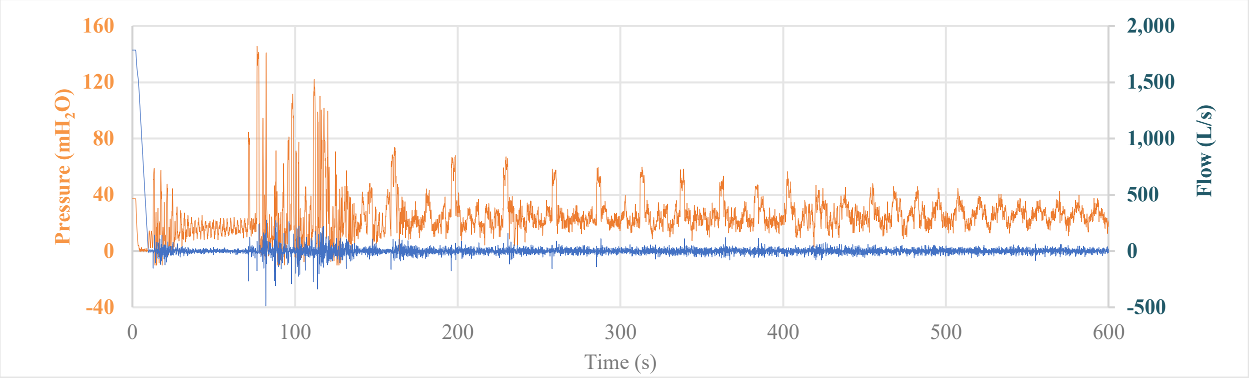


Figure 10 – Pressure and Flow during a Hydraulic Transient Event

Applying the generated equations of quality in the physical results of hydraulic transients, the Figure 11 is obtained.

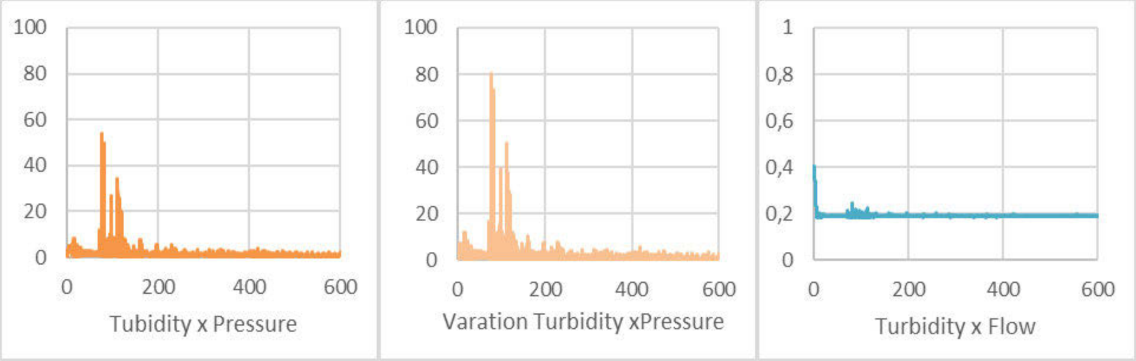


Figure 11 – Theoretical Turbidity During Hydraulic Transient

The result of the first water quality measurement campaign is presented in Figure 12 and the second one in Figure 13.

Analyzing the graphs is clear that the water quality change during the hydraulic transient. In the first test is observed that chlorine was zero in the begging of the test because the flow was zero for two hours before the test, and it take a time to increase, probably influenced by wall's pipeline effect. When the pumps were turned off the turbidity increase about 0,6uT.

In the second test is observed the turbidity increasement was about 0.7uT and 1.3 in one episode. The chlorine index had slight variation when the flow was different of zero, with a tendency to fall when the flow decreases to zero, probably influenced by wall's pipeline effect.

These preliminary results show in a real water system the relationship between hydraulic transients and the turbidity increase.

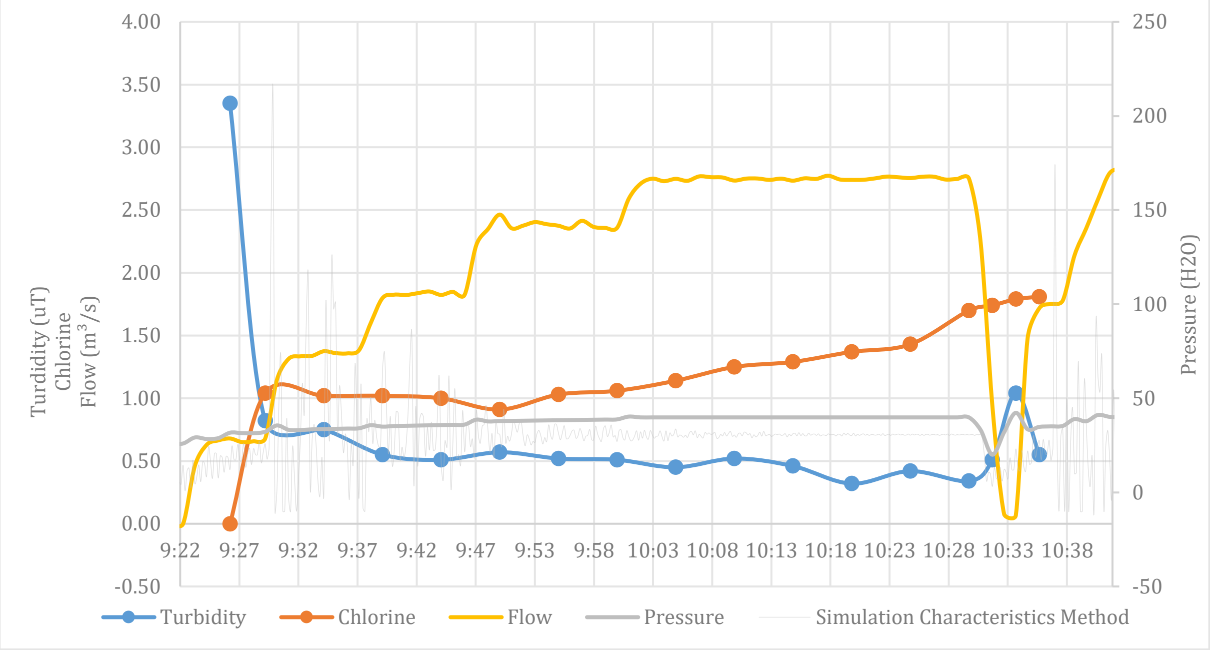


Figure 12 – First Campaign

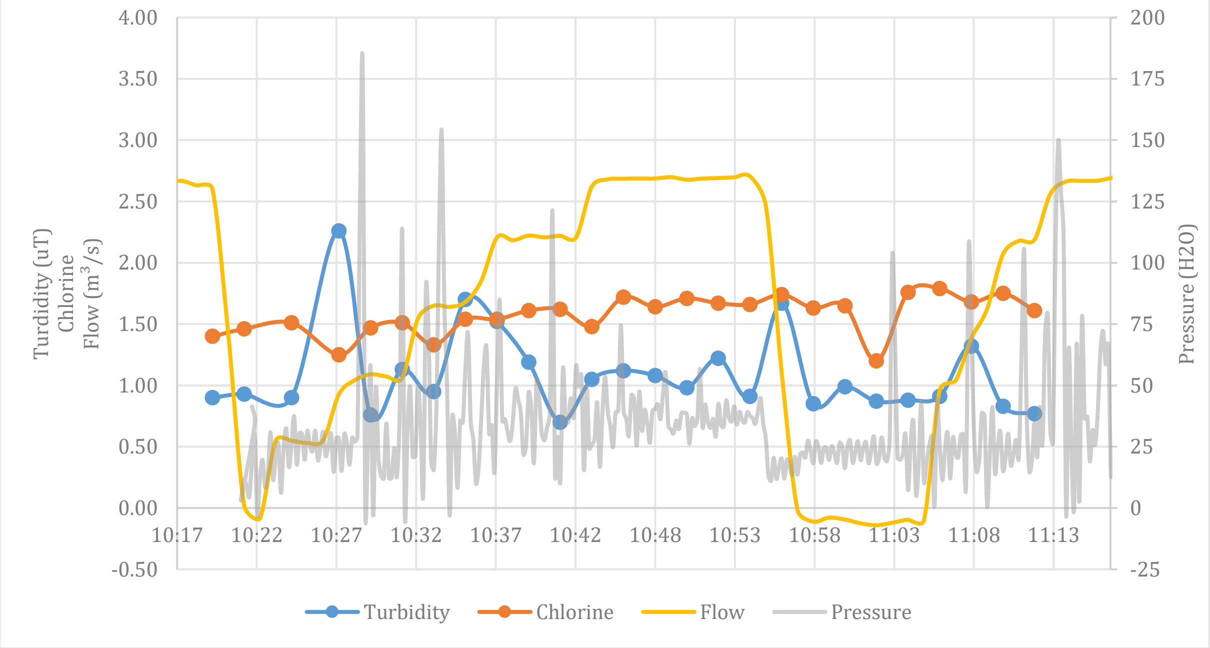


Figure 13 – Second Campaign

5 CONCLUSION

In addition, it is worth mentioning that monitoring and investigation of transients in operational systems is only occasionally done to investigate perceived problematic areas over a short period of time, and generally with physical focus. That happens because of the usually inability to manage the vast volumes of data generated from sampling with significantly higher temporal resolution than the state of art fifteen minutes [11].

The results prove in a real system the relationship between hydraulic transients and the turbidity increase. In this scenario, an indicator that correlates physical quantities with water quality has great relevance for the definition of projects and understanding of real hydraulic transient phenomena. It can be used to: compare different protect solutions; change material; get better flushing procedures or redefine any other condition. The studies are still in development and next steps is evaluate the preliminary results.

6 ACKNOWLEDGEMENTS

Part of the results already have been published in a Brazilian Nacional Seminary [51].

7 REFERENCES

[1] CHAUDHRY M. H. 1979. Applied Hydraulic Transients. Van Nostrand Reinhold Company. New York.

[2] WOOD, D.; LINGIREDDY, S; BOULOS, P.; KARNEY, B; MCPHERSON, D. 2005. Numerical methods for modeling transient flow in distribution systems. Journal AWWA, 97:7

[3] KARNEY, B.; MCINNIS, D. 1990. Transient analysis of water distribution systems. Journal of AWWA, R2 (7), 62-70.

[4] LANSEY, K.; BOULOS, P. (2005). Comprehensive Handbook on Water Quality Analysis for Distribution Systems. 448 pages.

[5] BOULOS, P.; KARNEY, B; WOOD, D; LINGIREDDY, S. (May, 2005). Hydraulic Transient Guidelines for Protecting Water Distribution Systems. Journal AWWA, pp. 111-124.

[6] FERNANDES, C.; KARNEY, B. 2004. Modelling the advection equation under water hammer conditions. Urban Water Journal. 1:2, 97-112, DOI: 10.1080/15730620412331290038.

[7] KARIM, M.R.; ABBASZADEGAN, M.; LECHEVALLIER, M. 2003. Potential for pathogen intrusion during pressure transients. Journal of AWWA 95 (5), 134-146.

[8] LECHEVALLIER, M.W. (1999). The Case for Maintaining a Disinfectant Residual. Journal of AWWA. 91: 86-94.

[9] KIRMEYER, G. J. et al. 2001. Pathogen intrusion into the distribution system. AWWA Research foundation, AWWA, Denver.

[10] FUNK, J. E.; WOOD, D. J.; VANVUUREN, S. J.; LCCHEVALLIER, M. 1999. Pathogen intrusion into water distribution systems due to transients. Proc., 3rd\ASME/JSME Joint Fluids Engineering Conference, July 18e23, San Francisco.

[11] Hoskins, A. & Stoianov, Ivan. (2014). InfraSense: A Distributed System for the Continuous Analysis of Hydraulic Transients. Procedia Engineering. 70. 823-832. 10.1016/j.proeng.2014.02.090.

[12] WALSKI, Tom; MINNICH, Kendric; SHERMAN, Corbin; STRAUSE, Lee; WHITMAN, Brian. 2017. Can There be a Law of Conservation of Turbidity. Procedia Engineering. 186. 372-379. 10.1016/j.proeng.2017.03.233.

[13] LINDLEY, T. R.; BUCHBERGER, S. G. 2002. Assessing intrusion susceptibility in distribution systems, J. Am. Water Works Assoc. 94, pp. 66-79.

[14] COLLINS, R.; BOXALL, J. 2013. Influence of Ground Conditions on Intrusion Flows through Apertures in Distribution Pipes, J. Hydraul. Eng.-ASCE 139, pp. 1052-1061.

[15] CLAYTON, C., AND VAN ZYL, J. 2007. “The effect of pressure on leakage in water distribution systems.” Proc. ICE Water Manage., 160(2), 109–114.

[16] FONTANAZZA, C. M. 2015. 13th Computer Control for Water Industry Conference. Contaminant intrusion through leaks in water distribution system: experimental analysis.

[17] WESTON, Sally; COLLINS, Richard; BOXALL, Joby. 2018. Transient mobilization of replicated biofilms. In: WDSA / CCWI Joint Conference 2018. WDSA / CCWI Joint Conference 2018, 23-25 Jul 2018, Kingston, Canada. Queen’s University and the Royal Military College, Canada.

[18] KERAMAT, A.; PAYESTEH, M.; BRUNONE, B.; MENICONI, S. 2020. Interdependence of flow and pipe characteristics in transient induced contamination intrusion: Numerical analysis. Journal of Hydroinformatics, 22(3), 473-490. https://doi.org/10.2166/hydro.2020.069.

[19] MCINNIS, D. 2004. A relative-risk framework for evaluating transient pathogen intrusion in distribution systems. Urban Water Journal, 1:2, 113-127, DOI: 10.1080/15730620412331290010.

[20] BOXALL, J.B.; SKIPWORTH, P.J.; AND SAUL, A.J. 2001. A novel approach to modelling sediment movement in distribution mains based on particle characteristics. In B. Ulanicki, B. Coulbeck, and J.P. Rance (Eds.), Water software systems: vol. 1: Theory and Applications (Water Engineering & Management) (pp. 263–273). Baldock: Research Studies Press. Vreeburg, J.H.G., and Van den Boomen, M. (2002). New concepts for self-cleaning networks and improved water quality. Water 21, 4(6), 43–45.

[21] WESTON, Sally; COLLINS, Richard; BOXALL, Joby. 2015. Do Transients Contribute to Turbidity Failures of Water Distribution Systems?. 10.13140/RG.2.1.4667.9122.

[22] WESTON, Sally; COLLINS, Richard; BOXALL, Joby. 2019. A novel demonstration of adhered material mobilization by hydraulic transients. Environmental Science Water Research & Technology. 2019.

[23] SPEIGHT, Vanessa; MOUNCE, Steve; BOXALL, Joby. 2019. Identification of the causes of drinking water discoloration from machine learning analysis of historical datasets. Environmental Science: Water Research & Technology. 5. 10.1039/C8EW00733K.

[24] WESTON, Sally; COLLINS, Richard; BOXALL, Joby. 2021. An experimental study of how hydraulic transients cause mobilization of material within drinking water distribution systems. Water Research, Volume 194. 2021. 116890, ISSN 0043-1354, https://doi.org/10.1016/j.watres.2021.116890.

[25] PRECIADO, Carolina Calero; HUSBAND, Stewart; BOXALL, Joby; OLMO, Gonzalo; SORIA-CARRASCO, Víctor; MAENG, Sung Kyu; DOUTERELO, Isabel. 2021. Intermittent Water Supply Impacts on Distribution System Biofilms and Water Quality, Water Research, Volume 201, 2021, 117372, ISSN 0043-1354, https://doi.org/10.1016/j.watres.2021.117372.

[26] CHEN, R; ZHUANG, Y; Yu Y; Shi B. 2021. Enhanced perfluorooctanoic acid (PFOA) accumulation by combination with in-situ formed Mn oxides under drinking water conditions. Water Res. 2021 Feb 15;190:116660. DOI: 10.1016/j.watres.2020.116660. Epub 2020 Nov 20. PMID: 33279743.

[27] WESTON, Sally; SHEPHERD, Will; COLLINS, Richard; BOXALL, Joby. 2018. Experimental Quantification of Intrusion Volumes Due to Transients in Drinking Water Distribution Systems. Journal of Pipeline Systems Engineering and Practice. 10. 10.1061/(ASCE)PS.1949-1204.0000348.

[28] BRAGA, Arthur Sass; SAULNIER, Rain; FILION, Yves; CUSHING, Alexandria. 2020. Dynamics of material detachment from drinking water pipes under flushing conditions in a full-scale drinking water laboratory system, Urban Water Journal, DOI: 10.1080/1573062X.2020.1800759.

[29] FISHER, Ian; KASTL, George; SATHASIVAN, Arumugam. 2017. New model of chlorine-wall reaction for simulating chlorine concentration in drinking water distribution systems. Water Research. 125. 427-437. 10.1016/j.watres.2017.08.066.

[30] ARMAND, Hooman; STOIANOV, Ivan; GRAHAM, Nigel. 2015. A holistic assessment of discoloration processes in water distribution networks, Urban Water Journal, DOI: 10.1080/1573062X.2015.1111912.

[31] ARMAND, Hooman; STOIANOV, Ivan; GRAHAM, Nigel. 2015. Investigating the Impact of Sectorized Networks on Discoloration. Procedia Engineering. 119. 407-415. 10.1016/j.proeng.2015.08.902.

[32] BLOKKER, E.; SCHAAP, P.G. 2015. Temperature Influences Discoloration risk. Procedia Engineering. 119. 10.1016/j.proeng.2015.08.887.

[33] BLOKKER, E.J.M.; SCHAAP, P.G. Particle Accumulation Rate of Drinking Water Distribution Systems Determined by Incoming Turbidity. Procedia Engineering. Volume 119. 2015. Pages 290-298. ISSN 1877-7058. https://doi.org/10.1016/j.proeng.2015.08.888.

[34] HUSBAND, Stewart; JACKSON, M.; BOXALL, Joby. 2015. Identifying Material Accumulation Processes in Drinking Water Distribution Systems with Extended Period EPANET MSX Turbidity Simulations. Procedia Engineering. 119. 398-406. 10.1016/j.proeng.2015.08.901.

[35] SUMMEREN, Joost; RATERMAN, Bernard; VONK, Erwin; BLOKKER, E.; ERP, Johan; VRIES, Dirk. 2015. Influence of Temperature, Network Diagnostics, and Demographic Factors on Discoloration-Related Customer Reports. Procedia Engineering. 119. 10.1016/j.proeng.2015.08.903.

[36] HUSBAND, Stewart; BOXALL, Joby. 2016. Understanding and managing discoloration risk in trunk mains. Water Research. Volume 107. 2016. Pages 127-140. ISSN 0043-1354. https://doi.org/10.1016/j.watres.2016.10.049.

[37] WESTON, Sally; COLLINS, Richard; BOXALL, Joby. 2017. Can a Valve Closure Transient Cause Mobilisation of Adhered Particles?. 10.15131/shef.data.5364496.

[38] AISOPOU, Angeliki; STOIANOV, Ivan; GRAHAM, Nigel. 2011. Modelling Discoloration in WDS Caused by Hydraulic Transient Events. Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference, WDSA 2010. 522-534. 10.1061/41203(425)49.

[39] KIM, Hyunjun; KIM, Sanghyun. 2017. Evaluation of chlorine decay models under transient conditions in a water distribution system. Journal of Hydroinformatics. 19. jh2017082. 10.2166/hydro.2017.082.

[40] XU J, HUANG C, SHI X, DONG S, YUAN B, NGUYEN TH. 2018. Role of drinking water biofilms on residual chlorine decay and trihalomethane formation: An experimental and modeling study. Sci Total Environ. 2018 Nov 1 5;642:516-525. doi: 10.1016/j.scitotenv.2018.05.363. Epub 2018 Jun 13. PMID: 29908510.

[41] LI, Weiying; TAN, Qiaowen; ZHOU, Wei; CHEN, Jiping; LI, Yue; WANG, Feng; ZHANG, Junpeng. 2019. Impact of substrate material and chlorine/chloramine on the composition and function of a young biofilm microbial community as revealed by high-throughput 16S rRNA sequencing. Chemosphere. 242. 125310. 10.1016/j.chemosphere.2019.125310.

[42] SABOE, Daniel; HRISTOVSKI, Kiril D.; BURGE, Scott R.; BURGE, Russell G.; TAYLOR, Evan; HOFFMAN, David A. 2021. Measurement of free chlorine levels in water using potentiometric responses of biofilms and applications for monitoring and managing the quality of potable water, Science of The Total Environment. Volume 766, 2021, 144424, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.144424.

[43] ZHANG, Di; WANG, Feifei; DUAN, Youli; CHEN, Shenghua; ZHANG, Aihong; CHU, Wenhai. 2020. Removal of trihalomethanes and haloacetamides from drinking water during tea brewing: Removal mechanism and kinetic analysis. Water Research. 184. 116148. 10.1016/j.watres.2020.116148.

[44] LEMUS-PÉREZ, M.F.; SUSA, M. Rodríguez. The effect of operational conditions on the disinfection by-products formation potential of exopolymeric substances from biofilms in drinking water, Science of The Total Environment, Volume 748, 2020, 141148, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.141148.

[45] ZHANG, Haiya; ZHAO, Lvtong; LIU, Dibo; WANG, Jun; ZHANG, Xiaojian; CHEN, Chao. 2020. Early period corrosion and scaling characteristics of ductile iron pipe for ground water supply with sodium hypochlorite disinfection. Water Research. 176. 115742. 10.1016/j.watres.2020.115742.

[46] WANG, Manna; ATEIA, Mohamed; AWFA, Dion; YOSHIMURA, Chihiro. 2020. Regrowth of Bacteria after Light-based Disinfection — What we know and where we go from here. Chemosphere. 268. 10.1016/j.chemosphere.2020.128850.

[47] DEVANE, Megan; MORIARTY, Elaine; WEAVER, Louise; COOKSON, Adrian; GILPIN, Brent. 2020. Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research. 185. 116204. 10.1016/j.watres.2020.116204.

[48] ATEIA, Mohamed; GAR ALALM, Mohamed; AWFA, Dion; YOSHIMURA, Chihiro. 2019. Modeling the Degradation and Disinfection of Water Pollutants by Photocatalysts and Composites: A Critical Review. Science of The Total Environment. 698. 10.1016/j.scitotenv.2019.134197.

[49] DOUTERELO, I; FISH, KE; BOXALL, JB. 2018. Succession of bacterial and fungal communities within biofilms of a chlorinated drinking water distribution system. Water Res. 2018 Sep 15;141:74-85. DOI: 10.1016/j.watres.2018.04.058. Epub 2018 May 3. PMID: 29778067.

[50] SORENSEN, JPR; VIVANCO, A; ASCOTT, MJ; GOODDY, DC; LAPWORTH, DJ; READ, DS; RUSHWORTH, CM; BUCKNALL, J; HERBERT, K; KARAPANO, I; GUMM, LP; TAYLOR RG. 2018. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Res. 2018 Jun 15;137:301-309. DOI: 10.1016/j.watres.2018.03.001. Epub 2018 Mar 3. PMID: 29554534.

[51] AGOSTINHO, M. S. P. ; FERNANDES, C. V. S. . Physical and Water Quality Indicators Assessment in main pipelines: The impact of Hydraulic Transient Events. In: XXIV Brazilian Symposium on Water Resources, 2021, Belo Horizonte. Água Múltiplas Dimensões. Porto Alegre: ABRHidro, 2021. v. 1. p. 1-10.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14733

PRESSURE SENSOR PLACEMENT IN WATER NETWORKS BASED ON SENSITIVITY MATRIX: A REAL INSTRUMENTATION PROJECT PLAN

Joan Van Eeckhout[1], Josep Cugueró[2] and Ramon Pérez[3]

[1,2,3] CS2AC, Universitat Politècnica de Catalunya. Rambla sant Nebridi 10, 08222, Terrassa, Spain
[1]Taigua, Aigua Municipal de Terrassa. Carrer de la Societat, 30, 08221 Terrassa, Spain

[1] joan.van.eeckhout@upc.edul,[2] josep.cuguero@upc.edu, [3] ramon.perez@upc.edu.

Abstract

The instrumentation and supervision of water distribution networks (WDN) are challenging tasks as the use of monitoring devices imply a high cost including installation and maintenance. One of the emerging issues is the definition of the best possible layout of the measurements to be recorded. Considering the sampling costs and budget limitations on the number of sensors to be distributed, to determine a trade-off is mandatory. The trade-off may highly depend on the final use of these sensors. This work presents a methodology to support practitioners to decide which is the optimal pressure sensor placement and the appropriate number of pressure sensors to install on a WDN. Two main applications for demand calibration and leak detection and localisation are considered. A sensor placement methodology is applied to find the most sensible locations with respect to demand groups. Additionally, as the sensor placement methodology does not define an optimal number of sensors to be distributed in the WDN, an approach to formulate a trade-off is developed. This approach generates an indicator that quantifies the degree of relevance of the information recorded by a given sensor layout. The methodology presented is applied on a real WDN to select the most appropriate pressure measurement layout as part of an instrumentation project plan.

Keywords
Water Distribution Networks, Sampling design, Sensors.

1 INTRODUCTION

Water Distribution Networks (WDNs) generally are large and complex systems formed by thousands of pipes and nodes and its instrumentation and supervision are challenging tasks as the use of monitoring devices imply a high cost including installation and maintenance. One of the emerging issues is the definition of the best possible layout of the measurements to be recorded, as the number of measurement devices is far smaller than the potential measurement points [1]. The use of models is crucial in this process, even if the final application using these measurements might not be based on WDN models. However, the suitability of the layout of the sensors may depend on the final application using these sensors [2].

Within the framework of an industrial thesis, the municipal water company of Terrassa (TAIGUA) intends to calibrate its WDN models. An industrial thesis consists in a strategic research project of a company where a doctoral student develops his research training in collaboration with a university. Following the seven-step general calibration procedure proposed by Ormsbee and Lingireddy [3], considering the intended use of the models, three model use types were identified whose differences arise in the precision of the demand model [4]. Models of type one are used for planning purposes within a ten year horizon. Models of type two are intended to perform short term predictions in peak days and need a more precise demand model. Models of type three would allow leakage detection and localization and need telemeter measurements to be fed constantly.

To generate a type three model a microcalibration process has to be performed. This microcalibration process requires more instrumentation to be installed in the WDN. As one objective is to generate and calibrate different groups of demand in the same DMA [5] and pressure monitoring is much less expensive than flow monitoring [6], it is chosen to install pressure sensors within the DMAs. Furthermore, results presented in [7] show that for meshed networks measuring pressure seems to be the best option when calibrating geographical groups of demand.

The pressure sensors that will be installed are the remote transmission units with data loggers ADT1 [8]. Despite the usual versatility of the pressure loggers which allows to place them at hydrants or blow off valves and move them to other locations without the need of a construction project, the company has decided to install the pressure sensors directly at the distribution pipes by means of small valve boxes avoiding possible vandalism damage. However, by means of pressure drilling, also referred to as hot tapping, line tapping or pressure tapping, the pressure sensors will be directly installed at pressurized pipelines without having to drain down the system or interrupt the service. Therefore, the pressure sensors will be installed at fixed locations. This fact gives relevance to carry out a sampling design (SD).

When considering the sampling costs and budget limitations on the number of sensors to be distributed to determine a trade-off is mandatory. Generally, the trade-offs analysed in literature the balance the need for better-calibrated models with the SD costs to justify the expenses on sampling efforts and data collection [9], [10], [11], [12], [13], [14], [15]. However, Kang and Lansey [16] point out that considering the trade-off between model accuracy and precision is more appropriate for the optimal meter placement than considering and minimize the metering costs.

There is not a definitive guidance to practitioners on how to balance the SD costs and there is not much evidence that SD methodologies are being used in practice [17]. Therefore, it is necessary to develop a methodology that supports practitioners to decide which is the appropriate number of sensors to install on a WDN.

In this paper, with the aim to perform demand calibration with pressure sensors, a sensor placement methodology is applied to find the most sensible locations with respect to demand groups [5].

Additionally, as the sensor placement methodology does not define an optimal number of sensors to be distributed in the WDN, an approach to obtain a trade-off is developed. This approach generates an indicator that quantifies the degree of relevance of the information recorded by a given sensor layout.

As a leak-detection and localization approach can be coupled with the demand calibration methodology [18], the proposed indicator does not only consider or depend on a single final application. The procedure to generate this indicator is based on analysing the magnitude of the pressure changes that the sensors would measure when variations in the demand occur. Thus, comparing the values of the indicator for several layouts of sensors with different numbers of sensors distributed in the network, the optimal number of sensors can be defined.

The methodology presented in this paper is applied on a real WDN to select the most appropriate pressure measurement layout as part of an instrumentation project plan. The studied WDN supplies the northern urban area of Terrassa and consists of four District Metered Areas (DMAs) gravity-fed from a single tank. The two lower DMAs are separated by a pressure reducing valve (PRV) from the two upstream DMAs.

2 PROBLEM FORMALIZATION

Given a WDN and its EPANET model with all the physical behaviour well defined and validated in a macrocalibration phase (Water Distribution Network Model calibration and Continuous Maintenance: Terrassa, a real application), the expected nodal pressures can be obtained simulating this model. The simulation of a WDN model depend on the demands and the boundary conditions, as for instance the levels of the tanks or the status of the pumps and valves status. Equation (1) represents the process of obtaining the estimated nodal pressures vector $\hat{\mathbf{p}} \in \mathbb{R}^{N_n}$ through the simulation of the WDN model at the boundary conditions B_{Cond} and nodal demands vector $\mathbf{d} \in \mathbb{R}^{N_n}$**.** Being N_n the number of nodes of the WDN model.

$$Simulate\ Model(B_{Cond}, \mathbf{d}) = \hat{\mathbf{p}} \tag{1}$$

The relationship between pressures and demands is non-linear. However, given a working point $(B_{Cond}, \mathbf{d})$ a linear approximation can be obtained as shown in equation (2).

$$\hat{\Delta}\mathbf{p} = \mathbf{S}(B_{Cond}, \mathbf{d}) \cdot \Delta\mathbf{d} \tag{2}$$

Where $\mathbf{S} \in \mathbb{R}^{N_n \times N_n}$ is the sensitivity matrix which contains information interrelating the variation of pressures in all the nodes caused by a variation of the demand.

A sensor placement approach, based on the analysis of the sensitivity matrix [5] must be applied to find the locations where the pressure is more sensible to the demand. Equation (3) represents a sensor placement process to find the optimal sensor ubications given the sensitivity matrix $\mathbf{S}$ and the number of sensors N_s.

$$Sensor\ Placement(\mathbf{S}(B_{Cond}, \mathbf{d}), N_s) = \mathbf{Id}^* \tag{3}$$

Where $\mathbf{Id}^* \in \mathbb{N}^{N_s}$ is the index of the nodes of the optimal sensor locations.

The number of sensors distributed on the WDN in these layouts must be below a budget limit defined as $N_s \leq N_{sMax}$.

Once the possible sensors layouts with different number of sensors N_s have been determined, the optimal number of sensors $N_s{}^*$ must be stablished. To determine the optimal number of sensors, an indicator must be defined and used to quantify the degree of relevance of the information recorded by a sensor layout. This indicator is defined as $\hat{\Delta}\mathbf{p}_{\boldsymbol{meas}} \in \mathbb{R}^{N_n}$ as presented in Equation (4). Its computation will be presented in the methodology section.

$$Information\ of\ Instrumentation\ Indicator(\mathbf{Id}^*) = \hat{\Delta}\mathbf{p}_{\boldsymbol{meas}} \tag{4}$$

To determine the optimal number of sensors $N_s{}^*$, the indicator $\hat{\Delta}\mathbf{p}_{\boldsymbol{meas}}$ will be computed and evaluated for all the feasible layouts of sensors such that $N_s \leq N_{sMax}$ until the increase of the number of sensors does not increase the indicator quantifying the relevance of the information recorded by them. Equation (5) represents the trade-off used to find optimal number of sensors $N_s{}^*$ from which the weighted increase of the indicator does not compensate the increase in the cost of the sampling design C_{Sensor}.

$$W(mean(\hat{\Delta}\mathbf{p}_{\boldsymbol{meas}}(N_s{}^* + 1)) - mean(\hat{\Delta}\mathbf{p}_{\boldsymbol{meas}}(N_s{}^*))) < C_{Sensor} \tag{5}$$

3 CASE STUDY

Terrassa's WDN is divided into eight pressure floors. The division is performed by minimizing the altitude variation within these pressure floors to ease the management and efficiency. As the altitude variations within these floors are reasonable, the floors can be supplied mainly by gravity from tanks. The company have generated eight hydraulic models of these pressure floors and intends to calibrate these models.

The results presented in [4](Water Distribution Network Model calibration and Continuous Maintenance: Terrassa, a real application) validated the physical model of the pressure floor of Sulleva to generate models of type one and type two. To generate a type three model a microcalibration process will be performed. This microcalibration process requires more instrumentation installed in the WDN.

The WDN of Sulleva consist of 32,2 km of pipes and supplies approximately 530.000 m3/year to 4.500 users. The pressure floor is gravity-fed from a single tank and is instrumented with five flow meters dividing the WDN in its four DMAs. The two lower DMAs are separated by a pressure reducing valve (PRV) from the two upstream DMA. Figure 1 presents the hydraulic model of the pressure floor of Sulleva. The water source is the reservoir signalled with a blue square. The boundaries of the four DMAs are defined by the flowmeters, signalled by green squares, and by closed valves, signalled by black squares. The PRV separating the two upper DMAs from the two lower DMAs is signalled with a red square. The pipes of the two upper DMAs are signalled in black and the pipes of the two lower DMAs are signalled in magenta.

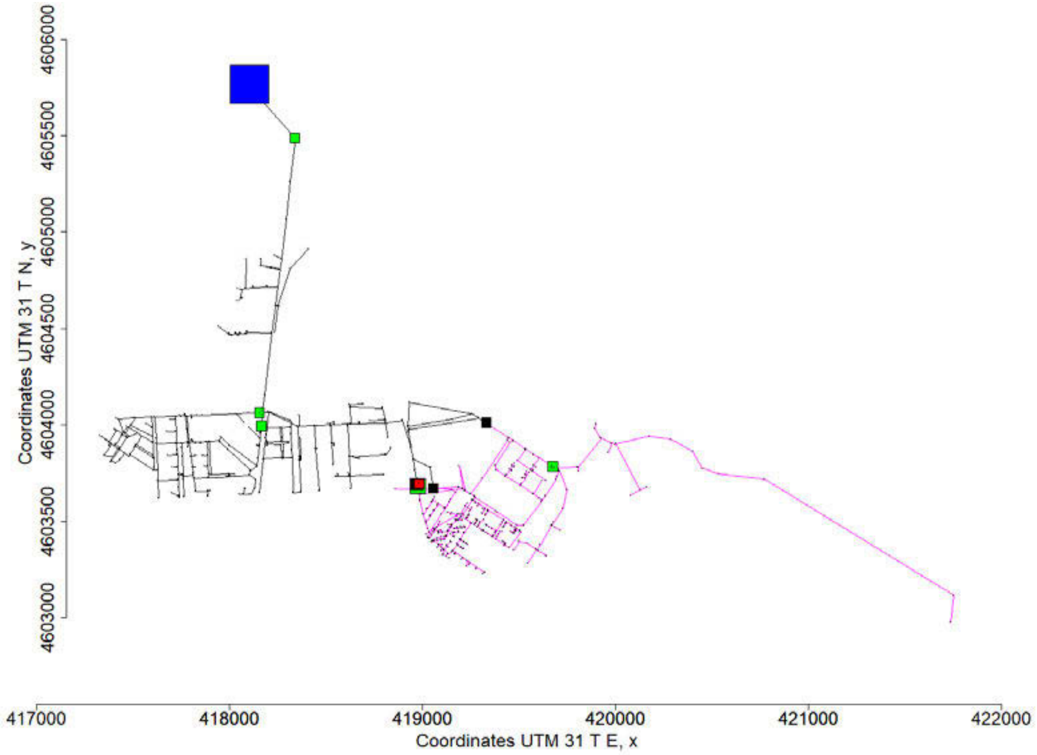


Figure 1. WDN model of Sulleva pressure floor.

As one of the flow meters and a pressure sensor are located at the outlet of the PRV, the WDN model is divided in two models to avoid including the PRV in the sensitivity matrices computation. In the upstream of the PRV model, the PRV is substituted by a consumption node and the demand of the node is fixed according to the flow measured by the flow meter located at the outlet of the PRV. In the downstream of the PRV model, the PRV is substituted by a reservoir whose hydraulic head is fixed according to the measurements of the pressure sensor located at the outlet of the PRV.

In this work the sampling design methodology is applied only to the upstream of the PRV model and the results obtained are displayed to illustrate the procedure for applying the methodology

CC BY NC SA

to a specific WDN model. This WDN model contains 18,7 km of pipes, 573 links, 543 nodes and 115 consumption nodes.

4 METHODOLOGY AND RESULTS

4.1 Simulation for the Sampling Design

Usually, to model the demands the users of the same type are assumed to have similar diurnal pattern and the consumption of each user is computed by multiplying the pattern coefficients with the baseline demand as presented in [19]. However, user diurnal patterns are not always available in practice. Quite often, water companies only manage information with low temporal resolution, the monthly or quarterly billing of each consumer. This information can be used to compute the baseline demand or base demand of each user. Additional information collected from flow-meters recording all the water entering and leaving the DMAs may be used to estimate each user consumption at every instant. The basic model presented in Equation (6) uses the nodal base demand, together with the total DMA consumption metered at the DMA inlets to calculate the demand of each node at each sample.

$$\boldsymbol{d}_i(t) = \frac{bd_i}{\sum_{j=1}^{N_n} bd_j} \cdot \boldsymbol{q}_{\boldsymbol{in},\boldsymbol{DMA\,k}}(t) \qquad (6)$$

Where bd_i is the base demand of node i; and $\boldsymbol{q}_{\boldsymbol{in},\boldsymbol{DMA\,k}}(t)$ is the total DMA water consumption metered at sample t.

As no prior information about the diurnal pattern of consumption of the users is available, the information of the flowmeters measuring the inflows and outflows of the DMAs is used to model the demands of the users by means of equation (6) and generate a simulation that will be analysed to compute the sensitivity matrix at each working point and carry on the sampling design based on the model simulation.

4.2 Sensor Placement

Due to the large number of unknown values or parameters, it is impossible to calibrate the model of a real system precisely [20]. Kang and Lansey [21], [22] have emphasized the importance to make the model parameters identifiable by reducing the number of calibrated parameters to make the system overdetermined or at least even determined, as not sufficient measurements are available. The same conclusion was drawn from the Battle of the Water Calibration Networks, organized in 2010, pointing out that calibration size problem reduction is an important factor to consider to avoid model overfitting and unnecessary simulations reducing the search space [23]. Too much aggregation leads to few parameters that are easy to calibrate, but it also entails worse model predictions with significant errors. Increasing the number of calibrated parameters improves the agreement between model predictions and measurements, but parameter uncertainty increases too. Thus, the trade-off between model error due to aggregation of parameters and parameter uncertainty should be taken into account [24].

As the demands are intended to be calibrated by means of pressure measurements, the Singular Value Decomposition (SVD) to generate the right-singular vectors **V** is applied to the sensitivity matrix of the nodal pressures versus the demands presented as **S** in equation (2). Equation (7) show the SVD of **S**, where $\boldsymbol{\Sigma} \in \mathbb{R}^{N_n \times N_n}$ is a diagonal matrix containing the singular values of matrix **S** and $\mathbf{U} \in \mathbb{R}^{N_n \times N_n}$ and $\mathbf{V} \in \mathbb{R}^{N_n \times N_n}$ are the left and right singular vectors of matrix **S**.

$$SVD(\mathbf{S}) = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathrm{T}} \qquad (7)$$

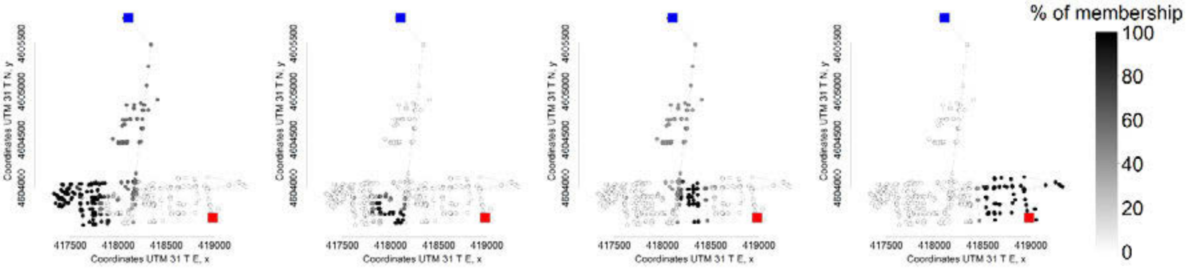
Using **V** the resolution matrix **R** can be defined as equation (8).

$$\mathbf{R} = \mathbf{V} \cdot \mathbf{V}^{\mathrm{T}} \tag{8}$$

Applying the delta vector generation process to the parameter resolution matrix **R** the demands can be grouped depending on their resolvability [25], [26], [5]. This demand grouping approach tends to generate geographical groups as topological information is included in the sensitivity matrix [5].

The sampling design is performed according to a budget limit that sets a bound N_{sMax} on N_s. Thus, the sensor placement methodology is applied for all the possible number of sensors satisfying $N_s \leq N_{sMax}$.

In the particular case of application $N_{sMax} = 6$. To illustrate the sensor placement methodology, the results for $N_s = 4$ are shown. To obtain an even determined system of equations guaranteeing the system identifiability the number of parameters must be equal to the number of sensors. Therefore, for $N_s = 4$ the demands are parametrized in four geographical groups of demands. Figure 2 represents the membership of each node to each of the demand groups obtained from applying the demand grouping approach at a particular working point of the EPS. Thus, the demand groups displayed are obtained analysing the sensitivity matrix computed at that specific working point (see Equation (2)).



a) 1st Demand group b) 2nd Demand group c) 3rd Demand group d) 4th Demand group

Figure 2. Nodal membership to demand groups.

Maximizing the sensitivity is important in inverse problems because if the measurements are insensitive to the parameters, a large change in those parameters will change very little the measurements, perhaps within the error of measurement, and the parameters will be determined with low confidence [27].

As the groups of demands are intended to be calibrated by means of pressure measurements, a new sensitivity matrix of the nodal pressures versus the demand groups $\mathbf{S}' \in \mathbb{R}^{N_n \times N_s}$ is computed. Thus, SVD to generate the left-singular vectors $\mathbf{U}' \in \mathbb{R}^{N_n \times N_n}$ is applied to the sensitivity matrix $\mathbf{S}'$ (equation (9)).

$$SVD(\mathbf{S}') = \mathbf{U}' \cdot \mathbf{\Sigma}' \cdot \mathbf{V}'^{\mathrm{T}} \tag{9}$$

Using $\mathbf{U}'$ information density matrix $\mathbf{I}_d$ can be defined as

$$\mathbf{I}_d = \mathbf{U}' \cdot \mathbf{U}'^{\mathrm{T}} \tag{10}$$

Applying the delta vector generation process to the parameter information density matrix $\mathbf{I}_d$ the most sensible sensor locations to each demand group can be determined [25], [5].

To make the sensor placement more robust considering the temporal changes of the demands the sensor placement process is performed under unsteady hydraulic conditions by applying the process to each working point of an EPS (see equation (3)) [7], [16]. Thus, applying the sensor placement process to k different working points, $k \cdot N_s$ possible locations of sensors are obtained, from which N_s sensos must be selected, being k the number of steps of the EPS analysed to perform the sensor placement. For the case presented in this paper $k = 96$. In most of the cases, the network topology has the highest impact on the sensitivity matrix and the sensors are placed at nearby locations when analysing different working points [7].

The procedure to select the N_s final sensors consists of 4 steps:

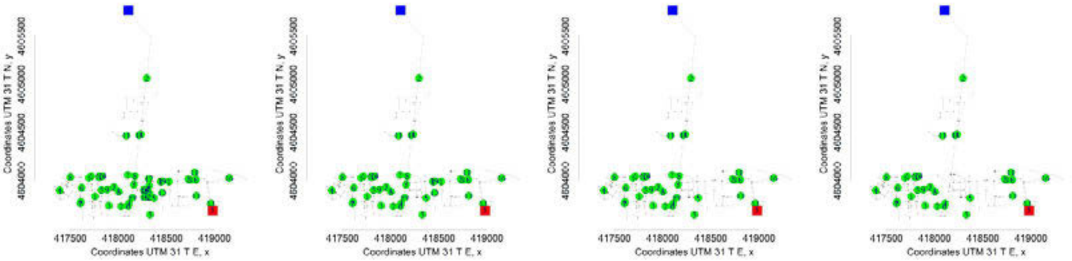
1. Select the set of possible locations of sensors that, having an hydraulic separation between them of less than a predefined threshold according to the WDN dimensions, contain the maximum number of repetitions. Repetitions are the number of times that a location has been selected as one of the $k \cdot N_s$ possible locations of sensors.
2. For each possible location of sensor s in the set, a weight w_s is calculated depending on its number of repetitions and the distance d to other possible locations of the set and repetitions r_i of other possible N_{Set} locations in the set (equation (11)).

$$w_s = \sum_{i=1}^{N_{Set}} \frac{r_i}{10^{\frac{d_{s,i}}{d_{Max}(s)}}} \tag{11}$$

   Where: $r_i$is the number of repetitions of the possible locations of sensors i; N_{Set} is the number of sensors in the set; $d_{s,i}$ is the distance between the possible locations of sensors s and i; and $d_{Max}(s)$ is the maximum distance between the possible location of sensor s and all other possible locations of sensors in the set.

3. The location of sensors in the set with highest weight is selected. All the other sensors are deleted from the possible locations of sensors and their number of repetitions is added to the selected sensor.
4. Repeat the process until N_s locations of sensors are selected.

Figure 3 presents the process of selection of the $N_s = 4$ sensor locations among the $k \cdot N_s$ possible locations obtained applying the sensor placement process to k different working points. The possible locations of sensors are signalled in green, the sets of possible locations of sensors are signalled in cyan and the selected locations of sensors are signalled in red.

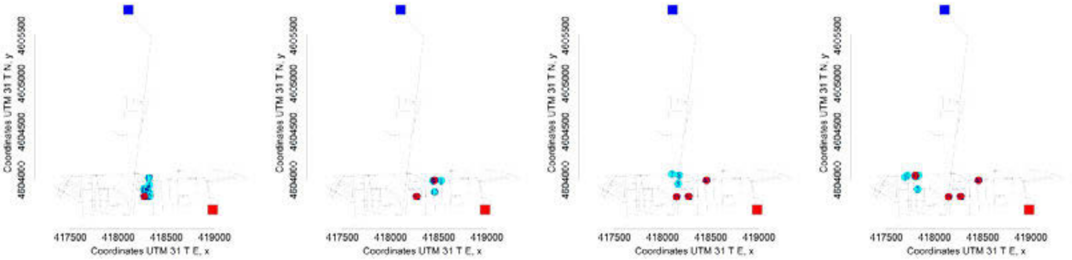


a) Possible locations of the sensors

b) Possible locations of the sensors once 1st sensor is selected

c) Possible locations of the sensors once 2nd sensor is selected

d) Possible locations of the sensors once 3rd sensor is selected



e) 1st sensors set and 1st sensor selected

f) 2nd sensors set and 2nd sensor selected

g) 3rd sensors set and 3rd sensor selected

h) 4th sensors set and 4th sensor selected

Figure 3. Process of selection of the locations of the sensors.

4.3 Optimal Number of Sensors

It has been stated that to determine the optimal number of sensors it is necessary to define an indicator $\hat{\Delta}\mathbf{p}_{meas}$ that quantifies the relevance of the information recorded by a pressure sensor layout (see equation (4)). An easily interpretable indicator can be based on the effect that a demand variation have on pressure measurements. The indicator $\hat{\Delta}\mathbf{p}_{meas}$ is defined as the maximum variation of pressure that can be measured in the network given a fixed demand variation Δd occurring at each node separately.

Once the indicator $\hat{\Delta}\mathbf{p}_{meas}$ is defined, it is necessary to define which must be the magnitude of these demand variation Δd and which is an acceptable value for the pressure variations measured. To determine these magnitudes, an optimal scenario is defined in which the WDN is fully instrumented with pressure sensors located at all the nodes of the WDN model. In such scenario Δd is defined as the minimum variation of demand that, occurring at any node of the WDN, would always be detectable. Detectable means that the produced pressure variation at any of the measurement locations is larger than the accuracy of the pressure sensors Δp_{min}

Pseudocode shown in Table 1 presents the procedure applied to determine Δd, where $\hat{\mathbf{p}}_{\mathbf{0}} \in \mathbb{R}^{N_n}$ is the nodal pressure vector at $\mathbf{d_0} \in \mathbb{R}^{N_n}$ reference working point; and $\hat{\Delta}\mathbf{p}_{meas}{}^{*} \in \mathbb{R}^{N_n}$ is the vector containing N_n elements corresponding to the maximum variation of pressure measured when varying the demand of each node separately at the optimal scenario of a fully instrumentated WDN $(N_s = N_n)$.

In line 3 of the pseudocode in Table 1, the first simulation at the $\mathbf{d_0}$ reference working point is performed and the nodal pressures obtained are stored in vector $\hat{\mathbf{p}}_{\mathbf{0}}$ (See Equation (1)). In lines 4 to 13 sets of N_n simulations are performed by increasing the demand of each node separately by Δd. In line 9 the nodal pressures obtained from these simulations are stored in vector $\hat{\mathbf{p}}_{\mathbf{var}}$. In line 10 the maximum variations of pressure produced at each of the N_n simulations varying the consumption of each node separately are stored in vector $\hat{\Delta}\mathbf{p}_{meas}{}^{*}$ at index of the node whose demand is varied. The while loop applied in line 4 sets a terminating condition which consist in breaking the loop if each of the maximum variation of pressure caused by each of the variation of

cc BY NC SA

demand is larger than the sensor accuracy Δp_{min} for all the N_n simulations. Therefore, considering the sensor accuracy, a variation of demand of magnitude Δd applied at any node would be detected in the hypothetical scenario of having sensors installed at all the nodes of the WDN model. If it is not the case, the magnitude of the variation of the demand Δd is increased and more set of N_n simulations are performed until the terminating condition is fulfilled.

The functions $min()$ and $max()$ in tables 1 and 2 return the minimum and maximum element of a given vector respectively. The function $abs()$ in tables 1 and 2 returns a vector containing the absolute value of each element of a given vector and the operator $\boldsymbol{vector}(j)$ is accessing to the j^{th} element of the given vector.

Table 1. Pseudocode to determine the magnitude of the variation of the demand and the optimal information of instrumentation indicator [Caption, Cambria, 10pt, Italic, centred]

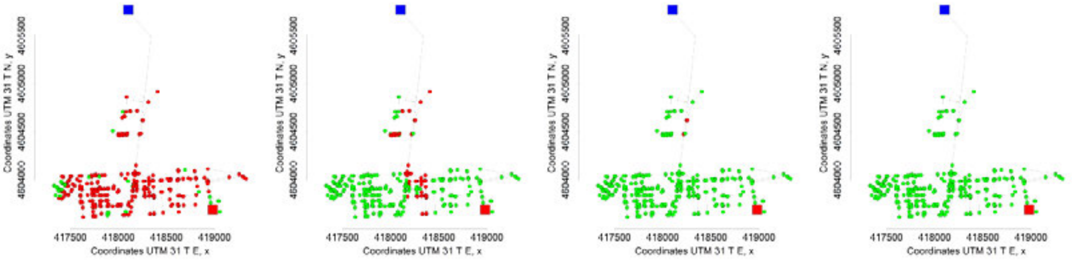
Pseudocode to determine the magnitude of the variation of the demand and the optimal information of instrumentation indicator

Require: WDN model, $\mathbf{d_0}$, Δp_{min}

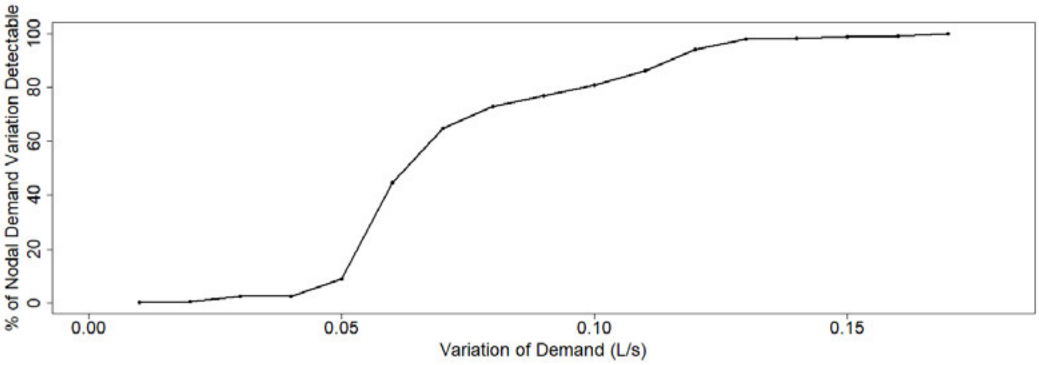
Return: Δd, $\Delta\hat{\mathbf{p}}_{meas}^{*}$

```
$i = 0$
$\Delta\hat{\mathbf{p}}_{meas}^{*} = \mathbf{0}$
$\hat{\mathbf{p}}_{\mathbf{0}} = Simulate\ Model(\mathbf{d_0})$
While $min(\Delta\hat{\mathbf{p}}_{meas}^{*}) \leq \Delta p_{min}$:
    $i = i + 1$
    $\Delta d = i \cdot \delta$
    For $j = 1 : N_n$ do
        $\mathbf{d_{var}} = \mathbf{d_0}$
        $\mathbf{d_{var}}(j) = \mathbf{d_0}(j) + \Delta d$
        $\hat{\mathbf{p}}_{\mathbf{var}} = Simulate\ Model(\mathbf{d_{var}})$
        $\Delta\hat{\mathbf{p}}_{meas}^{*}(j) = max(abs(\mathbf{p_0} - \hat{\mathbf{p}}_{\mathbf{var}}))$
    End for
End while
Return: $\Delta d$, $\Delta\hat{\mathbf{p}}_{meas}^{*}$
```

The pressure sensors that are installed by applying the sampling design approach presented in this paper have an accuracy of ± 1 cmH2O. Thus, the sensor accuracy Δp_{min} is set to 1 cmH2O. The procedure presented in pseudocode of table 1 to the studied WDN gives a magnitude of the variation of the nodal demand of $\Delta d = 0{,}17$ L/s. Figure 4 illustrates how increasing the magnitude of Δd, the number of nodes whose demand variation is detectable increases for the studied WDN. The nodes signalled by green represent the nodes whose demand variation causes a detectable pressure variation. The nodes signalled by red represent the nodes whose demand variation causes too small pressure variations for being detectable.



a) $\Delta d = 0.05$ L/s b) $\Delta d = 0.1$ L/s c) $\Delta d = 0.15$ L/s d) $\Delta d = 0.17$ L/s



e) Percentage of nodal demand variation detectable depending on the magnitude of Δd

Figure 4. Nodal demand variation detectable depending on the magnitude of Δd.

The second output of the pseudocode of table 1 is the nodal vector $\mathbf{\Delta\hat{p}_{meas}}^*$ corresponding to the maximum pressure variation caused by each nodal demand variation separately. Figure 5 is the plot of the $\mathbf{\Delta\hat{p}_{meas}}^*$. It associates each component of $\mathbf{\Delta\hat{p}_{meas}}^*$ to the node whose nodal demand variation is causing that pressure variation. This figure displays the indicator that quantifies the relevance of the information recorded at optimal situation in which the WDN is fully instrumented with pressure sensors at each node ($N_s = N_n$). The indicator computed for the optimal scenario is useful to appreciate which areas of the network have room for improvement when comparing with the indicator computed for feasible layouts of sensors.

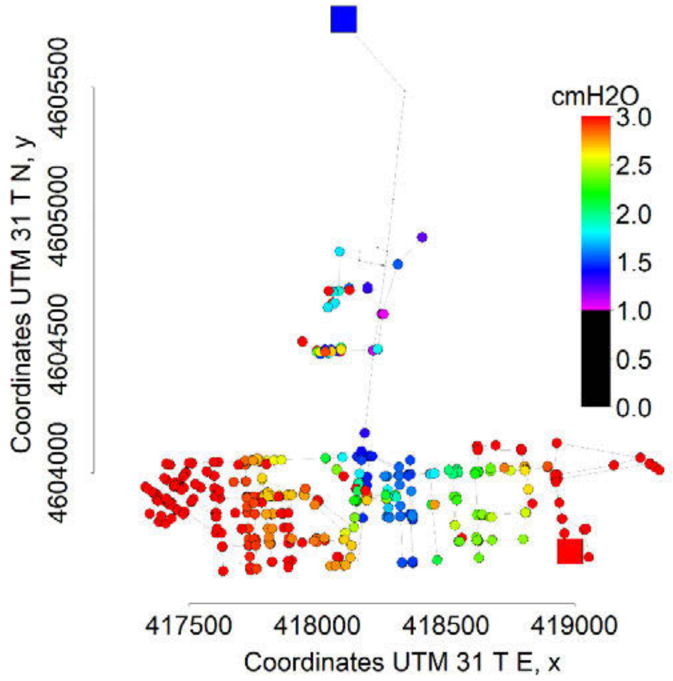


Figure 5. Maximum pressure variation measurable $\hat{\boldsymbol{\Delta}}\mathbf{p}_{meas}^{*}$ *for a fully instrumented WDN* ($N_s = N_n$).

To generate the indicator $\hat{\boldsymbol{\Delta}}\mathbf{p}_{meas}$, only the pressure variations produced at the locations of the sensors must be considered.

The process to generate the indicator $\hat{\boldsymbol{\Delta}}\mathbf{p}_{meas}$ consists in performing $N_n + 1$ simulations,. The first simulation corresponds to the working point $\mathbf{d_0}$ with respect to which the variation of pressures measured by the sensors will be analysed. The following N_n simulations represent a variation of demand applied on each node separately. Thus, comparing the pressures obtained from the N_n simulations with respect to the pressures obtained from the first simulation considering that only the pressures at the locations of the sensors will be known, the maximum variations of pressures that could be measured can be obtained.

Pseudocode presented in Table 2 corresponds to the process generating the indicator $\hat{\boldsymbol{\Delta}}\mathbf{p}_{meas}$ as the maximum pressure variations are searched only among the locations of the sensors. Notice that the execution of the first and the last N_n simulations of the pseudocode presented in table 1 are equivalent to the $N_n + 1$ simulations executed in pseudocode presented in table 2 if the magnitude of the variation of the nodal demand Δd is fixed according to the output of pseudocode presented in table 1.

BY NC SA

Table 2. Pseudocode to determine the maximum pressure variation detectable for a given layout of sensors caused by each nodal demand variation separately

$Information\ of\ Instrumentation\ Indicator$

Require: WDN model, $\mathbf{d_0}$, Δd, $\mathbf{Id}^*$

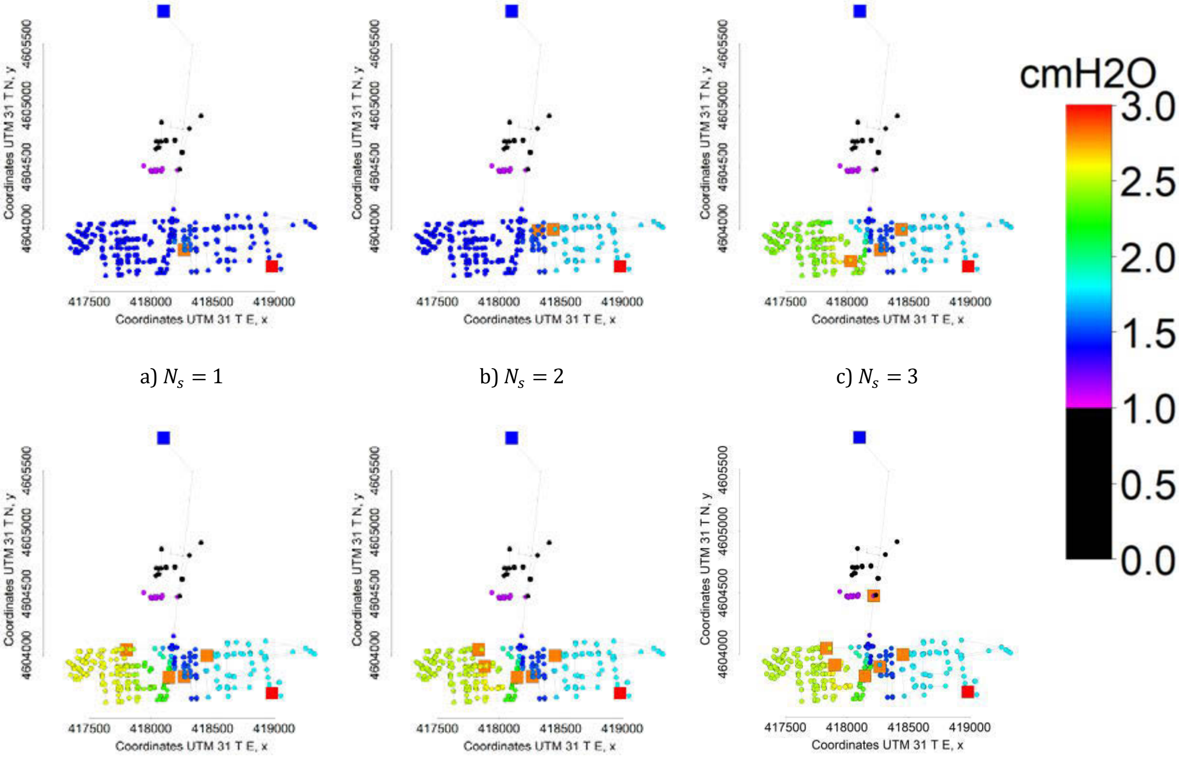
Return: $\Delta\hat{\mathbf{p}}_{meas}$

```
$\Delta\hat{\mathbf{p}}_{meas} = \mathbf{0}$;
$\hat{\mathbf{p}}_0 = Simulate\ Model(\mathbf{d_0})$;
For $j = 1{:}N_n$ do
    $\mathbf{d}_{var} = \mathbf{d_0}$
    $\mathbf{d}_{var}(j) = \mathbf{d_0}(j) + \Delta d$
    $\hat{\mathbf{p}}_{var} = Simulate\ Model(\mathbf{d}_{var})$;
    $\Delta\hat{\mathbf{p}}_{meas}(j) = max(abs(\hat{\mathbf{p}}_0(\mathbf{Id}^*) - \hat{\mathbf{p}}_{var}(\mathbf{Id}^*)))$;
End for
Return: $\Delta\hat{\mathbf{p}}_{meas}$
```

It is expected that for a reduced number of sensors distributed on the WDN increasing the number of sensors will significantly increase the magnitude of the maximum variation of pressures measured by the sensors. However, it is also expected to reach a certain number of sensors from which increasing the number of sensors distributed on the WDN does not significantly increase the magnitude of the maximum variation of pressures measured by the sensors.

Finally, to determine the optimal layout of pressure sensors both processes, the sensor placement and optimal number of sensors definition, must be applied. Initially, the first methodology locating the sensors is applied for different number of sensors to obtain several possible layouts of sensors. Thus, the second methodology quantifying the optimality of a layout of sensors is applied to decide which of the previously obtained layout is the most appropriate for the specific WDN being studied. These sampling design methodologies can adapt according to a budget, as the maximum number of sensors distributed on the network can be limited.

Figure 6 presents the results of applying the entire sampling design methodology. The sensor placement is applied for several predefined N_s number of sensors from one to six according to the budget limit. The locations of the sensors are signalled by orange squares. Thus, the optimal number of sensor process is applied to generate the indicator $\Delta\hat{\mathbf{p}}_{meas}$ given the six different layouts of sensors obtained from the sensor placement process.


Coordinates UTM 31 T N, y
4604000
4604500
4605000
4605500
417500
418000
418500
419000
Coordinates UTM 31 T E, x
cmH2O
3.0
2.5
2.0
1.5
1.0
0.5
0.0

a) $N_s = 1$ b) $N_s = 2$ c) $N_s = 3$

a) $N_s = 4$ b) $N_s = 5$ c) $N_s = 6$

Figure 6. Maximum pressure variation measurable $\boldsymbol{\Delta p_{meas}}$ *for feasible number of sensors* $N_s \leq N_{sMax}$.

Figure 7 displays the mean of all the maximum variation of pressures measurable by the sensors given the six different layouts of sensors. It can be appreciated that distributing five or six pressure sensors with the sensor placement procedure applied does not significantly increase the magnitude of the maximum variation of pressures that would be measured by the sensors with respect to the layout using four sensors. For this reason, the optimal number of sensors to be installed is set to four.

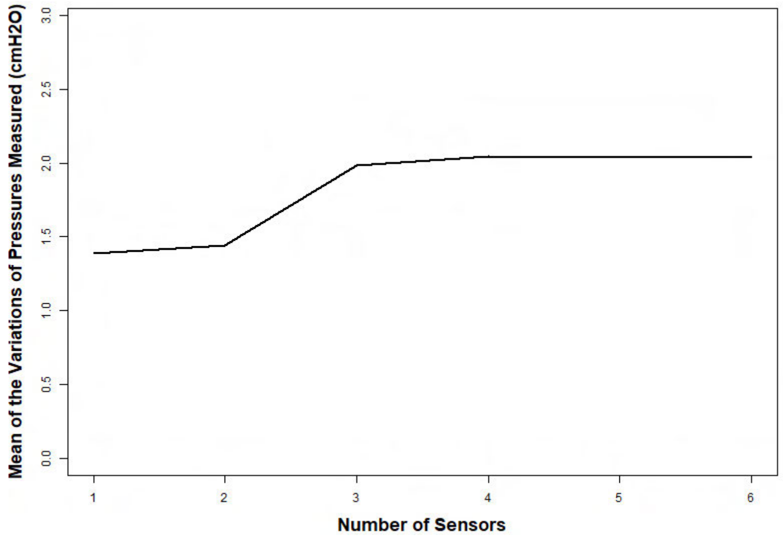


Figure 7. Mean of maximum pressure variation measured $\boldsymbol{\Delta p_{meas}}$ *for feasible number of sensors* $N_s \leq N_{sMax}$.

BY NC SA

5 CONCLUSIONS

This work has presented a complete model-based methodology to determine an optimal pressure sensor placement layout and the appropriate number of pressure sensors to install on a WDN.The methodology uses a sensor placement methodology from the literature that finds the most sensible locations with respect to demand groups. Additionally, to determine the optimal number of sensors a new methodology that can be interpretable by the company has been developed with the aim to overcome the gap between research and practice. The sampling design methodology is able to adapt to a budget as the maximum number of sensors distributed on the network can be limited.

This methodology has been developed within an instrumentation project plan of TAIGUA. That has been successfully applied to several real WDN and the company is installing pressure sensors according to the layouts of sensors obtained. Results illustrating this application have been displayed.

With the information recorded by the new sensors it is intended to microcalibrate the demand of the WDN models and implement a leak-detection and localization approach that can be coupled with the demand calibration methodology. Additionally, the calibrated WDN model and a precise demand model can be used to enhance the energetic and economic efficiency of the operational control of the WDN.

6 REFERENCES

[1] R. Wéber and C. Hős, "Efficient Technique for Pipe Roughness Calibration and Sensor Placement for Water Distribution Systems," J. Water Resour. Plan. Manag., vol. 146, no. 1, p. 04019070, Jan. 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001150.

[2] Perez, Ramon; Sanz, Gerard; Cuguero, Josep; and Cuguero, Miquel Angel, "Optimal Placement Of Metering Devices For Multiple Purposes", CUNY Academic Works, 2014.

[3] L. E. Ormsbee and S. Lingireddy, "Calibrating hydraulic network models," J. Am. Water Works Assoc., vol. 89, no. 2, pp. 42–50, Feb. 1997, doi: 10.1002/j.1551-8833.1997.tb08177.x.

[4] Ramon Pérez R., Cugueró J., Romera J., Van Eeckhout J., Cabot J., Franch O., "Water Distribution Network Model Calibration and Continuous Maintenance : Terrassa , A Real Application," Modern Environmental Science and Engineering., vol. 7, no. 9, pp. 891-897, 2021, doi: 10.15341/mese(2333-2581)/09.07.2021/003.

[5] G. Sanz and R. Pérez, "Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition," J. Water Resour. Plan. Manag., vol. 141, no. 10, p. 04015020, Oct. 2015, doi: 10.1061/(ASCE)WR.1943-5452.0000535.

[6] T. Walski, P. Sage, and Z. Wu, "What Does it Take to Make Automated Calibration Find Closed Valves and Leaks?," in World Environmental and Water Resources Congress 2014: Water Without Borders - Proceedings of the 2014 World Environmental and Water Resources Congress, 2014, pp. 555–565, doi: 10.1061/9780784413548.059.

[7] G. Sanz and R. Pérez, "Comparison of demand calibration in water distribution networks using pressure and flow sensors," Procedia Eng., vol. 119, no. 1, pp. 771–780, 2015, doi: 10.1016/j.proeng.2015.08.933.

[8] "ADT1 remote transmission units with data loggers from Keller manufacturer." https://keller-druck.com/en/products/wireless-solutions/remote-transmission-units/adt1-box.

[9] G. Yu and R. S. Powell, "Optimal design of meter placement in water distribution systems," Int. J. Syst. Sci., vol. 25, no. 12, pp. 2155–2166, 1994, doi: 10.1080/00207729408949342.

[10] R. W. Meier and B. D. Barkdoll, "Sampling Design for Network Model Calibration Using Genetic Algorithms," J. Water Resour. Plan. Manag., vol. 126, no. 4, pp. 245–250, Jul. 2000, doi: 10.1061/(ASCE)0733-9496(2000)126:4(245).

[11] W. B. F. De Schaetzen, G. A. Walters, and D. A. Savic, "Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms," Urban Water, vol. 2, no. 2, pp. 141–152, 2000, doi: 10.1016/s1462-0758(00)00052-2.

[12] Z. S. Kapelan, D. A. Savic, and G. A. Walters, “Multiobjective Sampling Design for Water Distribution Model Calibration,” J. Water Resour. Plan. Manag., vol. 129, no. 6, pp. 466–479, Nov. 2003, doi: 10.1061/(ASCE)0733-9496(2003)129:6(466).

[13] J. P. Vtkovský, J. A. Liggett, A. R. Simpson, and M. F. Lambert, “Optimal Measurement Site Locations for Inverse Transient Analysis in Pipe Networks,” J. Water Resour. Plan. Manag., vol. 129, no. 6, pp. 480–492, Nov. 2003, doi: 10.1061/(ASCE)0733-9496(2003)129:6(480).

[14] Z. S. Kapelan, D. A. Savic, and G. A. Walters, “Optimal Sampling Design Methodologies for Water Distribution Model Calibration,” J. Hydraul. Eng., vol. 131, no. 3, 2005, doi: 10.1061/(asce)0733-9429(2005)131:3(190).

[15] K. Behzadian, Z. Kapelan, D. Savic, and A. Ardeshir, “Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks,” Environ. Model. Softw., vol. 24, no. 4, pp. 530–541, Apr. 2009, doi: 10.1016/j.envsoft.2008.09.013.

[16] D. Kang and K. Lansey, “Optimal Meter Placement for Water Distribution System State Estimation,” J. Water Resour. Plan. Manag., vol. 136, no. 3, pp. 337–347, 2010, doi: 10.1061/(asce)wr.1943-5452.0000037.

[17] D. A. Savic, Z. S. Kapelan, and P. M. R. Jonkergouw, “Quo vadis water distribution model calibration?,” Urban Water J., vol. 6, no. 1, pp. 3–22, 2009, doi: 10.1080/15730620802613380.

[18] G. Sanz, R. Pérez, Z. Kapelan, and D. Savic, “Leak Detection and Localization through Demand Components Calibration,” J. Water Resour. Plan. Manag., vol. 142, no. 2, p. 04015057, 2016, doi: 10.1061/(asce)wr.1943-5452.0000592.

[19] G. Sanz and R. Pérez, “Demand Pattern Calibration in Water Distribution Networks,” Procedia Eng., vol. 70, pp. 1495–1504, 2014, doi: 10.1016/j.proeng.2014.02.164.

[20] T. M. Walski, “Model calibration data: The good, the bad, and the useless,” J. / Am. Water Work. Assoc., vol. 92, no. 1, pp. 94–99, 2000, doi: 10.1002/j.1551-8833.2000.tb08791.x.

[21] D. Kang and K. Lansey, “Real-Time Demand Estimation and Confidence Limit Analysis for Water Distribution Systems,” J. Hydraul. Eng., vol. 135, no. 10, pp. 825–837, Oct. 2009, doi: 10.1061/(ASCE)HY.1943-7900.0000086.

[22] D. Kang and K. Lansey, “Demand and Roughness Estimation in Water Distribution Systems,” J. Water Resour. Plan. Manag., vol. 137, no. 1, pp. 20–30, Jan. 2011, doi: 10.1061/(ASCE)WR.1943-5452.0000086.

[23] A. Ostfeld et al., “Battle of the Water Calibration Networks,” J. Water Resour. Plan. Manag., vol. 138, no. 5, pp. 523–532, Sep. 2012, doi: 10.1061/(ASCE)WR.1943-5452.0000191.

[24] K. N. Mallick, I. Ahmed, K. S. Tickle, and K. E. Lansey, “Determining Pipe Groupings for Water Distribution Networks,” J. Water Resour. Plan. Manag., vol. 128, no. 2, pp. 130–139, Mar. 2002, doi: 10.1061/(ASCE)0733-9496(2002)128:2(130).

[25] R. A. Wiggins, “The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure,” Rev. Geophys., vol. 10, no. 1, pp. 251–285, 1972, doi: 10.1029/RG010i001p00251.

[26] R. C. Aster, B. Borchers, and C. H. Thurber, Parameter Estimation and Inverse Problems. Elsevier, 2019.

[27] J. A. Liggett and L. Chen, “Inverse Transient Analysis in Pipe Networks,” J. Hydraul. Eng., vol. 120, no. 8, pp. 934–955, Aug. 1994, doi: 10.1061/(ASCE)0733-9429(1994)120:8(934).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14734

DISINFECTION BY PRODUCTS ESTIMATION IN A WATER DISTRIBUTION NETWORK

Laura Vinardell[1], Irene Jubany[2] and Ramon Pérez[3]

[1,2] Sustainability Area, Eurecat, Centre Tecnològic de Catalunya, Plaça de la Ciència, 2, 08243 Manresa, Spain;
[3]acs2ac, Universitat Politècnica de Catalunya, Rambla Sant Nebridi 10, 0822 Terrassa, Spain.

[1] laura.vinardell@eurecat.org, [2] irene.jubany@eurecat.org, [3] ramon.perez@upc.edu

Abstract

Even though disinfection is necessary to ensure water safety for human consumption, some disinfectants produce disinfection by-products (DBPs) that may be dangerous for human health. Current European legislation obligates water distributors to limit some DBPs concentration to final consumers. Then, water companies must control these compounds and are obligated to periodically monitor their network. DBPs modeling can be very useful for estimating online DBPs concentration throughout the network, increasing DBPs control and knowledge, but avoiding DBPs analytics time and resources consumption [1].

Trihalomethanes (THM), the first DBPs discovered, have long been the most studied and modeled. Previous studies have mostly used linear relations between variables and THM concentration, but also computational modelling, mechanistic and data driven models [2, 3]. Even though, there are still challenges to beat: most studies use a small database and laboratory-scale for model building, forgetting the impact of network pipelines and season. In addition, significant variables for DBPs' formation such as retention time are most of the time neglected due to its difficulty to measure. Finally, THMs are not the only DBPs generated from disinfection or even the most toxic: other DBPs must be studied, and their formation pathways along the network investigated.

In this study, data from a full-scale distribution network was used: online sensors and sampling campaigns. To include hydraulic conditions as retention time, EPANET software and R programming are used to simulate the network. Different models, mechanistic and data driven, have been used to estimate the chlorine decay and DBP formation within the network. Results of the calibration and validation of these models and the conclusions obtained are presented.

Keywords
Disinfection by-products, modelling, distribution water networks, hydraulic simulation, EPANET, HAA, THM.

1 INTRODUCTION

Disinfection by-products are an increasing concern for protecting human health. These compounds are formed due to a disinfection treatment: the disinfectant reacts with compounds present in water as natural organic matter and anthropogenic contaminants, forming halogenated molecules. This field was discovered in 1974 when chloroform and other trihalomethanes were found in drinking water supplies in the USA. Since then, more than 700 DBPs have been identified from all disinfection processes (chlorine dioxide, ozonation, chloramination, and chlorination) [4, 5]. Some of these compounds are carcinogenic and/or toxic at certain concentrations [6].

Since DBPs' discovery, some countries have included in their drinking water legislation a threshold for these compounds. Current European regulation about water intended for human consumption is (EU) 2020/2184 [7]. It considers microbiological, chemical, and physical parameters. Regarding DBPs, it includes TTHM (Total THM: chloroform, bromodichloromethane, dibromochloromethane, and bromoform), five HAA (HAA_5: monochloroacetic acid, dichloroacetic acid, trichloroacetic acid, monobromoacetic acid, and dibromoacetic acid), chlorate, chlorite, and bromate (Table 1:). This is relatively new legislation since it was implemented in December 2020. The previous one [8] is from 2003 and only included TTHM and bromate.

Table 1: European current DBPs legal thresholds for water for human consumption

	TTHM	HAA_5	Bromate	Chlorite	Chlorate
UE Legal threshold [µg/L]	100	60	10	250	250

To accomplish the European legislation, water distribution companies monitor water quality at different network locations. Water network operators are not forced to monitor their water quality online, and only need to analyze a few samples distributed along the network. For DBPs analysis it is required specific and expensive instrumentation: they are generally quantified with gas or liquid chromatography at laboratories, a time-consuming and expensive analytical technique.

Due to the high cost and maintenance of online sensors, only a few distribution networks have them installed. The alternative for DBPs online control is the use of mathematical models based on easily measurable data. Unfortunately, water quality modelling has some disadvantages: it is water source and treatment applied depending (highly specific), needs a large amount of data, and can require using complex mathematical tools. For all these reasons, DBPs online monitoring is a research area in development and needs to be deeply investigated.

This study aims to use a model from literature able to predict DBPs in a full-scale water distribution network, using water quality online sensors data, and hydraulic simulations from the pilot site. DBPs formation does not only depend on its precursors (organic matter, some inorganic compounds and type and dose of disinfectant), but also on the network's hydraulics (residence time, velocity) and environmental characteristics (temperature, pH, etc.). For that purpose, sampling and data from a full-scale network were used, and EPANET software tool for hydraulic simulations.

2 MATERIALS AND METHODS

The study site is in Tarragona (Catalonia, Spain). *Consorci d'Aigües de Tarragona* (CAT) is the entity managing water distribution in this province to the municipalities. Water comes from Ebro's River and is dispensed to the whole province after ozonation treatment at drinking water treatment plant (DWTP). Chlorination is applied in distribution network in different boosting stations.

Tarragona's network includes 1 DWTP, 23 pump stations (PS), 141 tanks, and online sensors in several locations (free chlorine, flowmeters, and other water quality parameters).

2.1 Sampling campaign definition

To decide the water network section for the study, historic data of the whole drinking water distribution network was analyzed. Data comprised TTHM, UV254, conductivity and total organic carbon (TOC) measurements from 2017 to 2019. In addition, distance and number of rechlorinations was also included in the study. The statistical analysis comprised correlations between the different variables (TTHMs and water quality) and principal components analysis (PCA) to identify the more significant variables that explain TTHM values dispersion.

This analysis was necessary to decide CAT network's section where the sampling campaign should be performed, in other to include full-scale data to the research. Data available from the sampling campaign is summarized in Table 2: study parameters.

Table 2: study parameters

Field measurements	**Laboratory measurements**
pH	THM_4 [µg/L]
Conductivity [µS/cm]	HAA_5 [µg/L]
Temperature [ºC]	Chlorate [µg/L]
Free Chlorine [mg/L]	Turbidity [NTU]
	TOC [mg/L]

Parameters measured in the field were measured using field sensors. THM were measured by HS / GC - µECD, HAA and chlorate with LC-MS/MS, turbidity with a turbidimeter and TOC with high temperature oxidation combustion equipment.

2.2 Water network hydraulic simulation

EPANET Software

EPANET is a free software for hydraulic network simulations, designed by the United States Environmental Protection Agency [9]. This software makes it possible to simulate hydraulic and quality behavior in a water distribution network: residence time, free chlorine concentration, storage tank pressure, and flow, among other parameters.

Figura 1 shows the complete hydraulic model of the study network. For this study, work was carried out on the A to C section corresponding to the water treatment plant effluent (A) and a pumping station (C) located 60 km away. Then the network was simplified preserving only the main pipeline. Pump curves, valve status, and pipeline diameters were revised and/or modified to balance flowmeter sensor measurements.

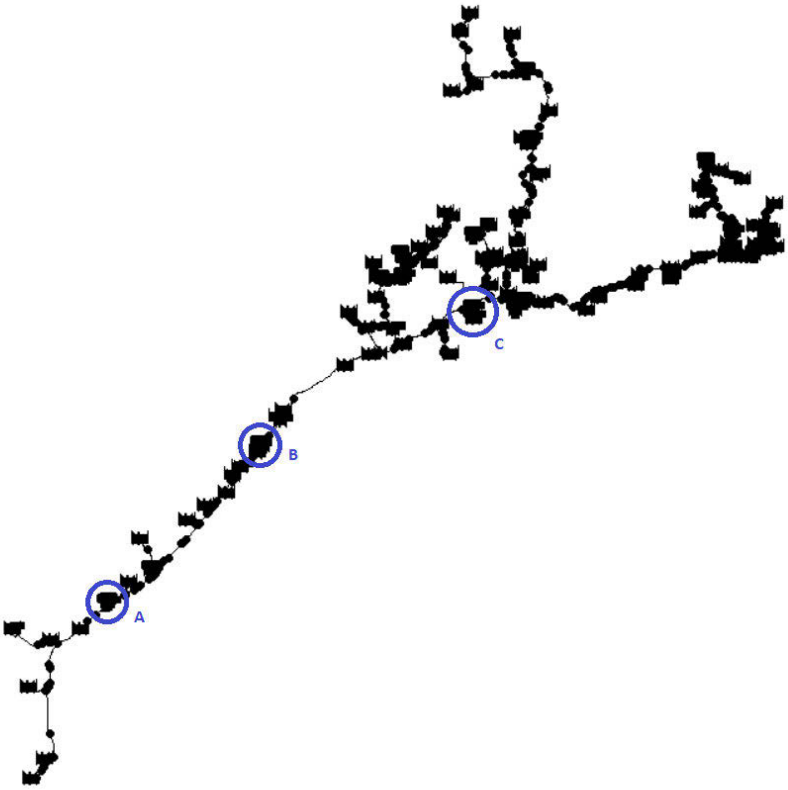


Figure 1: Case-study hydraulic model

The reason for this section's choice for hydraulic modeling was its simplicity for free chlorine modeling: there is only the chlorination at A (the following disinfection process is at C). Modeling this simple model is the first step in this research context.

For EPANET hydraulic simulation, flowmeter and free chlorine sensors data from October 2021 were used. The section chosen had different regimes depending on water demand: water can pass through intermediate tank storage (B in Figure 1: Case-study hydraulic model) or can advance directly from A to C. Due to this complexity, clustering analysis of flowmeters data was done to discretize the different regimes at tank storage B, using PAM clustering function in RStudio language.

PAM clustering [10] is an algorithm that organizes data around medoids (cluster centers), to minimize the dissimilarities between data in the same cluster and maximize between different clusters. *NBClust* R package was used for the optimal number of clusters decision.

Hydraulic and quality simulations were done for 72h, in the regime where water by-pass B tank:

- First 24h for hydraulic model stabilization
- 24h for model identification
- Last 24h for model validation

Chlorine decay model

Free chlorine behavior in water distribution networks is commonly described as first-order kinetics [11]:

$$Cl = Cl_0 e^{-kt} \tag{1}$$

Where *Cl* is the final chlorine concentration, Cl_0 is the initial chlorine concentration, *k* is the decay constant and *t* the residence time. The same eq. (1) can be written as:

$$\ln\left(\frac{Cl_0}{Cl}\right) = kt \tag{2}$$

Using eq. (2), *k* was estimated by linear regression method using residence time (*t*) and free chlorine concentrations at A (initial concentration Cl_0) and C (final concentration *Cl*). *k* constant was estimated using identification data and then quality simulation was performed for 48h (identification and validation data)

The mean squared error (MSE) was calculated to compare predicted values and measured values (from free chlorine sensors). MSE must not worsen significantly with the validation data compared to the identification data to rule out overfitting.

TTHMs prediction model

Clark et al. [12] used a mathematical model for TTHM formation prediction using kinetics relationships:

$$TTHM = T\left\{C_{A0} - \left[\frac{C_{A0}(1-R)}{1-Re^{-ut}}\right]\right\} + M \tag{3}$$

Where *M* is the initial TTHMs, *t* is the residence time, and *T, R* and *u* are parameters depending on kinetic relations between free chlorine, chlorine demand matter and DBPs generated. This model does not include rechlorinations. Therefore, TTHM prediction could not be extended from C booster station.

In the same study, Clark et al. used different water samples with different *T, R, u* and *M* values. The samples were defined by location and quality parameters (pH, Temperature, TOC and residual free chlorine). For this research, the values *T, R* and *u* from the sample with quality parameters closer to the study case (data from A) were used (Table 3).

Table 3: water sample quality characteristics and model parameters of [12] reference study

	Quality conditions				**Model parameters**			
	pH	**T [°C]**	**TOC [mg/L]**	**Free Cl [ppm]**	**T**	**R**	**u**	**M**
Reference sample from [12]	8.15	17.9	1.87	1.73	41.29	0.25	0.44	10.00
A	7.70	24.5	3.00	1.22	-	-	-	-

3 RESULTS

3.1 Sampling campaign definition

The statistical analysis showed that the highest correlations were found between TTHMs and the number of rechlorinations ("nº rechl.") (high correlations) and with conductivity (moderated correlation). Other moderated correlations were Temperature – TOC and year – TOC (Table 4**¡Error! No se encuentra el origen de la referencia.**).

Table 4: variables correlation

	Year	Temperature [ºC]	TTHM [µg/L]	Conductivity [µS/Cm]	TOC [mg/L]	UV$_{254}$	nº rechl.
Temperature [ºC]	-0.05021						
TTHM [µg/L]	-0.16646	0.17517					
Conductivity [µS/Cm]	-0.23006	-0.25826	0.30543				
TOC [mg/L]	-0.41349	-0.36898	-0.10633	0.17625			
UV$_{254}$	0.16498	-0.01802	-0.05994	-0.20935	0.07519		
nº rechl.	-0.22228	0.08059	0.56943	0.03430	-0.07588	-0.11523	
Distance [km]	0.00158	0.04319	0.02287	-0.02804	0.01283	-0.06475	0.09644

PCA study also showed a significative relation between TTHMs and the number of rechlorinations (Figure 2: PCA results).

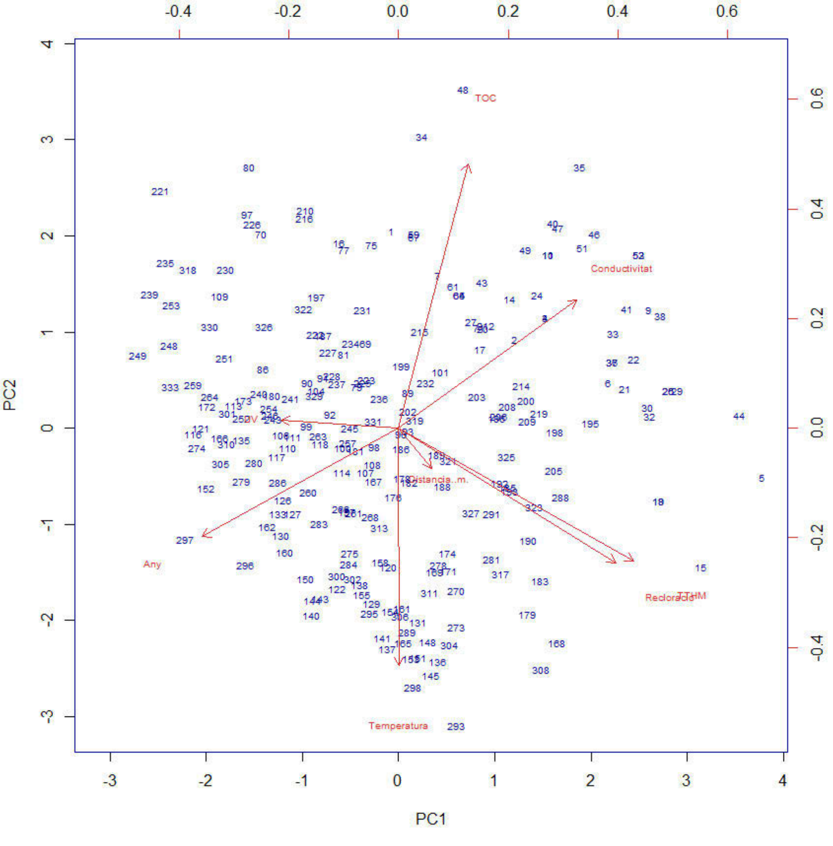


Figure 2: PCA results

These findings showed that rechlorinations may have a significant role in THMs formation in this network. For this reason, the main objective of the research is to study its effect on DBPs formation. Thus, the selected location for further studies (illustrated in red in Figure 3: study section) is composed of the principal pipeline and two final branches (with three and four boosters stations respectively). To include seasoning in the study for predicting DBPs' formation, the field study will last a whole year (October 2021 – October 2022), and a sampling campaign will be done biweekly. Sampling campaigns started in September 2021.

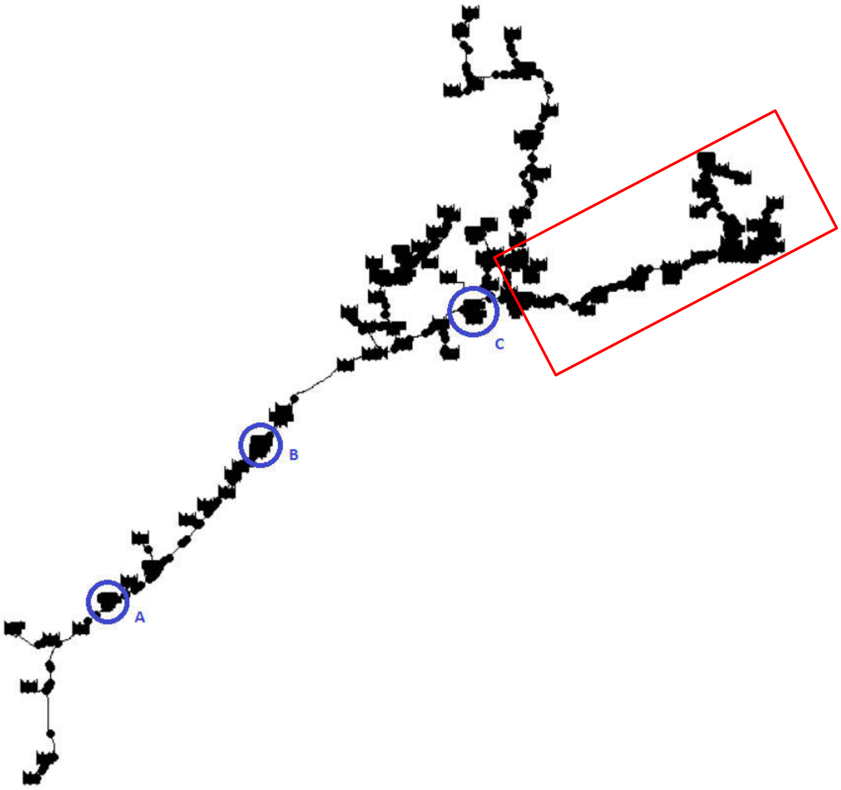


Figure 3: study section

The selected study section is posterior to the hydraulic modeled sector (A to C in Figure 3: study section**¡Error! No se encuentra el origen de la referencia.**). Further work will include extending the hydraulic simulation to the studied section using a kinetic predictive model for DBPs that includes rechlorinations.

Currently, TTHMs predicted at C storage tank intake (before rechlorination) with Clark et al. model [12] were compared to the full-scale TTHM measured from the sampling campaign at the sampling location closer to C.

3.2 Chlorine decay model

The free chlorine decay constant obtained by linear regression was k = 1.001 [h^{-1}] (Figure 4: linear regression model for identification data). Then, free chlorine decay model for this site is:

$$Cl = Cl_0 e^{-1.001t} \tag{4}$$

Free chlorine decay constant k was included as *Global kbulk* and *Global kwall* in EPANET. Then, quality simulation was performed for the same 72h as the hydraulic model. Results are presented at Figure 5: Comparison between measured and simulated free chlorine decay, where the difference between free chlorine simulation (using k decay obtained) and the measured from sensors can be seen.

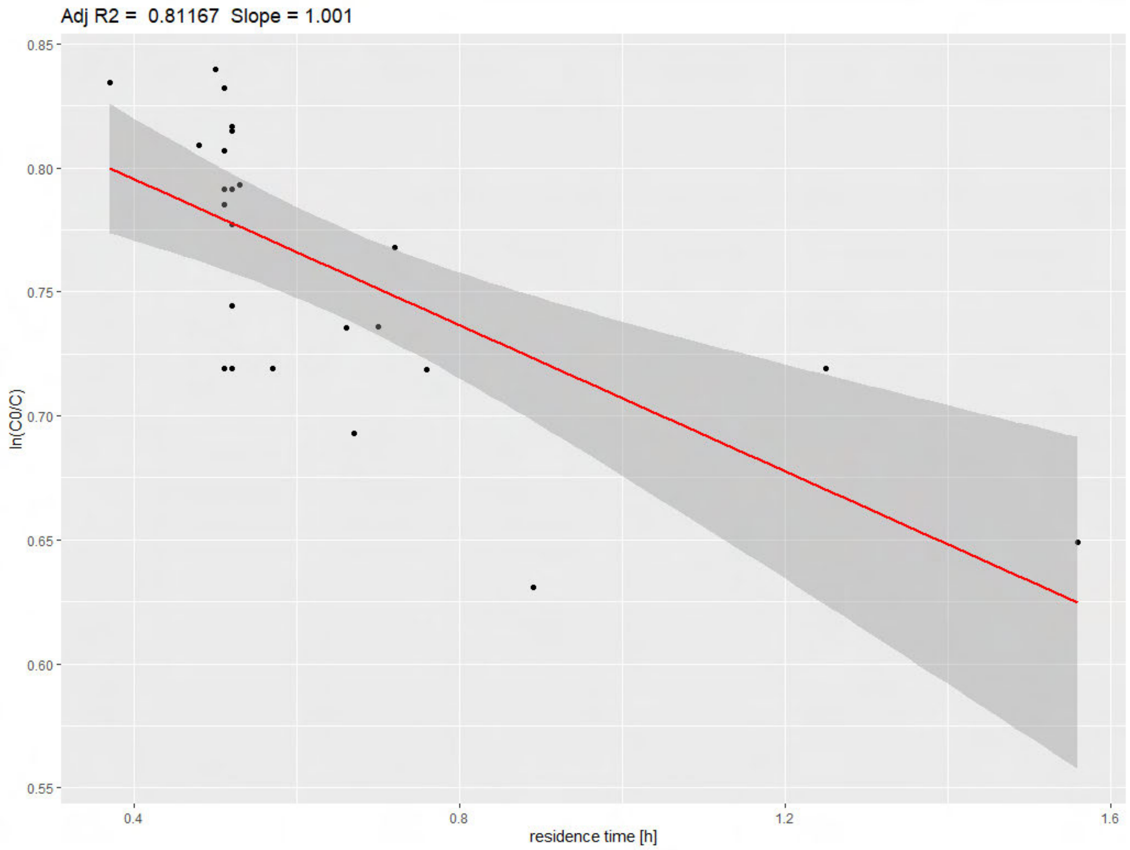


Figure 4: linear regression model for identification data

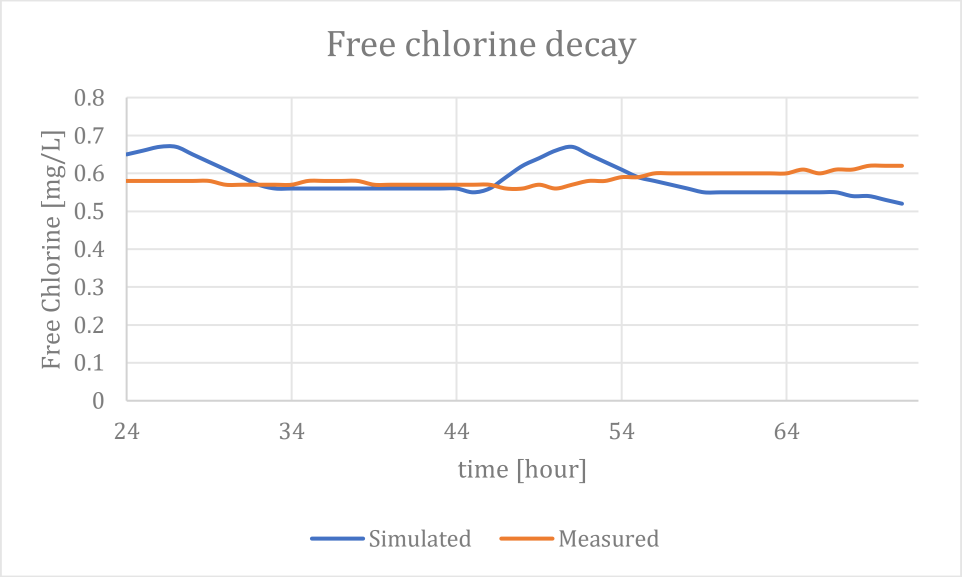


Figure 5: Comparison between measured and simulated free chlorine decay

MSE results (Table 5: MSE results) indicate the error made by the model for identification and validation data. Though both values are very small, the result for validation duplicates the one for identification. The overfitting of the model should be treated including more data when the different hydraulic regimes are reliably simulated.

Table 5: MSE results

Identification data MSE	Validation data MSE
0.001696	0.003888

3.3 TTHMS predictive model

Residence time obtained from the hydraulic simulation combined with available data of free chlorine available were used together with a TTHM prediction model from the literature [12]. The intention was to study if TTHM prediction model estimates a coherent value at C tank intake with TTHM analytically measured in the sampled location closer to C tank.

TTHM estimated by the Clark et al. model was calculated using eq. (3). T, R and u parameters used were obtained from the sample reference (Table 3); residence time t is from EPANET hydraulic model calculus, and free chlorine initial concentration C_{A0} and initial TTHMs concentration M from A (drinking water treatment plant output) data.

Table 6: TTHMs predicted and measured

TTHMs predicted using Clark et al. model at C intake	TTHMs analytically measured
11.35 µg/L	58.6 µg/L

Table 6 shows the TTHMs calculated by Clark et al. model [12] and the full-scale measurement. The difference between both values could be explained because there are two booster stations between both locations. Residence time at C tank output is close to 5h, seven times more than in C tank intake. Contact time increment at storage tanks between free chlorine and organic matter may explain TTHMs increase.

4 CONCLUSIONS

This study shows that kinetic models and software for hydraulic simulations are useful tools for water quality predictions. Drinking water network simulations allow to include complex hydraulic calculations (water demand, pressure, residence time, etc.), improving distribution network management.

In this study, residence time was identified using full-scale data and EPANET software simulation. Free chlorine constant decay was calculated and validated for a first-order decay model. Finally, using a kinetic model from the bibliography, TTHMs were estimated, obtaining a realistic value.

Even though, this study is the first step to obtain a more accurate and specific model for TTHM prediction. In further research, the model will use only site parameters calibrated for the site and must include rechlorination data. The hydraulic simulation must be extended to the study section, including all the different regimes and chlorination conditions. Finally, other DBPs must be also considered: HAA and oxyhalides because they are also dangerous for human health and need deeper research.

5 ACKNOWLEDGEMENTS

The authors want to acknowledge *Consorci d'Aigües de Tarragona* for providing hydraulic and water quality data and providing access to the drinking water network for sampling. Laura Vinardell is a fellow of Eurecat's "Vicente López" PhD grant program.

6 REFERENCES

[1] Ike IA, Karanfil T, Ray SK, Hur J (2020) A comprehensive review of mathematical models developed for the estimation of organic disinfection byproducts. Chemosphere 246:. https://doi.org/10.1016/j.chemosphere.2019.125797

[2] Clark RM, Sivaganesan M (1998) Predicting Chlorine Residuals and Formation of TTHMs in Drinking Water. J Environ Eng 124:1203–1210. https://doi.org/10.1061/(asce)0733-9372(1998)124:12(1203)

[3] Morrow CM, Minear RA (1987) Use of regression models to link raw water characteristics to trihalomethane concentrations in drinking water. Water Res 21:41–48. https://doi.org/10.1016/0043-1354(87)90097-2

[4] Brown D, Bridgeman J, West JR (2011) Predicting chlorine decay and THM formation in water supply systems. Rev. Environ. Sci. Biotechnol. 10:79–99

[5] John C. Crittenden, R.Rhodes Trussell, David W. Hand,Kerry J.Howe GT (2012) Water treatment principles and design

[6] Richardson SD, Plewa MJ, Wagner ED, et al (2007) Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. Mutat Res - Rev Mutat Res 636:178–242. https://doi.org/10.1016/j.mrrev.2007.09.001

[7] European Commission (2020) Directive (EU) 2020/2184 of the European Parlliament and of the Council of 16 December 2020 on the quality of water intended for human consumption. Off J Eur Union 2019:1–61

[8] Presidencia M De (2003) Real Decreto 140 / 2003 , de 7 de febrero , por el que se establecen los criterios sanitarios de la calidad del agua de consumo humano . 1–40

[9] Rossman L a. (2000) Epanet 2. 104

[10] Kaufman L, Rousseeuw PJ (1990) Partitioning Around Medoids (PAM). In: Finding Groups in Data: An Introduction to Cluster Analysis

[11] Johnson J. (1978) Measurement and persistence of chlorine residuals in natural waters. In: Environmental Impact of Water Chlorination. pp 43–71

[12] Clark RM (1998) Chlorine Demand and TTHM Formation Kinetics: A Second-Order Model. J Environ Eng 124:16–24. https://doi.org/10.1061/(asce)0733-9372(1998)124:1(16)

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14738

INFRASTRUCTURE BENCHMARKING FOR SEMI-REAL URBAN STORMWATER NETWORKS

Sina Hesarkazzazi[1], Mohsen Hajibabaei[2], Aun Dastgir[3] and Robert Sitzenfrei[4]

[1,2,3,4]Unit of Environmental Engineering, Faculty of Civil Engineering, Institute of Infrastructure, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria

[1] *sina.hesarkazzazi@uibk.ac.at,* [2] *mohsen.hajibabaei@uibk.ac.at,* [3] *aun.dastgir@uibk.ac.at,*
[4] *robert.sitzenfrei@uibk.ac.at*

Abstract

Attaining accurate information on the underground urban drainage/storm networks is a very difficult task, specifically in (developing) countries undergoing urban expansion. The problem of limited data availability gets even worse when considering that most case studies cannot made public due to security reasons. This, therefore, shed light on the lack of possessing enough case studies with different topological characteristics/attributes for the urban drainage research community to regularly explore and exploit their strategies/methodologies in a unified and integrated manner, allowing scholars to analyse and even compare their implications.

To cover this research gap, we developed a ready-to-use open-access database of a multitude of (semi) real urban stormwater case studies with different characteristics. Specifically, optimized centralized and decentralized stormwater networks are considered for creating the benchmarks while redundant flow pathways are introduced into these structures (by leveraging complex network analysis) to create more diverse characteristics. This allows that if for a real case study area, no further information is available (e.g., developing countries), with the proposed approach, a feasible model for principal evaluations can be created with very low efforts. We expect more scholars join this momentum in the future by adding more case studies with various attributes to form a comprehensive and rich database not only in terms of hydrodynamic models and case studies but also digital sensors and observation data (e.g., flow meter, rain gauge data, etc.).

Keywords
Urban drainage networks, benchmarks, optimization, redundancy, functional resilience.

1 INTRODUCTION

Urban drainage networks (UDNs), specifically urban stormwater networks (USNs), are essential assets within the cities, responsible for protecting citizens from potential urban flooding and ensuring preservation of the environment. Due to the complexities and detailed description of input information entailed in their modelling and simulation, various spectrums of approaches, methods, and technologies have been continuously deployed and tested by scientists to improve different aspects of UDNs performance. These include design [1], [2], resilience [3], [4], quality issues [5], rehabilitations [6], [7], etc. However, the implications from these studies are oftentimes case-specific, and therefore not allowing to draw generic conclusions. This is because these results were often derived from analysis of a single case study, which may not be a good representative of the underlying mechanism and the process. The reason is partly due to the fact that getting accurate and detailed structural/hydraulic information and model on the underground infrastructures. Specifically UDN/USN is a tedious task [8]–[10]. In addition, this data, including detailed layout or sewer/inlet locations, sewer dimensions, materials, sewer slopes, are often non-existent (due to aging, errors and missing information) or not publicly available by water utility companies due to security and publicity reasons. Thus, such a lack of universal and commonly accepted benchmark database hinders hydrological and hydraulic scholars for exploring, exploiting and comparing their developed methodologies/approaches to reach a unified consensus model.

Availability of open data and uniform benchmarks can be found more within the field of water distribution networks (WDNs) [11]–[16]. In contrast, such attempts have been rather limited in the domain of UDNs. For example, M Möderl et al., (2009) developed a MATLAB tool "Case Study Generator" for the generation of virtual combined sewer systems (CSSs) wherein the layout of the sewer systems is based on an adapted Galton-Watson branching process [17]. R Sitzenfrei et al., (2010) developed an innovative software called VIBe (i.e., Virtual infrastructure Benchmarking) for algorithmic generation of virtual case studies (VCSs) for UDNs (combined sewer systems) and WDNs [14]. Urich et al., (2010) developed an agent-based approach to generate virtual CSSs as an extra module integrated into the VIBe software [13], [18]. However, despite being pioneers in the domain of virtual case study generation, the aforementioned frameworks did not entirely represent and capture the real-world characteristics of these structures. For example, a very simple layout procedure was utilized in these studies without considering road networks. This is because road networks are often viewed as a proxy for sewer layout representation thanks to their strong correlation with drainage layouts [19]. Further, very simplified hydraulic dimensioning was used in these studies, such as Rational or time-area method [20]. Blumensaat, Wolfram, & Krebs, (2012); Duque, Bach, Scholten, Fappiano, & Maurer, (2022), further integrated street network information in addition to other freely available data for the creation of more realistic sewer systems [21], [22]. Also, Nedergaard Pedersen et al., (2021) presented a comprehensive and unique real-world data-access data set, with the provision of existing meters, sensors and gauges in the real-world network [23]. Although these studies provided alternative databases for the drainage community while capturing a clearer picture of the real-world infrastructures, most of them are either computationally and/or mathematically expensive to reproduce or not open access for community to easily and quickly harness their potentials. Moreover, none of the mentioned efforts provided a great number of (optimal) ready-to-use semi-real-world USNs with different topological characteristics. As a result, in this study, we publish an open-access database of a multitude of optimized (semi) real USNs with different properties (e.g., diverse number of redundant flow paths). These networks employ complex network analysis for the design and adding redundant flow paths. Then, one specific application/modelling aim of this database is investigated in terms of functional resilience performance (i.e., evaluating flood characteristics of the networks under low to heavy storm events). However, this database can be used for any other applications such as structural resilience (i.e., sewer/inlet blockage), green-blue infrastructure analysis, etc.

2 METHODS

2.1 Complex network analysis of UDNs

UDNs can be analysed using a mathematical branch called graph theory G wherein the sewers, weirs and pumps are represented by the edges (E) of the graph and inlets/manholes and outlets are represented by the vertices (V) of the graph. Graphs are often represented by adjacency matrix A, in which its entry $a_{ij} = 1$ if there exists an edge between the source and target nodes of the graph (i, j), and $a_{ij} = 0$ otherwise. Thus, matrix A is V by V dimensions. To reproduce the properties of USNs; first, the graph needs to be directed since UDNs are primarily gravity-driven by specific and defined directions (e.g., based on their slope or flow directions); and second, USNs-tailored graph weights must be defined and allocated to vertices and/or edges of the graph depending on the purpose of the modelling (e.g., sewer lengths, diameters, velocity, etc.). These weights are given to the graph to mimic the structural and/or functional characteristics of the USNs, facilitating the relevant analysis. In this work, we use Eigenvector centrality measure [24], [25], to identify the sensitive locations/inlet nodes for initiating the redundancy construction.

This measure is a generalization of node degree centrality, (i.e., the number of edges connected to a node). Eigenvector centrality (shown in Eq. 1) indicates that a node can gain high eigenvector centrality not by having a high degree of centrality but also by being connected to nodes which themselves have high centrality. This measure is used to rank and score the nodes for constructing redundant paths for creating our benchmark database. Sewer diameters are used as edge weights/importance to detect the nodes interacting directly and indirectly with large pipelines of the UDNs.

$$\lambda c_i^E = \sum_{j=1}^{\#V} a_{ij}\, c_j^E \tag{1}$$

where, $i = 1, 2, \ldots, V$, are the nodes, λ is a constant eigenvalue, c_i^E is the eigenvector centrality of node i, c_j^E is the eigenvector centrality of other neighbouring nodes relative to node i, and a_{ij} are the entries of adjacency matrix A.

2.2 Optimization of USNs

Starting with the base graph (e.g., street network), sewer layout generation is initiated to infer e.g., flow directions. Leveraging complex network analysis, a deterministic multi-objective optimization (brute force method), and a set of layout objective functions (based on ground slope, sewer length and cumulative runoff area), a great number of layouts were generated, pre-screened and discarded. Then, due to contradicting layout objective functions (e.g., maximizing negative lopes and minimizing runoff distribution), the Pareto front of the best performing layouts were obtained. These solutions lying on the Pareto front were then forwarded to the SWMM-based hydraulic optimization to finally infer the optimal (decentralized) layout as well as the best-centralized layout. To quantify the degree of topological centralization, the generic measure introduced in [1] is used shown in Eq. 2:

$$DC = 100\ \times (1 - \frac{log_{10}^{ON_s}}{log_{100}^{IN}})\ (\%) \tag{2}$$

This measure has a logarithmic relationship between the number of (selected) outlet nodes (ON_s, at most IN -1) and the total number of inlet nodes (IN). This measure indicates that e.g., if one outlet node is selected among the outlet candidates during the layout optimization process, the DC is equal to 100% or a fully centralized system.

Hydraulic design optimization is a process where after finding the optimal spatial layouts, the dimensions of sewers, slopes and whether a pump station is needed at a node, are addressed using an adaptive search approach. Meanwhile, a set of constraints are satisfied simultaneously such as minimum and maximum cover depths, minimum and maximum sewer slopes, minimum and maximum velocity and telescopic pattern for the sewers. In this work, a 2-year block rain event with a duration of 15 minutes is used in combination with the dynamic wave flow routine in SWMM5 [26] for the aforementioned processes in agreement with technical guidelines [27]. A full description regarding the layout creation procedure and hydraulic design optimization, can be found in [1].

2.3 Redundancy-promoting scenarios

In order to generate several case studies with different topological attributes, we applied herein a resilience-boosting measure called redundancy [28]. Redundancy can be enriched by adding extra storage tanks, providing extra capacity at certain locations, or introducing redundant flow pathways. It should be noted that redundancy, redundant flow paths, and loops are used interchangeably throughout this manuscript. In this work, the latter attribute (i.e., redundant flow

CC BY NC SA

paths) is introduced to the optimized centralized and decentralized layouts, which originally do not have any redundant paths (loops). The redundant paths are added based on three strategies. (1) introducing redundant paths from upstream sections towards downstream sections, (2) introducing redundant paths from downstream sections towards upstream sections, and (3) introducing redundant paths from certain locations determined by Eigenvector centrality (Eq. 1). These redundant pathways are introduced cumulatively to the designed centralized and decentralized solutions based on the mentioned three strategies until achieving the maximum number of loops. Based on the design characteristics and the candidate nodes to construct redundant paths, two situations may occur (Fig. 1). Note that if constructing redundant flow paths from a node imposes the decentralized solution connected, we exclude that node from the redundancy candidacies. For example, nodes T, S and R, are those whose connections turn the decentralization notion of the network into a centralized one. Therefore, only node Q is considered in the example below to construct a redundant path (Fig. 1).

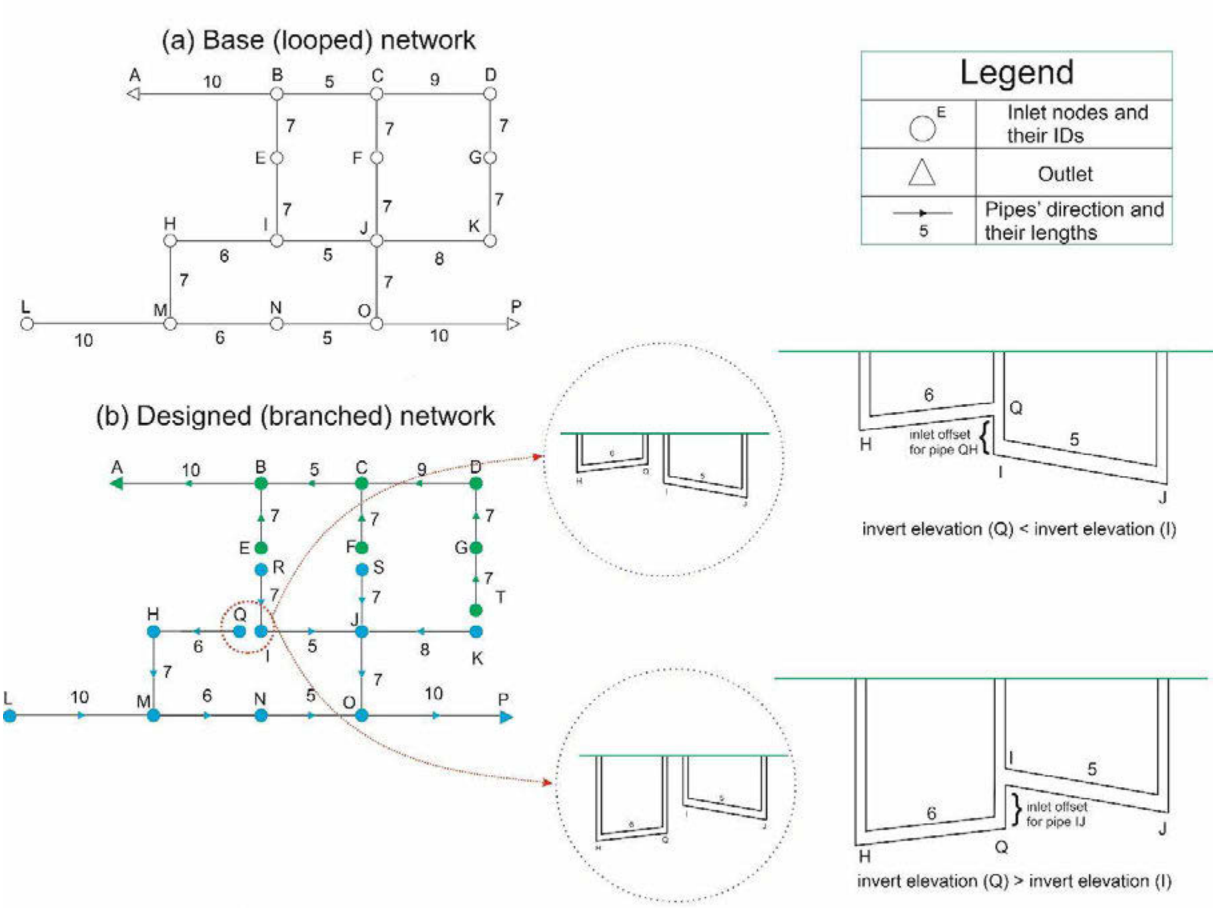

(a) Base (looped) network
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
10
5
9
7
8
6
Legend
Inlet nodes and their IDs
Outlet
Pipes' direction and their lengths
(b) Designed (branched) network
Q
R
S
T
inlet offset for pipe QH
invert elevation (Q) < invert elevation (I)
inlet offset for pipe IJ
invert elevation (Q) > invert elevation (I)

Figure 1. Pedagogical example of redundancy-promoting strategies [29] *(under review).*

2.4 Resilience evaluation

As mentioned earlier, a 2-year block rain event is used to design all USN solutions, achieving a limited/specific drainage capacity. However, due to the ongoing climate-change precipitation events, infiltration rate and drainage capacity of USNs are frequently exceeded, leading to flooding events out of the manholes. Thus, functional resilience of the USNs is evaluated in this study under medium (10-year) and extreme rainfall events (50-year) according to Eq. 2.

$$HPI = 100 \times (1 - \frac{V_{flooding}}{V_{runoff}}) \quad (\%) \qquad (2)$$

CC BY NC SA

Where, HPI represents the hydraulic performance indicator, $V_{flooding}$ is the total ponded flood volume [m^3], and V_{runoff} is the total runoff volume [m^3].

2.5 Case study

The investigated case study area is a part of Innsbruck (a steep region), an alpine city in Austria, as shown in Fig. 2 with around 15-meter difference in ground elevations. The free available data is an open street map from *https://www.geofabrik.de*, 30-m resolution digital elevation model (DEM) from *https://srtm.csi.cgiar.org*, and 20-m impervious layer map from *https://land.copernicus.eu* integrated into a base graph, upon which the (de)centralized design solutions are created and designed. It is worth mentioning that the subcatchments were delineated according to Voronoi diagrams or Thiessen polygons [30], [31]. This network with 11 potential outlet locations in the proximity of the river Inn is depicted in Fig. 2 with their characteristics as presented in Table 1. This case study is then designed based on two structures (centralized and decentralized) while redundant paths are added to them cumulatively. During this process (except for two branched centralized and decentralized networks), 200 networks are created for centralized system based on each redundancy-constructing strategy, and 139 networks for decentralized structure. In other words, 600 (200 × 3 strategies) networks are created for centralized solution with different number of loops, and 417 (139 × 3 strategies) networks are created for decentralized solution with different number of loops, leading to an overall 1019 USNs. More details regarding this case study can be found in [1].

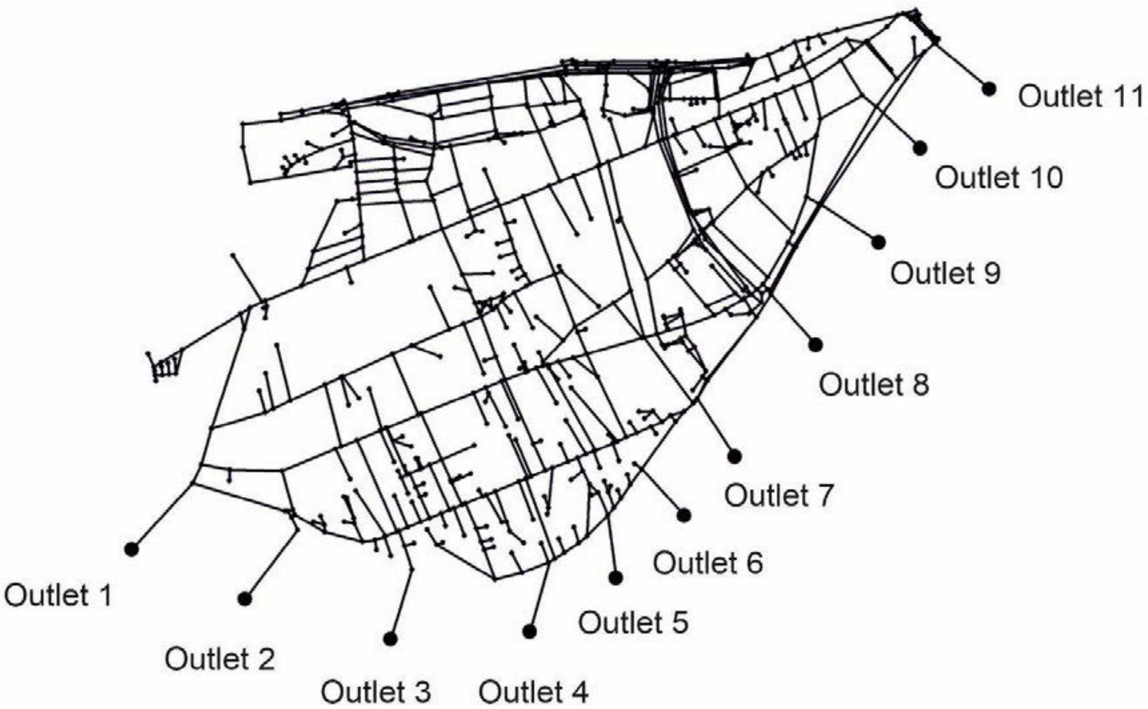


Figure 2. Case study (fully looped base graph/street network): a steep area within the city of Innsbruck [1].

Table 1. Characteristics of the case study [1].

	Total number of nodes	Total number of pipes	Total number of Subcatchment	Total lengths (m)	Average pipe lengths (m)	Total area (ha)	Total runoff area (ha)	Total number of outlet candidates	Elevation (m)
Case study	700	913	700	50,712	55.54	188	100	11	583 - 568

3 RESULT & DISCUSSION

Once the optimal centralized and decentralized layouts were chosen and hydraulically designed (shown in Fig. 3), they were benchmarked as the backbones of the database, to which redundant paths are introduced, thus, forming our whole database. As can be seen, the capital savings resulting from decentralization far outweighed the centralized design with construction cost (CC) as two times cheaper, emphasizing the significance of the decentralized solution. The DC measure of the optimal (decentralized) solution (Fig. 3b) is equal to 68%.

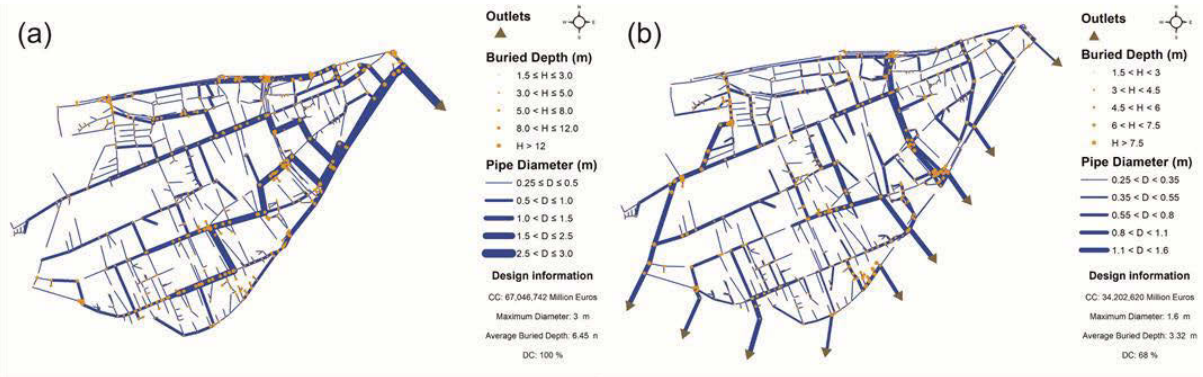

(a)
Outlets
Buried Depth (m)
1.5 ≤ H ≤ 3.0
3.0 < H ≤ 5.0
5.0 < H ≤ 8.0
8.0 < H ≤ 12.0
H > 12
Pipe Diameter (m)
0.25 ≤ D ≤ 0.5
0.5 < D ≤ 1.0
1.0 < D ≤ 1.5
1.5 < D ≤ 2.5
2.5 < D ≤ 3.0
Design information
CC: 67,046,742 Million Euros
Maximum Diameter: 3 m
Average Buried Depth: 6.45 n
DC: 100 %
(b)
Outlets
Buried Depth (m)
1.5 < H < 3
3 < H < 4.5
4.5 < H < 6
6 < H < 7.5
H > 7.5
Pipe Diameter (m)
0.25 < D < 0.35
0.35 < D < 0.55
0.55 < D < 0.8
0.8 < D < 1.1
1.1 < D < 1.6
Design information
CC: 34,202,620 Million Euros
Maximum Diameter: 1.6 m
Average Buried Depth: 3.32 m
DC: 68 %

Figure 3. Purely centralized design on the left and optimal (decentralized) design on the right for city of Innsbruck (steep area) [1].

Thereafter, the redundant flow paths are added to the two structures (based on three strategies explained above), forming our benchmark database. As mentioned before, the functional resilience of this dataset is analysed in this paper, however, any other applications can be performed, such as structural resilience (i.e., sewer/inlet blockage), green-blue infrastructure analysis, etc.

Fig. 4(a) and (b) show the resilience (via HPI) performance under low-medium rain (i.e., 10-year event with total rain amount equal to 53.94 mm, having 15 min durations) where (a) denotes its distribution with extra paths constructed from upstream and downstream (with the implication from the branched network), and (b) represents the HPI values against the percentage of loops with the implication from (eigenvector) centrality-based strategy. Regarding the resilience response of networks, the influence of flow redistribution achieved via redundancy is lessened due to the steepness of the catchment, thus, minor flood volume is expected. As seen in Fig. 4(a), the branched decentralized layout slightly outperformed its centralized counterpart.

CC BY NC SA

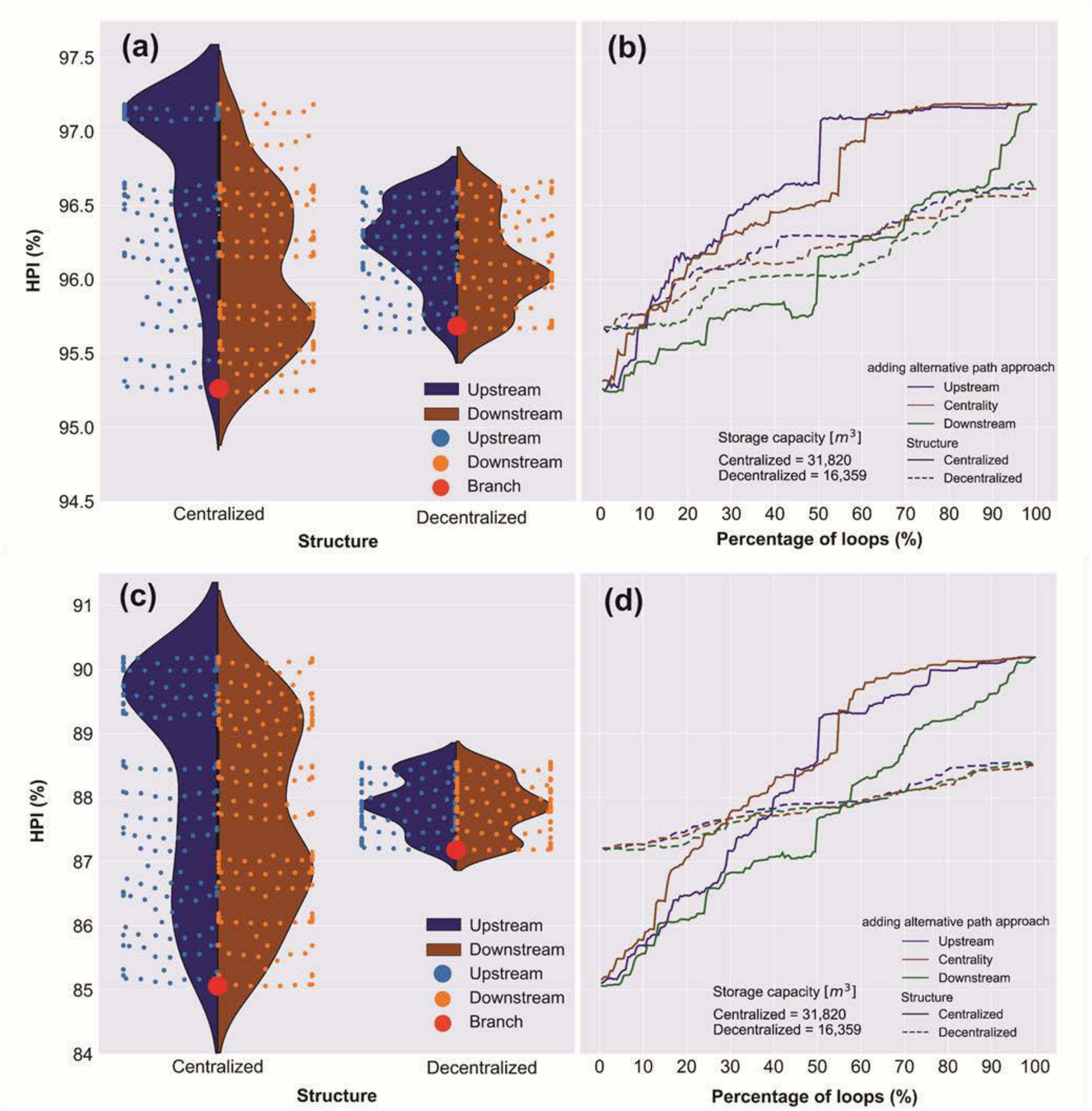


Figure 4. Resilience performance (HPI measure) of centralized and decentralized networks in the city of Innsbruck (case study) while redundancy (alternative water flow paths) is added using upstream, downstream and centrality approaches. Fig. 4(a) and (b) are the performances under the 10-year rain event, and Fig. 4(c) and (d) are the performances under the 50-year rain event.

Additionally, Fig. 4(b) shows that placing cumulative alternative water paths from upstream yielded better resilience with more concentration of HPI values around 97% for centralized and 96% for decentralized solutions. Considering Fig. 4(b), adding alternative pathways into the centralized layouts elevated the resilience (ranging by 2%) quicker than decentralized ones (ranging only by 1%), primarily thanks to its (centralized) bulky storage capacity as well as more effective flow direction achieved by redundancy. Furthermore, there existed a close competition between the resilience response achieved via upstream and centrality whereas, downstream loops underperformed the other two strategies.

However, concerning extreme precipitation (under 50-year event with a total rain amount of 72.75 mm with 15 min duration), shown in Fig. 4(c) and (d), it can be inferred that the range of variations for the decentralized network still remains very limited by only 1% resilience improvement once all extra paths were cumulatively constructed. This quantity, however, changed by 5% for the centralized layout. This fact demonstrates the relative inefficiency of introducing redundancy into decentralized layouts in steep terrains because the catchment steepness combined with decentralization (at the design stage) could sufficiently attenuate the extreme flood discharges out of the system.

4 CONCLUSIONS

When different applications and modelling purposes of UDNs are evaluated, the generalization of results is often feasible if methodologies are applied to a great number of case studies. However, attaining accurate and detailed structural and hydraulic information/model on the underground UDNs is not a trivial task. Moreover, water utility companies are often not interested in sharing their data due to security and publicity reasons. Thus, to take a small step to bridge this research gap, this study aims to develop a case study benchmarks for the urban drainage community by providing an optimized centralized and decentralized semi-real-world case study with various topological variations (i.e., existence of different number of redundant flow pathways). This database can be reached through (https://github.com/iut-ibk/Benchmark-CaseStudies-Innsbruck.git) for the steep region discussed herein. Then, this dataset was used as the benchmark case study to evaluate the networks' functional resilience under low and high precipitation events. However, any other applications can be also performed such as structural resilience (i.e., sewer/inlet blockage), green-blue infrastructure analysis, etc. The results indicated the efficacy of decentralized scenarios for mitigating volumetric flow discharges (especially under heavy storm events) compared to centralized scenarios. The results also showed the suitability of Eigenvector centrality to identify the sensitive locations for the introduction of redundant flow pathways.

Note that, the same framework conducted in this study, is already applied to an entirely flat region in city of Ahvaz/Iran, whose publication is currently under review [29]. The database used in [29] can be also found through (https://github.com/iut-ibk/Benchmark-CaseStudies-Ahvaz.git).

5 ACKNOWLEDGEMENTS

This study was funded by the Austrian Science Fund (FWF): P 31104-N29.

6 REFERENCES

[1] S. Hesarkazzazi, M. Hajibabaei, A. E. Bakhshipour, U. Dittmer, A. Haghighi, and R. Sitzenfrei, "Generation of optimal (de) centralized layouts for urban drainage systems: A graph-theory-based combinatorial multi-objective optimization framework," Sustain. Cities Soc., p. 103827, 2022.

[2] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savić, "Using complex network analysis for optimization of water distribution networks," Water Resour. Res., vol. 56, no. 8, p. e2020WR027929, 2020.

[3] Z. Xu, X. Dong, Y. Zhao, and P. Du, "Enhancing resilience of urban stormwater systems: cost-effectiveness analysis of structural characteristics," Urban Water J., pp. 1–10, 2021.

[4] X. Dong, H. Guo, and S. Zeng, "Enhancing future resilience in urban drainage system: Green versus grey infrastructure," Water Res., vol. 124, pp. 280–289, 2017.

[5] Y. Jia et al., "Water quality modeling in sewer networks: Review and future research directions," Water Res., vol. 202, p. 117419, 2021.

[6] S. N. Mugume, K. Diao, M. Astaraie-Imani, G. Fu, R. Farmani, and D. Butler, "Enhancing resilience in urban water systems for future cities," Water Supply, vol. 15, no. 6, pp. 1343–1352, Jul. 2015.

[7] D. Butler, B. McEntee, C. Onof, and A. Hagger, "Sewer storage tank performance under climate change," Water Sci. Technol., vol. 56, no. 12, pp. 29–35, 2007.

[8] N. Metje et al., "Mapping the underworld–state-of-the-art review," Tunn. Undergr. Sp. Technol., vol. 22, no. 5–6, pp. 568–586, 2007.

[9] J. M. Muggleton, M. J. Brennan, and Y. Gao, “Determining the location of buried plastic water pipes from measurements of ground surface vibration,” J. Appl. Geophys., vol. 75, no. 1, pp. 54–61, 2011.

[10] M. Hajibabaei, S. Hesarkazzazi, M. Lima, F. Gschösser, and R. Sitzenfrei, “How to Perform Life Cycle Assessment for Water Distribution Networks with Partly Unavailable Data,” in World Environmental and Water Resources Congress 2021, pp. 863–871.

[11] L. H. M. Costa and G. P. W. Rodrigues, “Automatic Generation of Water Distribution Networks from Streets Layout,” Water Resour. Manag., vol. 35, no. 4, pp. 1299–1319, 2021.

[12] M. Möderl, R. Sitzenfrei, T. Fetz, E. Fleischhacker, and W. Rauch, “Systematic generation of virtual networks for water supply,” Water Resour. Res., vol. 47, no. 2, pp. 1–10, 2011.

[13] R. Sitzenfrei, M. Möderl, and W. Rauch, “Graph-based approach for generating virtual water distribution systems in the software VIBe,” Water Sci. Technol. Water Supply, vol. 10, no. 6, pp. 923–932, 2010.

[14] R. Sitzenfrei, S. Fach, M. Kleidorfer, C. Urich, and W. Rauch, “Dynamic virtual infrastructure benchmarking: DynaVIBe,” Water Sci. Technol. Water Supply, vol. 10, no. 4, pp. 600–609, 2010.

[15] W. Rauch et al., “Modelling transitions in urban water systems,” Water Res., vol. 126, pp. 501–514, 2017.

[16] I.-S. Rehm, J. Friesen, K. Pouls, C. Busch, H. Taubenböck, and P. F. Pelz, “A Method for Modeling Urban Water Infrastructures Combining Geo-Referenced Data,” Water, vol. 13, no. 16, p. 2299, 2021.

[17] M. Möderl, D. Butler, and W. Rauch, “A stochastic approach for automatic generation of urban drainage systems,” Water Sci. Technol., vol. 59, no. 6, pp. 1137–1143, 2009.

[18] C. Urich, R. Sitzenfrei, M. Möderl, and W. Rauch, “An agent-based approach for generating virtual sewer systems,” Water Sci. Technol., vol. 62, no. 5, pp. 1090–1097, 2010.

[19] M. Mair, J. Zischg, W. Rauch, and R. Sitzenfrei, “Where to find water pipes and sewers?-On the correlation of infrastructure networks in the urban environment,” Water (Switzerland), vol. 9, no. 2, 2017.

[20] D. Butler and J. Davies, Urban drainage. 2003.

[21] F. Blumensaat, M. Wolfram, and P. Krebs, “Sewer model development under minimum data requirements,” Environ. Earth Sci., vol. 65, no. 5, pp. 1427–1437, 2012.

[22] N. Duque, P. M. Bach, L. Scholten, F. Fappiano, and M. Maurer, “A Simplified Sanitary Sewer System Generator for Exploratory Modelling at City-Scale,” Water Res., vol. 209, p. 117903, 2022.

[23] A. Nedergaard Pedersen, J. Wied Pedersen, A. Vigueras-Rodriguez, A. Brink-Kjær, M. Borup, and P. Steen Mikkelsen, “The Bellinge data set: open data and models for community-wide urban drainage systems research,” Earth Syst. Sci. Data, vol. 13, no. 10, pp. 4779–4798, 2021.

[24] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” J. Math. Sociol., vol. 2, no. 1, pp. 113–120, 1972.

[25] V. Latora, V. Nicosia, and G. Russo, Complex networks: principles, methods and applications. Cambridge University Press, 2017.

[26] L. A. Rossman, Storm water management model user’s manual, version 5.0. National Risk Management Research Laboratory, Office of Research and …, 2010.

[27] Ö. W. Abfallwirtschaftsverband, “Regelblatt 11-Richtlinien für die abwassertechnische Berechnung und Dimensionierung von Abwasserkanälen,” 2009.

[28] S. N. Mugume, D. E. Gomez, G. Fu, R. Farmani, and D. Butler, “A global analysis approach for investigating structural resilience in urban drainage systems,” Water Res., vol. 81, pp. 15–26, 2015.

[29] R. Hesarkazzazi, S, Bakhshipour, A, E, Hajibabaei, M, Dittmer, U, Haghighi, A, Sitzenfrei, “Battle of Centralized vs. Decentralized Urban Drainage Networks (Under Review),” Water Res.

[30] F. Aurenhammer and R. Klein, “Voronoi Diagrams.,” Handb. Comput. Geom., vol. 5, no. 10, pp. 201–290, 2000.

[31] K. E. Brassel and D. Reif, “A procedure to generate Thiessen polygons,” Geogr. Anal., vol. 11, no. 3, pp. 289–303, 1979.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14739

METHODOLOGY FOR ANALYSIS OF SGDS COMPLIANCE IN URBAN WATER SYSTEMS. IMPLEMENTATION IN CASE STUDIES IN THE VALENCIAN COMMUNITY

Camila Garcia[1], Pilar Conejos Fuertes[2], Jaime Castillo Soria[3], P. Amparo López-Jiménez[4] and Modesto Pérez-Sánchez[5]

[1,4,5]Hydraulic and Environmental Engineering Department, Universitat Politècnica de València, Valencia (Spain)
[2]Idrica, Valencia (Spain)
[3]Global Omnium, Valencia (Spain)

[1] cgarcia1@posgrado.upv.es, [2]pilar.conejos@idrica.com, [3] Jcastillo@globalommnium.com , [4] palopez@upv.es, [5] mopesan1@upv.es

Abstract

The current economic, social, and technological development demands high quantities of good quality water, putting in more stress water resources. However, if water resources could be managed in a sustainable way, it will be possible to satisfy the present needs without compromising its long-term capacity. Given the situation, it has become a challenge to manage water systems sustainably, achieving all the goals set by the SDGs. In this sense, the sustainability of hydraulic systems must be considered globally to integrate all the aspects that intervenes in the concept, such as the economic, social, technical, and environmental fields. Key performance indicators (KPIs) can help managers to audit and evaluate hydraulic systems over time and from a sustainable point of view. This contribution presents a methodology for the implementation of sustainability indicators that will allow to assess the compliance of all the SDGs goals that intervenes in urban water systems. The defined indicators are linked to the management of hydraulic systems, and it was developed a classification of use based on their usefulness, complexity and cost when measuring them in the systems. This study also proposes a classification by levels of sustainability of the water systems, such as the energy efficiency classification of electrical appliances. The methodology is applicable for all types of hydraulic systems (supply, sanitation, treatment) and for each of its processes (operation, maintenance, administration, infrastructure). For the implementation of the KPIs, the level of development of the country or region where the study is carried out is considered. Therefore, for each case different indicators are defined from those already established to consider the characteristics of the systems and for the calculation of each KPI. A description of the system must be carried out, to have a memory of the evolution of the complete system and to define the indicators that are relevant. Therefore, through the results of the analysis it is possible to establish a benchmarking on the sustainable aspects of the water distribution system would represent a design, diagnosis, and management tool, which will improve sustainability in hydraulic systems. By improving the performance and sustainability of the systems, it is not only possible to achieve the objectives of goal 6, but also other associated goals will benefit, which highlights the relevance of sustainability indicators in hydraulic systems.

Keywords
SDGs, sustainability indicators, performance indicators, urban water systems, sanitation, water networks.

1 INTRODUCTION

Nowadays the concept of sustainability is quite well known, and it can be defined as a process that seeks to meet the present needs without compromising the capacity of the resources in the long or short term [1]. It was implemented by the UN to establish goals that will help to reach a sustainable development specially in developing countries. The UN established the concept followed by a plan of action known as the Agenda 2030, in which it is included 17 global goals (Sustainable Development Goals, SDGs) and targets to reach those goals [1]. Those goals are based on the pillars of a sustainable development: social, economic, and environmental.

The social component aims to improve the quality's life of the society by eradicating poverty, achieving equity and access to basic services such as education, water supply, sanitation, among others. The economic perspective looks to produce goods and services with a tariff policy characterized by a reasonable price. This implies that the demand is satisfied at a price that will allow to recover the costs and stimulates the investment. The environmental component seeks to preserve the ecosystems through the reduction of emissions, managing waste and improving the energy efficiency.

One of the SDGs is dedicated to water resources and systems, covering: supply, sanitation, management, and water quality. The SGDs present some interrelation between them, some of these relations are more explicit than others. For example, 3.3 (end with water-borne and communicable diseases), 3.9 (reduce number of deaths and diseases from water contamination) and 12.4 (management of chemical and waste reducing their release to water) are related to SDG6 (clean water and sanitation). There are others SDG that have a connection with SDG6 like, SG7 (water-energy nexus), SDG11 and SDG13 [1].

Considering the importance that concept of sustainability has taken in recent time, the need and desire for measuring the different aspects that a sustainable development carry has increase, including urban hydraulics systems. Water is a renewable resource, crucial for human and social development. According to the guideline of a sustainable development, such resources can't be used at a rate greater than its natura generation. Water has the potential of run out due to the current social, economic, and technological development that demands higher quantity and quality of the resource.

The actual water demand punt in stress water resources, as well as supply managers by trying to satisfy the demand. Actions are taken to meet the requirements but without considering sustainable aspects or looking out for the future state of the resources. The importance of the resource in daily activities and the increasing demand highlight the need to incorporate sustainable solutions in water systems.

According to the last report of the UN, there is still a deficit more than 60% in the achievement of the SDG6. In developed countries, water networks are not managed from a sustainable point of view and in developing countries the problem settles in the fact that not even the basics sustainable requirement are fulfilled, like the lack of water supplies and sanitation. Besides, 80% of the wastewater are discharge in water bodies with an accurate treatment.

Given the current situation and pressure imposed on the resources, the administration of hydraulics systems becomes a challenge. It's important that all the stakeholders that intervene in those systems have reliable information of the state and management of the networks. To assess it, Key Performance Indicators (KPIs) that evaluates the different aspects of sustainability in the systems are needed. Such indicators have been evaluated and proposed in different research.

KPIs are used to monitor and evaluate hydraulic systems. There are many types of indicators such as volumetric, percentual, among others. However, it must be established a system of indicators that help to audit from a sustainable point of view. This contribution presents a methodology for the implementation of sustainability indicators (SIs) that will assess the compliance of all the SDGs goals that intervenes in urban water systems, such as supply, treatment, and sanitation.

2 METHODOLOGY

2.1 Establishment of indicators

Through KPIs it is possible to evaluate a level of performance in a period of time for a system. According to the IWA [3], indicators for water system should have the following characteristics be quantifiable, objectives, with concise meaning, reasonably calculable, auditable, measurable, and easy to understand. The IWA's indicators are currently well known and used because of its easy and universal implementation. Also, because they have proved the efficiency in the measurements. Therefore, the properties of indicators defined by them was important to follow in the establishment of the proposed indicators so that these can be implemented as well.

In Figure 1 it is plotted the methodology that was implemented for the establishment of the indicators. The first step was a review of different case studies in which Sis were implemented. Also, other important sources were analysed such as UN reports and others made by governments and organizations, indicators guidelines like IWA's. Indicators that were often used in different research and had great value, were also review, such as those related to energy and environmental aspects.

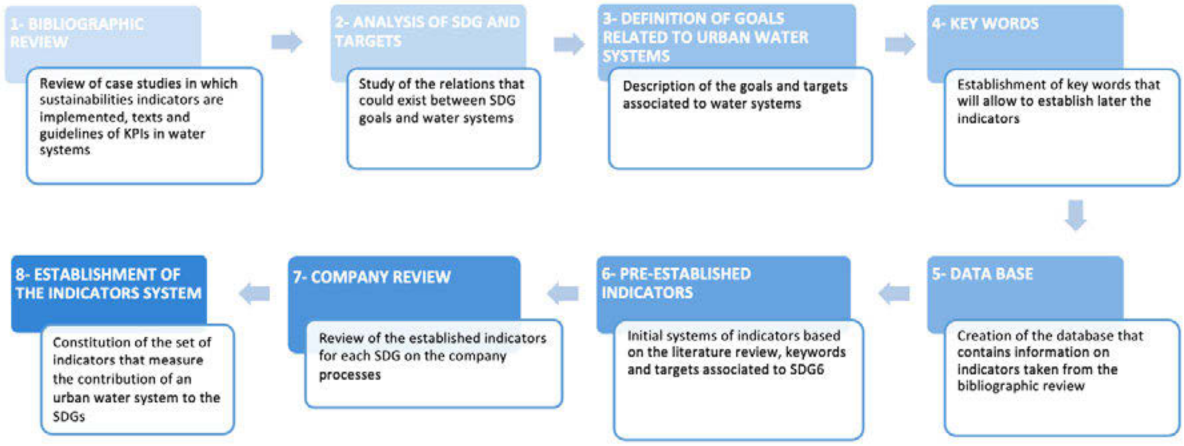

1- BIBLIOGRAPHIC REVIEW
Review of case studies in which sustainabilities indicators are implemented, texts and guidelines of KPIs in water systems
2- ANALYSIS OF SDG AND TARGETS
Study of the relations that could exist between SDG goals and water systems
3- DEFINITION OF GOALS RELATED TO URBAN WATER SYSTEMS
Description of the goals and targets associated to water systems
4- KEY WORDS
Establishment of key words that will allow to establish later the indicators
5- DATA BASE
Creation of the database that contains information on indicators taken from the bibliographic review
6- PRE-ESTABLISHED INDICATORS
Initial systems of indicators based on the literature review, keywords and targets associated to SDG6
7- COMPANY REVIEW
Review of the established indicators for each SDG on the company processes
8- ESTABLISHMENT OF THE INDICATORS SYSTEM
Constitution of the set of indicators that measure the contribution of an urban water system to the SDGs

Figure 1. Methodology followed for the establishment of indicators

The second step was to analyse in detail the goals based on the reviewed information. Important data about the concepts, definition of the key words and variables of each target was studied. Based on that, was described a relation between every SDG and the water urban systems. This step was important because to establish sustainable indicators that can capture all the variable and desire measures in water systems, the relationship must be described. Also, this will allow to consider all the interrelation between all the SDG and hydraulics systems or SDG6.

The third step was simultaneous with step 2 and consisted of defining and describing the related targets in which hydraulics systems can contribute in achievement of the sustainable development. As an example, one of the targets of SDG5 in which water systems can contribute defined was 5.1, about ending of discrimination against women. Water services providers can guarantee equal working conditions as well as increasing the number of female workers, if necessary. Also, by improving the coverage of the network, hygiene and sanitary facilities for women can be ensured.

CC BY NC SA

The fourth step was the establishment of key words that will allow to define the indicators. According to the concepts and interrelations found for every SDGs, some key words were defined. These key words allow to cover entirely the concept and measurements that each goal requires.

The fifth step was the creation of a database about the information collected from the bibliographic sources consulted, such as the indicators implemented. Since some were frequently used, reference values were stored to subsequently calculate levels of sustainability. Also, indicators that had not been implemented were included, because of their relevance in the measurements, such as those related to personnel, equality policies, among others.

The sixth step was the establishment of the initial system of indicators for each SDGs, based on the information previews reviewed and the key words. These indicators were intended to really focus on sustainable aspects rather than a more technical side of urban water systems. Continuously, in the next step a detailed reviewed process for each indicator was carry on. In this step the possible implementation and measurement of an indicator was studied in the different process that a company carries. This was done to be able to determine which indicators were able to be implemented but also to define its measurements and the contribution in the sustainable development of the company. During the review of the different processes carried out in the company, new indicators were established since they resulted relevant for the measurement of sustainability in water systems.

With the review of all the indicators and their implementation in different company's processes, it was possible to highlight the contribution of a hydraulic system for each SGDs. The last step was to define the indicators that could be used to measure the sustainability in water resources. The KPIs established, comply with characteristics mentioned before and are applicable to all types of hydraulic system (supply, sanitation, treatment) and for each of its processes (operation, maintenance, management, etc.).

2.2 Implementation of the system

The established KPIs system allows measuring the sustainability of urban water system (such as distribution, treatment plants, collection, and regeneration). When the methodology is carried out, the results obtained will allow to take measures that can improve a sustainable management of the system or at least comply with the basics sustainability principles that the UN exposes. For the implementation of the established system of indicators, it is proposed to follow the process presented in Figure 2.

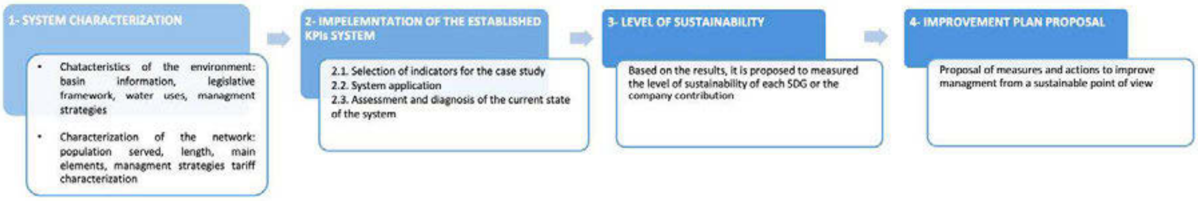

1- SYSTEM CHARACTERIZATION
• Chatacteristics of the environment: basin information, legislative framework, water uses, managment strategies
• Characterization of the network: population served, length, main elements, managment strategies tariff characterization
2- IMPELEMNTATION OF THE ESTABLISHED KPIs SYSTEM
2.1. Selection of indicators for the case study
2.2. System application
2.3. Assessment and diagnosis of the current state of the system
3- LEVEL OF SUSTAINABILITY
Based on the results, it is proposed to measured the level of sustainability of each SDG or the company contribution
4- IMPROVEMENT PLAN PROPOSAL
Proposal of measures and actions to improve managment from a sustainable point of view

Figure 2. Implementation process proposal for the KPIs system

To implement the methodology, the first step proposed is to identify the state of the environment where the system and the network are developed, so an evolutionary memory can be obtained and will allow to see the impact of the system in the sustainable development. Therefore, changes can be identified, and solutions can be defined. The characterization of the environment implies collecting information of the basin (e.g., area, characteristics of the water bodies and the land, uses of the water, legislative information, among others). On the other hand, for the characterization of the network it is important to obtain data on the population served, location of the treatment plant, outlet flow, length of the network, control elements and operational strategies.

The second step is the implementation of the established SIs system. For these, first should be selected the indicators defined by the type of systems. This means, that if the methodology will be applied to a treatment plant, only those indicators defined for this type of system would be selected from the list. The next step (2.2 in Figure) is to apply to the system the indicators by recollecting all the information and data relevant according to the variables needed. In this step, in addition, a list of the reasons was established for those cases in which the data for an indicator was not possible to measure. In Table 1are some of the arguments that were defined were that the measurement was the cost for assessing the information was high, not profitable or that it did not apply to the case since it referred to a different water cycle. Also, the level of difficulty in data collection was classified.

Table 1. Arguments implemented for those indicators without data

Difficulty level		**Arguments**	**Description of the arguments**
1	Very low difficulty	A	Cost of information: it’s not profitable to carry out the measurement
2	Low difficulty	B	Data protection
3	Medium difficulty	C	Measurement takes too long
4	High difficulty	D	Time scale of the data is not useful
5	Very high difficulty: it’s not possible to have access to the information	E	Third party data
		F	Doesn’t apply for the case
		G	Type of measured not performed
		H	Others

The step 2.3 is to process and calculate all the indicators to evaluate the status of the system from a sustainable point of view. Also, this will lead to a diagnosis from which it will be possible to take solutions or strategies that aim to improve a sustainable management of the water system and therefore raising the achievement of the SDGs. The solutions to be implement can be aimed at improving one or more SDGs, depending on which targets the company wants to improve first.

As a third step, it is proposed to calculate a level of sustainability. Initially, it is proposed that these levels have mobile ends since there is not much data yet. In this way, the same system (for example, a supply) will be able to feed its level database with more and more information and, said, the extreme values would be recalculated based on the amount of data.

The last step is to propose a plan based on the results obtained, in which a series of actions or measures can be defined to improve management from on a sustainable point of view. As mentioned before, the actions taken can be focused on improving the level of a single SDG or many

BY NC SA

of them. In the case of less developed countries, management can focus on fulfilling at least basic sustainability requirements.

3 RESULTS

3.1 Establishment of indicators

As described in the methodology, the second and third step for the establishment of the KPIs was the study and the goals and targets to be able to obtain an interrelation between the SDGs and urban water systems. Table 2 shows the result of the relations obtained. As mentioned before the defined targets established are those in which the urban hydraulic systems can contribute in certain way. The Table 3 shows an example of the described interrelation between SDG5 and SDG6, as well as the key words.

Table 2. Targets related to water systems for each SDGs

SDGs	Target related to urban water systems (SDG6)
SDG 1-No poverty	1.1; .1.4; 1.5; 1.a
SDG 2-Zero hunger	2.1; 2.2; 2.3; 2.4; 2.a
SDG3- Good health and well-being	3.1; 3.2; 3.3; 3.8; 3.9
SDG4- Quality education	4.4; 4.5; 4.7; 4.a; 4.b
SDG5- Gender equality	5.1; 5.4; 5.5
SDG7- Affordable and clean energy	7.1; 7.2; 7.3; 7.a; 7.b
SDG8- Decent work and economic growth	8.1; 8.2; 8.3; 8.4; 8.5; 8.6; 8.8
SDG9- Industry, innovation, and infrastructure	9.1; 9.2; 9.4; 9.5;9.a
SDG10- Reduced inequalities	10.1; 10.2; 10.3; 10.4; 10.5; 10.b
SDG11- Sustainable cities and communities	11.; 11.3; 11.4; 11.5; 11.6; 11.b
SDG12- Responsible consumption and production	12.2; 12.4; 12.5; 12.6; 12.8; 12.a
SDG13- Climate action	13.1; 13.2; 13.3
SDG14- Life on below water	14.1; 14.2; 14.3
SDG15- Life on land	15.1; 15.4; 15.5
SDG16- Peace, justice and strong institutions	16.1; 16.6; 16.7; 16.10
SDG17- Partnerships for the goals	17.3; 17.5; 17.7; 17.16

Table 3. Related targets of SDG5 and description of the relationship with SDG6

SDG	Target	Decription of the interrelation with SDG6	Keywords
SDG5 – Gender equality and women's empowerment	5.1	Asegurar instalaciones sanitarias y de higiene para mujeres y niñas en todos los lugares. Asegurar la igualdad de condiciones y cantidad de mujeres trabajando en los sistemas hidráulicos.	Género, desigualdad, discriminación, mujer, niña
	5.4	Fuentes de agua e instalaciones más cerca al hogar o zonas urbanas, para evitar el acoso en mujeres que transportan el agua (países en desarrollo)	Género, trabajo doméstico, servicios públicos, infraestructura, protección social, hogar, familia
	5.5	Asegurar la igualdad de condiciones y cantidad de mujeres trabajando en los sistemas hidráulicos.	Género, participación, mujer, igualdad, política, económica, pública

The system of KPIs was established to assess the achievement of all the SDGs related to water resources. For each SDG, was defined direct and indirect indicators. For the first, the relation between the SDG and the indicator is almost evident and the measurement is direct. Meanwhile, for the indirect indicators the relation even though the measurement does not contribute much to the tracking of the SDGs, there is still some link.

Also, for each indicator, the following information was defined and classified:

- ID
- SDGs's targets related to water
- Sustainable aspect: technical, economic, environmental, or social management (it could be related to more than one option)

- Water cycle: stage of the cycle for which the indicator is implemented (collection, treatment, distribution, collection, purification, and regeneration)
- Type of indicator: performance, financial, productive, energy and environmental efficiency

The KPIs system is conformed by 135 indicators. The Figure 3 shows that 110 can be implemented in supply systems (caption and distribution), 100 are for sanitation systems (collection and disposal) and finally 87 can be used for water treatment systems. Also, in Figure 4 is presented the number of indicators established for each SDG, according to the type of urban system. As an example, for SDG1, 10 indicators were defined for supply systems as well as for sanitation and 7 for treatment systems.

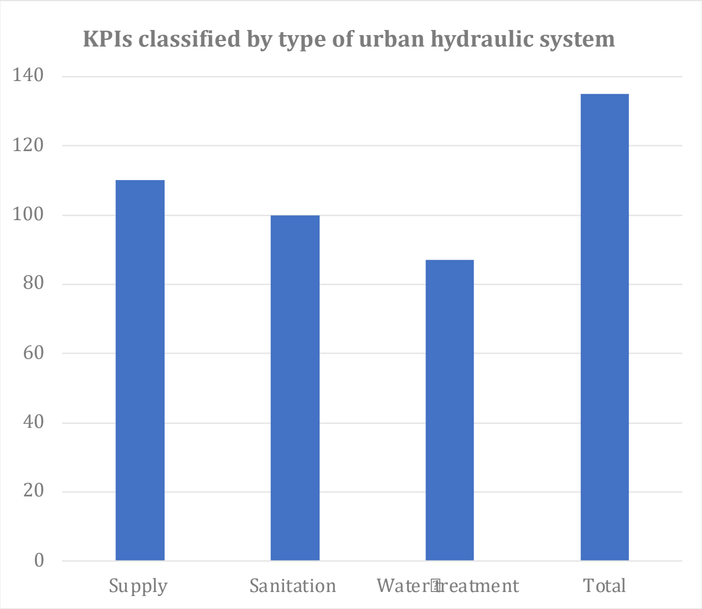


Figure 3. Distribution of the indicators classified by type of urban hydraulic system

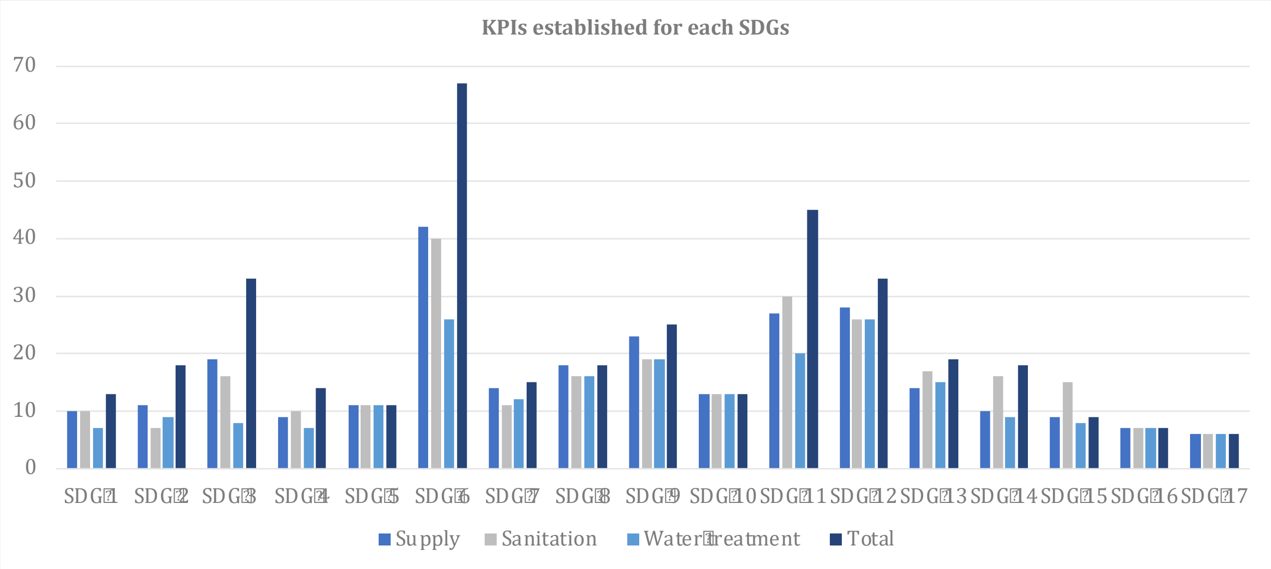


Figure 4. KPIs established for each SDGs, classified by type of urban water system

BY NC SA

3.2 Implementation of the system – case study

After having established the system of indicators, the implementation of it in a case study was next. The case study refers to the distribution network of the city of Valencia. As described in the methodology, once the characterization of the environment and the network was carried out, the indicators were selected to evaluate the system. Considering that the service to be evaluated was supply, there were 110 indicators for this type of system and only 97 could be applied. Figure 5 shows the distribution of the KPIs implemented in each SGDs for the case study respect to the total number of indicators of each SDGs established.

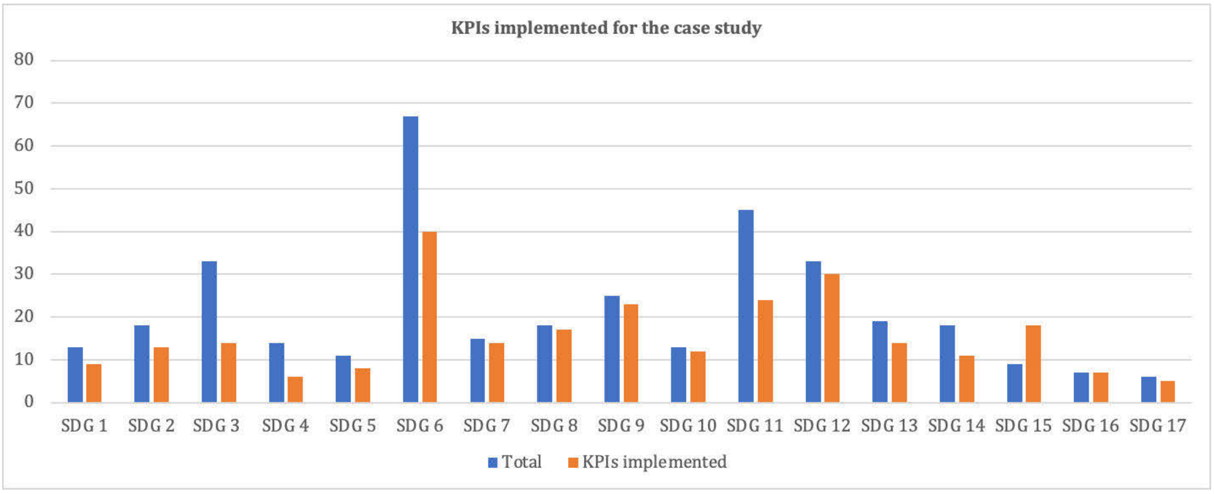
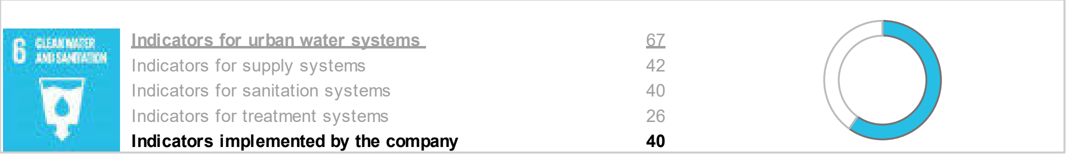


Figure 5. KPIs for each SDGs implemented for the case study


6 CLEAN WATER AND SANITATION

Indicators for urban water systems	67
Indicators for supply systems	42
Indicators for sanitation systems	40
Indicators for treatment systems	26
Indicators implemented by the company	**40**

Figure 6. Resume SDG6 results about indicators applied

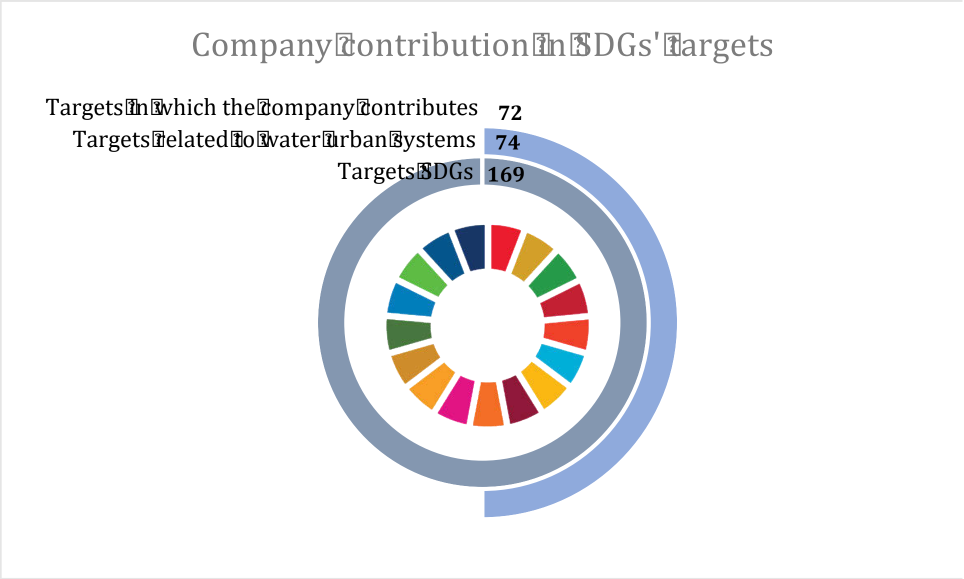


Figure 7. Company's contribution in SDGs' targets

In Figure 7 is represented the company's contribution in the SDG's targets. The UN established 169 indicators in total. It was found that from those, 74 targets have a interrelation with the SDG6 or urban water systems. Once the methodology was implemented and evaluated in the case study of Valencia, it was found that the water company contributes in 72 of those targets. Only in SDG 10 and SDG14, there are some indicators that were not implemented and therefore the contribution is not 100%. Nevertheless, the company's contribution is quite good.

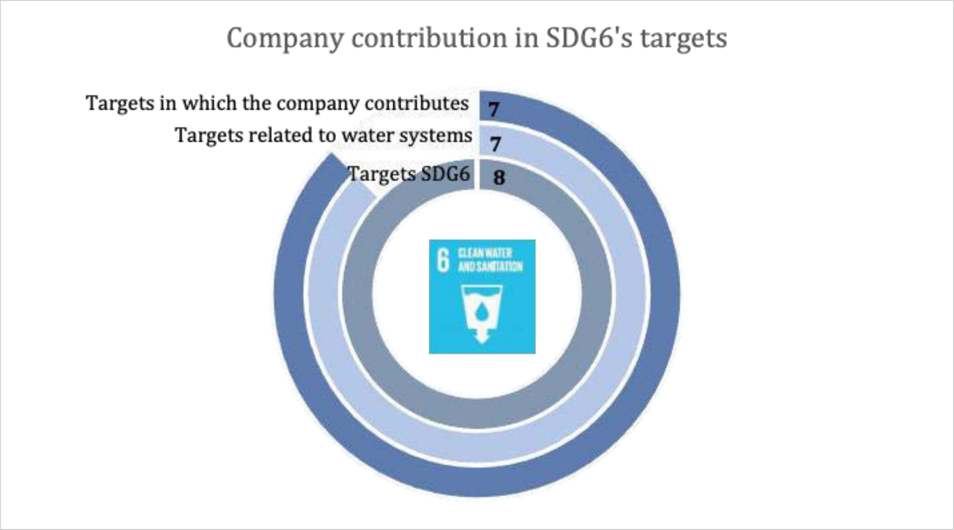


Figure 8. Company contribution in SDG's 6 targets

Figure 6 presents as an example the results of the indicators implemented specifically for the SDG6. It could be observed that of the 42 indicators established for supply systems, only 40 were applied. Alternatively, Figure 8 shows the contribution of the company specifically to the targets of SDG6. Of the 8 goals established by the UN for this SDG, 7 were related to the defined KPIs system and the company contributes in 7 of them. This mean that, the service provider has a considerable influence on this goal since it manages to contribute to almost all targets. If a sustainable management were followed by the company, all these targets could improve.

In general, there were 135 indicators, of which 32 were not applied. For supply systems, as mentioned before, 13 indicators couldn't be applied. The level of difficulty in data collection was classified, being 1 very low difficulty a very easy access to the information and 5 high difficulty indicates that there is no access to the data. In addition, certain arguments were defined to describe the reasons for the measurements that couldn't be performed. This was all represented in Table 1.

Of the 32 indicators for which information could not be obtained, the level of difficulty in collecting it mostly was 1, which represents a low rate. Only a few (less than 5) were classified as level 4 which indicates higher difficulty. However, even though the data was easy to obtain, the main reason for the lack of information was because it didn't apply or matches to the case of study. Besides that, for certain indicators the measurement of the variables was not made and in other cases it was not relevant to measure for the company. This can be seen in Figure 9 and Figure 10.

BY NC SA

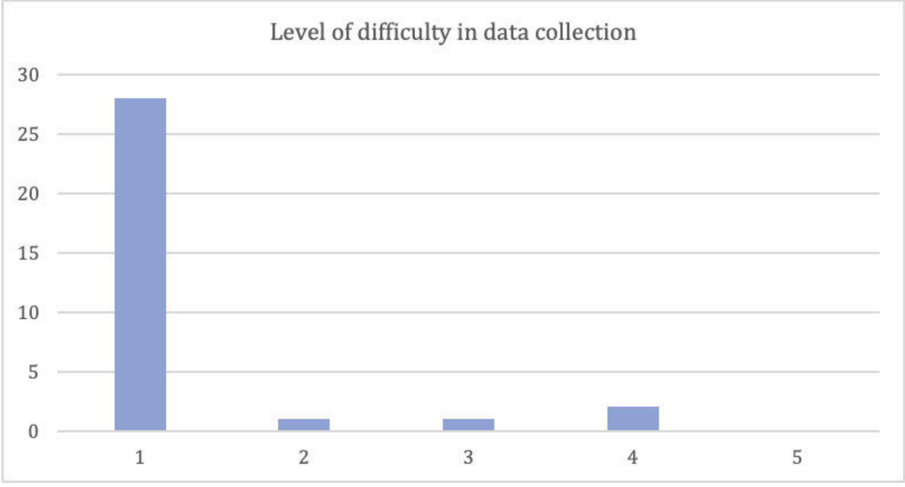


Figure 9. Level of difficulty in data collection

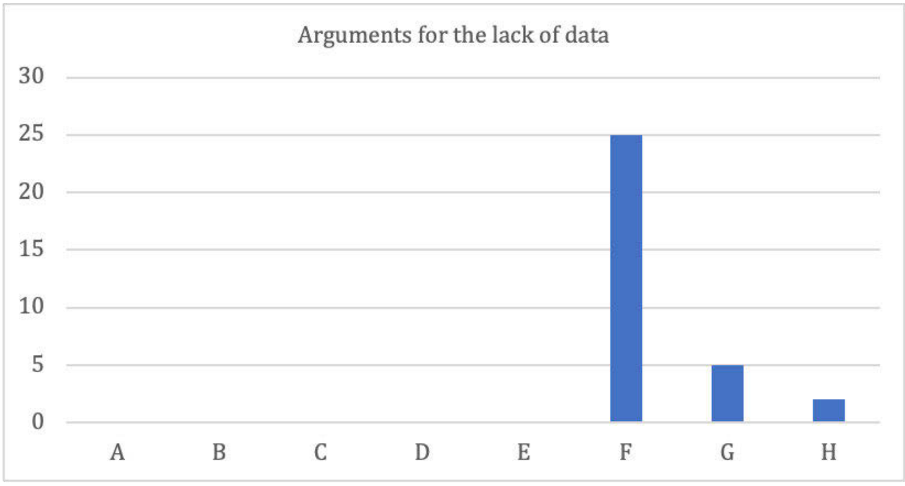


Figure 10. Arguments for the lack of data

4 CONCLUSIONS

The current social and technological development are putting under pressure water resourced due to the high demand (in quantity and quality). Therefore, the state of the resources is in risk for future generations. Also, such demand puts service managers under stress, making them to take decisions in which the sustainability or future status of the resources is not a priority. Besides that, the climate change also put pressure in the condition of water systems. Hence, its quite a challenge to manage urban water systems in a sustainable way that can contribute to the achievements of the SDGs.

With the implementation of a sustainable management, the present demand can be satisfied without compromising the capacity of the resources in the future and could lead to an optimal operation. In this research a methodology and a system of indicators was proposed that will allow to assess different hydraulics systems from a sustainable point of view. It will evaluate the contribution of the company for all the SDGs that the UN has defined to achieve a sustainable development.

As a first step of the methodology a bibliographic review was made, in which case study and guidelines were analysed. Then a detailed evaluation of the interrelation between the SDGs, targets and water systems was made. All the SDGs have a connection with urban water systems, some of these relations are more explicit than others. Based on the interrelations and some key words defined for each target, it was possible to establish a system of indicators.

The system is conformed by 135 indicators that will allow to evaluate water urban systems from a sustainable view. Of those, 110 can be implemented in supply systems, 100 are for sanitation and 87 can be used for water treatment systems. All of them comply with indicators' characteristics (IWA, 2000) such as being quantifiable, auditable, concise in meaning and universal. Also, each indicator was classified by its targets and SDGs, sustainable aspects, stage in water cycle and type of indicator.

An outline of the implementation process of the indicators was also proposed. It consisted of a series of steps such as: system characterization (information of the network and environment), assessment and diagnosis based on the sustainable indicators. Finally, an action plan can be made according to the results.

For the implementation of the methodology, two factors play a crucial role in the selection of the indicators from those already established. It will depend on factors such as the type of urban water system and its location. For networks in developing countries, the information may be a bit more complicated to obtain. In addition, some basic services indicators should be considered since they are not usually fulfilled, such as the service coverage. In developed countries this indicator could be not applied since the coverage of the services its already satisfied by a 100%. Therefore, the KPIs systems proposes implemented in a developed country could align the company's objective to the improvement of the SDGs instead of complying with the basics requirements.

The methodology was implemented in the supply system of Valencia, Spain. According to the type of system in which was used the KPIs system, only 110 indicators could be applied and only 97 could be measured. In general, 32 indicators were not applied. Of the 169 targets established by the UN, 74 are related to urban water systems. The company's contribution in the case study is from 72 targets, which is quite good.

The level of difficulty in the data collection was mostly low, meaning that was easy to access to the information. Only a few measures were classified as high difficulty. The main reason for the lack of data was because in most cases it didn't match the case of study and because for some indicators the measures were not made by the company or not relevant.

Based on the results, actions can be taken to improve the operation of the network for a more sustainable plan. Such actions can be focused on improving all the SDGs or only some of them, for example a company can choose if the target is to improve only SDG7 and create an energetic plan linked to sustainability or if should be align all the strategics plans of the company to sustainability. The proposed methodology will allow to evaluate the contribution of water services to the SDGs' targets. Therefore, it can be said that the systems can be global drivers for change in water systems. With more data, it will be possible to establish a degree of achievement the SDG to obtain sustainability levels and classified the system in a particular range. This could be a target for further studies.

5 ACKNOWLEDGMENTS

Authors greatly acknowledge financial support from the "Catedra Aguas de Valencia" in the grant "Assessing water distribution systems across indicators based on Sustainable Development Goals" for the first author of this contribution.

Grant PID2020-114781RA-I00 funded by MCIN/AEI/ 10.13039/501100011033

6 REFERENCES

[1] Romero, L., Pérez-Sánchez, M., López-Jiménez, P. (2017). Improvement of sustainability indicators when traditional water management changes: a case study in Alicante (Spain). AIMS Environmental science, 4, 502-522.

[2] Choulot, A., Denis V., Punys, P. (2012). Integration of small hydro turbines into existing water infrastructures. Hydropower-Practice and application.

[3] Cabrera E., Dane P., Haskins S., Theuretzbacher H. (2011). Benchmarking para servicios de agua. IWA publishing.

[4] Guio-Torres, D. (2008). Sustainability Indicators for Assessment of Urban Water Systems: The need for a common ground.

[5] Guppy L., Uyttendale P., Villolth K., Smakhtin V. (2018). Groundwater and Sustainable Development Goals: Analysis of Interlinkages. UNU-INWEH

[6] Hellstrom, D., Jeppson, U., Karman, E. (2000). A framework for systems analysis of sustainable urban water management. Environmental Impact Assessment Review, 20, 311-321.

[7] Hassing, J., Ipsen, N., Clausen, T. (2009). Integrated water resources management in action. DHI Water policy and UNEP-DHI center for water and environment.

[8] INE. (2021). Indicadores de la agenda 2030 para el desarrollo sostenible.

[9] Llacer-Iglesias, R., López-Jiménez, P., Pérez-Sánchez, M. (2021). Energy self-sufficiency aiming for sustainable wastewater systems: are all options being explored?. Sustainability, 13.

[10] Llacer-Iglesias, R., López-Jiménez, P., Pérez-Sánchez, M. (2021). Hydropower technology for sustainable energy generation in wastewater systems: learning from the experience. Water, 13.

[11] Loucks D., Gladwell J. (1999). Sustainability criteria for water resource systems. International Hydrology Series - UNESCO.

[12] Lundin, M. (2011). Indicators for Measuring the Sustainability of Urban Water Systems – a Life Cycle Approach.

[13] Lundin, M., Bengtsson, M., Molander, S. (2000). Life cycle assessment of wastewater systems: influence of system boundaries and scale on calculated environmental loads. Environment Science Tech, 34, 180-186.

[14] Lundin, M., Molander, S. Morrison, G. (1999). A set of indicators for the assessment of temporal variations in the sustainability of sanitary systems. Water Science Tech, 39 (5), 235-242.

[15] Lundin, M., Morrison, G. (2002). A life cycle assessment based procedure for development of environmental sustainability indicators for urban water systems. Urban water, 4. Elsevier, 145-152.

[16] Macías-Ávila, C., Sánchez-Romero,F., López Jiménez, P., Pérez-Sánchez, M. (2021) Leakage Management and Pipe System Efficiency. Its Influence in the Improvement of the Efficiency Indexes. Water, 199 (13)

[17] Maurya, S., Singh, P., Sing, A. (2020). Identification of indicators for sustainable urban water development planning. Ecological indicators, 108.

[18] Mercedes-García, A., Sánchez-Romero, F., Pérez-Sánchez, M., López Jiménez, P. (2021) Objectives, Keys and Results in the Water Networks to Reach the Sustainable Development Goals. Water, 1268 (13).

[19] Morrison, G. Fatoki, O., Zinn, E., Jacobsson, D. (2001). Sustainable development indicators for urban water systems: a case study evaluation of Williams Town, South Africa. Water SA, 27 (2).

[20] Pardo, M., Manzano, J., Cabrera, E., García-Sierra, J. (2002). Energy audit of irrigation networks. Biosystems engineering, 115. Elsevier, 89-101.

[21] Popawala, R., Shah N. (2011). Evaluation of sustainability index for urban water management system. 2nd International Conference on Environmental Science and Development, 4.

[22] Bagheri, A., Hjorth, P. (2006). A framework for process indicators to monitor sustainable development: practice to an urban water system. Development and Sustainability, 9, 143-161.

[23] Short A., Milman A. (2008). Incorporating resilience into sustainability indicators: An example for the urban water sector. Global Enviromental Change, 18 (4), 758-767.

[24] Spiller M. (2016). Adaptive capacity indicators to assess sustainability of urban water systems – Current application. Science of the total environment., 569–570 (1), 751-761.

[25] UN-Water. (2021). The UN water development report 2021, valuing water.

[26] UN-Water. (2021). Summary progress update 2021: SDG6 – water and sanitation for all.

[27] Van der Graaf J., Meester-Broertjes H., Bruggeman W. and Vles E. (1997). Sustainable technological development for urban water cycles. Water Science Tech, 35 (10),213-220.

WDSA CCWI 2022
2^{nd} International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14740

MODELLING CONSUMERS IN INTERMITTENT WATER SUPPLIES: A COMPARATIVE REVIEW OF EPANET-BASED METHODS

Omar Abdelazeem[1] **and David D. J. Meyer**[2]

[1,2] University of Toronto, Toronto, Ontario, (Canada)

[1] *o.abdelazeem@mail.utoronto.ca*, [2] *david.meyer@utoronto.ca*

Abstract

Intermittent Water Supply (IWS) networks, which pressurize for less than 24 hours/day, affect 1 billion people worldwide and are associated with increased water contamination and inequitable distribution. Due to these downsides, various methods to model consumers and understand their behaviour were proposed. We found more than 8 different methods of modelling IWS consumers, but we know of no comparative analyses of these methods nor their efficacy. This study comparatively reviews methods of modelling IWS consumers implemented in EPANET, due to their prevalence, reproducibility, and accessibility. Methods of modelling IWS consumers were found to fall into three groups based on their assumed consumer withdrawal behaviour: unrestricted, flow-restricted, and volume-restricted methods. We applied each method to three reference networks and compared the methods' performance after subjecting each reference network to common IWS improvement strategies, including changing the supply duration and/or source pressure. Flow-restricted methods assume consumers withdraw their demands at a constant rate, leading to unrealistic predictions when subjected to unexpected changes in supply conditions. Volume-restricted methods assume consumers withdraw at the highest, hydraulically feasible rate until their storage tanks fill. This assumption highlights pronounced inequality between consumers, as consumers advantaged by source proximity and/or elevation receive their demands faster and earlier. Our results demonstrate that the simulated behaviour of IWS depends substantially on the type of consumer model employed. Presented examples demonstrate that consumer model selection can change the simulation-predicted optimal strategies for coping with and improving IWS. IWS modelling methods should reflect the consumer behaviour in the modelled network and the model's intended use.

Keywords
Intermittent Water Supplies (IWS), Water Distribution Networks (WDN), Modelling, EPANET.

1 INTRODUCTION

Intermittent Water Supply (IWS) refers to water distribution networks that do not operate for 24 hours/day. This type of networks serves approximately 1 billion people around the world, mostly concentrated in the global south [1]. Intermittent operation is seldom in place by design but is often the result of utilities attempting to cope with water stress, lack of sufficient funds, and/or deterioration of infrastructure [2]. IWS networks distribute water inequitably among their consumers [3] and this inequality is exacerbated by their unplanned nature. IWS also degrades water quality during distribution [4]. Due to IWS' drawbacks, many IWS networks do not meet the specifications set for target 6.1 of the United Nations (UN) Sustainable Development Goals (SDG), which targets "universal and equitable access to safe and affordable drinking water for all" [5]. Thus, it is of high interest to improve service consistency and quality in IWS contexts if universal access to safely managed water supplies (SDG 6.1) is to be achieved.

To achieve SDG 6.1 and its national counterparts, utilities, researchers, and regulators have attempted to optimize intermittent operation to improve service and move towards continuous supply (e.g., [6]–[9]). But for IWS optimization efforts to yield practical value, optimizations must be founded on a modelling method that meaningfully represents consumer behaviours. Ideal

CC BY NC SA

representations of consumer behaviour would capture how consumers respond to IWS improvements such as changes in supply duration and/or pressure. Several methods have been proposed to capture the distinct features of IWS networks. These methods are heterogenous in their scope (steady-state operation vs. filling and draining of the network), their platform (open-source software like EPANET and SWMM vs. methods independent of available software) and their assumptions about consumer behaviour. Despite such diversity in modelling methods, we know of no qualitative or quantitative comparisons of these methods nor of their ability to usefully represent consumer behaviours and network conditions.

To provide guidance to utilities and researchers interested in modelling IWS, this paper aims to review and compare modelling methods that can be applied to IWS. We demonstrate each model's representation of consumers and their demand withdrawal behaviour in three test networks, subject to various practically motivated supply conditions. To do so, we:

1. Cluster similar methods of modelling IWS into groups based on their assumptions about the consumer (Section 2);
2. Summarize the test networks used, our methods and scenarios for comparison (Section 3);
3. Demonstrate how each method's assumptions about consumers affect the simulated network and the variations in consumer demands delivered (Section 4), and lastly;
4. Recommend methods best suited for utilities and researchers interested in improving IWS networks (Section 5).

2 LITERATURE REVIEW

The literature on IWS modelling, while relatively scarce, varies widely in scope, approach, and utilized tools. A majority of the reviewed literature (e.g., [10]–[13]) focuses on predicting the steady-state (post-pressurization) network conditions, while other efforts (e.g. [3], [14]) focus instead on describing the filling (pressurization) process. Several models of filling adapt the EPA's Storm Water Management Model (SWMM) (i.e., [14]–[17]) while others were constructed from scratch, providing greater flexibility in model construction often at the expense of reproducibility and ease of use (i.e., [3], [17], [18]). Of the steady-state modelling methods, most use EPANET due to its prevalence in modelling WDNs in general and its open-access nature. This review compares such EPANET-based methods of modelling IWS. A review of methods for modelling the filling of IWS networks – while important in ascertaining when consumers start receiving their demands (and therefore the total volume they can receive during a supply cycle)- is left for future work.

Some aspects of the steady-state performance of IWS can be modelled using methods that were not specifically formulated to represent IWS networks and/or consumers. We include such methods in our review, even though this inclusion may stretch such methods beyond their proposer's intent. One such category of methods aims to model pressure-dependent demands. In many IWS networks, consumer withdrawals are affected by pressures in the network [15]. This dependence of demand on adequate pressure distinguishes most IWS from their Continuous Water Supply (CWS) counterparts, where it is typically assumed that sufficient pressure is provided. Since EPANET was originally built to model CWS, its solver was not outfitted with a Pressure-Driven Analysis (PDA) option and operated exclusively on a Demand-Driven Analysis (DDA) basis, until 2020 [19].

Pressure-dependent demand in this context describes a method of modelling consumer withdrawals that includes the sensitivity of withdrawals to the available pressure in the network. In its general form, a pressure-dependent demand is typically defined by three regimes:

a) Below a minimum head threshold, no water flows to the consumer,

b) Above a desired head threshold, the consumer receives their desired demand,

c) Between the minimum and desired thresholds, flow to a consumer depends on the available head, and the minimum and desired head thresholds.

This pressure-dependence was first formulated as a relationship between head and flow rate by [20]. Wagner et al. [21] proposed the form used by the majority of later studies, an nth root relationship:

$$Q_j = 0, \quad if\ H_j \leq H_j^{\min} \tag{1a}$$

$$Q_j = Q_j^{\text{des}}\ , \quad if\ H_j \geq H_j^{\text{des}} \tag{1b}$$

$$Q_j = Q_j^{\text{des}}\ (\frac{H_j - H_j^{\min}}{H_j^{\text{des}} - H_j^{\min}})^{\frac{1}{n_j}},\ if\ H_j^{\text{des}} > H_j > H_j^{\min} \tag{1c}$$

Where Q_j is the actual demand withdrawn at node j. Q_j^{des} is the desired consumer demand, and H_j, $H_{\text{j}}^{\min}$ and H_j^{des} are the head available at node j, the minimum required head for flow to start at node j, and the minimum head required to deliver the consumer's desired demand at node j, respectively, and n_j is an exponent that is characteristic of node properties [21].

Earlier efforts to implement this pressure dependence in EPANET included iterative approaches (e.g., [22], [23]), and modifying EPANET source code or extending its solver (e.g., [24], [25]). Later efforts achieved comparable results by using model elements native to EPANET to modify demand nodes and account for pressure dependence. These later approaches provided higher accessibility and ease of use since they can be used within EPANET's Graphical User Interface.

All methods of modelling steady-state IWS networks include Pressure-Dependent Demand (PDD). Methods differ, however, in their formulation of this PDD relationship and its implementation in EPANET, as well as their assumed consumer behaviour in withdrawing demands. We group IWS modelling methods based on three types of demand withdrawal assumptions: Flow-restricted, Volume-restricted, and Unrestricted methods (Figure 1).

2.1 Flow-restricted methods

Flow-restricted methods of modelling networks assume that as long as pressure is greater than the desired pressure ($H_j > H_j^{\text{des}}$), withdrawal flow rates are independent of pressure (i.e., Equation 1b). When these methods are employed to model IWS, their predictions correspond to consumers that consciously restrain their withdrawal to receive only what the modeller expects, which is often what they need. When modelling IWS, the expected withdrawal rate (i.e., desired demand Q^{des}) is often defined as the total daily demand volume, spread out over the supply duration. We found three distinct ways in which flow-restricted methods have been implemented in EPANET (Figure 1).

EPANET-PDA: As of 2020, EPANET has been outfitted with a Pressure-Driven Analysis (PDA) option built natively in the source code and GUI. Thus, we include "EPANET-PDA" as one of the flow-restricted methods in our analysis. The EPANET-PDA method is used in the current conference's Battle of Intermittent Water Supply (BIWS).

Prior to the release of the native PDA option, however, flow-restricted methods were implemented in EPANET using Flow Control Valves (FCVs) to restrict the maximum flow (Q^{des} in Equation 1b). FCVs In EPANET restrict flow from the original (demand-driven node) to a new node, either a reservoir (FCV-Res) or an emitter (FCV-EM).

FCV-Res: When an FCV is connected to a downstream reservoir (FCV-Res) flow at intermediate pressure heads (i.e., between $H_{\text{j}}^{\min}$ and H_j^{des}) is governed by the reservoir's elevation and the friction between the original node and the reservoir. The reservoir elevation is set to be above the original node's elevation by exactly the minimum required head (H_j^{min}) [26], [27]. Gorev and Kodzhespirova [27] improved on [26]'s convergence by adjusting the resistance of the connection between the original node and the reservoir, by setting the connecting pipe's minor loss coefficient as:

$$k_j = (H^{des} - H^{min})\frac{g}{8}\left(\frac{\pi D^2}{Q^{des}}\right)^2 \tag{2}$$

where k_j is the minor loss coefficient, D is the pipe diameter, and other variables as defined before.

FCV-EM: FCVs have also been used in conjunction with emitters, rather than reservoirs. In this case, the nth root relationship (Equation 1c) is represented by the emitter's coefficient and the emitter's elevation is set to account for H_j^{min} similar to the reservoir [28]. The emitter's exponent is determined as:

$$C_d = \frac{Q^{des}}{(H^{des} - H^{min})^{\frac{1}{n_j}}} \tag{3}$$

Mahmoud et al. [29] modified this approach to selectively add the artificial elements only to pressure-deficient nodes to improve efficiency [29].

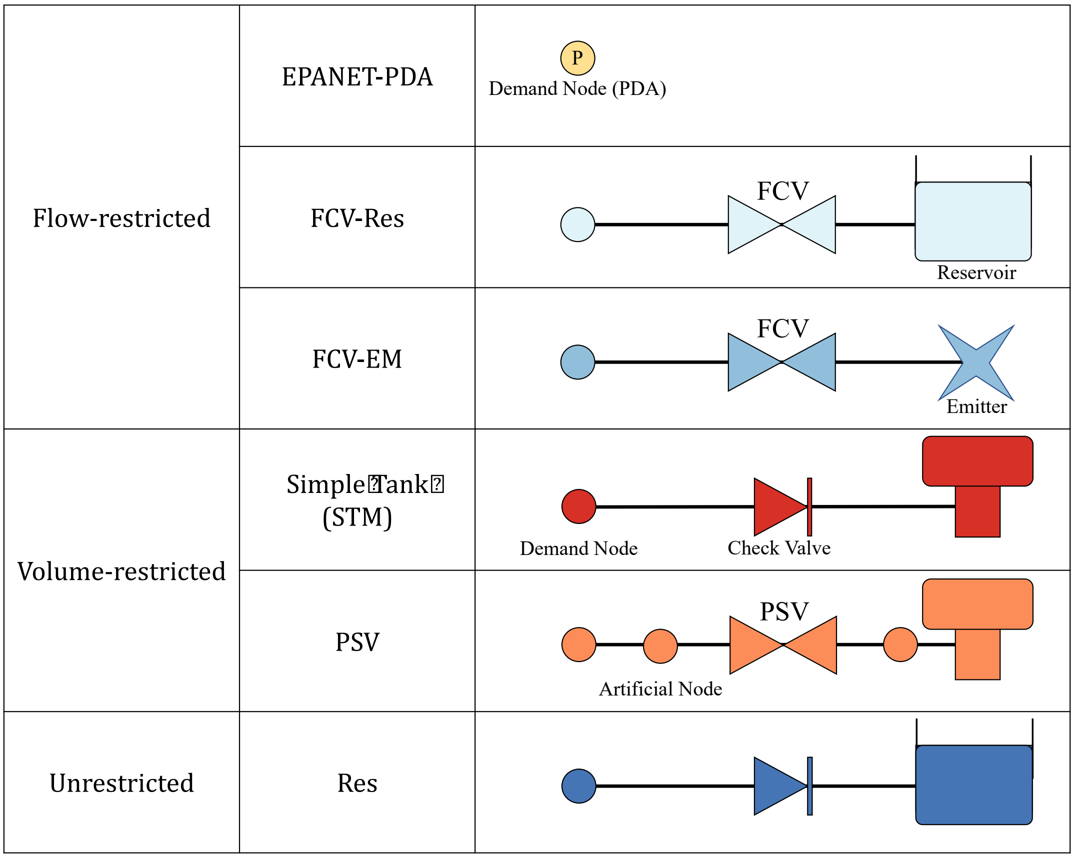

Flow-restricted
EPANET-PDA
P
Demand Node (PDA)
FCV-Res
FCV
Reservoir
FCV-EM
FCV
Emitter
Volume-restricted
Simple Tank (STM)
Demand Node
Check Valve
PSV
PSV
Artificial Node
Unrestricted
Res

Figure 1: EPANET representation of flow-restricted methods: EPANET's native PDA option (EPANET-PDA), FCVs with downstream reservoirs (FCV-Res), and FCVs with downstream emitters (FCV-EM); volume-restricted methods: Simple Tank Method (STM) and Pressure-Sustaining Valve method (PSV); and the unrestricted method using reservoirs (Res).

To test these methods, we implemented FCV-Res in accordance with [27] due to its improved convergence and we implemented FCV-EM in accordance with [28] since all nodes in an IWS network should be treated as pressure-deficient. Both FCV methods were compared with EPANET's (newly) native PDA option. The EPANET representation of each of these flow-restricted methods is shown in Figure 1.

2.2 Volume-restricted methods:

IWS consumers typically compensate for expected supply interruptions by storing water [30], [31]. Consumer storage in IWS networks transforms (integrates) their desired flow rate into an equivalent desired volume, V^{des}, over the supply cycle (from the start of one pressurization to the next) [11]. Consumers with storage need not withdraw water at the same rate as they desire to use it. Instead, volume-restricted methods assume that consumers withdraw water at the highest rate they can (hydraulically determined by the pressure in the network and service connection resistance) until they have received their demanded volume, after which they shut off their connection Volume-restricted methods have been implemented in EPANET using its tank element to represent this volume restriction. Specifically, the tank's volume corresponds to the desired demand volume, V^{des}. We found two EPANET-based methods of implemented volume-restricted flow.

Simple Tank: The Simple Tank Method (STM) was first proposed by Batterman and Macke [10] and later systematized and adapted by Taylor et al. [11]. The use of simple tanks was also observed in [32] and suggested by [15]. In STM, consumers are represented as a tank with a volume equal to V^{des}, The tank's elevation is set as the sum of the original node's elevation and $H_{\mathrm{j}}^{\mathrm{min}}$, equivalent to the reservoir and emitter elevations in the FCV-Res and FCV-EM methods. Consumer tanks are set to a uniform, nominal height (often of 1 metre to simplify postprocessing). The pipe connecting the original demand node to the consumer tank is equipped with a check valve to prevent backflow and the pipe's minor loss is adjusted as in Equation 2. As each tank fills, its pressure head slowly increases, which slows the withdrawal rate.

PSV: Sivakumar et al. [33] expanded the STM by adding a Pressure Sustaining Valve (PSV) upstream of the tank to ensure the upstream pressure is constant at atmospheric pressure. The PSV enables a flow rate expected of a tank filling from the top – as opposed to STM's filling from the bottom. The result of this distinction is that tanks are subjected to a constant head differential in the PSV method, therefore the withdrawal rate is constant until the tank is full. Compared to the STM, tanks in the PSV method should have an elevation that is lower by the tank's height.

2.3 Unrestricted method(s)

Res: Lastly, Mohapatra et al. [13] adopted an approach that is neither restricted by volume nor flow rate. This "Unrestricted" method assumes that the consumer will withdraw at the maximum possible rate and maintain it for the entire supply. This is achieved by replacing the tank element (volume-restricted storage) in the STM method with a reservoir element (unrestricted storage). This formulation would hold true if consumers were known to leave their taps open regardless of withdrawal volumes.

3 METHODS

To compare each method's predictions of steady-state behaviour in IWS networks, we implemented each method in three reference networks (Figure 2). Network 1 (6 demand nodes) is a single source, two-loop network introduced by [22]. Networks 2 (3 reservoirs, 64 demand nodes) and 3 (4 reservoirs, 245 demand nodes) are models based on the WDNs of Pescara and Modena, Italy, respectively introduced by [34]. The networks are numbered in order of increasing complexity. To consistently compare the predictions and performance of the modelling methods, we harmonized their assumptions about the minimum and desired pressures ($H_{\mathrm{j}}^{\mathrm{min}}$ and H_{j}^{des}).

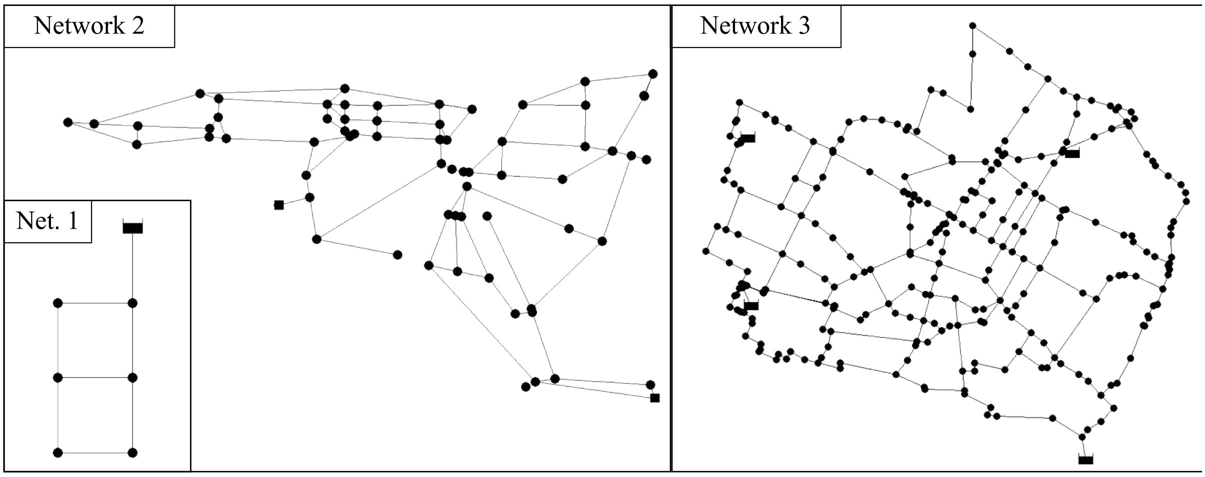


Figure 2: Networks used in method testing. Network 1 is a single-source, two-loop network with 6 Demand Nodes, Network 2 (Pescara, Italy) has 3 source reservoirs and 64 demand nodes, while Network 3 (Modena, Italy) has 4 source reservoirs and 245 demand nodes

In theory, the true value of these parameters depends on the degree of network skeletonization, and the resistance expected from the consumers' service connections. Literature on their appropriate value for various IWS contexts is starkly absent [35] and so these values are set at 0 and 10 metres for $H_{\mathrm{j}}^{\mathrm{min}}$ and H_j^{des} respectively. The exponent n_j was set at 0.5 for all nodes in all methods and in all networks. These assumptions about pressure head thresholds and exponents match those used for the BIWS network.

In EPANET-PDA, values for $H_{\mathrm{j}}^{\mathrm{min}}$ and H_j^{des} are directly input into EPANET's hydraulic options [36]. For the remaining methods, these values are represented in the connecting pipe's minor loss calculated as in Equation 2, except for FCV-EM where they are represented in the emitter's coefficient as in Equation 3. The value of Q^{des} used in equations 2 and 3 for all methods was set as the flow rate needed to satisfy consumer demands over the supply duration – which for most runs is 12 hours unless stated otherwise.

The analysis of all methods was conducted using the Water Network Tool for Resilience (WNTR) v. 0.4.1 [37] python package in Python 3 on a 2020 MacBook Pro M1. The speed of each method in each network was averaged across 1,000 timed runs. Instances where the difference in efficiency could potentially cause a considerable difference in user experience (time difference >50%) are noted below. To compare the predicted hydraulic behaviour of each method, we computed its predicted consumer demand satisfaction, S, defined as the ratio between the volume delivered to consumers and the total volume desired for all consumers during the supply cycle:

$$S = \frac{V^{\mathrm{sup}}}{V_T^{\mathrm{des}}} \tag{4}$$

where V^{sup} is the total volume supplied to consumers and V_T^{des} is the total demand volume for all consumers.

For volume-restricted methods, the total delivered volume was computed as the sum of the volumes stored in the tank at the end of supply, while for flow-restricted and unrestricted methods, it is computed as the product of the demand withdrawal rate and the supply duration. An equivalent satisfaction ratio for each node was computed in a similar fashion, to allow for an investigation of variance (if any) in demand delivered to consumers and subsequently, an investigation of supply equality. Consumers in a flow-restricted model are limited by their desired flow rate Q^{des}, which in imposes an upper limit on the average satisfaction ratio of the network defined by:

$$S \leq \frac{Q_T^{\mathrm{des}}}{V_T^{\mathrm{des}}} t \tag{5}$$

where Q_T^{des} is the sum of all desired demands and V_T^{des} is the sum of all desired demand volumes and t is the time since supply started (supply time).

A useful modelling method ought to be able to produce realistic results under "non-design" conditions, especially if modelling is to inform and guide the optimization of the operation and/or improvement of IWS networks (e.g., BIWS). We demonstrate and compare the methods' predictions under two such non-design conditions in which supply duration and pressure are varied.

Supply pressure is expected to be one such "non-design" condition of interest to utilities and/or researchers. Intermittent operation is often caused by water scarcity or insufficient funding [2], hence, low reservoir levels and/or energy shortages affecting pump operation are expected to frequently reduce source pressure. Thus, each method was also tested on each network under reductions of source pressure by 25, 50 and 75%, where source pressure was defined as the head differential between the highest reservoir and the highest demand node's elevation. Similarly, supply duration is expected to be a "non-design" condition of interest since variations in supply durations are common in IWS [38]. Here, while reference networks are configured assuming a supply duration of 12 hours, methods were tested under unexpectedly shortened supply (10 hours) and lengthened supply (14 hours) durations.

To exemplify the distinctions between the modelling methods, two scenarios of practical relevance to the operation/management of IWS are constructed. In the first scenario, a network (which satisfies consumer demands in 12 hours) is occasionally constrained to operate only for 10 hours a day (e.g., due to electricity blackouts). The utility is therefore considering a proposal to augment its source pressure during the 10 hours of supply (e.g., by installing an additional pump). In the second scenario, a network (which also satisfies in 12 hours) suddenly faces reduced source pressure (e.g., due to a pump failing). The utility is therefore considering a proposal to mitigate this reduced pressure by lengthening its supply duration (e.g., by running the operational pumps for longer).

Table 1: Description of the improvement scenarios used to compare modelling methods

Scenario	Problem	Improvement	Proposed Mitigation
1	Reduced to 10-hour supply 100% Source Pressure	Increase Source Pressure	10-hour supply 125% Source Pressure
2	12-hour supply Reduced to 25% Source Pressure	Lengthen Supply Duration	14-hour supply 25% Source Pressure

In each scenario, the hypothetical utility would evaluate the proposed mitigation based on the modelled gains (or lack thereof) in consumer demand satisfaction (Equation 4). Thus, we compared the decisions made by the utility in each scenario, depending on the modelling method they employed.

4 RESULTS & DISCUSSION

Based on the quantitative performance of the simulated modelling methods, we first compare method performance both within groups (e.g., flow-restricted methods) and between groups. Next, we inspect the predictions for the extreme consumers (10th and 90th percentiles) across different supply conditions (changed supply duration and pressure). Lastly, we investigate the implications of the different methods when used to model the improvement of IWS networks.

4.1 Types of IWS models

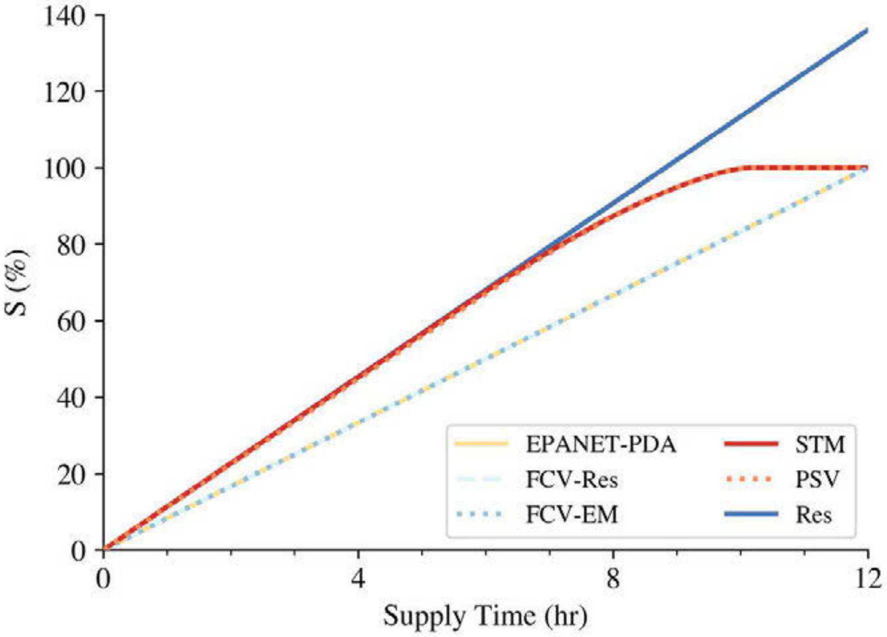


Figure 3: Demand satisfaction, S, in Network 3 vs. supply time for flow-restricted methods: EPANET-PDA (solid yellow), FCV-Res (dashed light blue) and FCV-EM (dotted light blue), and volume-restricted methods: STM (solid red) and PSV (dotted orange), and the unrestricted Res method (solid blue). The results of methods within the aforementioned groups are shown to be virtually identical on the aggregate level. A Significant difference is observed between groups as expected. Flow-restricted methods show a controlled rate bounded by the flow restriction; volume-restricted methods start withdrawing at a higher rate before levelling off due to tanks becoming full (volume restriction) while the unrestricted Res continues to withdraw at the maximum rate.

The average consumer satisfaction predicted by the simulated methods strongly supports our grouping of them (Figure 3). Flow-restricted methods exhibited controlled and constant withdrawal, as expected, since consumer withdrawals are bounded by the flow restriction. When supplied with ample pressure, flow-restricted consumer satisfaction follows a straight line defined by an intercept of zero and a slope equal to Q^{des}/V^{des} (e.g., see Figure 3). Flow-restricted methods produced nearly identical results (<0.01% difference) once assumptions around minimum pressure (H^{min}) and desired pressure and demand (H^{des} and Q^{des}) are harmonized. This close alignment between the historical methods of flow restriction (FCV-Res and FCV-EM) and the PDA option in EPANET 2.2 demonstrates the adequacy of earlier methods in representing pressure dependence. Of note, EPANET's PDA option is approximately 1.5 times faster in Network 1 and 2.5 times faster in Network 3. While the total runtime for all networks remains below 30 milliseconds per run, if a network with low skeletonization or a considerably larger scale is of interest (e.g., the BIWS network has 2,859 nodes compared to Network 3's 245), the computational cost would likely scale non-linearly, and the difference in modelling efficiency may become significant to the modeller.

Both unrestricted and volume-restricted methods predicted faster consumer withdrawals than flow-restricted methods while the average satisfaction in the network is relatively low (e.g., S <70% in Figure 3). Thereafter, some tanks start to become gull (i.e., the number of unsatisfied consumers dwindles), slowing average withdrawal rates. Both volume-restricted methods (STM and PSV) agreed closely on the average consumer satisfaction as long as pressures in the network were considerably higher than the tank heights as noted by [33], [35]. For example, their predictions were <0.7% different in Figure 3 where network pressures were 20-30 metres compared to tank heights of one metre.

The unrestricted Res method's predictions overshoot the desired demand for all consumers when supply durations are longer than hydraulically required to satisfy consumers. For example, when Network 3 was supplied for 12 hours, the average consumer satisfaction was 136% with a range of ~110-180% (Figure 3). The absence of restrictions on consumer storage may also lead to unrealistic results in networks where consumers have widely differing elevations and/or proximities to the source; hydraulically advantaged consumers could be predicted to receive much more than they can physically store, even if the average consumer satisfaction was <<100%.

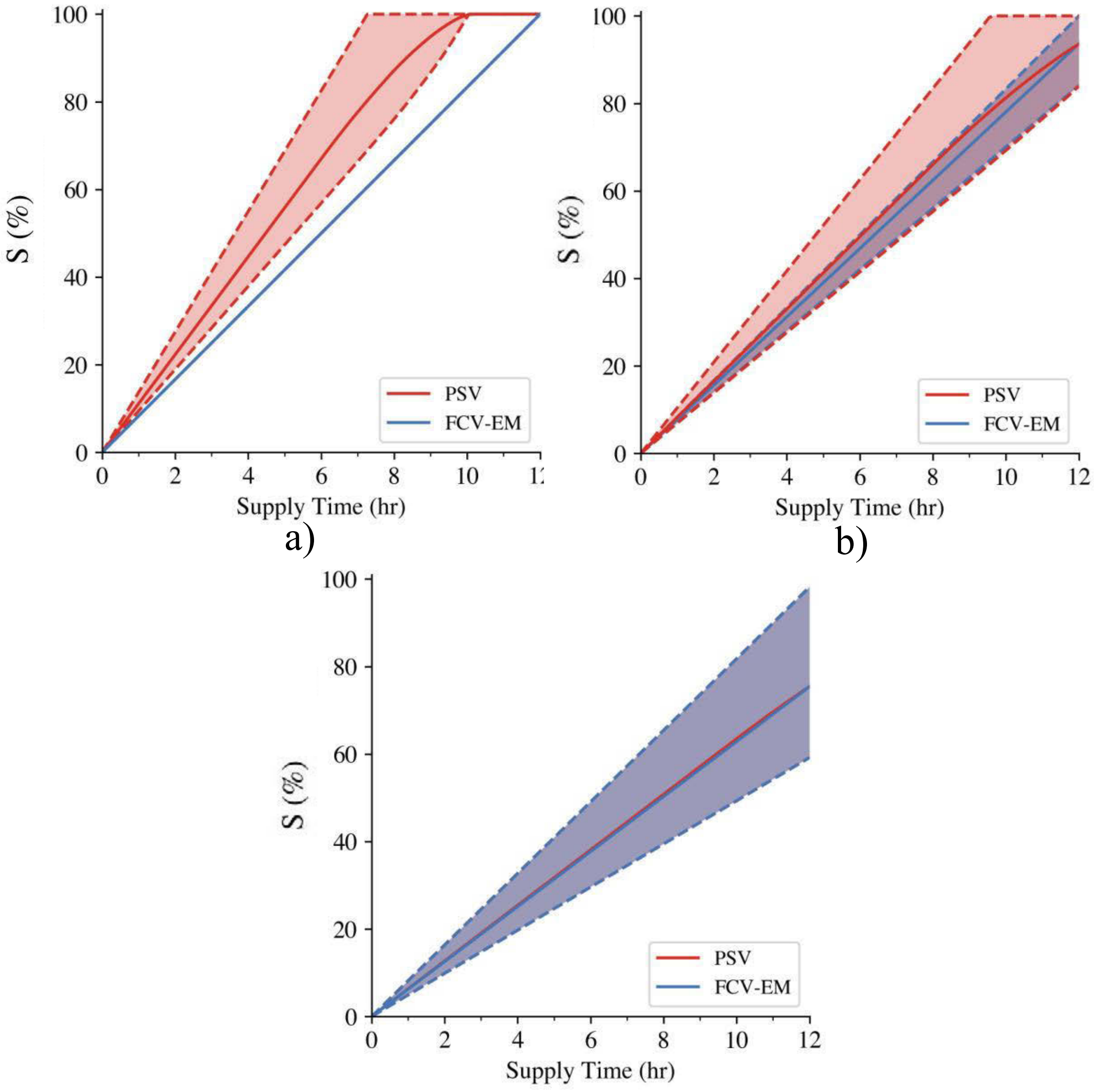
All flow- and volume-restricted methods on the total satisfaction ratio of the consumers at the end of the planned supply, provided water is supplied for the expected duration (12 hours in Figure 3). Before the end of the planned supply, however, flow- and volume-restricted methods differ in the predicted consumer satisfaction levels. In networks where some consumers are hydraulically advantaged over others, consumer satisfaction (100%) is achieved faster because demand is hydraulically staggered. Advantaged consumers, in volume-restricted models, fulfil their demands earlier, enabling higher pressures for disadvantaged consumers. More generally, variance in withdrawal rates modelled by volume-restricted methods has important implications for the equality of consumer withdrawals in IWS networks.

4.2 Supply distribution between consumers and inequality

Flow- and volume-restricted methods predict similar average levels of consumer satisfaction across a range of pressures when the network is supplied for the expected duration. At very low pressure, when all pressures are lower than the desired head H_j^{des}, both methods also agree on the distributions of consumer satisfaction since all consumers are hydraulically limited to a rate lower than the flow restriction (e.g., Figure 4c). As pressures increase, consumer withdrawals are increasingly restricted in flow-restricted methods, which imposes an upper limit on the predicted variance in consumer satisfaction (e.g., Figure 4b vs. 4a). Flow-restricted methods, by construction, exhibit complete uniformity between consumers who receive sufficient pressures ($>H_j^{\mathrm{des}}$). Studies of flow variation due to pressures above the desired pressure were not intended, and subsequently are not captured, by flow-restricted methods.

IWS consumers have been observed to actively seek measures that enable them to withdraw their demands as fast as possible. One estimate suggests that 25% of consumers in IWS use private (suction) pumps to actively pull water out of the network faster than it would otherwise flow [39]. In many IWS networks, and especially where suction pumps are prevalent, consumers tend to withdraw and fulfil their demands as quickly as possible. In such contexts, volume-restricted methods more accurately account for rapid withdrawals and therefore demand distribution between consumers than their flow-restricted counterparts.

Water networks, and especially IWS networks, should be evaluated based on more than their average performance. Equality of access to safe and affordable drinking water features prominently in many national and global policy goals. To realize such goals, tools that enable us to understand and mitigate inequalities under IWS are needed. Volume-restricted methods of modelling IWS capture more of this inequality than their counterparts and are therefore more useful for utilities and researchers seeking operational opportunities to maximize equality in IWS networks.



c)

Figure 4: Comparing a flow-restricted (FCV-EM) and a volume-restricted (PSV) method. Mean (solid), 10th and 90th percentile (shaded) of consumers' demand satisfaction for Network 3 under a) 100% source pressure, b) 50% source pressure, and c) 25% source pressure. As the source pressure decreases, the variance between consumers' satisfaction increases. By construction, flow- and volume-restricted methods agree when all consumers are pressure deficient. When at least some consumers have pressures higher than H_j^{des}, volume-restricted methods predict greater inequality since they are not restricted by the upper bound of flow-limited methods (defined in Equation 5).

4.3 Modelling the improvement of IWS

The choice of modelling method can have important practical implications for utilities using models to evaluate possible operational changes. To demonstrate, we contrast how differently flow- and volume-restricted methods would evaluate the same strategies to improve network performance in two scenarios.

BY NC SA

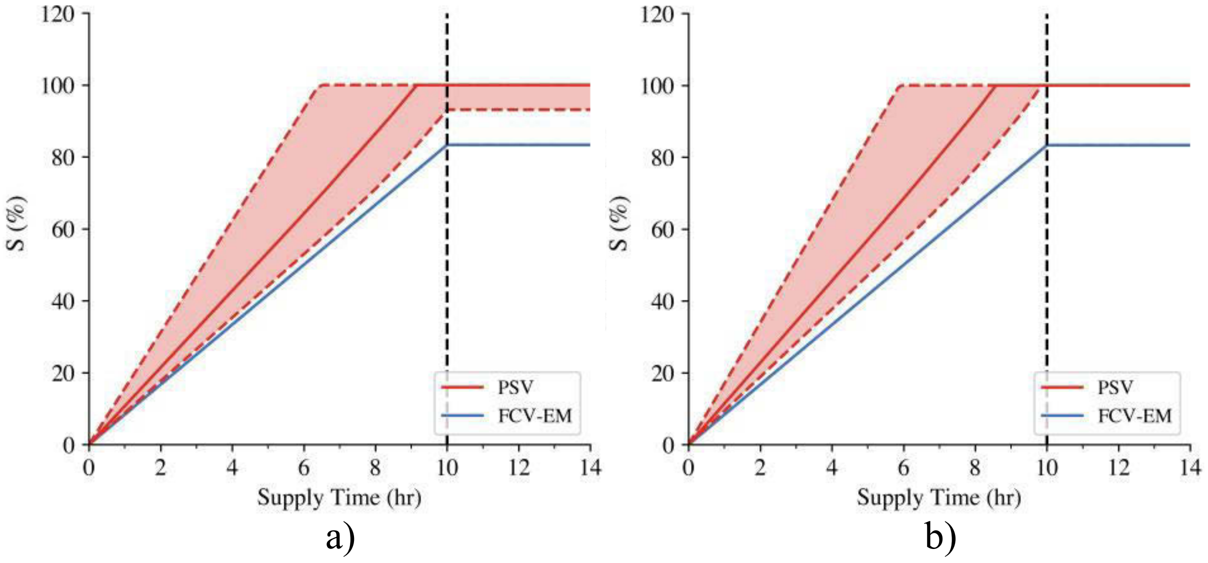


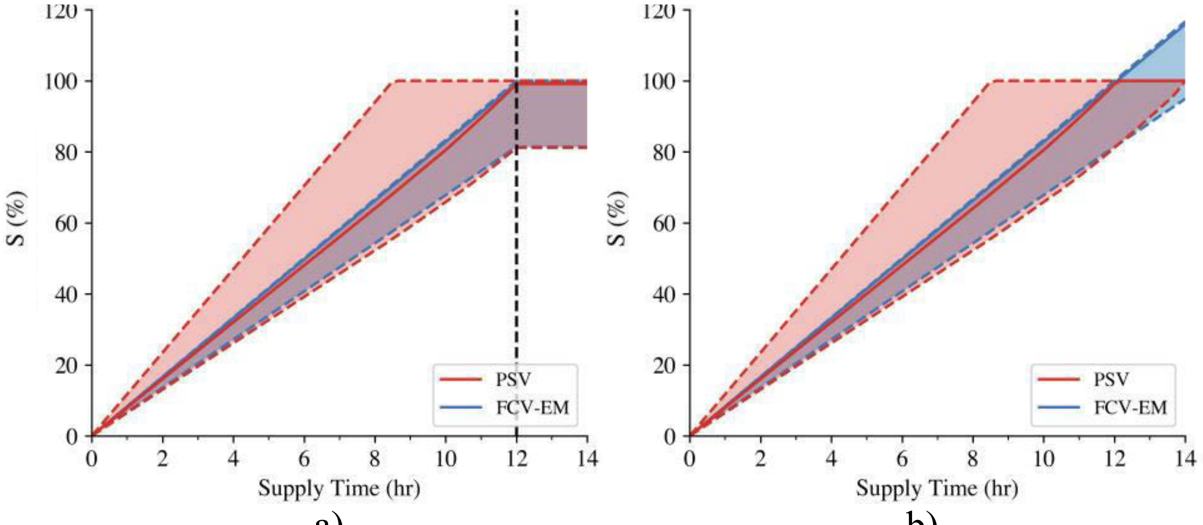
Figure 5: Comparing a flow-restricted (FCV-EM) and a volume-restricted (PSV) method. Mean (solid), 10th and 90th percentile (shaded) of consumers' satisfaction for Network 2 operating for a) 10 hours under normal source pressure and b) 10 hours under 125% source pressure. Dashed black line indicates the end of supply. Network 2 was configured assuming supply would last for 12 hours. Volume-restricted methods predict an increase in consumer satisfaction, while flow-restricted methods predict no change. The different assumptions about consumer behaviour made by the methods inform the utility towards opposing decisions.

In Scenario 1, a water utility is occasionally forced to operate their IWS network for 10 hours a day, instead of its normal 12 hours a day (Figure 5a). Both methods predict that after 10 hours of supply, the network is not yet satisfied. Flow-restricted methods predict that all consumers will be equally inconvenienced and receive only 83.3% (10/12) of their desired demand (Figure 5a). Contrastingly, volume-restricted methods predict that only a few disadvantaged consumers will bear the brunt of the shortage. To improve consumer satisfaction during these occasional restrictions in supply duration, the utility is considering augmenting its source pressure by 25%. Flow-restricted and volume-restricted methods differ drastically in their evaluation of this proposed pressure increase (Figure 5b). In this scenario, consumer withdrawals are constrained by supply duration, not pressure, since all consumers in the modelled network have pressures higher than their desired head H_j^{des}. As such, flow-restricted methods predict no improvement in consumer satisfaction (Figure 5b). Volume-restricted methods, however, predict that increased pressure would lead to higher consumer withdrawals and satisfaction (Figure 5b). Thus, the utility's choice of modelling method would change their evaluation of the proposed pressure increase; the proposed improvement could be deemed effective if a volume-restricted method was used, but ineffective if a flow-restricted method was used.

In the second scenario, a similar network that operates for 12 hours faces occasional pressure deficiencies (25% Source Pressure). Both flow- and volume-restricted methods agree on the unmitigated impact this pressure reduction would have on consumers, i.e., a portion of consumers cannot satisfy their demand (Figure 6a). To address this pressure deficit, the utility considers lengthening the supply to 14 hours per day to improve consumer satisfaction. Given the increased supply duration, volume-restricted methods predict that all consumers will be satisfied by the end of supply (Figure 6b). Flow-restricted methods, on the other hand, predict that some consumers will still see their demands unsatisfied, even when most consumers receive more than their desired demand (Figure 6b). Flow-restricted methods also predict that the total water supplied to consumers will be greater than their collective demand. In a water-scarce context, this predicted increase in total water supplied could discourage the utility from adopting the increase in supply duration. Similarly, any equality-focused utility might decide that supply

BY NC SA

durations longer than 14 hours are required based on a flow-restricted method when a volume-restricted method suggests otherwise.



a) b)

Figure 6: Comparing a flow-restricted (FCV-EM) and a volume-restricted (PSV) method. Mean (solid), 10th and 90th percentile (shaded) of consumers' demand satisfaction for Network 2 operating for r a) 12 hours under 25% source pressure and b) 14 hours under 25% source pressure. Dashed black line indicates the end of supply. Network 2 was configured assuming supply would last for 12 hours. While both methods predict a gain in consumer demand satisfaction, flow-restricted methods predict that some consumers remain unsatisfied and predict a higher outflow from the source.

In both scenarios, using a different type (group) of modelling method could lead the utility to a different decision. We suggest these scenarios highlight two major limitations of flow-restricted methods when modelling IWS. When most or all pressures in an IWS network are sufficient, flow-restricted methods artificially obscure inequality between consumer withdrawals (and potentially satisfaction). When IWS networks increase their supply duration beyond the duration used when setting flow0restriction settings, flow-restricted methods can predict withdrawn volumes larger than consumer demand. Thus, we recommend volume-restricted methods over flow-restricted ones in three IWS modelling circumstances:

1. When the equality of consumer withdrawals (rather than the average) is a key concern,
2. When network pressures are higher than (or expected to increase beyond) the desired pressure, and/or
3. When the effect of supply duration changes is of interest.

Conversely, we note that flow-restricted methods can enable the simulation of multiple supply cycles if set up with appropriate patterns and control rules (as seen in the BIWS network), which cannot be done in currently available volume-restricted methods. However, since flow-restricted methods do not model consumer storage, they are unable to capture how an unsatisfied demand in one cycle affects consumer behaviour over the following cycle (if such an interaction existed).

Hydraulic models of IWS can be used to evaluate and/or optimize opportunities to improve IWS network performance. Successful improvements to most IWS networks would result in increased pressure and supply duration. Hence, we strongly recommend that flow-restricted modelling methods be avoided when evaluating and/or optimizing IWS networks. If based on flow-restricted models, optimized IWS improvement strategies are likely to underestimate the withdrawal rates of advantaged consumers and underestimate inequality between consumers, and consequently may fail to correspond to optimal improvement strategies in physical IWS networks.

5 CONCLUSIONS & RECOMMENDATIONS:

The presented analysis emphasizes the importance of adopting a modelling method that fits both the modelling purpose and the nuances of consumers in IWS. Utilities and regulators should carefully assess the consumer behaviour in their context and select a modelling method that reliably portrays the behaviour. Based on the performance of the reviewed modelling methods, we suggest that:

- When IWS networks operate with supply durations shorter than needed for any consumers, unrestricted and volume-restricted methods agree and are best suited to simulate the hasty withdrawals we expect from unsatisfied consumers.
- When supply durations are perceived by consumers to be unreliable, flow-restricted methods are not recommended as consumers are unlikely to withdraw water slower than hydraulically possible.
- When network pressures are above the desired head for at least a few consumers, volume-restricted methods provide a more realistic prediction of consumer withdrawals. Otherwise, when network pressures are below the desired head for all consumers, all three methods are equivalent.
- When the phenomenon of interest occurs over multiple supply cycles, flow-restricted methods are the only EPANET-based method currently available
- When flow-restricted methods are preferred, the native PDA option supersedes the need for other FCV-based implementations

Research on the optimization and improvement of IWS networks is important and meaningful, especially when the modelling methods employed carefully represent the consumer's behaviour in IWS. We submit that flow-restricted methods are ill-suited to simulate IWS networks where pressures exceed the desired head for a considerable portion of consumers and thus are ill-suited to model an IWS network's transition to continuous supply.

Volume-restricted methods appear to allow for a wider range of non-idealized consumer behaviours. Unfortunately, to date, no EPANET-based volume-restricted approach can model an IWS network beyond the end of one supply duration. The development of volume-restricted methods that can simulate multiple cycles of intermittent operation, while conserving mass, would provide a more robust basis for optimizing and improving IWS networks and thus should be a priority for future work.

Future research should also be directed towards improving and validating assumptions about consumer behaviour, as well as developing modelling methods to capture these improved assumptions. Such research could significantly improve our ability to evaluate, improve and even optimize the performance of achieving SDG 6.1. Given that approximately 1 billion people (21% of piped water consumers) depend on intermittent supplies for their daily demands, research on IWS modelling is still relatively scarce. We commend the research done to date and call for more, urgently.

6 REFERENCES

[1] A. W. Bivins et al., "Estimating Infection Risks and the Global Burden of Diarrheal Disease Attributable to Intermittent Water Supply Using QMRA," Environmental Science & Technology, vol. 51, no. 13, pp. 7542–7551, Jul. 2017, doi: 10.1021/acs.est.7b01014.

[2] N. Totsuka, N. Trifunovic, and K. Vairavamoorthy, "Intermittent urban water supply under water starving situations," 2004.

[3] M. de Marchis, C. M. Fontanazza, G. Freni, G. la Loggia, E. Napoli, and V. Notaro, "Analysis of the impact of intermittent distribution by modelling the network-filling process," Journal of Hydroinformatics, vol. 13, no. 3, pp. 358–373, Jul. 2011, doi: 10.2166/hydro.2010.026.

[4] E. Kumpel and K. L. Nelson, “Comparing microbial water quality in an intermittent and continuous piped water supply,” Water Research, vol. 47, no. 14, pp. 5176–5188, Sep. 2013, doi: 10.1016/j.watres.2013.05.058.

[5] World Health Organization (WHO), “Progress on household drinking water, sanitation and hygiene 2000-2020: five years into the SDGs,” 2021.

[6] M. Solgi, O. Bozorg Haddad, S. Seifollahi-Aghmiuni, and H. A. Loáiciga, “Intermittent Operation of Water Distribution Networks Considering Equanimity and Justice Principles,” Journal of Pipeline Systems Engineering and Practice, vol. 6, no. 4, p. 04015004, Nov. 2015, doi: 10.1061/(ASCE)PS.1949-1204.0000198.

[7] E. E. Ameyaw, F. A. Memon, and J. Bicik, “Improving equity in intermittent water supply systems,” Journal of Water Supply: Research and Technology-Aqua, vol. 62, no. 8, pp. 552–562, Dec. 2013, doi: 10.2166/aqua.2013.065.

[8] P. P. Nyahora, M. S. Babel, D. Ferras, and A. Emen, “Multi-objective optimization for improving equity and reliability in intermittent water supply systems,” Water Supply, vol. 20, no. 5, pp. 1592–1603, Aug. 2020, doi: 10.2166/ws.2020.066.

[9] A. Ghorpade, A. K. Sinha, and P. Kalbar, “Multi-outlet storage tanks to improve water distribution networks in India,” Urban Water Journal, vol. 18, no. 7, pp. 570–578, Aug. 2021, doi: 10.1080/1573062X.2021.1914117.

[10] A. Batterman and S. Macke, “A Strategy to Reduce Technical Water Losses for Intermittent Water Supply Systems,” Fachhochschule Nordostniedersachsen, 2001. Accessed: Feb. 14, 2022. [Online]. Available: http://sdteffen.de/diplom/thesis.pdf

[11] D. D. J. Taylor, A. H. Slocum, and A. J. Whittle, “Demand Satisfaction as a Framework for Understanding Intermittent Water Supply Systems,” Water Resources Research, vol. 55, no. 7, pp. 5217–5237, 2019, doi: 10.1029/2018WR024124.

[12] P. Sivakumar, N. B. Gorev, T. T. Tanyimboh, I. F. Kodzhespirova, C. R. Suribabu, and T. R. Neelakantan, “Dynamic Pressure-Dependent Simulation of Water Distribution Networks Considering Volume-Driven Demands Based on Noniterative Application of EPANET 2,” Journal of Water Resources Planning and Management, vol. 146, no. 6, p. 06020005, Jun. 2020, doi: 10.1061/(asce)wr.1943-5452.0001220.

[13] S. Mohapatra, A. Sargaonkar, and P. K. Labhasetwar, “Distribution Network Assessment using EPANET for Intermittent and Continuous Water Supply,” Water Resources Management, vol. 28, no. 11, pp. 3745–3759, Sep. 2014, doi: 10.1007/s11269-014-0707-y.

[14] A. Campisano, A. Gullotta, and C. Modica, “Using EPA-SWMM to simulate intermittent water distribution systems,” Urban Water Journal, vol. 15, no. 10, pp. 925–933, Nov. 2018, doi: 10.1080/1573062X.2019.1597379.

[15] J. A. Cabrera-Bejar and V. G. Tzatchkov, “Inexpensive Modeling of Intermittent Service Water Distribution Networks,” in World Environmental and Water Resources Congress 2009, May 2009, pp. 1–10. doi: 10.1061/41036(342)29.

[16] M. Shrestha and S. G. Buchberger, “Role of Satellite Water Tanks in Intermittent Water Supply System,” 2012.

[17] A. M. Lieb, C. H. Rycroft, and J. Wilkening, “Optimizing Intermittent Water Supply in Urban Pipe Distribution Networks,” SIAM Journal on Applied Mathematics, vol. 76, no. 4, pp. 1492–1514, Jan. 2016, doi: 10.1137/15M1038979.

[18] S. Mohan and G. R. Abhijith, “Hydraulic Analysis of Intermittent Water-Distribution Networks Considering Partial-Flow Regimes,” Journal of Water Resources Planning and Management, vol. 146, no. 8, p. 04020071, Aug. 2020, doi: 10.1061/(asce)wr.1943-5452.0001246.

[19] L. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, “EPANET 2.2 User Manual,” Washington D.C., 2020. Accessed: May 12, 2022. [Online]. Available: https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=541156&Lab=CESER

[20] P. R. Bhave, “Node Flow Analysis Distribution Systems,” Transportation Engineering Journal of ASCE, vol. 107, no. 4, pp. 457–467, Jul. 1981, doi: 10.1061/TPEJAN.0000938.

[21] B. M. Janet Wagner, U. Shamir, and D. H. Marks, “WATER DISTRIBUTION RELIABILITY: SIMULATION METHODS.”

[22] W. K. Ang and P. W. Jowitt, “Solution for Water Distribution Systems under Pressure-Deficient Conditions,” Journal of Water Resources Planning and Management, vol. 132, no. 3, pp. 175–182, May 2006, doi: 10.1061/(ASCE)0733-9496(2006)132:3(175).

BY NC SA

[23] C. R. Suribabu and T. R. Neelakantan, "Balancing reservoir based approach for solution to pressure deficient water distribution networks," 2010.

[24] P. Ingeduld, A. Pradhan, Z. Svitak, A. Terrai, and D. Hydroinform as, "MODELLING INTERMITTENT WATER SUPPLY SYSTEMS WITH EPANET," 1100.

[25] C. Siew and T. T. Tanyimboh, "Pressure-Dependent EPANET Extension," Water Resources Management, vol. 26, no. 6, pp. 1477–1498, Apr. 2012, doi: 10.1007/s11269-011-9968-x.

[26] K. S. Jinesh Babu and S. Mohan, "Extended Period Simulation for Pressure-Deficient Water Distribution Network," Journal of Computing in Civil Engineering, vol. 26, no. 4, pp. 498–505, Jul. 2012, doi: 10.1061/(asce)cp.1943-5487.0000160.

[27] N. B. Gorev and I. F. Kodzhespirova, "Noniterative Implementation of Pressure-Dependent Demands Using the Hydraulic Analysis Engine of EPANET 2," Water Resources Management, vol. 27, no. 10, pp. 3623–3630, Aug. 2013, doi: 10.1007/s11269-013-0369-1.

[28] M. A. H. Abdy Sayyed, R. Gupta, and T. T. Tanyimboh, "Noniterative Application of EPANET for Pressure Dependent Modelling Of Water Distribution Systems," Water Resources Management, vol. 29, no. 9, pp. 3227–3242, Jul. 2015, doi: 10.1007/s11269-015-0992-0.

[29] H. A. Mahmoud, D. Savić, and Z. Kapelan, "New Pressure-Driven Approach for Modeling Water Distribution Networks," Journal of Water Resources Planning and Management, vol. 143, no. 8, p. 04017031, Aug. 2017, doi: 10.1061/(asce)wr.1943-5452.0000781.

[30] S. Galaitsi, R. Russell, A. Bishara, J. Durant, J. Bogle, and A. Huber-Lee, "Intermittent Domestic Water Supply: A Critical Review and Analysis of Causal-Consequential Pathways," Water (Basel), vol. 8, no. 7, p. 274, Jun. 2016, doi: 10.3390/w8070274.

[31] E. Kumpel, C. Woelfle-Erskine, I. Ray, and K. L. Nelson, "Measuring household consumption and waste in unmetered, intermittent piped water systems," Water Resources Research, vol. 53, no. 1, pp. 302–315, Jan. 2017, doi: 10.1002/2016WR019702.

[32] C. M. Fontanazza, G. Freni, and G. la Loggia, "Analysis of intermittent supply systems in water scarcity conditions and evaluation of the resource distribution equity indices," WIT Transactions on Ecology and the Environment, vol. 103, pp. 635–644, 2007, doi: 10.2495/WRM070591.

[33] P. Sivakumar, N. B. Gorev, T. T. Tanyimboh, I. F. Kodzhespirova, C. R. Suribabu, and T. R. Neelakantan, "Dynamic Pressure-Dependent Simulation of Water Distribution Networks Considering Volume-Driven Demands Based on Noniterative Application of EPANET 2," Journal of Water Resources Planning and Management, vol. 146, no. 6, p. 06020005, Jun. 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001220.

[34] C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth, "On the optimal design of water distribution networks: a practical MINLP approach," Optimization and Engineering, vol. 13, no. 2, pp. 219–246, Jun. 2012, doi: 10.1007/s11081-011-9141-7.

[35] D. Meyer, M. He, and J. Gibson, "Discussion of 'Dynamic Pressure-Dependent Simulation of Water Distribution Networks Considering Volume-Driven Demands Based on Noniterative Application of EPANET 2' by P. Sivakumar, Nikolai B. Gorev, Tiku T. Tanyimboh, Inna F. Kodzhespirova, CR Suribabu, and TR Neelakantan," Journal of Water Resources Planning and Management, vol. 147, no. 8, p. 07021009, 2021.

[36] L. A. Rossman, "EPANET 2: users manual," 2000.

[37] K. A. Klise et al., "Water network tool for resilience (WNTR) user manual," Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.

[38] T. Kumar, A. E. Post, I. Ray, M. Otsuka, and F. Pardo-Bosch, "From public service access to service quality: The distributive politics of piped water in Bangalore," World Development, vol. 151, p. 105736, Mar. 2022, doi: 10.1016/j.worlddev.2021.105736.

[39] D. D. J. Meyer, J. Khari, A. J. Whittle, and A. H. Slocum, "Effects of hydraulically disconnecting consumer pumps in an intermittent water supply," Water Research X, vol. 12, p. 100107, Aug. 2021, doi: 10.1016/j.wroa.2021.100107.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14741

ROBUSTNESS OF PROFILE SAMPLING IN DETECTING DISSOLVED LEAD IN HOUSHOLD DRINKING WATER

Amitosh Dash [1], Jip van Steen [2] and Mirjam Blokker [3]

[1,2,3] KWR Water Research Institute, Groningenhaven 7, 3433 PE, Nieuwegein (The Netherlands)

[1]amitosh.dash@kwrwater.nl, [2] jip.van.steen@kwrwater.nl, [3]mirjam.blokker@kwrwater.nl

Abstract

The norm for dissolved lead in drinking water is lowered in the Netherlands per 2023, from 10 to 5 μg/L. The effect is that: (a) homeowners and the Dutch water utilities want to find problematic lead components and remove them; and (b) the utilities want to know if copper service mains with lead soldering are expected to lead a norm exceedance. This calls for an improved sampling protocol (collected at the tap) to trace dissolved lead.

We help solve this problem using hydraulic and water quality simulations in EPANET. With short time steps, low flow rates, and short pipe lengths, a water quality time step of 1 second is not short enough. As EPANET cannot work with timesteps smaller than 1 second, we devised a work-around to ensure that the numerical solution does not result in a large error.

A household is defined as a system of pipes with various locations where water is used. The water demand patterns at each tap are generated using SIMDEUM. A few (sections of the) pipes are assigned to contain lead wherefrom dissolution occurs. The benchmark scenario involves a two-person household with average Dutch water consumption and with solely the service line containing lead. Next, fourteen scenarios are considered wherein either the geometry of the system, the water consumption, or location/extent of the lead releasing surface(s) are varied individually.

Of the sampling protocols considered, the results show that “profile sampling” is the most promising. In such a protocol, twenty consecutive samples of 300 ml are collected following a minimum of six hours of prolonged stagnation. Irrespective of the scenario (with a lead component), at least one of the twenty samples is guaranteed to possess a high concentration of lead, owing to our choice of a fine sample volume. Just two days of profile sampling can potentially unveil: (1) the location of the lead releasing component, (2) the volume of the lead releasing region, and (3) the saturation concentration of lead dissolution.

The key conclusions (for the current assumptions) are as follows. A lead service line potentially leads to an exceedance of the new norm, necessitating immediate action. The results also provide evidence that one week of proportional sampling is insufficient. Presently, it is considered to be the gold standard for measuring weekly intake of dissolved lead, however, it is susceptible to the stochastic nature of water demand leading to variations exceeding 50%. Lastly, a profile sampling protocol possesses the best opportunity to detect dissolved lead in the household drinking water network in a robust manner. In the future, a bespoke experimental facility (“PilotCity indoor installation”) will be used to validate the assumptions surrounding the lead dissolution and advection/diffusion model.

Keywords
Dissolved lead, EPANET water quality prediction, Households, Sampling protocol.

1 INCORPORATING LEAD TEMPORAL VARIABILITY WITH A STOCHASTIC DRINKING WATER DEMAND MODEL

Lead in drinking water has re-emerged as a problem of public interest and policy makers around the world are rushing to orchestrate actions in interest of public health. For example, in the Netherlands, the norm for weekly lead exposure per person is being reduced from 10 to 5 μg/L from 2023. The ultimate desire is to eradicate any lead releasing components, but information about their locations is seldom known. This is even more relevant for components present in premise plumbing, as it lies beyond the jurisdiction of the water utilities. To determine whether a household is exposed to lead in drinking water, sampling plays a crucial role.

Various sampling protocols exist to assess the extent to which lead in drinking water is a problem [1]. The most understudied contributor to lead variability is the impact of water use pattern (see Box 2 in [1]). How water is used has a significant effect on how dissolved lead propagates through the plumbing. The challenge often lies in simulating realistic water demand patterns, which is non-trivial due to its inherent stochasticity.

In this context, we aim to demonstrate how numerical simulations can contribute in estimating lead consumption under realistic water usage scenarios. We account for lead temporal variability by incorporating SIMDEUM [2,3], a stochastic drinking water demand model, in our framework. For this purpose, we perform simulations that are described in Section 2. In premise plumbing, water quality computations are afflicted by the presence of short pipes. To tackle this, in Section 3, we present how the combination of temporal deceleration and demand reduction can improve water quality calculations. In the context of the norm changes to be brought upon in the Netherlands, it is also important to calculate weekly intakes and its sensitivity to various factors (such as geometry of piping, household characteristics and location of lead releasing components). For that purpose, weekly intakes computed from the simulations are used to comment on proportional sampling in Section 4. Thereafter, in Section 5, we discuss the robust performance of profile sampling under varying scenarios. Finally, in Section 6, we summarize our main findings and offer an outlook on how we plan to build up on the present work.

2 SIMULATIONS IN EPANET TO CALCULATE DISSOLVED LEAD CONSUMPTION

While sampling protocols have to be eventually implemented in the real world, in order to design and test the effectiveness of sampling protocols, water quality simulations were deployed on EPANET. The advantages hereof include: perfect control and knowledge of the location of the lead releasing components and lead dissolution behaviour, the ability to control water demand patterns as well as the ability to accurately measure lead exposure. The framework is illustrated in Figure 1, the individual components hereof will be described in the forthcoming sections.

2.1 Framework and components of the numerical simulations

The starting point of the simulations is the definition of an indoor premise plumbing system. For this purpose, the lengths and diameters of the piping as well as the points-of-use need to be defined. We consider a typical Dutch household with water usage spread out over three storeys. The water enters the ground floor via a service line and continues into the indoor plumbing via the water meter. A toilet and kitchen is located on the ground floor, a toilet and shower on the first floor and a washing machine on the second floor. Moreover, on the second floor the boiler serves to supply hot water. Details about this geometry is available elsewhere [4]. The present work appends a service line to the existing plumbing. This service line with a diameter of 32 mm and a length of 1 m is added between the water source and the water meter.

Information about the points-of-use, together with the composition of the household (number of adults and children, attitude towards water consumption) are fed to SIMDEUM [2,3], a stochastic drinking water demand model. This program generates realistic water consumption patterns at each usage point. These generated demand patterns are then added to the EPANET model. The SIMDEUM demand patterns are not normalized (thus, temporal mean is not equal to unity), and thus, the base demand factor in EPANET can be set to unity.

The final aspect of the modelling involves the selection of lead releasing piping/components in the installation. These components are deemed to release dissolved lead into the water according to a lead dissolution model. In EPANET, we make use of a first order saturation growth model for the bulk reaction [5]. The lead dissolution model is described by equation (1).

$$\frac{dC}{dt} = \frac{4M}{DE}(E - C) \tag{1}$$

Here, the increase in dissolved lead concentration, C, in time, t, is governed by the pipe diameter, D, the dissolution rate, M, and equilibrium lead concentration or plumbosolvency, E. The solution to this differential equation is $C = E(1 - e^{-t/T})$, where $T = \frac{DE}{4M}$. The parameter T represents a timescale on how quickly the equilibrium lead concentration is achieved. To be accurate, the lead dissolution model is dependent on the ratio between the surface area and the volume (related to hydraulic diameter). In the case of a cylindrical pipe, this is reduced to $4/D$.

In our simulations, the plumbosolvency is assumed to be 110 µg/l while the lead dissolution rate is assumed to be 0.115 µg/(m^{2}s). These values are known to be dependent on parameters such as water chemistry and temperature. However, these factors are not included in the present study. Moreover, it is assumed that the water from the distribution network contains no lead.

One minor shortcoming of modelling the lead dissolution model as a bulk reaction instead of a wall reaction is that the effect of wall roughness cannot be trivially accounted for. This choice had to be made since the EPANET does not facilitate first order saturation growth models for wall reactions. EPANET has built-in methods to include the effect of wall roughness on the wall reaction rate. While we do not present the results here, we did verify that the EPANET bulk reaction approach gives comparable results with EPANET-MSX with a wall reaction. Since simulations with MSX are much slower, we did not use it further.

Of course, the process of lead dissolution has a lot more intricacies like the slow diffusion of lead from the walls towards the axis [6]. Moreover, the transport of lead is only governed by advection in our simulations. The inclusion of dispersion will influence our results [5]. The inclusion of this phenomenon in future EPANET versions [7] will undoubtedly be useful for reassessing the findings presented herein.

We limit ourselves to the analysis of dissolved lead. It is known that particulate lead can also contribute to lead consumption. However, incorporating the hydraulics of particulate lead goes beyond the scope of the present work. Moreover, we do not consider the additional influence of flow velocity on mechanical degradation of pipe scales which could potentially accelerate mass transfer of dissolved lead [8].

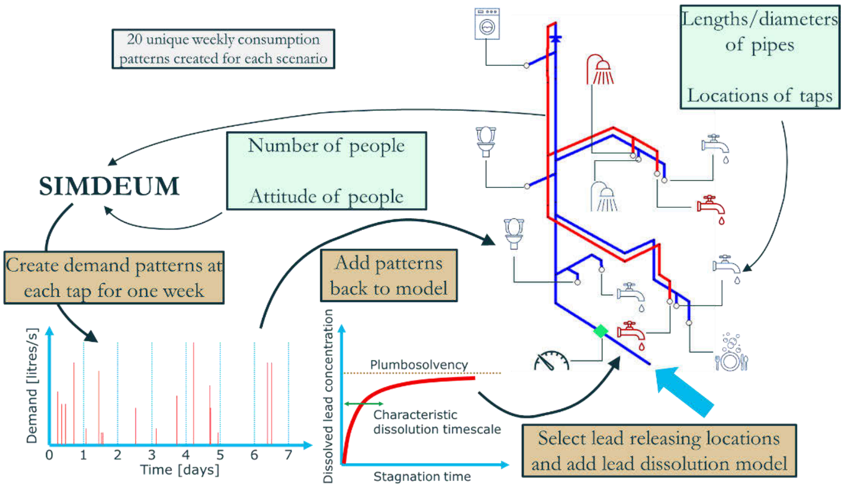

20 unique weekly consumption patterns created for each scenario
Lengths/diameters of pipes
Locations of taps
SIMDEUM
Number of people
Attitude of people
Create demand patterns at each tap for one week
Add patterns back to model
Demand [litres/s]
0
1
2
3
4
5
6
7
Time [days]
Dissolved lead concentration
Plumbosolvency
Characteristic dissolution timescale
Stagnation time
Select lead releasing locations and add lead dissolution model

Figure 1: Framework and components of the numerical simulations.

2.2 Temporal deceleration and demand reduction to counteract short pipes

From the perspective of hydraulics, the extended period simulations of EPANET can be seen as a composition of multiple steady state hydraulic simulations. This, however, does not apply to water quality simulations, where a Lagrangian approach is implemented. Thus, it is common practice for the water quality timestep to be much lower than the hydraulic timestep. If the timestep for the water quality calculations exceeds the time needed for water to be advected through a pipe, mass imbalance errors will result in inaccurate water quality predictions [9].

This conundrum is especially relevant for simulations in premise plumbing wherein piping is much shorter than distribution networks, and maximum velocities are comparable to pipe lengths. Inaccurate water quality predictions resulting from mass imbalance errors has been illustrated for EPANET 2.0 [9]. In EPANET 2.2 (released in 2020), mass balance errors are smaller. However, time steps in EPANET may only be long integers and thus cannot be smaller than 1 second [10].

In order to overcome issues that might arise therefrom, a workaround was devised. As a first step, all relevant timescales for the simulations (duration and hydraulic, quality, pattern, report time steps) are increased by a factor of ten. We refer to this as temporal deceleration. The next step is demand reduction wherein the water demand at each usage is reduced by a factor of ten to conserve volumetric intake. The key parameters are summarized in Table 1. Of course, this solution is expected to cause issues in the hydraulics as the Reynolds numbers are varied (variation in velocities but not in pipe diameters). In any case, the purpose of this workaround is to reduce errors that might arise in the water quality simulations.

Table 1: Summary of simulation properties.

Demand pattern generated on SIMDEUM [days]	7 (70 with temporal deceleration)
Duration [days]	140
Hydraulic timestep [s]	10
Quality timestep [s]	1
Report timestep [s]	100
Pattern timestep [s]	100
Base demand factor [-]	0.1
Data used for further analysis	Days 71-140

The inclusion of this workaround also necessitates a corresponding modification in the lead dissolution model. Given that temporal deceleration is a key aspect of the workaround, this must be also applied to the lead dissolution model. Thus, we reduce the parameter T by a factor of ten as well. This ensures dynamic similarity for the lead dissolution model.

The single biggest disadvantage of this workaround is that it leads to a huge rise in computational time. In the present simulations, the chemical reactions in question are relatively simple and can be implemented directly in EPANET. However, if more complex reactions are involved (such as in temperature modelling), the simulations would have to be performed on EPANET-MSX, wherein simulations are slower to begin with.

2.3 Scenarios considered

With the foundation for the simulations set, we now describe the fifteen scenarios we have simulated. These scenarios are shown in Table 2. A benchmark scenario is established, on basis of which variations are brought upon in either the geometry of the premise plumbing, the water consumption patterns, or the location of the lead releasing components. Variations in premise

BY NC SA

plumbing and location of the lead releasing components was implemented via WNTR [11] whereas changes in the water consumption pattern were implemented using SIMDEUM. Across all these cases, the lead dissolution model constants for plumbosolvency and mass dissolution rates were held constant. However, changes in pipe diameter would affect the characteristic timescale for lead dissolution.

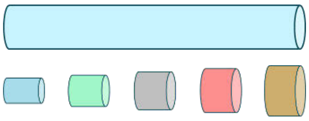
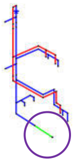
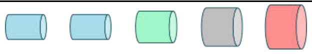
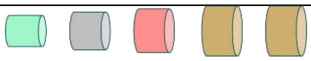
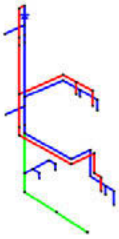
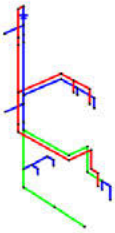
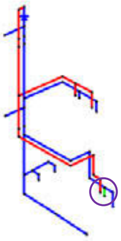
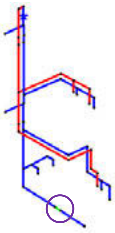
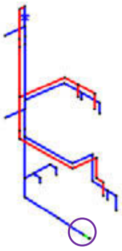
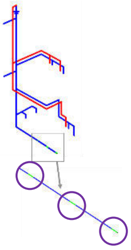
1. Benchmark case: The geometry as defined in Section 2.1 is considered. The water consumption patterns are generated using SIMDEUM with the assumption that it is inhabited by two adults with average water usage characteristics (by Dutch standards). The service line is the sole lead releasing component.
2. Four variations in geometry: The geometry is defined by a system of pipes with attributed lengths and diameters. In two scenarios, all pipes (lead releasing or not) are either shortened or lengthened by 10% each. To vary diameters, the various diameters in the premise plumbing are first inventoried and sorted. Thereafter, to reduce the diameters, the diameter of each pipe was reduced to the next lower possible diameter from the inventory. An exception are pipes with the smallest diameter which are unaffected. A similar approach is used for increasing the pipe diameters, where all pipe diameters are increased to the next possible larger value while the pipes with the largest diameters remain unaffected.
3. Four variations in water consumption patterns: To bring about variations in this category, either the number of inhabitants or their attitude towards water is changed. In this manner, we create water consumption patterns for the house being inhabited by either one adult or two adults with two children. Moreover, we consider the household being inhabited by two adults who either utilize water frugally or extravagantly (either due to habits or appliances) with respect to average Dutch standards. For the high usage, it is assumed that luxurious showers are adopted. For frugal usage, innovative sanitation concepts are applied in addition to efficient showers, washing machines and dishwashers [12].
4. Six variations in location of lead releasing components: First, the system of pipes responsible for lead dissolution is exaggerated. In one case, all pipes from the shutoff valve to the kitchen are responsible for lead dissolution. In another case, all pipes from the shutoff valve to the kitchen tap (excluding the kitchen tap itself) release lead. Then, the length of the lead releasing pipe is reduced to a piece of 20 cm, which is used for three cases - the kitchen tap itself, immediately upstream of the water meter and immediately downstream of the shutoff valve. In the final case, the service line is modified to have a diameter of 25 mm and a length of 3 m. The most upstream, the most downstream, and the central 20 cm of the service line release lead (an exaggerated form of lead soldering).

2.4 Sampling protocols

On top of the generated water demand patterns, sampling protocols were added on using WNTR. For each of the aforementioned fifteen scenarios, the following protocols were simulated:

1. “Proportional Sampling” – no additions/modifications were made to the water consumption patterns. Every moment of water usage at the kitchen tap is treated as a potential sample and no distinction is made whether the water is consumed or not (for example, drinking versus washing hands).
2. Profile sampling – Following a prolonged stagnation of six hours (between 0100-0700), twenty samples of 300 ml are collected consecutively at 0700 on each day of the simulation. These samples are appended as demand (300 mL in 10 seconds). During the prolonged stagnation, any water usage generated in the house by SIMDEUM is annulled.

Table 2: Schematic description of scenarios. Lead releasing locations are shown in green and are encircled.

<table>
<tr><td colspan="2">Benchmark</td><td>
(Pipes of certain lengths and diameters)</td><td colspan="2">
(Two adults with average Dutch consumption characteristics)</td><td>
(Lead service line)</td></tr>
<tr><td rowspan="2">Variations in geometry</td><td>Length</td><td colspan="2">
(10% shorter)</td><td colspan="2">
(10% longer)</td></tr>
<tr><td>Diameter</td><td colspan="2">
(One size smaller)</td><td colspan="2">
(One size larger)</td></tr>
<tr><td rowspan="2">Variations in household consumption characteristics</td><td>Number of people</td><td colspan="2">
(One adult)</td><td colspan="2">
(Two adults, two children)</td></tr>
<tr><td>Attitude of people</td><td colspan="2">
(Frugal usage)</td><td colspan="2">
(High usage)</td></tr>
<tr><td colspan="2" rowspan="3">Variations in lead releasing part(s)</td><td colspan="2">
(Shutoff valve to kitchen)</td><td colspan="2">
(Shutoff valve to kitchen tap)</td></tr>
<tr><td colspan="2">
(Kitchen tap only)</td><td colspan="2">
(Piece at water meter)</td></tr>
<tr><td colspan="2">
(Piece at shutoff valve)</td><td colspan="2">
(Lead solder in service line)</td></tr>
</table>

CC BY NC SA

3 EFFECTIVENESS OF TEMPORAL DECELERATION AND DEMAND REDUCTION IN TACKLING SHORT PIPES

The step involving temporal deceleration and demand reduction was specifically devised to tackle short piping and its effect on water quality calculations. In Figure 2(a), it can be seen that in several pipes, the product of maximum velocity and quality time step of 1 second exceeds the pipe length. Pipe lengths vary from 0.1-2 m, whereas maximum velocities vary from 0.29-2.12 m/s. This is a source of numerical errors when advecting plugs of dissolved lead. In order to prove that short piping is an issue and that temporal deceleration + demand reduction are effective, simplified numerical experiments were performed. These numerical experiments entailed water quality calculations using water age, thus bypassing the need for defining reactions.

Since the premise plumbing system shown in Figure 2(a) is quite extensive, a simpler version was created to test the efficacy of temporal deceleration. All points-of-use besides the kitchen tap for cold water were removed. Correspondingly, all plumbing besides the piping between the shutoff valve and the kitchen tap were removed. This is shown in Figure 2(b) as “with short pipes”. This route of piping includes segments of shorter lengths as well.

An alternative route is illustrated as “without short pipes” in Figure 2(b). Basically, a straight pipe directly connects the shutoff valve and the kitchen tap. To create this route, the following approach was devised:

1. The system “with short pipes” was skeletonized using the built-in function of WNTR to reduce to a single pipe. This creates one pipe whose diameter is the maximum value from all the pipe diameters before skeletonization (32 mm). The length of the piping equals the sum of all pipe lengths prior to skeletonization (11.89 m).
2. In the process of skeletonization, the total volume of the plumbing between the shutoff valve and the kitchen tap is increased. In order to enable a one-to-one comparison between the two routes, the diameter of the single pipe in “without short pipes” was reduced (20.65 mm) to conserve volume. Conservation of volume is necessitated since the water demands are left untouched.

Hereafter, for each of the two plumbing routes, two simulations are executed. The properties of the two types of simulations are summarized in Table 3. In the post-processing stage, while comparing results, the water age needs to be corrected for the simulations with temporal deceleration. In Figure 2(c), the differences in water ages are compared for the two routes of plumbing. When simulations are performed normally, there are extended periods of several hours wherein the differences (or errors) in the water ages exceeds 100 seconds. This insinuates that shorter pipes are indeed problematic in water quality calculations, especially in the context of premise plumbing. However, it can be seen that the implementation of temporal deceleration + demand reduction does reduce the propensity for errors in water quality calculations. Comparisons between velocities and pressures were performed but are not shown here. In short, velocities were unaffected by temporal deceleration/demand reduction while the pressures were. Since the Reynolds number is not held constant (reduction in velocities but not in diameters), it is unsurprising that the hydraulic calculations are affected. However, these are not expected to affect water quality calculations.

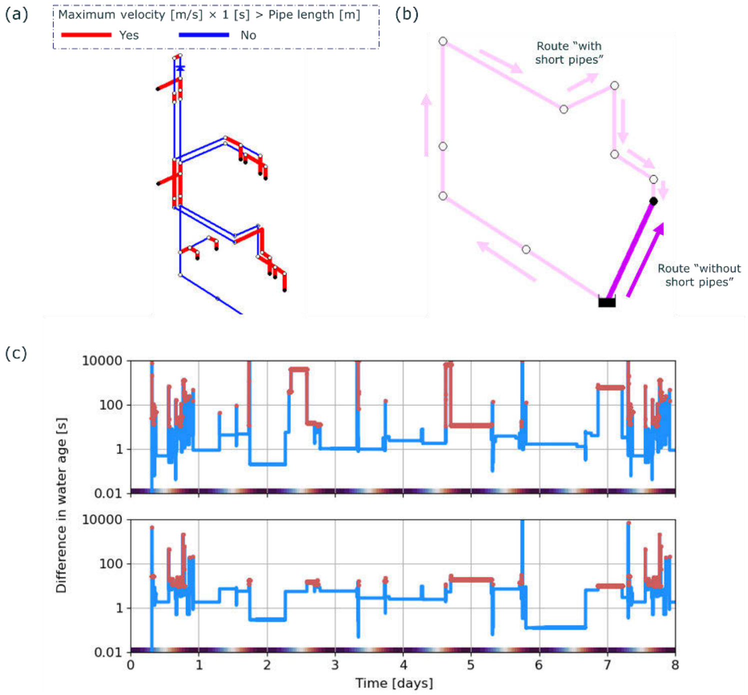


Figure 2: (a) Pipes wherein the product of maximum velocity and quality timestep exceeds pipe length (b) Two routes of plumbing from the shutoff valve to the kitchen tap. (c) Differences in water ages at the kitchen tap between simulations across the two routes, without/with (top/bottom respectively) the implementation of temporal deceleration + demand reduction respectively. Red markers are periods where the difference exceeds 10 seconds.

Table 3: Characteristics of the simulations with(out) temporal deceleration + demand reduction for Figure 2(c).

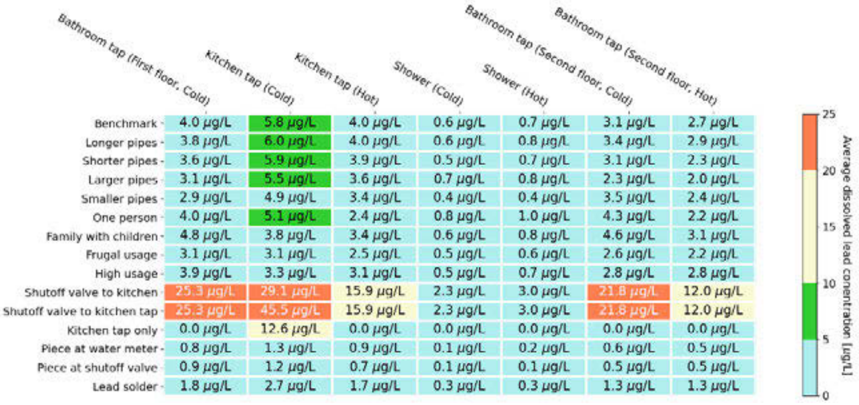
Simulation type → Property ↓	Normal	Temporal deceleration + demand reduction
Factor of deceleration and reduction	1	10
Hydraulic timestep [s]	10	100
Quality timestep [s]	1	1
Duration [days]	8	80
Report timestep [s]	10	100
Pattern timestep [s]	10	100
Base demand factor [-]	1	0.1
Post-processing of data to compare results		
Velocity	No action	×10
Pressure	No action	No action
Water age	No action	÷10

4 WEEKLY INTAKE AND IMPLICATIONS FOR PROPORTIONAL SAMPLING

One of the parameters that needs to be assessed is the weekly lead consumption via drinking water. This is especially of interest to the Dutch water utilities who wish to know whether copper mains with lead soldering can lead to norm exceedance. In this section, we present typical weekly lead consumption trends and how these are affected by the scenarios illustrated in Table 2. To clarify, the simulations considered here do not have any sampling protocol added onto the consumption patterns generated by SIMDEUM.

We consider the lead concentrations encountered at various points-of-use in the household in Figure 3. The key observations are as follows:

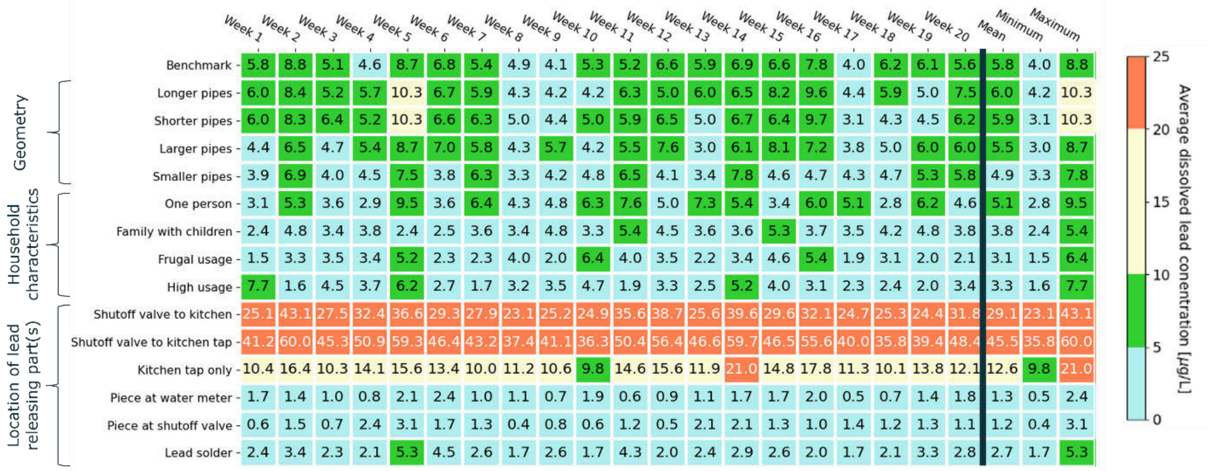
1. In the benchmark scenario, the presence of lead service line leads to exceedance of the to-be-sharpened norm at the kitchen tap, but not at the other points of usage. In fact, for the majority of the scenarios, the highest lead concentrations are encountered at the kitchen tap. This implies that the kitchen tap is the best point to perform proportional sampling measurements as it represents the worst-case scenario, next to the fact that most consumption is assumed to take place via the kitchen tap.
2. Variations in the geometry are expected to affect the total volume of lead in the system, propagation towards various nodes, and the timescale at which the equilibrium lead concentration is achieved. All of this leads to minor fluctuations in average consumption, but the conclusion is that a lead service line leads to unsafe conditions and needs immediate action.
3. Variation in household consumption characteristics brings upon reduced lead concentrations than the benchmark scenario. Moreover, in a couple of scenarios the kitchen tap is not the point where the highest lead concentrations are found. This demonstrates the sensitivity of the numerical simulations to water consumption characteristics, and how challenging it would be to generalize average lead consumption without actual measurements.
4. The scenarios wherein extensive lead plumbing is present unsurprisingly leads to norm exceedance. The presence of a small lead component (length 20 cm) can lead to breaches depending on its location relative to the point-of-use. When located at the kitchen tap itself, the norm is breached since the lead cannot be transported elsewhere. However, when these pieces are located further away (for example, in the service line), the lead discharge gets spread out. Our results suggest that the presence of copper service lines with lead solder might not lead to norm exceedance and need not be prioritized by the water utilities.



	Bathroom tap (First floor, Cold)	Kitchen tap (Cold)	Kitchen tap (Hot)	Shower (Cold)	Shower (Hot)	Bathroom tap (Second floor, Cold)	Bathroom tap (Second floor, Hot)
Benchmark	4.0 μg/L	5.8 μg/L	4.0 μg/L	0.6 μg/L	0.7 μg/L	3.1 μg/L	2.7 μg/L
Longer pipes	3.8 μg/L	6.0 μg/L	4.0 μg/L	0.6 μg/L	0.8 μg/L	3.4 μg/L	2.9 μg/L
Shorter pipes	3.6 μg/L	5.9 μg/L	3.9 μg/L	0.5 μg/L	0.7 μg/L	3.1 μg/L	2.3 μg/L
Larger pipes	3.1 μg/L	5.5 μg/L	3.6 μg/L	0.7 μg/L	0.8 μg/L	2.3 μg/L	2.0 μg/L
Smaller pipes	2.9 μg/L	4.9 μg/L	3.4 μg/L	0.4 μg/L	0.4 μg/L	3.5 μg/L	2.4 μg/L
One person	4.0 μg/L	5.1 μg/L	2.4 μg/L	0.8 μg/L	1.0 μg/L	4.3 μg/L	2.2 μg/L
Family with children	4.8 μg/L	3.8 μg/L	3.4 μg/L	0.6 μg/L	0.8 μg/L	4.6 μg/L	3.1 μg/L
Frugal usage	3.1 μg/L	3.1 μg/L	2.5 μg/L	0.5 μg/L	0.6 μg/L	2.6 μg/L	2.2 μg/L
High usage	3.9 μg/L	3.3 μg/L	3.1 μg/L	0.5 μg/L	0.7 μg/L	2.8 μg/L	2.8 μg/L
Shutoff valve to kitchen	25.3 μg/L	29.1 μg/L	15.9 μg/L	2.3 μg/L	3.0 μg/L	21.8 μg/L	12.0 μg/L
Shutoff valve to kitchen tap	25.3 μg/L	45.5 μg/L	15.9 μg/L	2.3 μg/L	3.0 μg/L	21.8 μg/L	12.0 μg/L
Kitchen tap only	0.0 μg/L	12.6 μg/L	0.0 μg/L	0.0 μg/L	0.0 μg/L	0.0 μg/L	0.0 μg/L
Piece at water meter	0.8 μg/L	1.3 μg/L	0.9 μg/L	0.1 μg/L	0.2 μg/L	0.6 μg/L	0.5 μg/L
Piece at shutoff valve	0.9 μg/L	1.2 μg/L	0.7 μg/L	0.1 μg/L	0.1 μg/L	0.5 μg/L	0.5 μg/L
Lead solder	1.8 μg/L	2.7 μg/L	1.7 μg/L	0.3 μg/L	0.3 μg/L	1.3 μg/L	1.3 μg/L

Figure 3: Dissolved lead concentrations at various points-of-use across all scenarios (mean of twenty weeks).

Proportional sampling is rightly perceived as the gold standard as far as measurement of lead consumption at household scales is concerned. However, the sample is typically collected over anywhere between a few days to a week. In Figure 4, we present how week-to-week variations in water consumption affects the dissolved lead concentrations at the kitchen tap across all the scenarios. Note that demand patterns in all scenarios are the same, except when household characteristics are altered. Thus, for a certain week, consumption pattern across the scenarios are correlated (aside from changes in household characteristics) – for example, from week 4 to week 5, increased lead exposure is seen across all scenarios. In the final three columns, the average, minimum, and maximum values across the twenty weeks are shown. In brief, it is irrefutable that week-to-week variations in water consumption patterns trigger variations in dissolved lead concentrations which, in turn, can affect the plan-of-action. Thus, it is recommended to implement proportional sampling over a longer timeframe – for example, five non-consecutive weeks spread out over a year.



		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18	Week 19	Week 20	Mean	Minimum	Maximum
	Benchmark	5.8	8.8	5.1	4.6	8.7	6.8	5.4	4.9	4.1	5.3	5.2	6.6	5.9	6.9	6.6	7.8	4.0	6.2	6.1	5.6	5.8	4.0	8.8
Geometry	Longer pipes	6.0	8.4	5.2	5.7	10.3	6.7	5.9	4.3	4.2	4.2	6.3	5.0	6.0	6.5	8.2	9.6	4.4	5.9	5.0	7.5	6.0	4.2	10.3
	Shorter pipes	6.0	8.3	6.4	5.2	10.3	6.6	6.3	5.0	4.4	5.0	5.9	6.5	5.0	6.7	6.4	9.7	3.1	4.3	4.5	6.2	5.9	3.1	10.3
	Larger pipes	4.4	6.5	4.7	5.4	8.7	7.0	5.8	4.3	5.7	4.2	5.5	7.6	3.0	6.1	8.1	7.2	3.8	5.0	6.0	6.0	5.5	3.0	8.7
	Smaller pipes	3.9	6.9	4.0	4.5	7.5	3.8	6.3	3.3	4.2	4.8	6.5	4.1	3.4	7.8	4.6	4.7	4.3	4.7	5.3	5.8	4.9	3.3	7.8
Household characteristics	One person	3.1	5.3	3.6	2.9	9.5	3.6	6.4	4.3	4.8	6.3	7.6	5.0	7.3	5.4	3.4	6.0	5.1	2.8	6.2	4.6	5.1	2.8	9.5
	Family with children	2.4	4.8	3.4	3.8	2.4	2.5	3.6	3.4	4.8	3.3	5.4	4.5	3.6	3.6	5.3	3.7	3.5	4.2	4.8	3.8	3.8	2.4	5.4
	Frugal usage	1.5	3.3	3.5	3.4	5.2	2.3	2.3	4.0	2.0	6.4	4.0	3.5	2.2	3.4	4.6	5.4	1.9	3.1	2.0	2.1	3.1	1.5	6.4
	High usage	7.7	1.6	4.5	3.7	6.2	2.7	1.7	3.2	3.5	4.7	1.9	3.3	2.5	5.2	4.0	3.1	2.3	2.4	2.0	3.4	3.3	1.6	7.7
Location of lead releasing part(s)	Shutoff valve to kitchen	25.1	43.1	27.5	32.4	36.6	29.3	27.9	23.1	25.2	24.9	35.6	38.7	25.6	39.6	29.6	32.1	24.7	25.3	24.4	31.8	29.1	23.1	43.1
	Shutoff valve to kitchen tap	41.2	60.0	45.3	50.9	59.3	46.4	43.2	37.4	41.1	36.3	50.4	56.4	46.6	59.7	46.5	55.6	40.0	35.8	39.4	48.4	45.5	35.8	60.0
	Kitchen tap only	10.4	16.4	10.3	14.1	15.6	13.4	10.0	11.2	10.6	9.8	14.6	15.6	11.9	21.0	14.8	17.8	11.3	10.1	13.8	12.1	12.6	9.8	21.0
	Piece at water meter	1.7	1.4	1.0	0.8	2.1	2.4	1.0	1.1	0.7	1.9	0.6	0.9	1.1	1.7	1.7	2.0	0.5	0.7	1.4	1.8	1.3	0.5	2.4
	Piece at shutoff valve	0.6	1.5	0.7	2.4	3.1	1.7	1.3	0.4	0.8	0.6	1.2	0.5	2.1	2.1	1.3	1.0	1.4	1.2	1.3	1.1	1.2	0.4	3.1
	Lead solder	2.4	3.4	2.3	2.1	5.3	4.5	2.6	1.7	2.6	1.7	4.3	2.0	2.4	2.9	2.6	2.0	1.7	2.1	3.3	2.8	2.7	1.7	5.3

Figure 4: Weekly variations and statistics thereof for dissolved lead concentrations at the kitchen tap.

While judging the results presented in this section, it is important to realize that the absolute values of lead concentration are valid for the assumptions made by us about the lead dissolution model, namely, the values for plumbosolvency and lead dissolution rate. For example, there are numerous factors concerning water chemistry that will lead to variations in lead leaching rates (see Box 1 in [1]). Moreover, the results are specific to this household. An aspect that has not been considered is the layout of the premise itself (for example, the length of the service line or the location of the kitchen in the household). Nevertheless, the results herein show the value of such simulations in assessing lead consumption at household scales.

5 ROBUSTNESS OF PROFILE SAMPLING

Numerous sampling protocols have been devised over the course of history. It is acknowledged that there is no universal protocol that answers all questions (for example, exposure at communal/household scale, presence/location of lead releasing component) and that each protocols answers a specific question [1]. In this section, we specifically consider profile sampling which is capable of localizing the lead releasing component(s).

In our simulations, the measurement protocol is implemented as follows. A prolonged standstill is imposed on each of the 140 days of the simulation between 0100-0700. This means that any water consumption created by SIMDEUM, everywhere in the household, in this timeframe is removed. Immediately following the prolonged standstill, twenty consecutive samples with a volume of 300 ml are collected at the kitchen tap. An example of this collection strategy is shown in Figure 5(a). Please note that the addition of a sampling protocol is intrusive and will affect the numerical values reported in Figure 3 and Figure 4.

The corresponding lead concentrations computed at the kitchen tap are shown in Figure 5(b). The pink line shows what would have been measured by a hypothetical lead concentration sensor inside the tap. From the lead concentrations corresponding to normal usage, it can be seen that there are instances wherein the lead concentration at the tap exceeds the (future) norm of 5 μg/L. This clearly illustrates that the Random Daytime approach (collecting a sample of water at a random time of the day) may be susceptible to these outliers and planning actions on basis thereof

BY NC SA

is unreliable. In the inset of Figure 5(b), it can also be seen that three of the twenty samples yield strikingly high lead concentrations (close to imposed plumbosolvency).

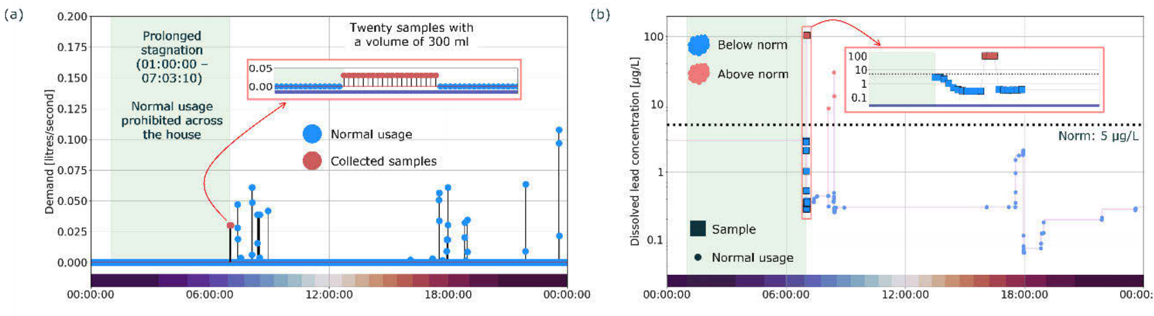
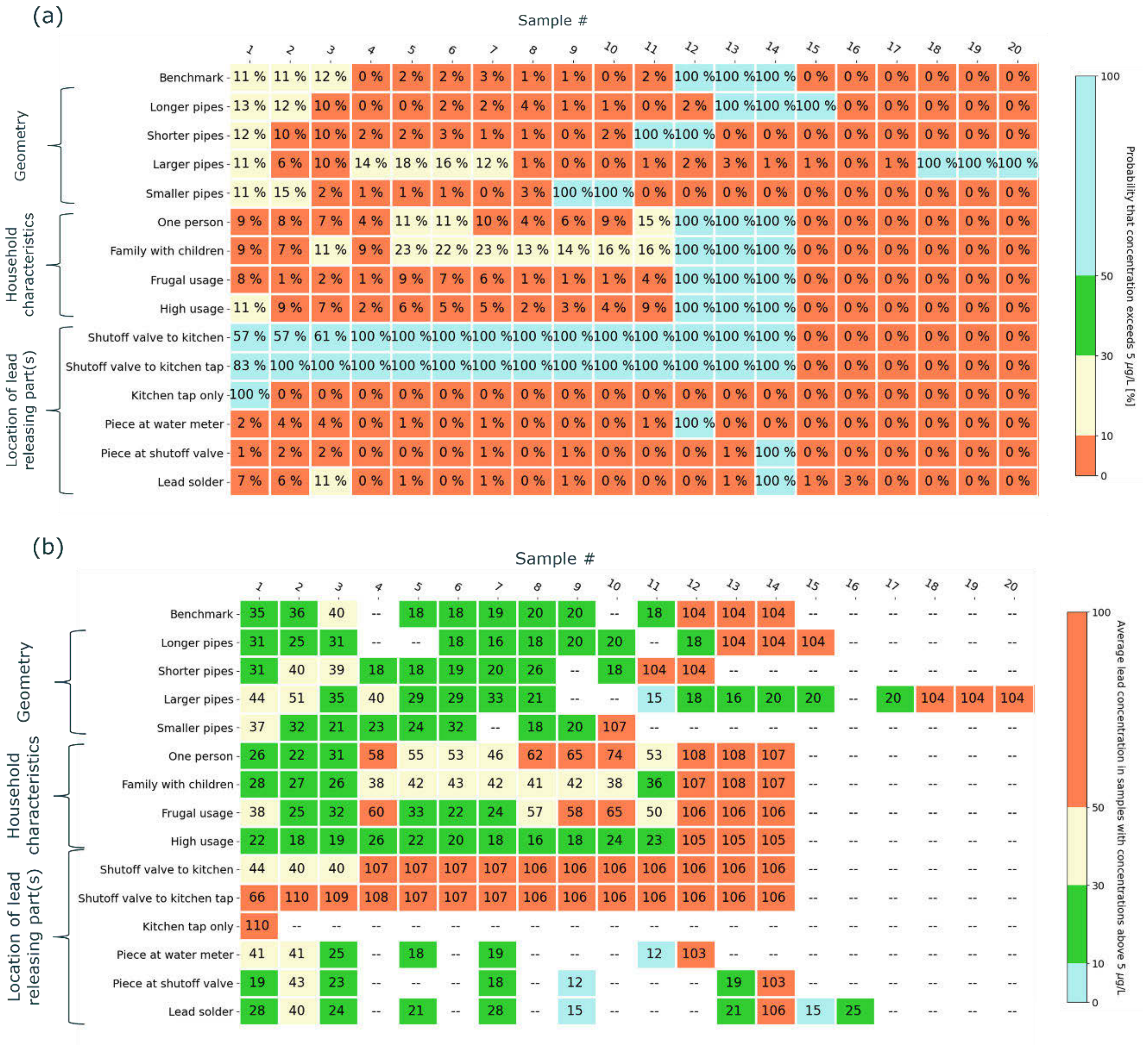


Figure 5: Example of a typical input/output in the simulations over one day at the kitchen tap for (a) water usage and samples for profile sampling (b) dissolved lead concentrations. The gradient colorbar at the bottom is a visual representation of the time of the day.


Probability that concentration exceeds 5 µg/L [%]
100
50
30
10
0
Average lead concentration in samples with concentrations above 5 µg/L
100
50
30
10
0

(a)

Sample #

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Benchmark	11 %	11 %	12 %	0 %	2 %	2 %	3 %	1 %	1 %	0 %	2 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
Geometry	Longer pipes	13 %	12 %	10 %	0 %	0 %	2 %	2 %	4 %	1 %	1 %	0 %	2 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %
	Shorter pipes	12 %	10 %	10 %	2 %	2 %	3 %	1 %	1 %	0 %	2 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Larger pipes	11 %	6 %	10 %	14 %	18 %	16 %	12 %	1 %	0 %	0 %	1 %	2 %	3 %	1 %	1 %	0 %	1 %	100 %	100 %	100 %
	Smaller pipes	11 %	15 %	2 %	1 %	1 %	1 %	0 %	3 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Household characteristics	One person	9 %	8 %	7 %	4 %	11 %	11 %	10 %	4 %	6 %	9 %	15 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	Family with children	9 %	7 %	11 %	9 %	23 %	22 %	23 %	13 %	14 %	16 %	16 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	Frugal usage	8 %	1 %	2 %	1 %	9 %	7 %	6 %	1 %	1 %	1 %	4 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	High usage	11 %	9 %	7 %	2 %	6 %	5 %	5 %	2 %	3 %	4 %	9 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
Location of lead releasing part(s)	Shutoff valve to kitchen	57 %	57 %	61 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	Shutoff valve to kitchen tap	83 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	Kitchen tap only	100 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Piece at water meter	2 %	4 %	4 %	0 %	1 %	0 %	1 %	0 %	0 %	0 %	1 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Piece at shutoff valve	1 %	2 %	2 %	0 %	0 %	0 %	1 %	0 %	1 %	0 %	0 %	0 %	1 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %
	Lead solder	7 %	6 %	11 %	0 %	1 %	0 %	1 %	0 %	1 %	0 %	0 %	0 %	1 %	100 %	1 %	3 %	0 %	0 %	0 %	0 %

(b)

Sample #

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Benchmark	35	36	40	--	18	18	19	20	20	--	18	104	104	104	--	--	--	--	--	--
Geometry	Longer pipes	31	25	31	--	--	18	16	18	20	20	--	18	104	104	104	--	--	--	--	--
	Shorter pipes	31	40	39	18	18	19	20	26	--	18	104	104	--	--	--	--	--	--	--	--
	Larger pipes	44	51	35	40	29	29	33	21	--	--	15	18	16	20	20	--	20	104	104	104
	Smaller pipes	37	32	21	23	24	32	--	18	20	107	--	--	--	--	--	--	--	--	--	--
Household characteristics	One person	26	22	31	58	55	53	46	62	65	74	53	108	108	107	--	--	--	--	--	--
	Family with children	28	27	26	38	42	43	42	41	42	38	36	107	108	107	--	--	--	--	--	--
	Frugal usage	38	25	32	60	33	22	24	57	58	65	50	106	106	106	--	--	--	--	--	--
	High usage	22	18	19	26	22	20	18	16	18	24	23	105	105	105	--	--	--	--	--	--
Location of lead releasing part(s)	Shutoff valve to kitchen	44	40	40	107	107	107	107	106	106	106	106	106	106	106	--	--	--	--	--	--
	Shutoff valve to kitchen tap	66	110	109	108	107	107	107	106	106	106	106	106	106	106	--	--	--	--	--	--
	Kitchen tap only	110	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Piece at water meter	41	41	25	--	18	--	19	--	--	--	12	103	--	--	--	--	--	--	--	--
	Piece at shutoff valve	19	43	23	--	--	--	18	--	12	--	--	--	19	103	--	--	--	--	--	--
	Lead solder	28	40	24	--	21	--	28	--	15	--	--	--	21	106	15	25	--	--	--	--

Figure 6: (a) Probability (in percentage) that a sample in a certain scenario contains dissolved lead with concentration above 5 µg/L. (b) Conditionally averaged lead concentrations in all samples wherein the dissolved lead concentration exceeds 5 µg/L.

CC BY NC SA

To investigate the statistics behind the performance of profile sampling, Figure 6 is considered. Basically, in Figure 6(a), for each sample under each scenario, the probability that the lead concentration therein exceeds 5 μg/L is shown.

Samples 1-3 (0.9 litres) in the benchmark case have chances of ~10% to possess concentrations above 5 μg/L. In the Netherlands, the protocol presently used to confirm the presence/absence of lead releasing components involves collecting a one litre sample following a six hour stagnation (corresponding to the average of samples 1-3 in Figure 5/Figure 6). This protocol is recommended by the Dutch National Institute for Public Health and the Environment (RIVM). This suggests that the present approach is inadequate in its goal of detecting the presence/absence of lead releasing components and warrants revisiting.

For the benchmark case, samples 12-14 have a 100% chance of meeting the criterion. From a practical perspective, this is attractive. This means that irrespective of the day/week/month on which this sampling protocol is performed, samples 12-14 are guaranteed to have lead concentration above 5 μg/L. Further reassuring is the fact that irrespective of the scenario, at least one of the twenty samples is guaranteed to possess lead concentrations above 5 μg/L. Even the small components of 20 cm long are picked up by this protocol. This partly demonstrates the robustness of profile sampling under realistic water consumption patterns. It is important to realize that the choice of 20 samples of 300 ml is instrumental. Had sampling been coarser (for example, 10 samples of 600 ml or 5 samples of 1200 ml), samples with high lead concentrations could have gotten diluted, diminishing the robustness of this sampling protocol.

To firmly establish the robustness of profile sampling, Figure 6(b) is considered. The numbers herein are conditionally averaged lead concentrations in samples wherein the concentration exceeds 5 μg/L. For the benchmark case, samples 12-14 have average concentrations above 100 μg/L, while for the other samples, it varies between 20-40 μg/L. In fact, for every sample/scenario combination in Figure 6(a) with 100%, the corresponding conditionally averaged lead concentration is in the vicinity of the plumbosolvency of 110 μg/L. This means that the lead from the lead releasing components will not only invariably end up in the same sample but will also end up in a high concentration, facilitating easy and unambiguous detection.

Figure 7 summarizes the principle behind profile sampling. The plumbing between the distribution network and a single faucet is shown. The water meter is preceded by a service line. The plumbing is partitioned into equal volumes representing the volume of each sample that is to be collected following a prolonged standstill. Based on sample collection over two days, a lot of information can be extracted from the profiles. The first day of measurements returns predictions for potential locations of the lead releasing components. The second day of measurements help classify the true and false positives. For example, in the schematic, sample 5 from day 1 is a false positive since it does not possess high concentrations on day 2. However, samples 12, 13 are true positives and correspond to the lead releasing component. We refer to samples 12, 13 as the peak. Following the peak, the lead is flushed away and there is no chance that a sample will contain any dissolved lead. Based on this peak, the following information can be gathered and can be useful in prioritizing which households need to be investigated first.

1. Location of the peak: If a drawing of the premise plumbing is available, based on the location of the peak and the volume of the individual samples, it should be theoretically possible to estimate where in the plumbing system the lead originates from. The location can be relevant in determining whether it is the responsibility of the premise owner or the water utility. The further away the lead releasing component is from the tap, the better it is for the safety of the inhabitants, since the dissolved lead gets spread out.
2. Breadth of the peak: The total volume of the peak indicates the severity of the problem. Naturally, the narrower the peak, the better it is for the inhabitants.

BY NC SA

3. Height of the peak: This contains information about the local level of plumbosolvency which will likely be related to the water chemistry and temperature, specific to the distribution area. The lower the peak, the better it is for the inhabitants.

It is known that each sampling protocol has its own objective. In the case of profile sampling, it is to detect and localize the lead releasing components, and is not expected to provide any insights into the lead consumption by the inhabitants. It would be interesting to see whether an empirical correlation can be established between the above three parameters and the weekly lead consumption. Fully flushing the plumbing prior to stagnation will improve the robustness of the technique by eliminating any false positives. In the present study, not flushing prior to stagnation allows us to assess the performance of the protocol recommended by the Dutch National Institute for Public Health and the Environment (through the first three samples).

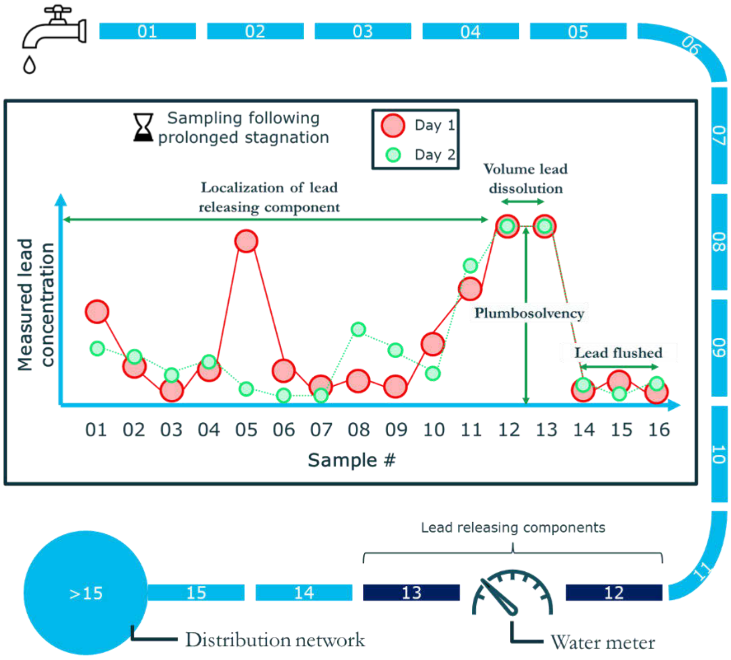

01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
>15
Sampling following prolonged stagnation
Day 1
Day 2
Volume lead dissolution
Localization of lead releasing component
Plumbosolvency
Lead flushed
Measured lead concentration
Sample #
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16
Lead releasing components
Distribution network
Water meter

Figure 7: Schematic of the functionality of profile sampling and the information that can be derived therefrom.

Please note that a similar conceptual drawing has been presented elsewhere [13]. However, it primarily illustrates how the slug of high lead concentration is generated and advected over the course of normal use and how it affects the composition of a one litre sample.

Questions can be raised on whether the robustness of profile sampling as demonstrated here can withstand real conditions. Field measurements in Flint [14], showed the value of this protocol in identifying the sources of lead. Similarly, it has been demonstrated [5] that the essence of this protocol can be captured in a laboratory setup (Home Plumbing Simulator of EPA) as well as EPANET based modelling (of the Home Plumbing Simulator). In their comparison, however, the concentration peaks in the experiments were far lower than that seen in their numerical simulations.

Another caveat is that the propagation of dissolved lead in our EPANET simulations are driven by advective forces. In reality, diffusion will also affect the propagation of dissolved lead. The addition of one dimensional mass dispersion to the EPANET framework led to a better match with the profiles obtained in controlled experiments [5]. The addition of this mass dispersion model broadened the concentration peak obtained by profile sampling while also lowering the peak itself. This means that in our simulations, the peak heights are overestimated while the peak widths are underestimated.

6 DISCUSSION: NUMERICAL MODELLING OFFERS AN ADDITIONAL AVENUE TO SUPPORT DECISION-MAKING SURROUNDING LEAD IN DRINKING WATER

The major objective of our work was to demonstrate the added value provided by EPANET hydraulic and water quality simulations in tackling the "lead in drinking water" problem. By leveraging SIMDEUM in generating realistic water usage patterns, we incorporate the understudied aspect of temporal lead variability in premise plumbing. Our numerical simulations combined the geometry of a typical Dutch household, a lead dissolution model (first order saturation growth) and realistic water consumption patterns. Moreover, we studied a total of fifteen scenarios, wherein either the geometry of the piping, household water consumption characteristics, location of lead releasing component was tweaked. For each scenario, 140 days of unique daily water usage patterns were simulated.

A shortcoming of combining EPANET water quality simulations with premise plumbing is that the timescale over which water advects through short pipes is shorter than 1 second, which can lead to numerical errors. To overcome this, and the fact that EPANET cannot work with timesteps shorter than one second, we simultaneously implemented temporal deceleration (increasing timescales of simulations, hydraulic/pattern/report timesteps by a certain factor, ten in our case) and demand reduction (decreasing water demand at each point-of-use by the same factor). This ascertains that the shortcomings caused by short piping is circumvented (reduction in errors).

Establishing a solid framework for our numerical simulations allowed us to estimate average lead exposure via drinking water at several points-of-use. It is seen that the average lead concentrations are highest at the kitchen tap for cold water, for majority of the scenarios. This, in combination with the fact that most consumption is assumed to occur at the kitchen tap, implies that performing proportional sampling should be continued at the kitchen tap itself. Typically, during proportional sampling, the composite sample is collected over one week. Based on our simulations of twenty weeks, we are able to demonstrate how the results of proportional sampling is prone to the week-to-week variations in water usage. Even though we do not make a distinction between the type of water consumption or consider the hot water tap, our results suggest that one week of sampling is insufficient. Thus, to get a better picture of average lead exposure, it is recommended to analyse samples collected over five non-contiguous weeks.

Finally, we demonstrated the strength of profile sampling in localizing the source of dissolved lead. The protocol involves collecting consecutive samples (in our simulations, 20 samples of 300 ml) immediately following a prolonged standstill. Basically, the following three parameters can be derived from the concentration profile across the samples – the location of the lead releasing component (location of the peak), volume of lead release (breadth of the peak), plumbosolvency (height of the peak). In the samples corresponding to the lead releasing location(s), it is expected that the lead concentrations will always exceed 5 µg/l and achieve values close to the plumbosolvency, in contrast to the other samples. This trend is visible across all scenarios, demonstrating the robustness of profile sampling. Flushing the plumbing prior to the prolonged stagnation will likely make the technique more robust (by eliminating any previously stagnated plugs of dissolved lead). It would be interesting to see whether any correlation exists between the above three parameters and the weekly lead consumption.

Our simulations also include several assumptions. Two of the key assumptions are: (i) the parameters for the lead dissolution model (plumbosolvency, mass dissolution rate) have been assumed, (ii) propagation of dissolved lead across the plumbing is driven purely by advection while the influence of diffusion is overlooked. In the future, we intend to utilize a bespoke experimental facility (PilotCity Indoor Installation, whose functionalities resemble the Home Plumbing Simulator of EPA) to probe the extent to which our assumptions are (in)valid.

Accordingly, we can re-do the simulations presented herein with the adjusted assumptions and perform a one-to-one comparison with the results obtained in a controlled laboratory setting.

Nevertheless, the framework presented in this work provides an avenue for future simulations surrounding the topic of dissolved metals in drinking water. Of course, simulations entail assumptions and will always be detached from reality. However, it provides an ideal playing ground to test sampling protocols and prioritize households which need immediate action. We would eventually like to expand our methodology to include locations frequented by children (for example, schools), multi-storeyed housing with common (lead or copper with lead solder) service line. Finally, our approach offers a convenient environment to test sampling protocols that are being proposed by Dutch water utilities, in wake of the stringent norms surrounding lead in drinking water.

7 REFERENCES

[1] Triantafyllidou, S., Burkhardt, J., Tully, J., Cahalan, K., DeSantis, M., Lytle, D., & Schock, M. (2021). Variability and sampling of lead (Pb) in drinking water: Assessing potential human exposure depends on the sampling protocol. Environment International, 146, 106259.

[2] Blokker, E. J. M., Vreeburg, J. H. G., & van Dijk, J. C. (2010). Simulating residential water demand with a stochastic end-use model. Journal of Water Resources Planning and Management, 136(1), 19-26.

[3] Blokker, M., Agudelo-Vera, C., Moerman, A., van Thienen, P., & Pieterse-Quirijns, I. (2017). Review of applications for SIMDEUM, a stochastic drinking water demand model with a small temporal and spatial scale. Drinking Water Engineering and Science, 10(1), 1-12.

[4] Moerman, A., Blokker, M., Vreeburg, J., & van der Hoek, J. P. (2014). Drinking water temperature modelling in domestic systems. Procedia Engineering, 89, 143-150.

[5] Burkhardt, J. B., Woo, H., Mason, J., Shang, F., Triantafyllidou, S., Schock, M. R., Lytle, D. & Murray, R. (2020). Framework for modeling lead in premise plumbing systems using EPANET. Journal of water resources planning and management, 146(12), 04020094.

[6] van der Leer, D., Weatherill, N. P., Sharp, R. J., & Hayes, C. R. (2002). Modelling the diffusion of lead into drinking water. Applied Mathematical Modelling, 26(6), 681-699.

[7] Shang, F., Woo, H., Burkhardt, J. B., & Murray, R. (2021). Lagrangian method to model advection-dispersion-reaction transport in drinking water pipe networks. Journal of Water Resources Planning and Management, 147(9), 04021057.

[8] Xie, Y. (2010). Dissolution, formation, and transformation of the lead corrosion product PbO2: Rates and mechanisms of reactions that control lead release in drinking water distribution systems. PhD thesis, Washington University in St. Louis.

[9] Davis, M. J., Janke, R., & Taxon, T. N. (2018). Mass imbalances in EPANET water-quality simulations. Drinking Water Engineering and Science, 11(1), 25-47.

[10] Private communication with Juneseok Lee, Jonathan Burkhardt, Feng Shang

[11] Klise, K. A., Murray, R., & Haxton, T. (2018). An Overview of the Water Network Tool for Resilience (WNTR). Presented at 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25, 2018

[12] Agudelo-Vera, C., Blokker, M., Vreeburg, J., Vogelaar, H., Hillegers, S., & van der Hoek, J. P. (2016). Testing the robustness of two water distribution system layouts under changing drinking water demand. Journal of Water Resources Planning and Management, 142(8), 05016003.

[13] Lytle, D. A., Formal, C., Cahalan, K., Muhlen, C., & Triantafyllidou, S. (2021). The impact of sampling approach and daily water usage on lead levels measured at the tap. Water Research, 197, 117071.

[14] Lytle, D. A., Schock, M. R., Wait, K., Cahalan, K., Bosscher, V., Porter, A., & Del Toral, M. (2019). Sequential drinking water sampling as a tool for evaluating lead in flint, Michigan. Water Research, 157, 40-54.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14742

ROADMAP DRINKING WATER DISTRIBUTION: AN INVENTORY OF THE CHALLENGES OF THE DUTCH WATER COMPANIES AND RESEARCH NEEDS TOWARDS A FUTURE PROOF WATER SUPPLY

Mirjam Blokker[1] PhD MSc, Karel van Laarhoven2 PhD MSc, Petra Holzhaus[3] MSc, Jan Vreeburg[4] PhD, MSc

[1,2,3,4] KWR water research Institute
PO Box 1072, 3430 BB Nieuwegein; Groningenhaven 7, 3433 PE Nieuwegein, The Netherlands
[1]Technical University Delft, Faculty of Civil Engineering and Geosciences, Stevinweg 1, 2628 CN Delft.

[3] *petra.holzhaus@kwrwater.nl*

Abstract

New developments bring challenges for the water sector. While the world quickly changes, the distribution network is relatively inert and can only slowly change. The drinking water companies in the Netherlands perform a joined research program (BTO, BedrijfsTak Onderzoek) to the impact of rapid world changes to the drinking water supply and prepare measures to cope with this. In 2020 broad expert panels of all drinking water utilities participating in the BTO discussed their expected and desired future developments, ultimately leading to a new roadmap for distribution research for the period till 2050. This roadmap is the backbone for the new research period for the BTO and simultaneously serves as a reflection on the present research program. The most important societal challenges in which Dutch drinking water companies want to invest are found in in (a) cooperation in the underground, (b) customer satisfaction, (c) water quantity and quality and (d) future-proofing of the distribution network. For the research program this indicates that knowledge needs to be expanded in three directions: (1) water quality and temperature of the drinking water in the distribution network, (2) justified decisions on maintenance and replacements and (3) transition toward strategic network design.

Keywords
Water quality, asset management, research programming.

1 A STRUCTURAL APPROACH TO CHALLENGES FOR SAFE AND HEALTHY DRINKING WATER SUPPLY: THE BTO (*B*EDRIJFS*T*AK *O*NDERZOEK) RESEARCH PROGRAM

Social developments change the use of public space in short time and in new directions. Energy transition, artificial intelligence, urbanisation, climate change ICT- developments that generate massive amounts of data to mention a few of the drivers. The impact of these drivers on public drinking water supply will be substantial. There is a need to develop knowledge to analyse the possible impact and to develop a strategy to answer to these challenges and remain the high level of drinking water supply in the Netherlands.

The publicly owned Dutch drinking water companies jointly formulate six-year research plans to develop knowledge that is needed. KWR Water Research Institute facilitates the formulation of the research plans and is responsible for execution of these plans. Partly in own research and partly in close cooperation with scientific institutions as universities, laboratories, start-ups, private companies and international research institutes.

KWR is owned by the ten Dutch public water companies and as such is a not-for-profit, independent research institute serving to the best interest of the Dutch drinking water companies.

The funding for the BTO program is a solid base to develop universal knowledge that is tailored and implemented to the needs of individual companies/shareholders and further developed and challenged within an international arena of scientific and commercial stakeholders.

This structural approach to the challenges in the form of the joint research program BTO for basic knowledge in combination with the tailored company-wide implementation leads to a high performance public drinking water supply. The annually published statistics on the performance (the VEWIN factsheet) show the core figures: chlorine free drinking water supplied with non-revenue water of 3-5%, failure rate of 0.05 per km per year to an average tariff of €1.31/m^3 based on full cost recovery (figures 2020).

The BTO has six year cycle in which the headlines for research are established. On a yearly base the results of research are communicated and evaluated to and with the water companies. If needed topics are adjusted or changed gradually, keeping in mind the headlines.

In this paper the process and results of the preparation for the new cycle for the topic of distribution of chlorine free drinking water are described. The result is the Roadmap Distribution 2050.

2 THE PROCESS: WORKSHOPS WITH WATER COMPANIES

To develop the Roadmap 2050 while a pandemic is going on was a challenge in itself. As general set up we have chosen for a series of workshops with individual water companies in the period March-December 2020. The first workshop was on the 3rd of March and turned out to be the only 'old-fashioned' non-digital one and started from a common base. Each following workshop was updated with information of the previous one following a 'itterative concept'. Eventually, all retrieved visions and expectations on the development of the drinking water sector in the Netherlands were brought together in a final workshop on December 16, with all the pro's and con's of a massive digital workshop.

In 2008 the first roadmap for distribution research was set up, in which the planning period was 5 years (Vloerberg 2008). In 2017 the research plan for the guiding committee for the Distribution research was set up in close collaboration with a group of experts, the so-called Quartermasters. This plan looked 6 years ahead (Blokker, 2017). During the research periods in 2010/2011 various initiatives were undertaken to understand what is needed to get to the core of "operational excellence" and what research is needed to get there. Various policy relevant drivers were taken into account such as lowest (societal) costs, best performance, adaptivity, transparency towards end user, etc (see also Blokker et al. 2015).

For the present series of workshops we used several theories to develop future visions, some brought in by the water company, some developed by KWR, resulting in the universal core set up for the workshops. With each specific water company the approach was tailored to the developments within the company towards strategic approach of future scenario's. With one company the workshop was based on their own vision towards 2040, another workshop concentrated on four future scenario's based on urban/societal development. In several workshops we worked with thematic scenario's (drought, foreseeable future, pandemic, energy transition, etc). Each new workshop was adapted to the results of the previous one, resulting in the snowball effect and a maturing insight.

The horizon for the roadmap was set at 2050 as a suitable date. In that year society plans to be 'climate adaptive', 'energy neutral' and natural gas as energy source is banned and replaced with solar and wind energy. Basic question is "How can we see to it that in 2050 the drinking water network is equipped for task in line with contemporary societal expectations?"

3 THE RESULTS OF THE WORKSHOPS

Most of the participants in the workshops had a technical background in the distribution of drinking water. In total over 80 people participated in 10 workshops, including the Belgian Watergroep as ‘foreign’ participant in the BTO (one of workshops was a combination of two companies; one of the benefits of digital working) in which a good mix could be found of researchers and scientists as well as practical operators. All being well familiar with distribution issues, sometimes discussions went very in depth on technical impact and consequences of various developments. But thanks to the broad set up, also the societal and other aspects were contemplated.

The closing workshop in December 2020 was visited by 59 participants. The results of the company workshops and the earlier roadmaps were consolidated into a number of statements that could be confirmed or denied, followed by lively discussion. This technique enforces clear statements on how to formulate the outcome of the very elaborate discussions. It also channels the conclusions into clear and comprehensive outcomes.

The first set of outcomes make clear in what topics water companies are prepared to invest:

a) cooperation in the underground or urban subsurface space;

b) customer satisfaction;

c) water quantity and quality;

d) future-proofing of the distribution network;

This results into research need in three directions:

- water quality and temperature of the drinking water in the distribution network;
- justified decisions on maintenance and replacements;
- transition toward strategic network design.

Not surprisingly this highly coincides with the outcomes of the earlier outcomes for the period 2018-2023 that resulted in three research directions: Decision making on maintenance and replacements networks; Design of futureproof networks and Ensuring good water quality in the network.

This shows the complexity of the topics mentioned and the validity of the results already obtained in the various research projects conducted in the period 2018-2023.

The Roadmap Distribution 2050 is schematically draw in Figure 1

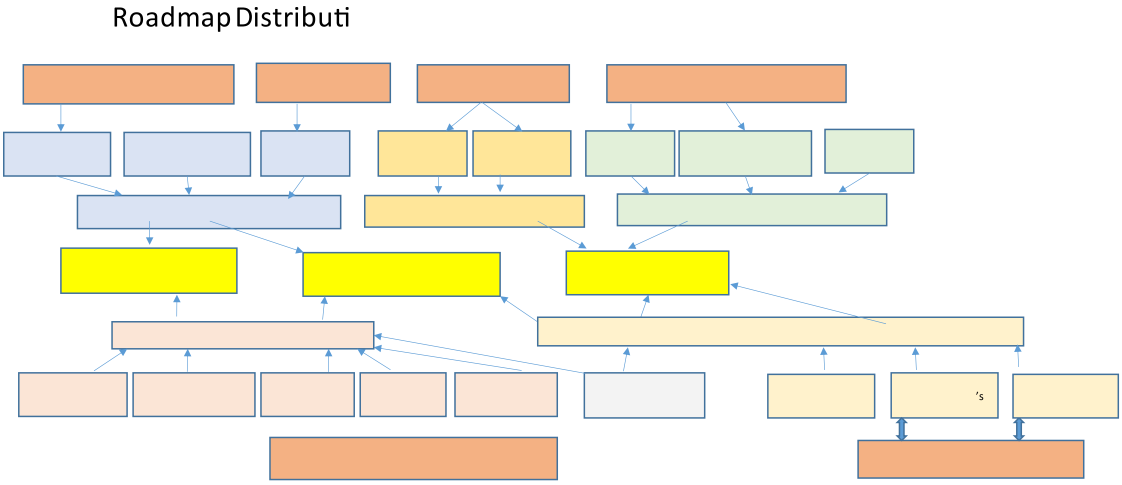


Figure 1 Distribution Roadmap 2050

The yellow boxes in the centre form the projected outcome: knowledge to be used for achieving the main goals. The brown boxes at the top and bottom of the roadmap give the models that were developed within earlier research programs. Compared to the earlier roadmap these modes are not new, but are entering in a new phase. All the other boxes give the developments that need to be clarified and quantified and can be translated into research projects.

CC BY NC SA

4 RESEARCH PROGRAM AND COLLABORATION WITH STAKEHOLDERS; OBSERVATIONS AND RECCOMANDATIONS

Relevant research in drinking water networks can only be done in very close cooperation with water companies. Modelling has proven to be a very strong method to do research in a network and calculate effects of many scenario's. Useful data needed as input for the models for verification and validation can be obtained when researchers and technicians of the water companies work closely together . Only than reliable simulations can be made and scenario's calculated or optimisations performed.

In the intensive series of workshops several effects are reached. During the presentations it became clear that even despite the numerous meetings and dissemination activities in the past it takes more time to adapt to new knowledge on a broad level in companies. A lot of the research questions and topics raised in the workshops were already dealt with in theoretical studies and reported on. This brings to mind the statement "the power of education is repetition" but also that there is still a gap to be bridged to bring the theory to the day to day practice. This implies an obligation of both the water company as the researchers.

Data on all aspects of the drinking water distribution are needed for the research. Though much is expected from sensors and sensor networks, most relevant data must be generated 'in the field'. This identifies the need for a culture change at water companies. The culture of water companies is to measure performance in quantities: km pipe constructed, cubic metres water produced, continuity of supply, etc. In such a culture a failure is almost an insult that needs to be corrected as soon as possible, also encouraged by performance indicators on continuity of supply. However, a lot of information can be gathered by carefully observing the failure in it's circumstances, salvage of the material and thorough analysis of that. This demands for a culture in which the analysis of events and safeguarding data is at least equally important as rapid repair and re-instalment of the supply service.

Together with this culture change on the relevance of careful data capturing in conventional operation (next to the promising sensor data on water quality) pilots and demonstrators should be part of the research process. In a 'true pilot', technology or concepts proved in laboratories, bench scale testing or modelling are applied in practice to see if they are scalable and still effective. Only in a 'demonstrator' the scalable technology is tested for factors that are essential for a commercial application. In the BTO- research 'true pilots' are often confused with demonstrators and it is highly recommended to align the goals and expectations before starting a pilot or demonstrators and communicate consequently with respect to these goals and expectations.

With increasing possibilities of modelling and optimisation in combination with a rapid growing knowledge base and data availability, the fundamentals of drinking water distribution came up as a topic in the workshops: What is resilience, robustness, continuity, capacity, accessibility, maintainability, sensor usage, interaction with environment, societal costs, stewardship, etc. These discussion require a broad interaction with all (technical) aspects of drinking water supply system with the societal wishes and demands. Especially this is interesting and involves a lot of new skills and knowledge.

BY NC SA

5 REFLECTION

Direct, non intended, result of this series of workshops is a broadening of the knowledge of the results of previous research within all the drinking water companies. Also the realisation that implementation needs constant attention and care. Very helpful in this has been the overview of 20 years of research on distribution, and the robust approach on main lines. All the research fitted within the three main topics both identified in this roadmap as in the previous research roadmaps and plans. The development is consistent and logical and gives confidence that this are the most important topics.

The clusters of knowledge needs and questions are sharpened from a practical perspective. It gives a better insight in the coherence with external factors and stakeholders and results in more explicit wording of the various approaches and steps that are needed. Effectively, the overarching questions on which the new research program will focus are:

- On the field of temperature and residence time: what is the microbiological risk of high temperature in the network, given the existing (spread in) residence times; how can these risks be contained; what role is played by the house installation.
- On the field of collaboration in the underground (with other ‘inhabitants’): How can partners in the underground align their replacement or construction plans to optimise the societal benefit; how do we get specific knowledge on our networks enabling us to make well contemplated decisions on maintenance and replacement.
- On the field of design of networks for the future: which goals do we formulate for the future networks and what are the drivers for networks design for the next 100 years? Which network structures will help us realise these goals? Which measuring technologies and which models need to become ‘parts of the network’ to help realising these goals.

6 REFERENCES

[1] Agudelo-Vera, C., Blokker, E. J. M., Büscher, C. en Palmen, L. (2015). "Drinkwaterinfrastructuur van de Toekomst - Interactieve sessies met experts van de drinkwaterbedrijven." BTO 2015.052, KWR, Nieuwegein.

[2] Beuken, R. H. S., Koop, S. H. A., Brouwer, S., van Alphen, H. J. en Eijkman, J. (2020). "Onderbouwing van integrale besluitvorming voor assetmanagement." H2O-Online.

[3] Blokker, E. J. M. (2017). "Zesjaren- onderzoeksplan; thema Distributie." BTO 2018.003, KWR, Nieuwegein.

[4] Blokker, E. J. M., Büscher, C., Palmen, L. en Agudelo-Vera, C. (2015). "Strategische planning van drinkwaterinfrastructuur: een conceptueel kader en bouwstenen voor drinkwaterbedrijven – Nederlandstalige samenvatting bij BTO 2015.048." BTO 2015.049, KWR, Nieuwegein.

[5] Brouwer, S. en Sjerps, R. M. A. (2018). "Klantperspectieven." BTO 2018.083, KWR, Nieuwegein.

[6] VEWIN (2020). “Drinking water fact sheet 2020”, https://www.vewin.nl/SiteCollectionDocuments/Publicaties/Cijfers/Vewin-Drinking-Water-Fact-Sheet-2020.pdf

[7] Vloerbergh, I. N. (2008). Waterdistributie van de toekomst; op tijd voorbereid, Kiwa Water research, Nieuwegein. BTO 2008.022.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14748

DATA DRIVEN MONITORING OF IOT ENABLED WATER DISTRIBUTION NETWORKS

Rohit Raphael[1] and Sridharakumar Narasimhan[2]

[1]Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India)
[2]Dept. of Chemical Engineering and Robert Bosch Centre for Data Science and Artificial Intelligence,
Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India)

[1] *rohitraphael@gmail.com*, [2] *sridharkrn@iitm.ac.in*

Abstract

Real time monitoring capabilities of many domestic water networks are still very limited. This could lead to several direct and indirect issues such as inability to detect leaks leading to wastage of water, failure to identify over usage and exploitation of water by certain end nodes in the system, non-equitable supply of water and abnormalities like pipe volume reduction due to scaling and deposition of silt in pipelines. Since water networks are expected to operate for several decades, these issues lead to large uncertainty in the quantity of the water being supplied and utilised by any community. The present work proposes a system for monitoring the water supply system with the help of Internet of Things (IoT) where sensors and free spectrum communication techniques measure, transmit, store and analyse parameters of interest like flow, level and vibration. Here we make use of easily available and economically feasible sensing techniques like ultrasonic level sensors, hall effect based flow sensors and 3 axis gyroscope based vibration sensors. Suitable communication modules based on free spectrum transmission techniques like HC-12 and LoRa are utilised for transmitting the data from the sensor nodes. This work emphasizes the use of non-intrusive sensing methods, thereby reducing the difficulties arising due to disruption in water supply during installation and maintenance of the system. Low-cost monitoring capabilities and scalability are inherent advantages of the proposed work.

Keywords
Water distribution networks (WDN), Internet of Things (IoT), Telemetry, Free spectrum communication.

1 INTRODUCTION

1.1 WDN monitoring using IoT based sensor systems

Access to clean water is critical for the existence of life in the planet. Thus monitoring of water distribution networks using basic parameters like flow and level to identify the quantitative variation of the water supplied at different nodes is of great importance. In majority of the water distribution systems we come across in and around us, there is near to none quantitative feedback taken from the network to measure the amount of water being utilised, wasted or circulated in the system. Being a system that is meant to run for an indefinite period of time, this leads to a large uncertainty in the quantity of the water being supplied and utilised by any community. Proper quantisation of the water supply system with the help of sensors which can measure and store the parameters of interest like flow and level in this case, can bring down this amount of uncertainty to a large extend.

1.2 Related work

Water networks and efficient usage of water resources are actively studied due to the decrease in quantity and quality of water available to consumers. Deployment of monitoring systems [1] in water distribution networks can enhance the quality of service by balanced supply and reduction

BY NC SA

of water wastage. It has been estimated that the per capita water usage of a developed nation like USA is around 330 lpcd [2] while that of a country like India is 135 lpcd [3].

Internet of things (IoT) represents a network of physical objects or things which are inter-connected and exchange data between them. It has been gaining popularity in a variety of fields in the recent years due to its extensive applications and possibilities in low power, inexpensive and reliable data collection and transmission capabilities. In any sensor based systems for remote used cases, there is always a concern regarding the range of the nodes to transmit the collected data [4]. Application of new technologies including free spectrum transmission networks can rectify such concerns and thus help devise a practical system for WDN monitoring. Adige, a LoRa based WDN monitoring system communicates to several hundreds of meters with much lower energy expenditure [5] when compared to traditional GSM/GPRS based systems, hence can be used in sensing data from remote location to a receiver gateway far from the sensor node. Making use of tools like Arduino IDE, SQL workbench, Grafana, Python IDE etc. for programming, data collection, storage and visualisation provides an efficient platform for developing a novel WDN monitoring system.

2 SYSTEM ARCHITECTURE

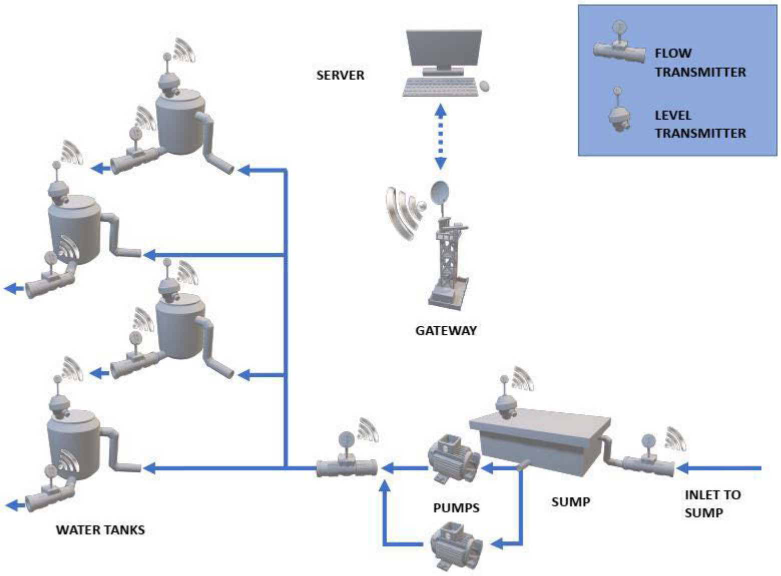


Figure 1. Domestic water distribution network

2.1 Components of parameter monitoring system in WDN

A typical domestic water distribution system consists of a sump, pumps, pipes and overhead tank as shown in Figure 1. Main components of a typical WDN monitoring system are the sensors (flow and level), the communication module, the sensor interface module and the gateway. The gateway pushes the data to the server which stores the data and visualises it on the dash board. The basic requirements for a sensor module for large scale deployments are the economic feasibility for acquisition and installation in large numbers, ruggedness and readiness for functioning environmental conditions with high heat and moisture, measurement accuracy and ease of electrical interfacing with the electronic system used for the module.

Most of the flow sensors available in market are very expensive since they make use of electromagnetic flow sensing methods or ultrasonic flow sensing methods. The high cost ranging from approximately 150$ (140 €) to a few thousand dollars per sensor render it practically difficult for implementation in large numbers. The more commonly used cheaper alternatives are the turbine based or vane based flow sensors. Here for our requirements, after testing multiple flow sensors available in the market, we decided to use turbine type flow sensors with Hall Effect magnetic pickup from Sea YF-S series sensors. These are readily available and considerably cheaper than the ultrasonic or electromagnetic alternatives.

Majority of the tanks in the water distribution networks have depths ranging from 5-10 feet. To monitor the level of water in these we need level sensors which were cheap and could continuously measure the level variation. The cheapest option was to use traditional conductivity probes which were immediately found unsuitable since they had discrete level measurements. Continuous measurement with a resolution of at least a few cm was necessary for data driven monitoring. In this perspective we had to look for alternatives which were at a reasonable price difference and support our requirements. After checking the available level measurement sensors like ultrasonic and LIDAR type sensors, we decided to finalise on ultrasonic level sensors which were comparatively cheaper and reliable. The added advantage of water proof design was ideal for outdoor applications.

For communication we tested multiple communication protocols like LoRa, HC-12, HC-05, BLE, etc. Depending on the topology the a typical domestic water distribution system we have chosen HC-12 modules. While LoRa is good for long range outdoor transmission with reasonable line of sight, HC-12 is ideal for places where obstructions like buildings are present. These are quite easily available, cheaper than other alternatives, reasonably low powered and ideal for indoor as well as outdoor application.

For interfacing the sensors, collecting and transmitting the data through the communication module, we need a sensor module. This includes a microcontroller which collects the data from the sensors and transmits them to desired destination nodes. Here we use an Atmel ATmega 328 embedded into an Arduino Nano module mounted on a custom PCB with necessary interfacing options with both sensor and communication modules. These PCBs were designed in house and fabricated from third party PCB fabrication companies. These also include solar panel integration for powering the whole system, Li Ion cell for backup during night time, charging and over discharge protection circuits for the batteries and standard interface pin outs for I2C and SPI interfaces.

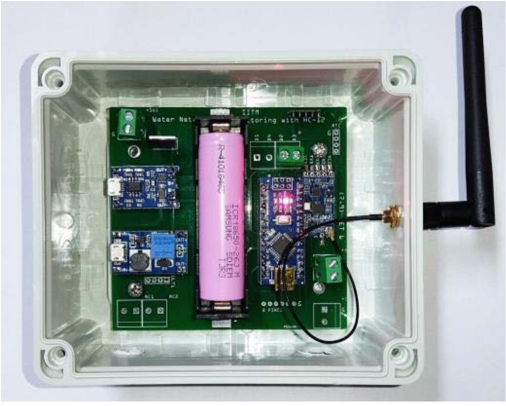


Figure 2. Sensor interface module

For each cluster we needed a gateway node which collects data from the sensor nodes in the cluster and pushes it to the database. The obvious and tested option for this was to use a Raspberry Pi single board computer which met all the requirements including the ability to interface with a HC-12 based Arduino receiver node, LAN and Wi-Fi capabilities to use intranet/internet for transferring the data into the database and remote access feature for troubleshooting using tools like VNC server.

CC BY NC SA

2.2 Hardware design of WDN monitoring system

WDN systems usually have large number of end nodes which collect level and flow data from tanks and pipelines associated with these. Using HC-12 modules for data telemetry data transmission have been found to be very efficient, low cost and reliable [6]. Here we make use of waterproof level sensors like Maxbotix MB series and low cost JSN-SR04 level sensors for level sensing in tanks. The Maxbotix sensor has a range of 10m at a resolution of 1 cm while the JSN-SR04 sensor provides a range of 4.5m at a resolution of 1 cm. The JSN-SR04 has a dead zone of 20 cm which can be tolerated. An ATmega 328 microcontroller based Arduino Nano provides the backbone for the interface unit. ATmega 328 has been found to be an ideal microcontroller for applications similar to ours by many researchers and hardware developers [8]. The Arduino uses two digital pins for interfacing with the JSN-SR04, one for the trigger pin and the other for the echo pin. No special libraries are required for interfacing the sensor since it makes use of the time of travel of the ultrasonic wave to determine the depth/ height of water in the tank.

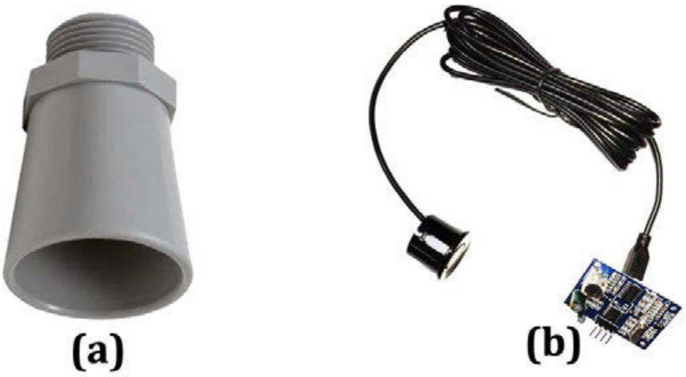


Figure 3. (a)Maxbotix MB series sensor (b) JSN-SR04 level sensor

We used flow sensors ranging from 3/4 inch size to 2 inch size for majority of the installations. These sensors are using a Hall Effect type pulse counter mechanism which utilises the number of pulses to quantify the flow through the pipeline. They have a three wire connection two of which are for the Vcc and Ground and the third wire is the signal wire. These were available in both brass metal build as well as ABS plastic build depending on the size of the pipeline. The smaller sizes from 1/4 inch to 1 inch were brass made and 1.5 inch and above were brass. The installation was fairly straight forward since the sensors came with threaded ends which were using standard thread types and sizes which were compatible with PVC, uPVC and cPVC fitting.

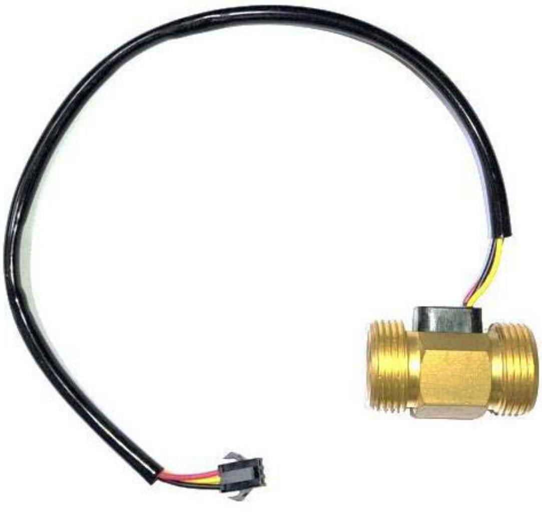


Figure 4. Hall Effect type flow sensors

2.3 Data management and IoT

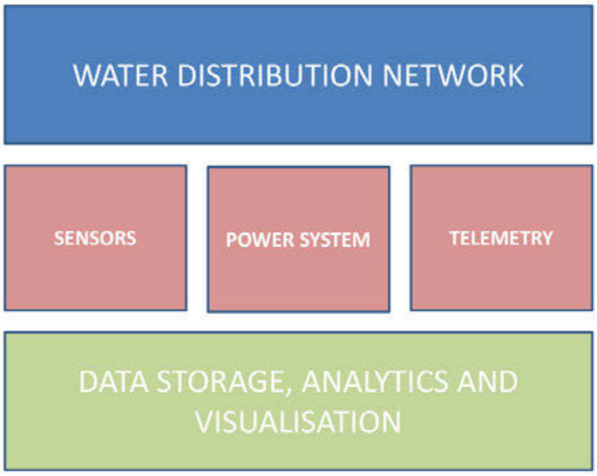

WATER DISTRIBUTION NETWORK
SENSORS
POWER SYSTEM
TELEMETRY
DATA STORAGE, ANALYTICS AND VISUALISATION

Figure 5. Components of an IoT system

Internet of things links objects around us with a central management system for seamless data exchange [8]. In the case of a WDN, the objects are the tanks and pipeline systems and associated sensors monitoring parameters like level and flow. The wireless telemetry provides the link between the objects and the central system comprising of the storage, analytics and visualisation. These will be done using free spectrum technologies like HC-12 or LoRa.

The data acquired from the sensor is then transmitted by the node using an HC-12 wireless transmission module which is interfaced to the Arduino using a software serial based UART interface. HC-12 supports 100 channels from C001 to C100 which can be set using AT commands. It can transmit at 8 different power levels from -1 dBm to 20 dBm [7]. This can also be set using AT commands. These features are quite useful when we are dealing with multiple clusters in nearby areas and chances of interference are quite high. An external antenna connected using a UFL to SMA Cable is also used for better transmission. Because of these customisation options in transmission, HC-12 is very versatile and useful. Isolation of clusters is also possible by assigning different channels for different clusters.

Figure 6. HC-12 module

The gateway node also includes an HC-12 module which is programmed to receive data from different sensor node using time based reception. The data is then transmitted to the server where it is stored and visualised.

A server is deployed to host the data collected from the water distribution systems. This is stored in the database for further analytics. The data is also displayed on a web based dashboard for real time monitoring.

Powering the individual components is also crucial when it comes to IoT systems as these are meant to be low power systems which should run on batteries. The end nodes might be located in remote locations where mains power may not be available and thus need to be powered using rechargeable batteries with solar recharging mechanisms. Gateway systems can be powered using mains power since these cater to a large number of end nodes and can be installed in locations with power availability. Gateway systems also require networking capabilities thus wired or wireless network access also needs to be provided. Data management in the central system is carried out in a well-equipped system, typically a server, PC or industrial computer.

3 RESULTS

We consider a system consisting of intermittent water supply. Since water is supplied intermittently, a ground level storage reservoir is used to store the water supplied and pumped to overhead tanks (as shown in Figure1). Current operation procedure for pumping of water from sumps into overhead tanks is purely time based and there is no feedback from the tank given to the pump operator regarding the level of water in the tank. Thus there is complete uncertainty regarding the amount of water being pumped into the tank and how long the pump has to run for maintaining optimum water in it. All the tanks have a ball float arrangement to prevent overflow while the pump continues to run at low flow rates.

3.1 Operation schedule

Figure 7 shows the variation of flow rate of water pumped into a tank vs. the level of water (depth from surface) in one of the installations.

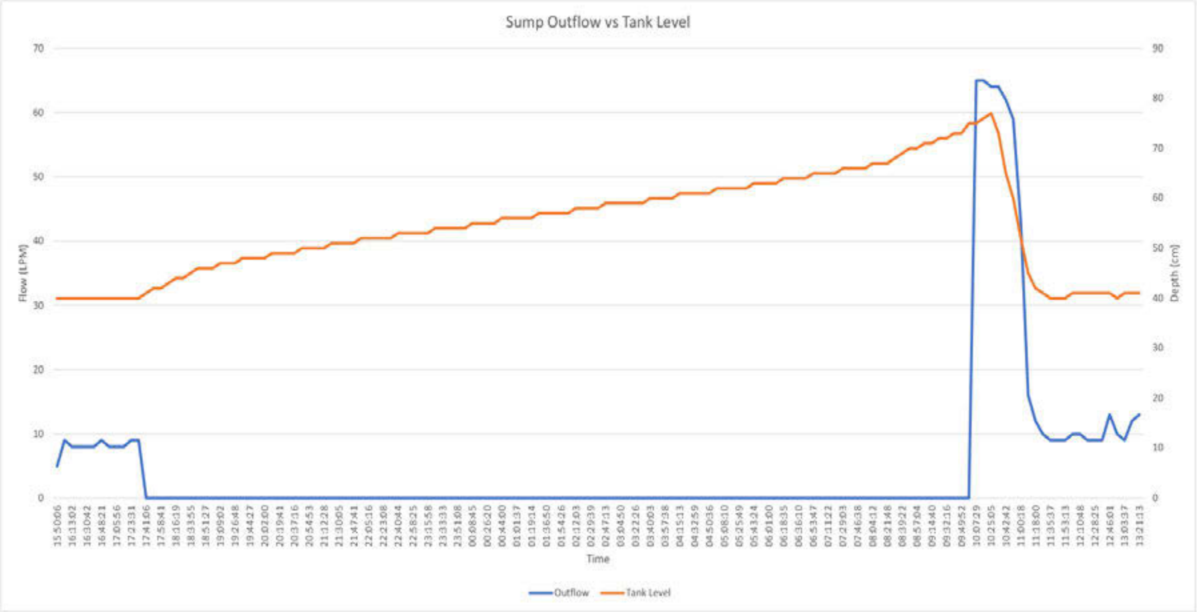


Figure 7. Flow rate of water pumped (LPM) vs. depth of water from top of tank (cm)

We can see that same pump operation schedule followed on a daily basis. Here it is evident that the pumping starts at around 10:00 am and goes on till 5:30 pm. The pump is pumping process fills the tank at around 65 litres per minute (LPM) and the tanks get filled by around 11: 30 am. The ball floats valve closes the input to the main tank and the flow rate into the tank reduces to around 10 LPM. But unaware of this the pump is still running for a much longer time till 05:30 pm. This is caused because of the absence of a feedback system to alert the pump house when the tank is full. This leads to wastage of power by running the pump beyond the required time to fill the tank. The 3 hp pump at the pump house consumes around 2kW of power from our observation. Daily the pump is running for an extended period of 5 hours, which results in an energy loss of 10 kWh. Approximate estimate of water usage in the building monitored was found to be 10275 LPD (litres per day).

BY NC SA

3.2 Estimation of flow from level and associated measurements

One of the main advantages of having a data driven monitoring system is the derivation of unknown parameters from known ones. Flow is one of the most important parameter we need for WDN monitoring system. But flow meters are mostly intrusive and require more maintenance when compared level sensors. Here we make use of the level data to derive the flow and then compare this derived value with actual measured value in the network.

We look into a number of parameters in this estimation technique as follows:

Table 1. Relevant parameters for flow estimation and verification

Parameter	Character	State / Unit
State of pump	A	ON/OFF
State of inflow	B	ON/OFF
Level in tank	C	cm
State of ball float valve	D	Open / closed
Flow into sump	F_1	LPM
Flow from sump	F_2	LPM
Volume of sump	V	L

Among these parameters, F_1 and F_2 are measured to verify the accuracy of the estimated flow rates to the actual values. Considering the other parameters, it is possible to estimate the water consumed from minimal measurements. Here we consider a network which consists of an underground sump which receives water from the main supply line from the main sump. The water is then pumped to the main water tank on top of the building from where it is distributed to the end nodes.

Here we can consider 4 cases as follows:

Case 1: A on; B on; C known,

In this case,

$$F_1 - F_2 = \frac{dV}{dt} \tag{1}$$

In order to compute the volume of water supplied or consumed, when there is simultaneous filling and pumping of water, we need to determine F_1 or F_2 as shown above in equation (1). Under some conditions, it is possible to determine the water supplied or consumed if F_1 or F_2 follow certain trends and can be estimated independently.

Case 2: A off; B on; C known

In this case,

$$F_1 = \frac{dV}{dt} \tag{2}$$

This relation can be used to estimate the filling rate to the sump or equivalently the water supplied. If the filling rate does not vary over a time period the average flow rate can be calculated using a linear fit as described below.

In Figure 8 we considered roughly 400 data points collected from level sensor which has been time stamped using an RTC (real time clock) module. We use a linear trend line to determine the rate of change of volume in the sump. Fig. 1 shows the plot of Level vs. time in sump corresponding to case-2. The slope of the trend line determines an approximate value of rate of change of volume. This gives the inlet flow rate at normal conditions. It was found to be approximately 7.99 LPM. This is very close to the actual reading of the flow meter which was approximately 8 LPM. If the filling rate varies, the volume supplied can be obtained by numerical integration.

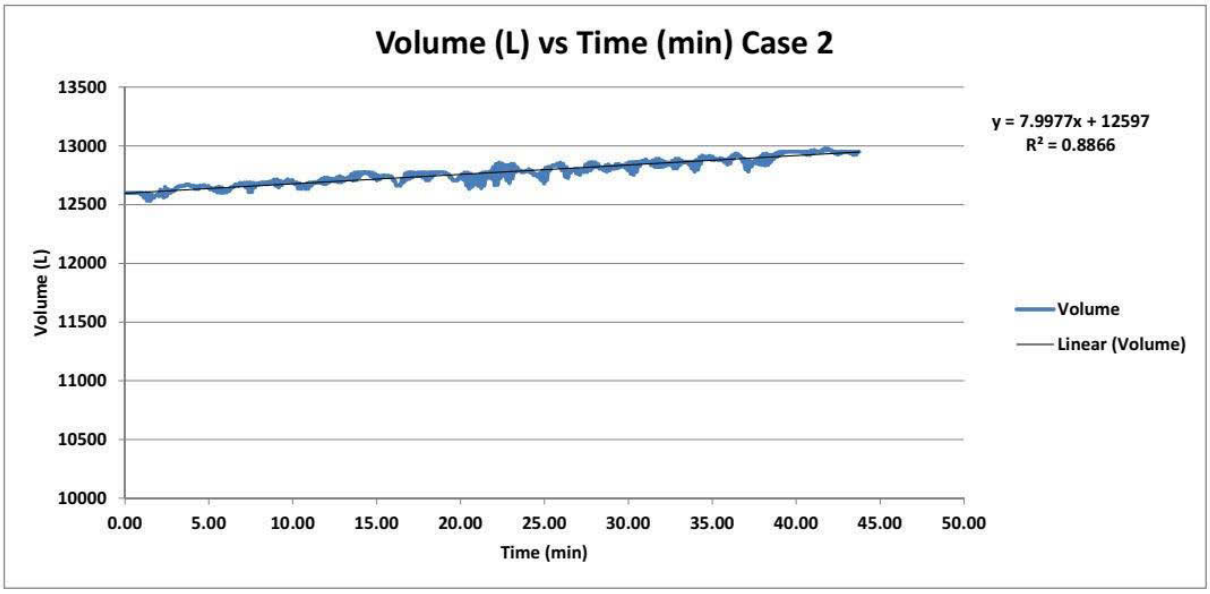


Figure 8.Volume vs. Time (Case 2)

Case 3: A on; B off; C known

In this case,

$$F_2 = -\frac{dV}{dt} \tag{3}$$

Equation (3) gives the pumping rate of the sump. This value is constant since the pump runs at identical conditions all the time. Here also we considered roughly 400 data points collected from the level sensor in the sump. In Figure 9, the slope of the linear trend line gave a value of around -67.45 LPM. This is approximately equal to the 65 LPM shown by the flow sensor installed in the line. The negative shows that the flow is from the tank and thus the volume varies with a negative slope.

Case 4: A on; B off; C known; D closed

This condition occurs when the float valves in the tanks gets closed when the water level reaches the max value. This causes a high resistance in the system and thus the flow rate decreases from the pump although the pump is running. This drastically reduces the flow rate which can be monitored from the level feedback from the supplied tank. In Figure 10, the slope of the trend line was found to be -9.39 which is very close to the 10 LPM value observed in the flow sensors.

BY NC SA

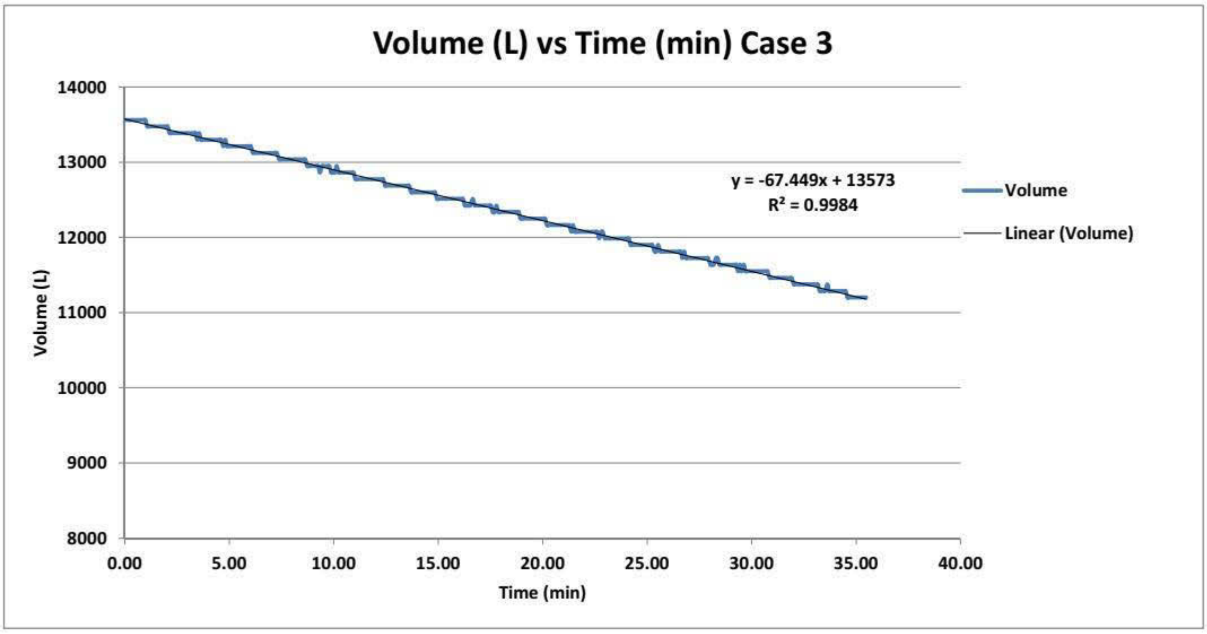


Figure 9.Volume vs. Time (Case 3)

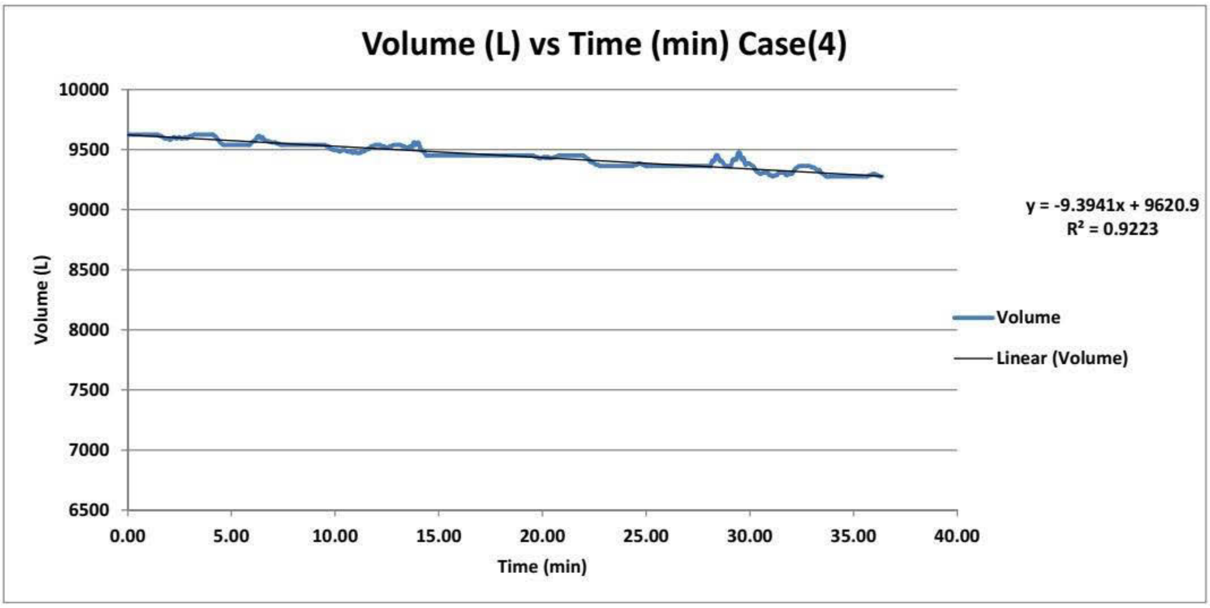


Figure 10.Volume vs. Time (Case 4)

4 CONCLUSIONS AND FUTURE SCOPE

In this paper we present an efficient, easy to implement IoT based wireless system using free spectrum technologies for monitoring water distribution networks. Uncertainties arising both from operation of the network from the operator side and utilisation of resource from the end user side is minimised. The disruption of water supply arising due to installation and maintenance of flow sensors is avoided processing the data appropriately to determine the flow from level and associated parameters.

Data collected from the IoT base WDN monitoring system is stored for further analysis. Real time data from the monitoring system can be visualised for checking the performance and operation schedule of the WDN. Further developments in both software and hardware side to implement a dashboard for mobile application is underway. The proposed system can be scaled for larger networks in various communities and organizations.

5 ACKNOWLEDGEMENT

This work is funded by Robert Bosch Centre for Data Science and Artificial Intelligence, IIT Madras, Chennai, Tamil Nadu, India.

6 REFERENCES

[1] Chinnusamy Saravanan, Prasanna Mohandoss, Partha Paul, R. Rohit, N. Murali, S Murty Bhallamudi, S. Narasimhan and S. Narasimhan, "IoT enabled monitoring and control of water distribution network," in WDSA/CCWI Joint Conference Proceedings, Ontario, Canada, vol. 1, 2018.

[2] Maupin, Molly A., Joan F. Kenny, Susan S. Hutson, John K. Lovelace, Nancy L. Barber, and Kristin S. Linsey, “Estimated use of water in the United States in 2010,” in US Geological Survey, vol. 1405, 2014.

[3] Jaladhi, Vavaliya, Bhavsar Dhruv, Kavadi Utkarsha, and Mohammad Mahroof. "Online performance assessment system for urban water supply and sanitation services in India," in Aquatic Procedia 6, 2016, pp. 51-63.

[4] Shahmirnoori, Alireza, Motahareh Saadatpour, and Amin Rasekh. "Using Mobile and Fixed Sensors for Optimal Monitoring of Water Distribution Network under Dynamic Water Quality Simulation," in Sustainable Cities and Society, 2022.

[5] Cattani, Marco, Carlo Alberto Boano, David Steffelbauer, Stefan Kaltenbacher, Markus Günther, Kay Römer, Daniela Fuchs-Hanusch, and Martin Horn. "Adige: an efficient smart water network based on long-range wireless technology," in Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, 2017, pp. 3-6.

[6] Jose, Dylan Philip, Alton Lavin D'Souza, Athina Ann Thomas, and D. Daniel. "IoT based water management using HC-12 and Django," in 2019 IEEE International Conference on Data Science and Communication (IconDSC), 2019, pp. 1-6.

[7] Robert Rozee, “HC-12 Wireless Serial Port Communication Module,” http:/datsi.fi.upm.es/docencia/Informatica_Industrial/DMC/HC-12_v2.3A.pdf, Jan. 2016.

[8] Perumal Thinagaran, Md Nasir Sulaiman, and Chui Yew Leong. "Internet of Things (IoT) enabled water monitoring system," in 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE), 2015, pp. 86-87.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14750

COUPLING AGENT-BASED MODELING WITH WATER DISTRIBUTION SYSTEM MODELS TO SIMULATE SOCIAL DISTANCING AND WATER INFRASTRUCTURE PERFORMANCE DURING COVID-19

Brent Vizanko[1], Leonid Kadinski S.M.ASCE[2], Avi Ostfeld F.ASCE[3], Emily Berglund M.ASCE[4] and Christopher L. Cummings[5]

[1]Dept. of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, North Carolina, USA
[2]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, Israel
[3]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, Israel
[4]Dept. of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, North Carolina, USA
[5]US Army Engineer Research and Development Center [Contractor]; Senior Research Fellow, Genetic Engineering and Society Center, North Carolina State University, Raleigh, NC, USA; Gene Edited Food Program, Iowa State University, Ames, Iowa, USA

[1] *bvizank@ncsu.edu, [2]kleonid@campus.technion.ac.il, [3]ostfeld@technion.ac.il,*
[4] *emily_berglund@ncsu.edu, [5]christopherlcummings@gmail.com*

Abstract

Beside the immense impacts on public health, the COVID-19 pandemic also disrupted daily routines for people around the globe due to the adoption of social distancing measures, such as working from home and restricted travel in order to minimize viral exposure and transmission. Changes in daily routines created new water demand patterns, and the spatial redistribution of water demands in urban water distribution system networks affects water age, nodal pressures, and energy consumption. A range of factors influence individuals' social distancing decisions including demographics, risk perceptions, and prior experience with infectious disease. This presentation reports a comprehensive modeling framework to capture decisions to social distance, the effect of social distancing on water demands, and the effects on the performance of water infrastructure. First, new Bayesian Belief Network (BBN) models are developed to simulate social distancing decision-making based on publicly available survey data describing COVID-19 risk perception, social distancing behaviors, and demographics. Data were collected in March and April of 2020 and included over N=6,991 participants from 11 countries in North America, Europe, and Asia. Feature sets are developed from participant characteristics using forward selection and Naïve Bayes classifiers to predict behaviors, including working from home. BBN model output is used within an agent-based modeling (ABM) framework to simulate how individuals interact within a community and dynamically adopt social distancing behaviors based on communication and transmission of infection. Agents represent individuals who transmit COVID-19, communicate with each other, decide to social distance, and exert water demands at residential and non-residential locations. COVID-19 transmission among agents is modelled using a susceptible-exposed-infected-removed (SEIR) model. Finally, the ABM is coupled with a water distribution model to simulate how changes in the location of demands affect water distribution metrics. The model is applied for a virtual city, Micropolis, to explore how varying population characteristics can affect water infrastructure. This research provides a new framework to develop and evaluate water infrastructure management strategies during pandemics.

Keywords
Bayesian Network, Water Distribution System, Modelling, COVID-19, Agent-based Modelling.

1 INTRODUCTION

The COVID-19 pandemic, caused by the novel coronavirus (SARS-CoV-2), has caused immense public health concerns and impacted communities around the world. Governmental mitigation efforts including lockdowns and mask mandates have been widely instituted, and one third of the global population lived under some form of restriction in April 2020 [1]. By January 1, 2021, more than 80% of countries had a mask mandate in effect, and approximately 50% had mandated one or more social distancing measures, such as working from home, workplace closures, school closures, or international travel controls [2]. These social and economic restrictions were implemented to mitigate the transmission of the coronavirus but also had profound impacts on the daily lives of the people living with restrictions [3]. Many individuals adapted their daily routines of commuting to work and visiting places of interest to shop, dine out, or socialize. As a result, communities changed their interactions with infrastructure and their consumption of resources and services provided by infrastructure. One impact of these behavioral changes is the spatial and temporal change to domestic water demand, with an overall increase in residential water demand, reduction in the overall demand, and shifting of the common bimodal daily pattern [4]. Changes in demand subsequently impact the operation and management of the water distribution systems (WDSs), which are designed to deliver water to meet demands and expectations around the levels of service, including pressure. Water utilities reported a range of operational and management challenges due to changes in customer behaviours and observed noticeable differences in water demands during the COVID-19 pandemic [5].

To address water management challenges that arise during pandemics, water utilities need comprehensive modelling tools for demand, pipe flow, and pressure prediction that account for consumer behaviors, pandemic coping strategies, mobility, and changes in demands. Integrated modelling frameworks have been developed to simulate consumer behaviors to estimate changes in water demands and the associated effect on network performance. Modelling frameworks were developed to simulate demand changes and changes in infrastructure performance based on consumer decisions to adopt water reuse technology and alternative water sources [6-8] and consumer responses to contamination of drinking water [9-11]. The tool developed in this research builds on these previously developed frameworks, listed above, that couple agent-based models (ABMs) with hydraulic modeling. Agents represent individual water consumers that exert demands at nodes in a water distribution system and travel among nodes using diurnal patterns. As agents become aware of disease transmission, they make decisions to social distance, including working from home and cooking meals at home. A Bayesian Belief Network (BBN) modeling approach is used to represent agent decision-making around social distancing behaviors. The BBN model was developed using survey data that was collected to explore psychological predictors, risk perception, and coping strategies during the COVID-19 pandemic (N=6,991). A susceptible-exposed-infected-removed (SEIR) model is integrated to simulate COVID-19 transmission, using parameter settings that are specific to the transmission of COVID-19 [12]. The ABM framework is applied to simulate coping strategies that are taken by consumer agents to avoid exposure to COVID-19 and the emergent shifts in water distribution system performance metrics, including pipe flows and nodal pressures.

2 MATERIALS AND METHODS

An ABM framework is developed to integrate three modules, including a BBN model for agent decision making, a SEIR model for COVID-19 transmission, and a hydraulic model for simulation of the water distribution system flows and pressures (*Figure 1*). Individual water consumers are represented as agents that move between work and home nodes based on predetermined patterns. Agents make decisions to work from home based on the posterior probability of the BBN based on their individual state and parameters at each time step (*Table 1* lists agent parameters).

COVID-19 is transmitted between agents when a healthy agent occupies the same node as an infected agent and a threshold exposure probability is attained. Information about other infected agents is used to update an agent's understanding of the environment and inform its decision at the next time step. Once agent mobility and COVID-19 transmission are complete, the demand at each water network node is calculated based on the number of individual agents at each node compared the node capacity. This framework is shown in *Figure 1*, and the modules of the framework are described in the following sections.

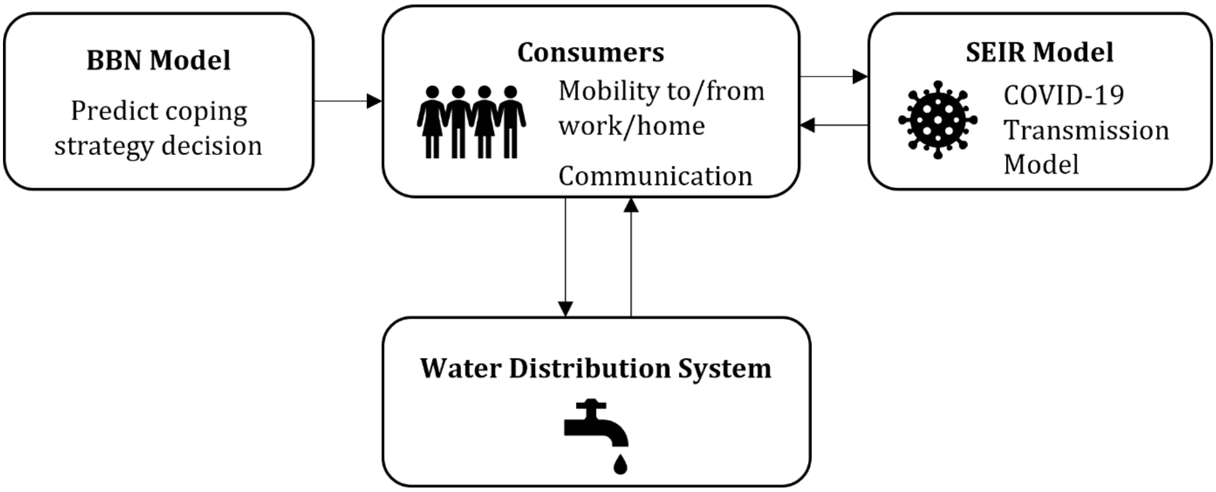

BBN Model
Predict coping strategy decision
Consumers
Mobility to/from work/home
Communication
SEIR Model
COVID-19 Transmission Model
Water Distribution System

Figure 1. Agent-based modelling framework including BBN model for agent coping strategy decision making, SEIR model for COVID-19 transmission dynamics, and a hydraulic model for water distribution simulation.

2.1 COVID-19 Risk Perception Dataset

A new dataset was collected and made publicly available to explore how individuals around the world responded to the coronavirus and perceived information about protective behaviors [12]. Responses were collected from N=6,991 participants in 11 countries (Australia, Canada, Germany, Italy, Japan, Mexico, Spain, Sweden, South Korea, United Kingdom, and United States), between mid-March and mid-April, 2020. The timing was specifically chosen to capture a subset of countries before governmental mandates and others after mandates were put into place. This also induced differences in the number of infected individuals in each country, increasing the complexity of the survey cross-section. Participants were selected as representative based on age, gender, and ethnicity with approximately 700 participants selected from each country. The dataset was used to train the BBN model, which predicts agent decisions on coping strategy based on interactions with other agents and the environment. Development of the BBN using this dataset is described below.

2.2 Agent-based Model Framework

The ABM implements consumer agents that communicate with each other, exert water demands at their current location, transmit COVID-19, and employ personal social distancing measures. The ABM was developed using Mesa, a Python package specifically designed for ABM creation and data collection. Agents are instantiated as objects using the Mesa framework and assigned parameters to describe specific attributes of interest (*Table 1*). Agents then move between home and work nodes according to predetermined patterns and spread COVID-19 through contact with other agents. Each agent's decision to work from home is updated each time they are potentially exposed. The coupling of the ABM with the hydraulic simulation is based on frameworks that were developed in previous research [13, 14].

Each simulation was run with an hourly time-step and continued for a total of 90 days. The model reports agent location, disease state, and work-from-home status and nodal demand and pressure at each hourly time-step.

Table 1. Agent parameters and state variables. All COVID-19 related time data are reported by Kerr et al. [12].

Attribute	Value
Work node	All work nodes
Home node	All residential nodes
Age	[0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+]
COVID-19 status	[Susceptible, exposed, infected, removed]
Time in exposed compartment	~ log-normal(4.5, 1.5)
Time in infectious compartment	All time spent in symptomatic, severe, critical states
Time in symptomatic state	~ log-normal(1.1, 0.9) (to severe state) ~ log-normal(8.0, 2.0) (to removed compartment)
Time in severe state	~ log-normal(1.5, 2.0) (to critical state) ~ log-normal(18.1, 6.3) (to removed compartment)
Time in critical state	~ log-normal(10.7, 4.8) (to dead state) ~ log-normal(18.1, 6.3) (to removed compartment)
Symptomatic	[Symptomatic, asymptomatic]
WFH decision	[Not WFH, WFH]
Predictors*	All predictors in *Table 2*

2.2.1 Coping Strategy Decision Model

Each agent uses a Bayesian Belief Network (BBN) model to select work-from-home (WFH) decisions, expressed as WFH and Not WFH in *Table 1*. BBN models were constructed using forward selection and the Naive Baye's classifier. Previous work showed little difference between forward selection and backward elimination, and forward selection is a more efficient model-building approach [15]. Models were evaluated and selected using accuracy and F_1. Accuracy is defined as the ratio of the number of true predictions made to the total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

The F_1 metric is the harmonic mean of the recall and precision metrics. Recall and precision are defined as the proportion of true positives to the total correct values and the total true values, respectively.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + TN} \tag{3}$$

$$F_1 = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \tag{4}$$

A series of cross-validation steps were performed to reduce systematic error in the selection of responses used for the training and validation datasets. Cross-validation was completed with 10 runs of 10 folds with each run, using nine folds for training and one fold for validation. Each run used a different fold for validation.

2.2.2 COVID-19 Transmission Model

Disease transmission of COVID-19 is modeled using a SEIR model called Covasim, which was developed by Kerr et al. [11]. At the beginning of each simulation, 5% of the population was assumed as infected, while the remaining agents were susceptible. As agents move between nodes, susceptible agents have an age-progressive (increasing with age) probability of becoming exposed when an infected agent occupies the same node. For nodes with capacities greater than 10 agents, the maximum number of agents that was potentially exposed was restricted to 10 to reflect realistic contact dynamics. Once an agent is flagged as exposed, they can become exposed and then pre-symptomatic or asymptomatic, based on an age-progressive probability. Time spent in each stage is log-normally distributed with mean and variance calculated using a range of sources, as reported by Kerr et al. [12]. If an agent is asymptomatic, they progress to the removed stage where they stay until the end of the simulation. If an agent is symptomatic, they progress through increasing stages of disease severity from mild to severe to critical, based on age-progressive probabilities. Agents in the mild and severe stage move to the removed stage after a recovery period, and agents in the critical state have an age-progressive probability of entering a death state. Agents that are in the removed stage no longer contract or transmit the disease.

2.2.3 Hydraulic Model

Hydraulic simulation is modeled using the Python package Water Network Tool for Resilience (WNTR) which utilizes EPANET, version 2.2 [16, 17]. Each agent exerts water demand at the node they occupy at each hourly time step, and demands are aggregated at each node and passed to the EPANET simulation using WNTR. The WNTR package is built in Python, which allows for direct communication between the agents in the ABM and the hydraulic simulation. Results from the hydraulic simulation for each node and for each hourly time step were recorded.

3 CASE STUDY

The ABM framework was developed and applied for Micropolis, a virtual city developed by Brumbelow et al. [13] for the purpose of modelling a small, realistic city for water distribution system security research. The network consists of 458 terminal nodes (434 residential, 15 industrial, and nine commercial nodes), which represent 4,606 residents and a daily demand of 4.54 ML/day. Diurnal demand patterns are defined for each node type to simulate hourly changes in demand throughout the network. Each node is initialized with a base demand that is updated by the ABM based on node capacities and the number of agents at each node.

4 RESULTS

4.1 BBN Model Performance

A naïve model was constructed using work-from-home (WFH) as the predictant. The predictors were added using forward selection. The model with the largest accuracy was chosen, and included the predictors shown in *Table 2*. The model is defined as a naïve Bayes model, where the predictant is the only parent node, and all selected predictors are children nodes. The percentage of 'yes' responses in the dataset used for this model was 39%, and a WFH decision of 'yes' is labeled true, and 'no' is considered false. The accuracy and F_1 of this model were 63% and 51%, respectively.

Table 2. Predictors selected for inclusion in the BBN model for agent WFH decision making.

Predictor Selected	**Question**
Prosociality	To what extent do you think it's important to do things for the benefit of others and society even if they have some costs to you personally?
Anticipating personal financial problems	How likely do you think it is that you will be directly and personally affected by the following in the next 6 months: Financial problems?
Exposure to COVID-19 media through place of work of education.	Have you come across information about coronavirus/COVID-19 from: Official messages from your place of work or education?
Exposure to COVID-19 media from the World Health Organization	Have you come across information about coronavirus/COVID-19 from: World Health Organization?
COVID-19 worry, 2 months ago	Thinking back, how worried were you about coronavirus/COVID-19: 2 months ago?
Personal worry about terrorism	How worried are you personally about the following issues at present: Terrorism?
Healthcare worker	Are you a healthcare provider (e.g. doctor, nurse, paramedic, pharmacist, carer)?
Trust in immigrants	How much do you trust each of the following: Immigrants?
Trust in neighbors	How much do you trust each of the following: People in your neighborhood?
Ethnic Minority	Do you consider yourself to be part of a minority group within the country you are currently living in?
Previously affected by SARS epidemic	Have you personally been affected by a previous similar epidemic such as SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome) or Ebola?
COVID-19 worry: 1 month ago	Thinking back, how worried were you about coronavirus/COVID-19: 1 month ago?
Effect of COVID-19 pandemic	To what extent have you been affected by the coronavirus/COVID-19 in the following ways: I have experienced social difficulties as a result of the pandemic?
Sought information about COVID-19	Have you sought out information specifically about coronavirus/COVID-19?
Education qualification	Highest educational qualification
General trust in society	Generally speaking, would you say most people can be trusted, or that you can't be too careful in dealing with people?
Exposure to COVID-19 through media mass media	Have you come across information about coronavirus/COVID-19 from: Journalists and commentators in the media (TV, radio, newspapers)?

4.2 Agent-based Modelling Results

Two scenarios were tested. In the base case, agents visit nodes using their mobility patterns and become infected through disease transmission. In the second case, WFH, agents are mobile, become infected, and decide to work from home. The SEIR model and hydraulic model were used for both scenarios, and the WFH scenario include an active BBN model. In the base scenario, agents followed mobility patterns throughout the 90-day simulation and do not work from home, and they spread COVID-19 and progress through disease severity states. In the WFH scenario, agents follow the established mobility patterns unless they had previously decided to work from home based on the BBN model, in which case, that agent would stay at their residential node and not travel to work for the remainder of the simulation.

The daily maximum and mean system water demand for both scenarios are shown in *Figure 2*.Also demonstrated in *Figure 2* is the increase in max water demand as more agents work from home but shows an overall drop in water demand across the system, corroborating previous work showing decreasing water demand as a result of social distancing [4]. The cumulative number of infected agents and the current number of agents working from home is shown in *Figure 3*. The disease dynamics are shown in *Figure 3* where early decisions to work from home prevented wide-spread transmission of COVID-19. To understand the impact of working from home on the hydraulic system, the system-wide pressure was compared for both scenarios at hour 12 during day 45 (near the peak number of agents working from home from *Figure 3*). These plots are shown in *Figure 4*, which exhibits the differences in system-wide pressure when agents are working from home and exerting demand at their residential node rather than their work node. No pressures changed alarmingly, but the main trunk leading from the reservoir at the north end of the system saw a 50% reduction in demand at this time point. This could lead to changes in water delivery stability or flow changes, causing downstream disruptions.

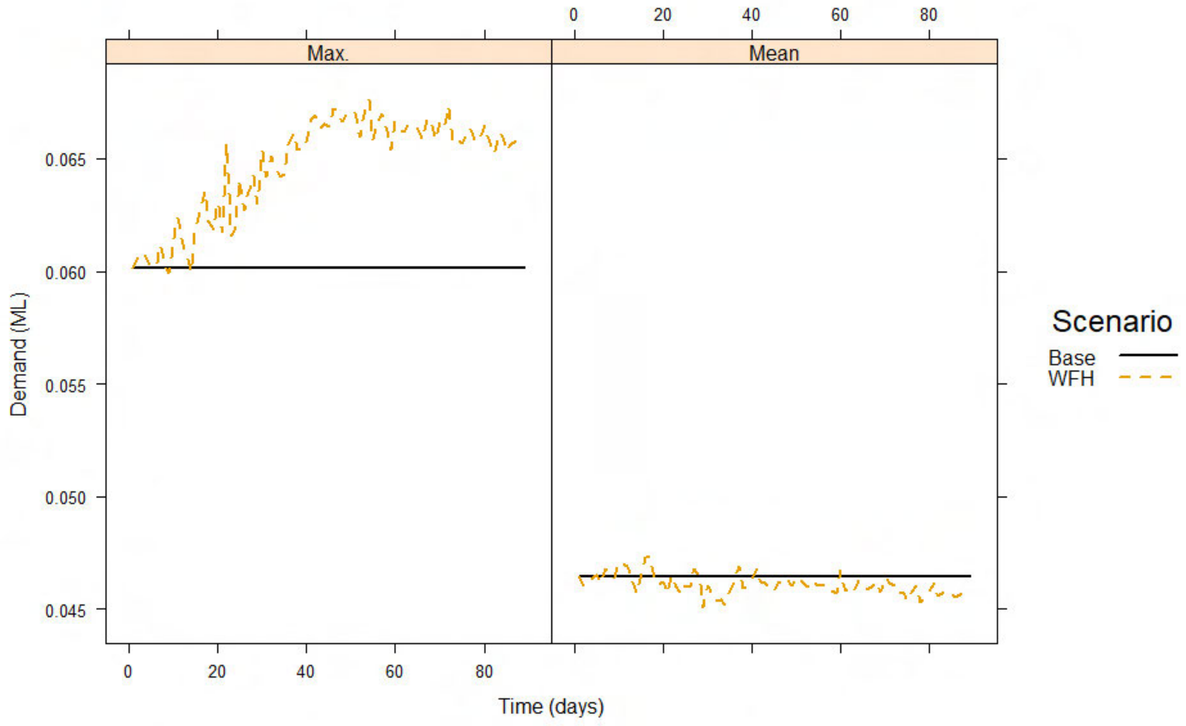


Figure 2. Daily maximum and mean system demand during both scenarios.

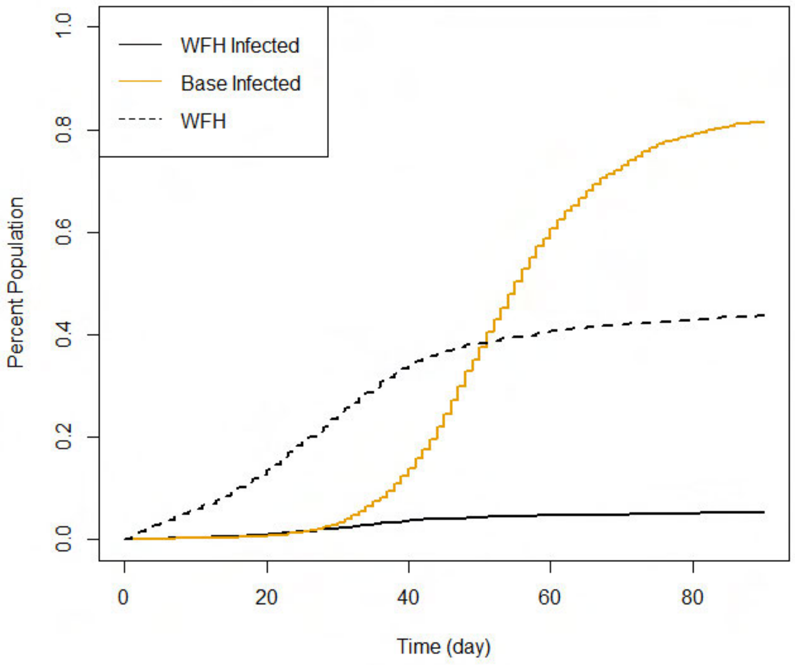
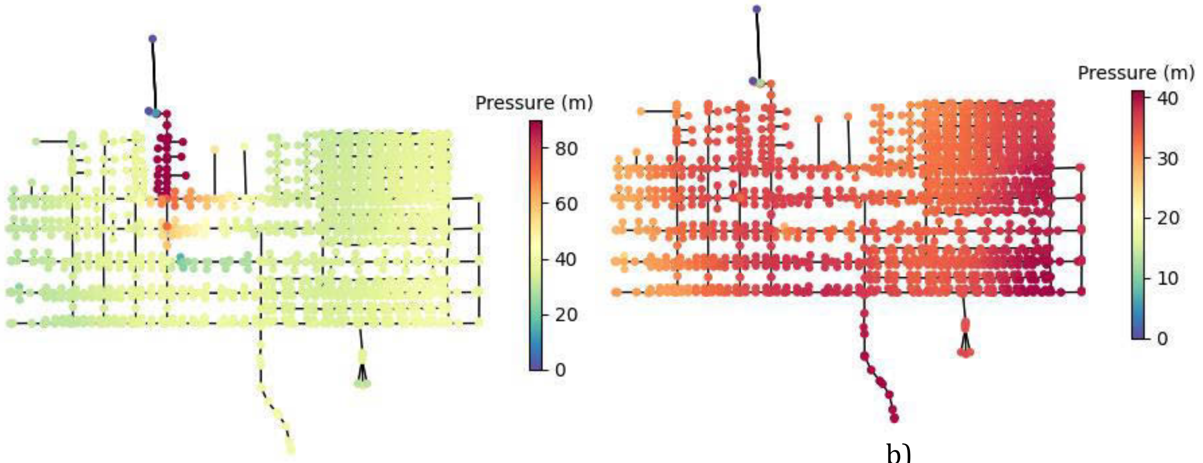


Figure 3. Cumulative percent infected for both scenarios plotted along with the percentage of agents working from home.



a)

b)

Figure 4. Node pressure comparison between base (a) and WFH (b) scenarios.

5 CONCLUSION

A Bayesian Belief Network was trained using COVID-19 centered survey data from N=6,991 participants to produce a naïve Bayes predictive model for work-from-home prediction. The accuracy and F_1 for the model were 63% and 51%. The BBN was used to simulate agent decisions to work from home within the ABM framework. Trends in overall water demand and system-wide pressure were analyzed and indicated overall changes in system dynamics due to agents working from home. Demand patterns mirrored real-world quantitative and qualitative results, and changes in nodal pressures demonstrate system-wide impacts from agent social distancing.

Future work will continue to explore other BBN models and predictors to tune the decision making model to better match observed patterns, and further analysis will evaluate the range of change in hydraulic performance due to social distancing behaviors during pandemics.

6 REFERENCES

[1] D. Koh, “COVID-19 lockdowns throughout the world,” Occup Med (Lond), p. kqaa073, May 2020, doi: 10.1093/occmed/kqaa073.

[2] T. Hale et al., “What have we learned from tracking every government policy on COVID-19 for the past two years?” University of Oxford, Mar. 2022. Accessed: Apr. 01, 2022. [Online]. Available: https://www.bsg.ox.ac.uk/sites/default/files/2022-03/What-have-we-learned-from%20tracking-two-years-of-COVID-responses-BSG-research-note-March-2022.pdf

[3] N. Ayati, P. Saiyarsarai, and S. Nikfar, “Short and long term impacts of COVID-19 on the pharmaceutical sector,” DARU J Pharm Sci, vol. 28, no. 2, pp. 799–805, Dec. 2020, doi: 10.1007/s40199-020-00358-5.

[4] J. Cahill, C. Hoolohan, R. Lawson, and A. L. Browne, “COVID-19 and water demand: A review of literature and research evidence,” WIREs Water, vol. 9, no. 1, p. e1570, 2022, doi: 10.1002/wat2.1570.

[5] E. Z. Berglund et al., “Effects of the COVID-19 Pandemic on Water Utility Operations and Vulnerability,” Journal of Water Resources Planning and Management, vol. 148, no. 6, p. 04022027, Jun. 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001560.

[6] V. K. Kandiah, E. Z. Berglund, and A. R. Binder, “Cellular Automata Modeling Framework for Urban Water Reuse Planning and Management,” Journal of Water Resources Planning and Management, vol. 142, no. 12, p. 04016054, Dec. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000696.

[7] V. K. Kandiah, E. Z. Berglund, and A. R. Binder, “An agent-based modeling approach to project adoption of water reuse and evaluate expansion plans within a sociotechnical water infrastructure system,” Sustainable Cities and Society, vol. 46, p. 101412, Apr. 2019, doi: 10.1016/j.scs.2018.12.040.

[8] E. Ramsey, J. Pesantez, M. A. K. Fasaee, M. DiCarlo, J. Monroe, and E. Z. Berglund, “A Smart Water Grid for Micro-Trading Rainwater: Hydraulic Feasibility Analysis,” Water, vol. 12, no. 11, Art. no. 11, Nov. 2020, doi: 10.3390/w12113075.

[9] M. E. Shafiee, E. Z. Berglund, and M. K. Lindell, “An Agent-based Modeling Framework for Assessing the Public Health Protection of Water Advisories,” Water Resour Manage, vol. 32, no. 6, pp. 2033–2059, Apr. 2018, doi: 10.1007/s11269-018-1916-6.

[10] M. E. Shafiee and E. Z. Berglund, “Complex Adaptive Systems Framework to Simulate the Performance of Hydrant Flushing Rules and Broadcasts during a Water Distribution System Contamination Event,” Journal of Water Resources Planning and Management, vol. 143, no. 4, p. 04017001, Apr. 2017, doi: 10.1061/(ASCE)WR.1943-5452.0000744.

[11] H. Strickling, M. F. DiCarlo, M. E. Shafiee, and E. Berglund, “Simulation of containment and wireless emergency alerts within targeted pressure zones for water contamination management,” Sustainable Cities and Society, vol. 52, p. 101820, Jan. 2020, doi: 10.1016/j.scs.2019.101820.

[12] C. C. Kerr et al., “Covasim: An agent-based model of COVID-19 dynamics and interventions,” PLoS Comput Biol, vol. 17, no. 7, p. e1009149, Jul. 2021, doi: 10.1371/journal.pcbi.1009149.

[13] M. E. Shafiee and E. M. Zechman, “An agent-based modeling framework for sociotechnical simulation of water distribution contamination events,” Journal of Hydroinformatics, vol. 15, no. 3, pp. 862–880, Jul. 2013, doi: http://dx.doi.org/10.2166/hydro.2013.158.

[14] E. M. Zechman, “Agent-Based Modeling to Simulate Contamination Events and Evaluate Threat Management Strategies in Water Distribution Systems: Agent-Based Modeling to Simulate Contamination Events,” Risk Analysis, vol. 31, no. 5, pp. 758–772, May 2011, doi: 10.1111/j.1539-6924.2010.01564.x.

[15] M. A. K. Fasaee, E. Berglund, K. J. Pieper, E. Ling, B. Benham, and M. Edwards, “Developing a framework for classifying water lead levels at private drinking water systems: A Bayesian Belief Network approach,” Water Research, vol. 189, p. 116641, Feb. 2021, doi: 10.1016/j.watres.2020.116641.

[16] K. A. Klise et al., “Water Network Tool for Resilience (WNTR) User Manual,” p. 50, Sep. 2017.

[17] L. A. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, “EPANET 2.2 User Manual,” p. 190, 2020.

WDSA CCWI 2022
2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14751

ITERATIVELY TUNING THE REGULARISATION PARAMETER IN AN INVERSE METHOD FOR LOCALISING LEAKS IN WATER DISTRIBUTION NETWORKS

Yuanyang Liu[1], Ivan Stoianov[2] and Filippo Pecci[3]

[1,2,3]InfraSense Labs, Department of Civil and Environmental Engineering, Imperial College London, London, UK

[1] *yuanyang.liu16@imperial.ac.uk*, [2] *f.pecci14@imperial.ac.uk*, [3] *ivan.stoianov@imperial.ac.uk*

Abstract

Novel methods for detecting and localising leaks in water distribution networks are being increasingly investigated as water utilities face unprecedented financial and environmental challenges in reducing water losses. A promising method includes the solution of the regularised inverse problem that minimises the difference between simulated and measured data in addition to the regularisation term. However, the results of leak localisation are sensitive to the choice of the regularisation parameter. In this paper, we propose and investigate a method for iterative tuning of the regularisation parameter to improve the leak localisation performance in case of multiple time steps measurements. The numerical investigation utilises a benchmarking network model, and different metrics are applied to evaluate the results of the leak localisation performance, such as the performance and the distance metric.

Keywords
Inverse Problem, Leakage localisation, Water Distribution Networks.

1 INTRODUCTION

Water distribution networks are used to satisfy water demand at all planned locations with an adequate pressure head, deliver the necessary volume of water during firefighting, minimise water loss and supply disruptions, and preserve water quality in the distribution pipes. However, 126 billion cubic metres of water are wasted each year globally, resulting in a loss of 39 billion US dollars [2]. Moreover, water leaks from pipes also contribute to contamination and health problems as pollutants are involved in the water and flow back to the pipe through the breaks due to negative pressure [3].

Detecting and localising leaks in the water distribution networks is challenging and usually consists of three steps. First, an anomaly event is detected in the system, and then the localisation methods are applied to reduce the research area. In the end, pinpoint techniques are used to precisely locate the leaks.

Leak detection technologies are used to signal a burst event in the water supply system without providing any specific information about the location of the burst event in the system. In the last 20–25 years, it has been increasingly common to use transient analysis approaches for leak detection because advances in metering technology have made it possible to gather near real-time data from pressure and flow sensors. For example, Liggett and Chen [4] developed an inverse transient analysis that models the pressure responses using least squares regression between measured and computed pressure responses. Leaks can be detected when there are deviations in the pressure response of the system. Numerous transient analysis-based approaches are described in the literature [5]–[8]. However, due to the significant effect of system uncertainties, these approaches are limited to a single node or grouped pipelines [9]. More recently, data-driven approaches based on the pressure sensor and flow metre readings have been developed to identify a leak as a divergence from normal observations [10]–[12]. Jung and Lansey [13] use a nonlinear Kalman filter in conjunction with a hydraulic model to address the issue that data-driven approaches cannot distinguish between real demand changes and leakage, allowing for a known change in the system's operating conditions accommodated by the model.

Leak localisation is a process of narrowing the search region for leaks to make pinpointing approaches more effective [9]. Detecting and localising leaks using pinpointing techniques is practical and exact [9]. For example, using leak noise correlators, at least one sensor is placed in contact with the pipe on both sides of the suspected leak to capture the sound data. The sound data is then analysed mathematically to determine the specific location of the leak on the pipe, which is accomplished by correlating the noises that reach both sensors and computing the time difference between each sensor's leak site. The trace gas technique includes pumping a non-toxic, lightweight combination of hydrogen and nitrogen gas into the pipe where the leak is thought to be occurring. The gas will then escape from the leak location in the pipe and be detected by the gas-sensitive sensors installed in the pipe [14]. Khulief, et al., [15] employed hydrophones to conduct sound measurements within the pipeline. The leak signal can be readily detected when the hydrophone passes over the leak location. While specific leak pinpointing methods are the most precise technology for leak localisation, they are also time-consuming, labour-intensive, and costly, particularly when a broad region has to be searched [9].

Numerous leak localisation approaches rely on hydraulic models, pressure sensor measurements and customer demand information . The leak localisation strategy proposed by Pudar and Liggett [16] is based on addressing an inverse problem by minimising residuals between simulated and measured data. Although the inverse problem may be used to identify multiple leaks that occur simultaneously, the inverse problem is often under-determined in water supply networks since the network's sensor count is far smaller than the number of probable leak locations. Pudar and Liggett [16] minimised the l_2-norm of the leak parameters to solve the under-determined problem. Sanz, et al., [17] addressed the under-determined problem by using a grouped node technique to minimise the number of unknowns. The leak location is determined as part of the demand calibration process. Demand components representing each node's state in a zone with grouped nodes are computed during the calibration process [18]. The six leakage detection indicators calculate the variance in the present and historical demand components. If the sum of scores for the grouped nodes exceeds the global threshold, this indicates the existence of a leak and its approximate location [17]. The capability to identify leaks is dependent on the relationship between the leak flow and the consumption of demand components. Small leaks that occur in zones with high consumption components are not detected because the small differences generated by leaks are not observable [17]. Most recently, Chew, et al., [19] first calibrated the net demand profile, rectified the observed offsets in the sensors and calibrated the physical parameters such as roughness coefficients and valve setting, followed by the localisation of anomaly events.

Casillas, et al., [20] and Perez, et al., [21] proposed the sensitivity matrix method that compares the difference of the pressure measurement when a leak occurs at a single location to the sensitivity of pressure measurements to a leak flow from every node. However, the Sensitivity Matrix Method is limited to detecting the existence of a single leak since it compares the measurement sensitivity of a leak at one node to the residual [21]. Moreover, although the Sensitivity Matrix Method only generates a single leak candidate, the distance between the leak candidate and the true leak node is sensitive to the uncertainty of the measurements, model accuracy, and assumed leak flow while generating the leak sensitivity matrix [20].

To overcome the constraints of sensitivity-based techniques and solve the under-determined inverse problem. Blocher, et al., [1] formulated and solved a regularised inverse problem for leak localisation by minimising the least squares of residuals between simulated and measured data. By including a regularisation term, this technique replaces the ill-posed original problem with a well-posed and stable neighbouring problem, enabling the localisation of a leakage hotspot area. However, the results of leak localisation are sensitive to the leak location because the single value of the regularisation parameter cannot guarantee good performance for every leak scenario.

Moreover, Romero-Ben, et al., [22] applied both hydraulic model-based and data-based methods to detect and localise leaks. The hydraulic model-based method relies on the accuracy of the model, demand pattern and sensor information, and the full data-based method is based on graph interpolation and candidate selection criteria.

This article extends the method developed by Blocher, et al., [1] by investigating a method for iterative tuning of the regularisation parameter to improve the leak localisation performance assuming exact model and data. The paper has 4 sections. Section 2 describes the methodology of implementing the iterative tuning approach. Section 3 tests the proposed method for the Net25 network and compares it to the original regularised inverse method. Finally, section 4 concludes the changes made by the iterative tuning approach and future works.

2 METHODOLOGY

2.1 Problem Formulation

In this article, the proposed leak localisation method aims to improve the leak localisation performance by iteratively tuning the regularisation parameter and solving the regularised inverse problem. The leaks in the network are assumed to be detected, and the effects of uncertainties in the hydraulic model, sensor measurements, and customer demand information on the leak localisation results are assumed to be neglected.

The network is modelled as a directed graph with n_n junctions (demand nodes), n_p pipes (links), and n_0 source nodes (reservoirs/tanks). The link-junction incidence matrix is $A_{12} \in \mathbb{R}^{n_p \times n_n}$, which shows the connectivity of demand nodes to the pipes, and the link-source node incidence matrix $A_{10} \in \mathbb{R}^{n_p \times n_0}$ shows the connectivity of source nodes to the pipes. The entries of the incidence matrix can be defined by the relationship between a node $i \in \{1, \dots, n_n + n_0\}$ and a pipe $j \in \{1, \dots, n_p\}$. If there is no relationship between the node i and the pipe j, the entry is 0; if the pipe j leaves node i, the entry is -1; if the pipe j enters the node i, the entry is +1.

The hydraulic model is modelled based on the conservation laws of energy and mass, and the hydraulic equations are shown below:

$$A_{12}h + A_{10}h_0 + \phi(q) = 0 \tag{1}$$

$$A_{12}^T q - d - d_L(c, h) = 0 \tag{2}$$

where $h \in \mathbb{R}^{n_n}$ is the hydraulic head at demand nodes, $q \in \mathbb{R}^{n_p}$ is the pipe flow, $h_0 \in \mathbb{R}^{n_0}$ is the hydraulic head at source nodes, $\phi(q) \in \mathbb{R}^{n_p}$ is the head loss across the pipe and $d_L(c, h) \in \mathbb{R}^{n_n}$ is the leakage at each node. In the present work, the leakage is modelled as pressure-dependent flow, and the leak flow at each node i is defined by a modified orifice equation [23]:

$$d_{L,i}(h) = c_i(h_i - z_i)^{0.5} + c_{var,i}(h_i - z_i)^{1.5} \tag{3}$$

where h_i is the hydraulic head, z_i is the elevation, c_i and $c_{var,i}$ are the unknown coefficients depending on the orifice area and the slope of pressure head to orifice area, respectively. The modified orifice equation indicated in equation (3) has two terms: one for the original flow and another for the enlarged area flow. The first term dominates when the pressure head is small or the pressure head slope to the orifice area is small, whereas the second term dominates when the pressure head is big or the pressure head slope to the orifice area is large. In this article, the $c_{var} = 0$ for all nodes.

The Darcy-Weisbach (DW) or Hazen-Williams (HW) equations are typically used to determine the head losses across the pipe that are caused by friction. However, one of the challenges in using Darcy-Weisbach or Hazen-Williams equations to model head loss is that the fractional exponent in Hazen-Williams equations or explicit approximations for Darcy-Weisbach equations may have a singularity at zero flow [24]. Therefore, a quadratic approximation approach proposed by Pecci, et al., [25] is used to calculate the head loss across the pipe j:

$$\phi(q_j) = q_j(a_j|q_j| + b_j) \tag{4}$$

where a_j and b_j are the unknown coefficients and computed in [25] Moreover, the hydraulic heads and flows are computed by solving equations (1) and (2) as outlined in [26]

The leak localisation problem in this article is similar to the one formulated by Blocher, et al., [1] as a regularised inverse problem and solved by minimising the difference between simulated and measured hydraulic states and the regularisation term. However, the regularisation parameter is not a fixed value, and it will update for each iteration.

$$\begin{aligned} \min_{h,q,c} \quad & w_h \sum_{m\in M} \left(h_m - \overline{h}_m\right)^2 + w_q \sum_{o\in O} \left(q_o - \overline{q}_o\right)^2 + \rho\|c\|_2^2 \\ \text{subject to} \quad & A_{12}h + A_{10}h_0 + \phi(q) = 0 \\ & A_{12}^T q - d - d_L(c,h) = 0 \\ & h \ge z \\ & 0 \le c \le c_{max} \end{aligned} \tag{5}$$

Where M is a set of pressure sensor locations, O is a set of flow meter locations, ρ is the regularisation parameter, W_h and W_q are the weights assigned to the pressure and pipe residuals, respectively, to guarantee that residuals are scaled equally.

Hydraulic head h, flow q, and leak coefficients c are the variables in the optimisation problem. The objective is to minimize the weighted sum of squared residuals and the weighted regularisation term. Equality constraints are based on hydraulic mass conservation and energy conservation laws. The third constraint specifies that the hydraulic head must be larger than the elevation head in order to maintain a non-negative pressure head. The fourth constraint specifies the range of leak coefficients.

The leak localisation results can be evaluated by the performance metric [1] and distance metric [20]. The former metric considers both the presence of the true leak node in the leak candidate set and the distance between the true leak node and the leak candidate set. Once the problem (5) is solved, the normalised attribute $u_i = \frac{c_i}{\max(c)}$ as an indicator of the leakage, the large value corresponds to a higher possibility of leakage. The performance metric is defined as the sum of the reward component and the penalty component:

$$\begin{aligned} \beta &= \gamma - \lambda \\ \gamma &= \sum_{i\in K} \frac{u_i}{|K|} \\ \lambda &= \frac{r^T u}{\sum_{i=1}^{n_n} r_i} \end{aligned} \tag{6}$$

where K is an index set containing the locations of true leak nodes, $|K|$ is the total number of true leak nodes, r_i is the shortest distance between node i and the closest true leak node, γ is the reward component, and λ is the penalty component. In the distance metric, the node i with the highest normalised attribute u_i is selected as the main leak candidate, and the distance between the main leak candidate and the true leak location D is computed as the measure for the leak localisation results.

2.2 Iterative tuning approach

The proposed approach iteratively solves the problem (5) with the updated regularisation parameter and search area to improve the leak localisation performance. The algorithm terminates when the relative change in the search area is smaller than a given threshold.

First, we simulate single leaks at each network node to determine the optimal regularisation parameter for each location. This is done by solving problem (5) with a set of regularisation parameters and selecting the parameter that results in the highest beta or the shortest distance, depending on the metric used. The regularisation parameter that has the highest frequency of occurrence across the entire network is chosen as the initial regularisation parameter.

When a leak is detected, we use the initial regularisation parameter to solve the regularised inverse problem (5). We consider network nodes for which the optimization procedure has assigned leak coefficients that exceed the threshold. The regularisation parameter is then updated to the value with the highest frequency of occurrence among nodes with leak coefficients above the threshold. Then, using the updated regularisation parameter, we solve a regularised inverse problem and obtain a new set of nodes with leak coefficients greater than the threshold. The iterative procedure is repeated until the relative change in the search area is smaller than 0.1.

3 CASE STUDY

Net25 is a publicly available water distribution network consisting of 22 demand nodes, 37 pipes, and 3 reservoirs. The network topology is depicted in figure (1), where nodes 23, 24, and 25 represent reservoir locations and nodes 1, 8, 13, and 21 represent sensor locations. Pipe properties are given in [27] and nodal characteristics are provided in [28]. To evaluate the Iterative Tuned Regularised Inverse Method (ITRIM), 22 single leak scenarios are simulated, and ITRIM is compared to the Regularised Inverse Method (RIM). Both methods are solved in the case of the different number of time steps measurements, including 1 time step measurements obtained at 12 pm, 6 and 12 time steps measurements obtained over 24 hours in 4 hours and 2 hours intervals, respectively. Leak flow is simulated using equation (3) with the parameters $c_i = 0.6 \times 10^{-3}\ [m^{2.5}s^{-1}]$ and $c_{var,i} = 0\ [m^{1.5}s^{-1}]$. Hydraulic equations (1) and (2) are solved to generate the hydraulic head and flow measurements. The following section implements the performance and distance metrics to compute the optimal regularisation parameter for each location and evaluate the leak localisation results.

3.1 Performance Metric

For each assumed single leak event in the Net25, a set of regularisation parameters ranging from [0, 0.1, 1, 10, 100, 1000, 10000] is implemented to solve the problem (5) with multiple time steps measurements. The optimal regularisation parameter for each location under multiple time steps measurements is the one results in the highest β. The leak localisation results are then computed by the procedure described in section 2.2. Finally, the leak localisation results from ITRIM are compared to those from RIM with the regularisation parameter=1000. Figure (2) shows the performance profiles of both ITRIM and RIM with 1 time step, 6 time steps and 12 time steps measurements, the dotted lines represent the RIM and the solid line represent the ITRIM. The performance profile illustrates the percentage of leak scenarios with performance greater than or equal to a threshold value:

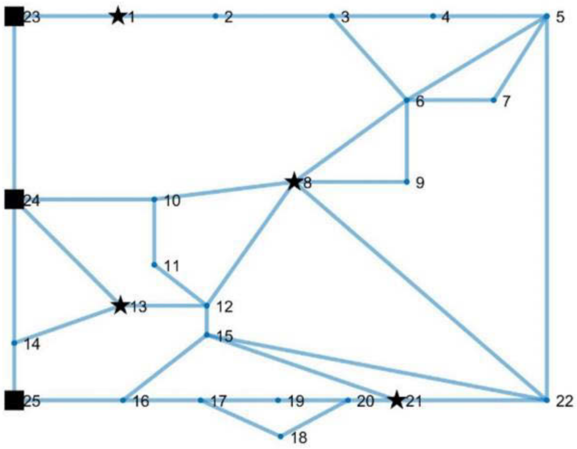


Figure 1. Layout of water distribution network: Net25.

$$P(\tau) = 100\frac{|v \in \ V: \beta_v \geq \tau|}{|V|} \tag{7}$$

where V is a set of leak scenarios, β_v is the performance for leak scenario $v \in V$ and β_v is the threshold value. In Figure (2), the performance profiles obtained by RIM with the different number of time step measurements are similar to one another. However, a comparison with the profiles obtained by ITRIM indicates clearly that the iterative tuning approach improves the leak localisation performance. For example, in the case of one time step measurements obtained at 12 pm, 68% of ITRIM leak localisation results yield a performance $\beta \geq 0.8$, and only 34% achieve the same performance if RIM is used. Additionally, if 6 or 12 time steps measurements are used, around 40% of ITRIM leak localisation results yield a performance $\beta = 1$, which means the leak localisation results correctly identify the true leak as the only main leak candidate. Moreover, the average performance for all leak scenarios increases from 0.73 to 0.89 when ITRIM is used.

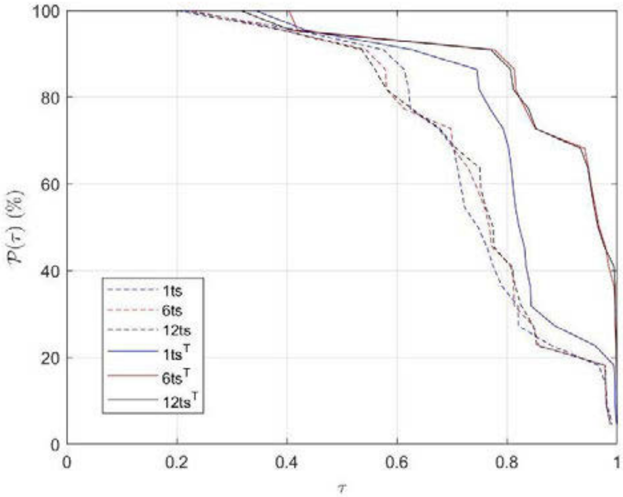
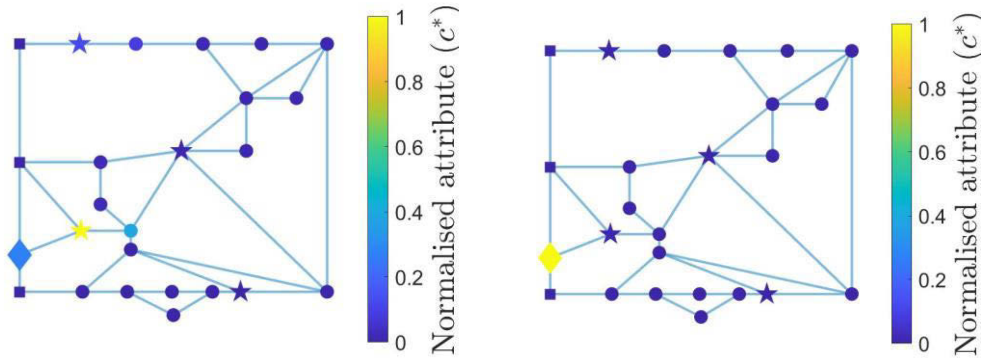


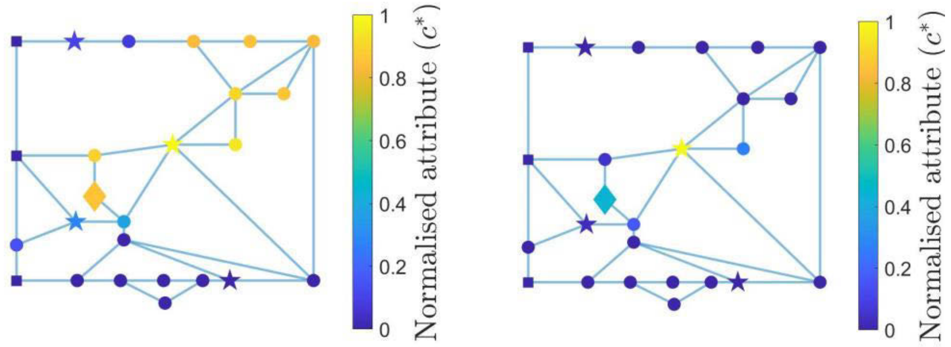
Figure 2. Performance profiles of the Regularised Inverse Method (RIM) and the Iterative Tuned Regularised Inverse Method (ITRIM) with the different number of time step measurements. Dotted lines with legend 1ts, 6ts and 12 ts represent the performance profiles of RIM, whereas solid lines with legends $1ts^T$, $6ts^T$ and $12\ ts^T$ represent the performance profiles of ITRIM.



(a) (b)

Figure 3. Leak localisation results for (a) Regularised Inverse method and (b) Iterative tuned Regularised method when a leak occurs at node 14 (diamond shape).

The largest positive performance change between RIM and ITRIM occurs when a leak is simulated at node 14. Figure (3) shows the leak localisation results of both RIM and ITRIM, and RIM assigns the largest normalised attribute to the neighbour of the true leak node and a small value of the normalised attribute to the true leak node, which results in a lower performance $\beta = 0.22$. However, ITRIM yields the true leak node as the only leak candidate, significantly improving leak localisation performance to $\beta = 1$.



(a) (b)

Figure 4. Leak localisation results for (a) Regularised Inverse method and (b) Iterative tuned Regularised method when a leak occurs at node 11 (diamond shape).

The largest negative performance change between RIM and ITRIM is achieved when a leak occurs at node 11. As shown in Figure (4), RIM assigns the high normalised attribute to a set of leak candidates, including the true leak node. In contrast, ITRIM only identifies one main leak candidate but is far from the true leak node. As a result, it leads to the reduction of performance β from 0.61 to 0.41.

ITRIM yields a small set of leak candidates and identities the true leak node as the main leak candidate in most cases. On the other hand, RIM identifies the true leak node as one of the main leak candidates, which results in a large set of leak candidates.

3.2 Distance Metric

Similar to the performance metric outlined in the previous section, the distance metric is implemented in this section to compute the optimal regularisation parameter for each location in the network that yields the shortest distance between the main leak candidate and the true leak node. The results are shown in table (1) for ITRIM and RIM with the different number of time step measurements. The distance between the main leak candidate and the true leak node for RIM does not change when multiple time steps measurements are used. However, in most leak scenarios, the distance is decreased when ITRIM is implemented. For instance, in a single-time step measurement, 10 out of 22 leak scenarios result in a shorter distance when ITRIM is utilised, and the average distance decreased from 1024 m to 535 m. In addition, if 12 time steps measurements are used, ITRIM correctly identifies the true leak node as the main leak candidate with a distance of zero in 17 leak scenarios, and the average distance for all leak scenarios is 273 m, which is reduced by more than 75% compared to RIM.

Table 1. Distance between the main leak candidate to the true leak node for both RIM and ITRIM with different number of time step measurements

Leak ID	1TS Distance [m]		6TS Distance [m]		12TS Distance [m]	
	RIM	ITRIM	RIM	ITRIM	RIM	ITRIM
1	0	0	0	0	0	0
2	1930	0	1930	0	1930	0
3	2017	0	2017	0	2017	0
4	2343	326	2343	0	2343	0
5	1858	1858	1858	0	1858	0
6	743	743	743	0	743	0
7	1358	500	1358	500	1358	500
8	0	0	0	0	0	0
9	443	443	443	300	443	0
10	249	249	249	0	249	0
11	791	791	791	791	791	791
12	762	0	762	0	762	0
13	0	0	0	0	0	0
14	1014	1014	1014	0	1014	0
15	832	0	832	0	832	0
16	1746	914	1746	0	1746	0
17	1823	1736	1823	1736	1823	0
18	1412	1412	1412	701	1412	2147
19	1575	864	1575	2407	1575	0
20	711	0	711	0	711	0
21	0	0	0	0	0	0
22	931	931	2252	2578	2252	2578
Average Distance [m]	1024	535	1084	409	1084	273

4 CONCLUSION

This article proposed the method to localise the leaks in the water distribution networks by iteratively solving the regularised inverse problem with the updated regularisation parameter. It determines the optimal regularisation parameter for each location in the network by solving the problem (5) with a set of regularisation parameters. The parameter that yields the highest beta or shortest distance is the optimal regularisation parameter at this location, depending on the choice of metric.

The proposed method has been evaluated with both the performance and the distance metric using a benchmark network Net25 and compared to the standard Regularised Inverse Method in the case of the different number of time step measurements. When the performance metric is applied with multiple time steps measurements, the average leak localisation performance rises from 0.73 to 0.89, representing a 21.3% increase. In the scenario where the distance metric and 12 time steps measurements are used, the average distance between the main leak candidate and the true leak node is reduced by 75%, going from 1084 metres to 273 metres.

In the present work, the effects of uncertainties in the hydraulic model, sensor measurements, and customer demand information on leak localisation performance are not taken into consideration. However, future work can explore the effects of uncertainties on the proposed leak localisation methods. Moreover, the network Net25 only consists of 22 demand nodes, which is relatively small compared to the real operational networks. Therefore, it also recommends evaluating the performance of the proposed leak localisation method in a larger operational network with near-real-time data.

5 REFERENCES

[1] C. Blocher, F. Pecci, and I. Stoianov, “Localizing Leakage Hotspots in Water Distribution Networks via the Regularization of an Inverse Problem,” Journal of Hydraulic Engineering, vol. 146, no. 4, p. 04020025, Apr. 2020, doi: 10.1061/(ASCE)HY.1943-7900.0001721.

[2] R. Liemberger and A. Wyatt, “Quantifying the global non-revenue water problem,” Water Supply, vol. 19, no. 3, pp. 831–837, May 2019, doi: 10.2166/ws.2018.129.

[3] Q. Xu, R. Liu, Q. Chen, and R. Li, “Review on water leakage control in distribution networks and the associated environmental benefits,” Journal of Environmental Sciences, vol. 26, no. 5, pp. 955–961, May 2014, doi: 10.1016/S1001-0742(13)60569-0.

[4] J. A. Liggett and L. Chen, “Inverse Transient Analysis in Pipe Networks,” Journal of Hydraulic Engineering, vol. 120, no. 8, pp. 934–955, Aug. 1994, doi: 10.1061/(ASCE)0733-9429(1994)120:8(934).

[5] J. Vítkovský, P. Lee, M. Stephens, M. Lambert, A. Simpson, and J. Liggett, “Leak blockage detection in pipelines via an impulse response method,” 2003, pp. 423–430.

[6] X.-J. Wang, A. Simpson, M. Lambert, J. Vítkovský, and J. Liggett, “Simulation of transients in a looped laboratory pipe network,” 2003, pp. 791–798.

[7] M. Ferrante and B. Brunone, “Pipe system diagnosis and leak detection by unsteady-state tests. 1. Harmonic analysis,” Advances in Water Resources, vol. 26, no. 1, pp. 95–105, Jan. 2003, doi: 10.1016/S0309-1708(02)00101-X.

[8] D. Covas and H. Ramos, Hydraulic transients used for leak detection in water distribution systems. 2001.

[9] R. Puust, Z. Kapelan, D. A. Savic, and T. Koppel, “A review of methods for leakage management in pipe networks,” Urban Water Journal, vol. 7, no. 1, pp. 25–45, Feb. 2010, doi: 10.1080/15730621003610878.

[10] G. Ye and R. A. Fenner, “Kalman Filtering of Hydraulic Measurements for Burst Detection in Water Distribution Systems,” Journal of Pipeline Systems Engineering and Practice, vol. 2, no. 1, pp. 14–22, Feb. 2011, doi: 10.1061/(ASCE)PS.1949-1204.0000070.

[11] S. R. Mounce, J. B. Boxall, and J. Machell, “Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows,” Journal of Water Resources

Planning and Management, vol. 136, no. 3, pp. 309–318, May 2010, doi: 10.1061/(ASCE)WR.1943-5452.0000030.

[12] M. Romano, Z. Kapelan, and D. A. Savić, "Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems," Journal of Water Resources Planning and Management, vol. 140, no. 4, pp. 457–467, Apr. 2014, doi: 10.1061/(ASCE)WR.1943-5452.0000339.

[13] D. Jung and K. Lansey, "Water Distribution System Burst Detection Using a Nonlinear Kalman Filter," Journal of Water Resources Planning and Management, vol. 141, no. 5, p. 04014070, May 2015, doi: 10.1061/(ASCE)WR.1943-5452.0000464.

[14] M. A. M. Haniffa and F. M. Hashim, "Recent developments in in-line inspection tools (ILI) for deepwater pipeline applications," in 2011 National Postgraduate Conference, Sep. 2011, pp. 1–6. doi: 10.1109/NatPC.2011.6136416.

[15] Y. A. Khulief, A. Khalifa, R. ben Mansour, and M. A. Habib, "Acoustic Detection of Leaks in Water Pipelines Using Measurements inside Pipe," Journal of Pipeline Systems Engineering and Practice, vol. 3, no. 2, pp. 47–54, May 2012, doi: 10.1061/(ASCE)PS.1949-1204.0000089.

[16] R. S. , and J. A. Liggett. Pudar, "Leaks in pipe networks.," Journal of Hydraulic Engineering, vol. 118, no. 7, pp. 1031–1046, 1992.

[17] G. Sanz, R. Pérez, Z. Kapelan, and D. Savic, "Leak Detection and Localization through Demand Components Calibration," Journal of Water Resources Planning and Management, vol. 142, no. 2, p. 04015057, Feb. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000592.

[18] G. Sanz and R. Pérez, "Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition," Journal of Water Resources Planning and Management, vol. 141, no. 10, p. 04015020, Oct. 2015, doi: 10.1061/(ASCE)WR.1943-5452.0000535.

[19] A. Chew, Z. Wu, T. Walski, X. Meng, J. Cai, J. Pok, and R. Kalfarisi, "Daily Model Calibration with Water Loss Estimation and Localization Using Continuous Monitoring Data in Water Distribution Networks," Journal of Water Resources Planning and Management, vol. 148, no. 5, May 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001546.

[20] M. v. Casillas, L. E. Garza-Castanon, and V. Puig, "Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks: Application to the Barcelona network," in 2013 European Control Conference (ECC), Jul. 2013, pp. 401–409. doi: 10.23919/ECC.2013.6669568.

[21] R. Perez, G. Sanz, V. Puig, J. Quevedo, M. Cuguero-Escofet, F. Nejjari, J. Meseguer, G. Gembrano, J. Mirats-Tur and R. Sarrate, "Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in barcelona [applications of control]," IEEE Control Systems, vol. 34, no. 4, pp. 24–36, 2014, doi: 10.1109/MCS.2014.2320336.

[22] L. Romero-Ben, D. Alves, J. Blesa, G. Cembrano, V. Puig, and E. Duviella, "Leak Localization in Water Distribution Networks Using Data-Driven and Model-Based Approaches," Journal of Water Resources Planning and Management, vol. 148, no. 5, May 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001542.

[23] J. E. van Zyl and C. R. I. Clayton, "The effect of pressure on leakage in water distribution systems," Proceedings of the Institution of Civil Engineers - Water Management, vol. 160, no. 2, pp. 109–114, Jun. 2007, doi: 10.1680/wama.2007.160.2.109.

[24] B. J. Eck and M. Mevissen, "Quadratic approximations for pipe friction," Journal of Hydroinformatics, vol. 17, no. 3, pp. 462–472, May 2015, doi: 10.2166/hydro.2014.170.

[25] F. Pecci, E. Abraham, and I. Stoianov, "Quadratic head loss approximations for optimisation problems in water supply networks," Journal of Hydroinformatics, vol. 19, no. 4, pp. 493–506, Jul. 2017, doi: 10.2166/hydro.2017.080.

[26] S. Elhay, O. Piller, J. Deuerlein, and A. R. Simpson, "A Robust, Rapidly Convergent Method That Solves the Water Distribution Equations for Pressure-Dependent Models," Journal of Water Resources Planning and Management, vol. 142, no. 2, p. 04015047, Feb. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000578.

[27] K. Vairavamoorthy and J. Lumbers, "Leakage Reduction in Water Distribution Systems: Optimal Valve Control," Journal of Hydraulic Engineering, vol. 124, no. 11, pp. 1146–1154, Nov. 1998, doi: 10.1061/(ASCE)0733-9429(1998)124:11(1146).

[28] P. D. Dai and P. Li, "Optimal Localization of Pressure Reducing Valves in Water Distribution Systems by a Reformulation Approach," Water Resources Management, vol. 28, no. 10, pp. 3057–3074, Aug. 2014, doi: 10.1007/s11269-014-0655-6.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14752

DETERMINATION OF THE COST OF EXTRACTION AND SALE PRICE OF WATER FOR DRINKING USE: A CASE STUDY

Sánchez Astello M. M.[1] and Bravo Anaya L.A.[2]

[1]Research Professor of the Department of Irrigation of the Universidad Autónoma Chapingo located in Carretera México Texcoco km. 38.5, Chapingo, Mexico. C.P. 56230. Tel. (52) 595 9521500 ext. 5698.
[2]Irrigation Engineer

[1] mastello83@hotmail.com, [2]msancheza@chapingo.mx.

Abstract

In Mexico, the Political Constitution of the United Mexican States, in relation to the human right for water, establishes the fulfillment of five essential characteristics: sufficient, healthy, acceptable, accessible and affordable. Article 4 recognizes that every person has the right to access, disposal and sanitation of water for personal and domestic consumption. The State will guarantee this right and the law will define the bases, supports and modalities for access and equitable and sustainable use of water resources, establishing the participation of the Federation, the states and the municipalities. And, on the other hand, article 115 indicates the functions of the municipality, among them "they will be in charge of the functions and public services of drinking water, drainage, sewage, treatment and disposal of their wastewater" (CPEUM, 1917). The problem in the hydraulic sector in Mexico, and specifically in drinking water from deep wells, is, on the one hand, the lag in the provision of this service, and, on the other, the high costs of extracting and conveying the resource. The objective of this research is to calculate the cost of extracting water for drinking use from a deep well in the town of San Diego, municipality of Texcoco, State of Mexico; and estimate a sale price through the unit price procedure described in the Regulations of the Law on Public Works and Related Services (RLOPSRM), in such a way that the Operating Agency of this well has accurate information on the different costs they incur and get a price that can cover these, plus make a profit. This town has 5,676 inhabitants and 1,381 inhabited dwellings, of these 1,242 have piped water and there are 2,356 users registered as beneficiaries of drinking water, with a well that costs 30 lps. To estimate the sale price, direct costs, indirect costs, financing costs and utility costs were calculated, this price is $0.37 euros/m3, direct costs equal 44%, indirect costs 14% and the financing cost represents 34.2%, because the people who do not pay for the service exceed 50% of the users. The sale price obtained per cubic meter is 228% more than the total cost of extraction, mainly influenced by the cost of financing. The current rate established by the Operating Body is 0.07 euros/m3, 18% of the rate obtained from what should be charged considering all the aspects involved in providing the service. The rate charged is equivalent to 41% of the extraction cost, that if, it does not cover what is spent on water extraction. Even when the Operating Agencies are autonomous for the collection of rates, they must comply with the guidelines set by each State, this work aims to disseminate a complete way of paying for the drinking water service and that the sale price obtained from the service covers all the costs of this.

Keywords
Extraction costs, Sale price water.

1 INTRODUCTION

One of the great water challenges that we face at a global level is to provide drinking water, sewage and sanitation services to the population, due, on one hand, to accelerated demographic growth and, on the other, to the increasing technical difficulties that this entails (CONAGUA 2015). In Mexico, CONAGUA (2018) reports that the coverage of piped water services (95.3%) includes the population that has piped water inside the home or land, from a public faucet or hydrant, or in another home.

One of the most important sources to provide the population with water for urban use is groundwater; the water extracted from the aquifer is mainly used for public-urban use, with a total of 376.92 million m3 per year (CONAGUA, 2002). The problem in the hydraulic sector and specifically in drinking water is, on one hand, the lag in the provision of services that mainly affect rural areas and, on the other hand, the high costs of extraction and conduction.

The objectives of this research are to calculate the cost of extracting water for drinking use from a deep well in the town of San Diego in Texcoco, Mexico; and estimate a sale price for the drinking water service, through the unit price procedure described in the Regulations of the Law on Public Works and Related Services (RLOPSRM, 2010), in such a way that the Operating Agency of this well has with accurate information on the different costs they incur and know a price that can cover them and also allows them to have a profit to capitalize the Agency.

2 GENERAL OBJECTIVE

Determine the cost of extraction and estimated sale price of groundwater for human use and consumption.

2.1 Particular objectives

a. Determine the fixed and variable costs in establishing a deep well.

b. Estimate a sale price of groundwater for human use and consumption, through the unit price analysis procedure established in the Regulations of the Law on Public Works and Related Services (RLOPSRM, 2010).

3 MATERIALS AND METHODS

The study area is the town of San Diego, a population belonging to the municipality of Texcoco, in the State of Mexico, located at latitude 19.4667 and longitude -99.6°, see figure 1.

This population has 5,676 inhabitants and 1,381 inhabited houses, 1,242 have piped water and 1,280 with drainage (INEGI, 2015).

The well has a depth of 250 m, a drilling diameter of 18 inches and a casing diameter of 12 inches, providing an output of 30 liters per second. It is equipped with a submersible pump and its civil works consist of sidewalks, flattening of the perimeter fence and the induction of the line to the main network. The electrical substation is carried out in accordance with the regulations of the Federal Electricity Commission.

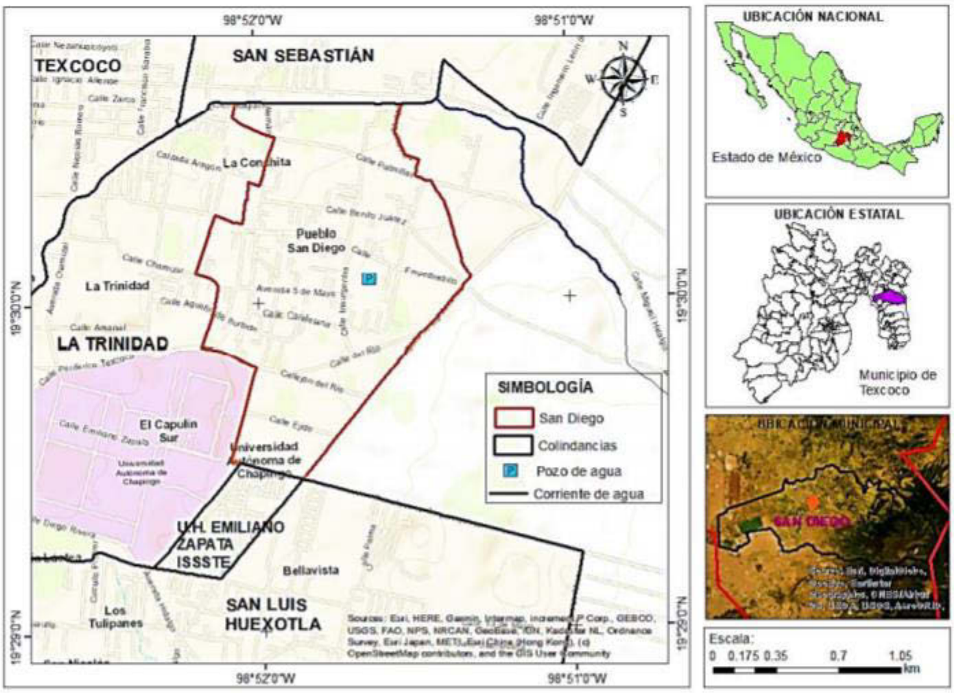

98°52'0"W
98°51'0"W
19°30'0"N
19°29'0"N
TEXCOCO
SAN SEBASTIÁN
N
W
E
S
La Conchita
Pueblo San Diego
La Trinidad
LA TRINIDAD
El Capulin Sur
Universidad Autónoma de Chapingo
U.H. EMILIANO ZAPATA ISSSTE
Bellavista
SAN LUIS HUEXOTLA
Los Tulipanes
SIMBOLOGÍA
San Diego
Colindancias
Pozo de agua
Corriente de agua
UBICACIÓN NACIONAL
Estado de México
UBICACIÓN ESTATAL
Municipio de Texcoco
UBICACIÓN MUNICIPAL
SAN DIEGO
Escala:
0 0.175 0.35 0.7 1.05
km

Figure 1 Localization of the municipality of San Diego

4 METHODOLOGY

4.1. Extraction cost

To calculate the cost of extracting groundwater from the deep well, the hourly cost is calculated and divided by the extracted volume, to obtain the cost per cubic meter:

$$CAS = \frac{CT}{Vex} \tag{1}$$

Where:

CAS = Volumetric cost of groundwater extraction, $/m3.

CT = Total cost of extraction, $.

Vex = Exploitation volume, m3.

The total cost is equal to the fixed costs plus the variable costs, and to calculate these, the methodology set forth in the Law of Public Works and Related Services and its Regulations is taken; where fixed costs are equal to depreciation, investment, insurance, and maintenance costs. And the variable costs are equal to the consumptions.

4.2. Fixed costs

Depreciation costs $$\mathrm{D} = \frac{\mathrm{Vm\text{-}Vr}}{\mathrm{Ve}} \tag{2}$$

Investment costs $$\mathrm{Im} = \frac{(\mathrm{Vm} + \mathrm{Vr})\mathrm{i}}{2\mathrm{Hea}} \tag{3}$$

Insurance costs $$Sm = \frac{(Vm+Vr)\, s}{2Hea} \tag{4}$$

where:

D: Hourly cost for depreciation of machinery or construction equipment.

Vm : Value of the machine or equipment considered as new on the date of presentation, discounting the price of tires and equipment, accessories or special parts, if any.

Vr: Salvage value of the machine or equipment that the contractor considers recovering for its sale, at the end of its economic life.

Ve: Economic life of the machine or equipment estimated by the contractor and expressed in effective working hours.

Im: Hourly cost for the investment of machinery or construction equipment, considered as new.

Hea: Number of effective hours that the machine or equipment works during the year

i: Represents the annual interest rate expressed as a decimal fraction.

Sm: Hourly cost for construction machinery or equipment insurance.

s: Average annual insurance premium, set as a percentage of the value of the machine or equipment and expressed as a decimal fraction.

Maintenance costs

$$Mn = Ko * D \tag{5}$$

where:

Mn: Represents the hourly cost for major and minor maintenance of machinery or construction equipment.

Ko: Represents a coefficient that considers both major and minor maintenance. it varies according to the type of machine or equipment and the characteristics of the work and is set based on statistical experience.

D: Represents the depreciation of the machine or equipment.

4.3. Variable costs

These are the consumption for electrical energy, labour, maintenance and oil consumption.

a) Cost for electricity consumption. To calculate the cost for electricity consumption, rate 6 called "Public service potable and black water pumping service" of the Federal Electricity Commission (CFE) was considered.

b) Cost per operation for labour. In this concept, an operator was contemplated who works 6 days a week operating the equipment, the salary that is paid to the operator oversees the Operating Agency (O.O.).

c) Equipment maintenance cost. They are all the activities carried out such as: maintenance of the electrical substation, blade manoeuvres, change of fuses and panel maintenance, all of them carried out every 6 months.

d) Cost for oil consumption. This cost is due to the consumption of oil necessary to keep the shaft that conducts the mechanical energy lubricated, to later be transformed by the pump impellers into hydraulic energy. Oil consumption was taken constant, equation (6) is normally used to determine the feasibility of projects involving these consumptions (FIRA, 1985).

$$Ca=0.0004 * HP \tag{6}$$

Where:

Ca = Hourly oil consumption (l/h).

HP = Rated power of the motor that is part of the pumping equipment, HP

Since the previous calculation is an hourly cost, it is necessary to multiply it by the annual operating time:

$$CA=(Ca*TO) * Pa \tag{7}$$

Where:

CA = Annual cost of oil.

Ca = Hourly oil consumption (l/h).

TO = Annual operating time, hour.

Pa = Oil price, \$/l

4.4. Operating volume

To calculate the exploitation volume (Ve) equation (10) will be used:

$$Ve = \frac{Q*3600}{1000}TO \tag{8}$$

Where:

Ve = Exploitation volume, m3 /year.

Q = Extraction rate, l/s.

3600 = Conversion factor from seconds to hours.

1000 = Conversion factor from litters to m3.

TO = Annual operating time, hours.

4.5. Groundwater sales price estimate

The total cost obtained for the concept of groundwater extraction is equivalent to the direct cost, and to this are added the indirect costs, for financing and utility:

$$PV = \frac{(CD + CI + CF + CU)}{Vex} \tag{9}$$

Where:

PV = Selling price, \$.

CD = Direct cost, \$.

CI = Indirect cost, \$.

CF = Cost of financing, \$.

CU = Cost per utility, \$.

Vex = Exploitation volume, m3

a) Calculation of indirect costs. These will be made up of the administration costs involved in the operation of the well, as well as the expenses of the offices of the O.O. and it is contemplated to add a cost for education and water culture, it is a proposal for workshops within the community for the awareness of the rational use of water and care for the environment at different

BY NC SA

educational levels within the community. Indirect costs are expressed as a percentage of the direct cost, (RLOPSRM, 2010).

b) Calculation of the financing cost. In the internal regulations of the A. C. it is contemplated "…the delay in the payment of the consumption will apply the following sanctions: the delay for more than 5 days in the payment of the consumption of drinking water will cause an interest of 10% for a single occasion. In the case of debt collection by judicial means, an annual interest of 20% will be applied on the debt"…., so this is taken as a financing cost. This cost must be represented by a percentage of the sum of the direct and indirect costs (RLOPSRM, 2010).

c) Calculation of cost per utility. The utility corresponds to an interest rate equivalent to an economic indicator that represents a profit without risk for the saver, in this case the interest rate of the Treasury Certificates of the CETES Federation consulted in February 2019 is taken.

5 RESULTS

5.1. Calculation of direct costs.

5.1.1. Fixed costs.

The cost of drilling and equipping the well is US$ 171,992.88, see Table 1.

Table 1. Cost of drilling the well and its equipment

Concept	Amount US$ (2019)	
Well drilling	$	81,061.69
Equipment Cost	$	41,066.70
Civil work	$	35,400.21
Electric connection	$	14,464.28
Total	$	171,992.88

The costs for depreciation and investment were calculated with equations (2) and (3), respectively, the effective hours are 12 hours for 6 days a week for 52 weeks, yielding an annual total of 3,744 hours. For the investment cost, the interest rate of 14.14% per year was taken, the annual depreciation cost is US$ 9,989.45 and the investment cost per hour is US$3.29, see table 2.

The cost of insurance is not considered. The maintenance cost was set at 4% of the value of the drilling, so:

Maintenance cost= US$81,061.69 * 0.04 = 3,249 US$/year

Table 2. Direct cost of depreciation and inversion of the well drilling and its equipment

Concept	Machine value	Salvage value (Vr)		Economic life	Depreciation costs	Investment cost
	Vm	% Vm	US$	Ve	US$/year	US$/hour
Well drilling	$81,061.69	0		25	$3,242.47	$ 1.53
Equipment Cost	$41,066.70	4%	$1,642.67	10	$3,942.40	$ 0.81
Civil work	$35,400.21	0		25	$1,416.01	$ 0.67
Electric connection	$14,464.28	4%	$578.57	10	$1,388.57	$ 0.28
				Total	$9,989.45	$ 3.29

The annual fixed cost (CF) is equal to depreciation costs + investment costs + maintenance costs:

$$CF = 9{,}989.38+12{,}315.25+3{,}242.44 = 25{,}547.10 \text{ US\$/year}$$

BY NC SA

5.1.2. Variable Cost Calculation

The cost of energy was obtained directly from the information obtained from the O.O. for 2019, according to the consumption per month and the rate 6 of the CFE, there is a cost per year of $958,243.45, see table 3.

Table 3. Cost of electricity for the year 2019

Mes	Total consumption	Price per kW	Cost	Cost with tax
	kW	US$	US$/kW	US$/kW
January	35897	0.0849	$ 3,046.35	$ 3,533.76
February	38962	0.0744	$ 2,900.40	$ 3,364.46
March	33469	0.0774	$ 2,591.15	$ 3,005.73
April	39730	0.0824	$ 3,273.04	$ 3,796.73
May	32715	0.0903	$ 2,954.90	$ 3,427.69
June	35319	0.0968	$ 3,417.97	$ 3,964.84
July	34631	0.1077	$ 3,729.49	$ 4,326.21
August	27902	0.1186	$ 3,309.47	$ 3,838.98
September	25250	0.1330	$ 3,358.31	$ 3,895.64
October	24299	0.1469	$ 3,569.48	$ 4,140.60
November	30217	0.1444	$ 4,363.84	$ 5,062.06
December	33295	0.1345	$ 4,477.89	$ 5,194.35
			Total	US$ 47,551.06

For the cost of labor for the operation of the pumping equipment, an operator whose salary is US$7.44 per day, works 6 days a week while the well works:

Cost per trade = US$ 2,282.92/year

The maintenance cost is made every 6 months with an average cost of US$ 744.41.

Maintenance cost = US$ 744.41* 2 = 1,488.83 US$/year

For the cost of oil consumption, equations (6) and (7) are used:

CA hour = 0.05 l/hour*2.71 US$/l = 0.0135 US$/hour

CA year= 0.0135 US$/hour * 3,744 hour/year = 501.67 US$/year

The variable cost (CV) is equal to the cost of electrical energy + cost of maintenance + cost of oil consumption:

$$CV = 47{,}551.06 + 1{,}488.83 + 501.67 = 49{,}541.56 \text{ US\$/year}$$

5.1.3. Total direct cost

The total direct cost is equal to the sum of fixed costs (CF) plus the sum of variable costs (CV):

$$CT = 25{,}547.10 \text{ US\$/year} + 49{,}541.56 \text{ US\$/year} = 75{,}088.66 \text{ US\$/year}$$

5.2. Extraction volume and groundwater extraction cost to direct cost

To calculate the extraction volume (Ve) in m^3, equation (8) is used, and the annual volume of exploitation is 404,352 m^3, the cost of groundwater extraction (CAS) in the well is:

$$Vex = \frac{30 * 3600}{1000} 3744 = 404{,}352.00\, m^3$$

For the cost of groundwater extraction, equation (1) is used

$$CAS = \frac{\text{US\$ } 75{,}088.66}{404{,}352.00\, m^3} = 0.1857\ \frac{US\$}{m^3}$$

5.3. Calculation of indirect costs

There are two indirect costs, those of the administration of the offices and the costs of education and water culture. Of the first, the O.O. was investigated, and the costs were classified according to the provisions of the RLOPySRM (2010), so this O.O spends US$ 24,653.43 per year on these costs. It is proposed to hold workshops within the community to raise awareness of the rational use of water and care for the environment. For the workshops, it is proposed that 12 be held at a cost of US$ 74.44 each one, yielding a total of US$ 893 per year.

Table 4. Administration costs of the San Diego well operator, Texcoco.

Concept		Amount US$
I.	**Fees, salaries and benefits of the following concepts:**	
a	Technical staff	$ 6,369.23
b.	Tickets and per diem for staff	$ 335.78
II.	**Depreciation, maintenance and income of the following concepts:**	
a	Equipment, furniture and fixtures	$ 377.17
b.	Depreciation or rent, and operation of vehicles	$ 383.74
III.	**Services of the following concepts:**	
a.	Consultants, advisors, services and laboratories	$ 9,487.39
IV.	**Office expenses of the following concepts:**	$ -
a.	Stationery and office supplies	$ 761.95
b.	Computer equipment	$ 42.08
c	Copies and duplicates	$ 875.04
d	Electricity, gas and other consumptions	$ 62.78
V.	**Previous and auxiliary works of the following concepts:**	
b.	Assembly and dismantling of equipment	$ 5,958.26
	TOTAL:	$ 24,653.43

The indirect costs are equal to:

CI = US$ 893 + US$ 24,653.43 = 25,546 US$/year

5.4. Financing cost calculation

The O.O. reported an income of US$ 27,903.40 for the concept of payment for water extraction during the second half of 2018 and January 2019, equivalent to a monthly income of US$ 2,325.28; the monthly rate for drinking water service is US$ 3.47, 1,148 users paid for the service during this period. Of the 2,356 users registered in the user registry, 1,208 did not pay for the service;

Taking what is stipulated in the internal regulations of the A. C., 20% is charged on the total debt, and given that the debt amounts to an annual amount of US$ 41.48, it is obtained that: US$41.48 *1.20 = US$49.77/year:

Financing cost = US$ 49.77*1,208 users (delinquent) = US$ 60,129.41/year.

5.5. Calculation of cost per utility

For the utility, the interest rate of CETES (government bonds) consulted in February 2019 was considered, estimating 8.20% and it is calculated on the CD + CI + CF:

Cost per utility = (75,088.66 + 25,546 + 60,129.41) * 8.20% = 13,182.65 US$/year

5.6. Estimated sale price

The sale price (PV, equation 9) is equal to the sum of the direct cost (of extraction), indirect cost, financing cost and utility cost per cubic meter extracted:

$$\mathrm{PV} = \frac{75{,}088.66 + 25{,}546 + 60{,}129.41 + 13{,}182.65}{404{,}352.00\ m^3} = \frac{173{,}946.72}{404{,}352.00\ m^3} = 0.43\ US\$/m^3$$

6 CONCLUSIONS

The direct cost per cubic meter extracted from the deep well is 0.1857 $\frac{US\$}{m^3}$. In the fixed costs of the cost of drinking water extraction, the investment cost represents 48.2%, in the variable costs, electricity represents 91.7% of this cost. Of the total cost, the variable cost represents 67%. In the sale price, direct costs are equivalent to 44%, indirect costs 14% and financing cost represents 34.2%. The sale price per cubic meter is 0.43 $US\$/m^3$, 228% more than the extraction cost, mainly influenced by the cost of financing. The current cost established by the O.O. is US$0.078 per cubic meter and what should be charged considering all the aspects involved in providing the service is US$ 0.43 per cubic meter, which represents 5.54 times more than the current rate. This cost is equivalent to 41% of the extraction cost, that is, it does not cover what is spent on water extraction.

7 REFERENCES

[1] Constitución Política de los Estados Unidos Mexicanos [México], 5 Febrero 1917, available at this address: https://www.diputados.gob.mx/LeyesBiblio/index.htm.

[2] CONAGUA, C. (2015). Manual de Agua Potable, Alcantarillado y Saneamiento (MAPAS). Conducciones. México: México.

[3] CONAGUA. (2018). Estadísticas del Agua en México. México: CONAGUA.

[4] CONAGUA. (2002). Determinación de la disponibilidad de agua en el acuífero Texcoco, Estado de México. México: CONAGUA.

[5] Reglamento de la Ley de Obras Públicas y Servicios Relacionados con la Misma, 28 de julio de 2010, available at this address: https://www.diputados.gob.mx/LeyesBiblio/index.htm.

[6] Idem (5)

[7] INEGI (2015). Prontuario de información geográfica municipal de los Estados Unidos Mexicanos. available at this address: http://www3.inegi.org.mx/contenidos/app/mexicocifras/datos_geograficos/15/15099.pdf

[8] FIRA (1985). Consumo de aceites en motores. Boletín de divulgación técnica. México. D.F.

[9] Idem (5)

[10] Idem (5)

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14753

WHICH FLOW AND PRESSURE CONSTRAINTS ARE REQUIRED FOR SUSTAINABLE OPERATION OF WATER DISTRIBUTION SYSTEMS?

Olivier Piller[1], Jochen W. Deuerlein[2], Sylvan Elhay[3] and Angus R. Simpson[4]

[1]INRAE/UR ETTIS, Bordeaux regional centre, Cestas F-33612, France
[2]3S Consult GmbH, Albtalstrasse 13, 76137 Karlsruhe, Germany
[3]School of Computer Science, University of Adelaide, Adelaide SA 5005, Australia
[4]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide SA 5005, Australia

[1] *olivier.piller@inrae.fr*, [2]*deuerlein@3sconsult.de*, [3] *sylvan.elhay@adelaide.edu.au*,
[4] *angus.simpson@adelaide.edu.au*

Abstract

Water quantity and quality modelling tools are often used to understand the working and operate properly water distribution systems. In this paper, we discuss how to choose flow and pressure constraints at nodes and links in the distribution network graph for sustainable operation of these systems. Using practical and concrete examples, we show how the problem of choosing the appropriate flow and pressure constraints amounts to verifying that a certain algebraic condition of maximum rank holds (a constraint qualification condition), or equivalently that there is a corresponding spanning tree with unsaturated links and demand nodes. Some situations of non-convergence in the solution path are discussed.

Keywords
Sustainable operations, Convex programming, Interior point methods, Active set methods, Flow and pressure constraints, Primal and dual constraints, Pressure-driven analysis, Design, Water distribution systems.

1 BACKGROUND

Water Distribution Systems (WDSs) are complex and aging infrastructures that need to be operated properly, protected, and made more resilient to natural disasters. There are considerable preservation, health & safety, and sustainability issues at stake in being able to properly manage and understand the functioning of water networks. Sustainable management of the distribution system may involve redesigning the system to better control its hydraulic state (i.e., pressure and velocity) [1,2].

Recently, several authors [3,4] have proposed ad hoc formulation and active set or interior point methods to deal with linkflow constraints for Demand Driven Modelling (DDM) and Pressure Driven Modelling (PDM) steady-state problems. To guarantee a non-empty feasible solution set and the existence of a solution, they choose to constrain the flows, $\boldsymbol{q}$, only on a cotree of the network graph and zero flow must belong to the constraint interval. In this way, there is always an unconstrained path to supply the demand nodes, and zero flow in a disconnected component is possible. Nevertheless, it may be advantageous to consider certain flow and pressure constraints also in the spanning tree, while imposing zero flow in the cotree (as in the self-cleaning network concept), [1]).

For PDM modelling, it is classical to extend the graph of the network, and produce a virtual control equivalent network, for example, by adding additional reservoirs, Flow Control Valves (FCVs), check valves and a throttle-control valves connected to each demand node as in [5]. The flow rates on these additional links correspond to the outflows at junction nodes.

In the very general case, the spanning tree of interest and the corresponding cotree are not known beforehand or given. Theoretically, there could be flow and pressure constraints on any link, from which only a subset belonging to a cotree (outflow, **c**, and linkflow constraints) and spanning tree (pressure constraints) can be active at the same time. The final decomposition/partitioning in a spanning tree and cotree is no longer a property of the network graph. It is rather driven by the actual consumption and boundary conditions.

2 GRAPH REPRESENTATION OF PRESSURE AND FLOW CONSTRAINTS

As in [3], flow rate controls, check valves and pumps are modelled as min-max box constraints for links, $q_j^{min} \leq q_j \leq q_j^{max}$: 1) Choosing an upper limit, which is smaller than the link's unconstrained flow, models the action of a flow control valve; 2) Choosing a lower limit, which is larger than the link's unconstrained flow, models the action of a pump; 3) Fixing flow direction models check valves. Additionally, fixing the flow to zero models a closed valve, and choosing the capacities of the links amounts to determining linkflow characteristics in network design.

When calculating a DDM or PDM steady state where there are flow constraints, the flow rate can be assigned to its lower or upper limit in the active set method [3]. In such a case, we would say the link is saturated. In the path to the solution, the assignment (saturated or inactive) may change, causing a situation where some demand nodes are isolated (not connected to a fixed head node). If a linkflow constraint is saturated, its link is no longer relevant to the energy equation and the link is removed from the graph, and the assigned flow rate at its min or max bounds is subtracted or added at both end nodes of the link. The saturation of links at the origin of some isolated nodes is illustrated in Figure 1.

In the interior point method [4] as in other penalty methods (e.g. [6]), a penalty head loss is added to the total head loss link j, so the linkflow constraints are never saturated and removed from the network graph. In this paper, for the sake of simplicity, we choose not to illustrate with a penalty method.

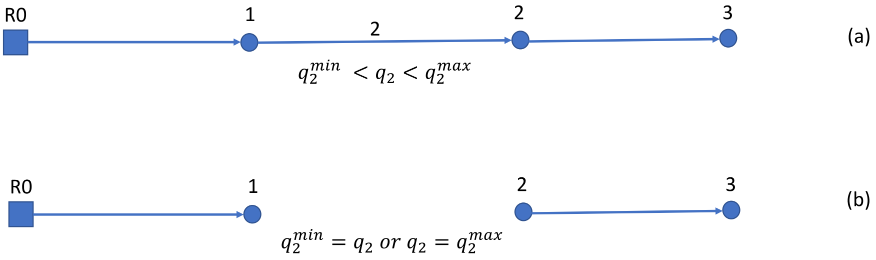

R0
1
2
2
3
$q_2^{min} < q_2 < q_2^{max}$
(a)
R0
1
2
3
$q_2^{min} = q_2$ or $q_2 = q_2^{max}$
(b)

Figure 1. Graph representation of linkflow constraints; at top (a), linkflow constraint is inactive; at bottom (b), the constraint is saturated causing nodes 1 and 2 to be disconnected from source node R0.

In PDM simulation, c_i positive and bounded by above outflows are also modelled as min-max box constraints, $0 \leq c_i \leq d_i$, where d_i is the nominal demand at node i. In this paper, we also extend the network graph, but in simpler manner than in [5]. The purpose of this extension is to facilitate the analysis of existence and uniqueness of the solution. It is not used in finding the solution [3, 4, 7]. Here one additional virtual reservoir and a link are added to each demand node as in Figure 2. Figures 2a and 2b illustrate the cases when the outflow constraint at note 3 is inactive. Nodes 2 and 3 are not isolated, even if linkflow constraint 2 is saturated and link 2 removed from the network graph. In Figure 2c, nodes 2 and 3 are isolated; the heads at nodes 2 and 3 are not uniquely defined and if the flow bounds at link 2 and node 3 are not compatible, there is no flow solution.

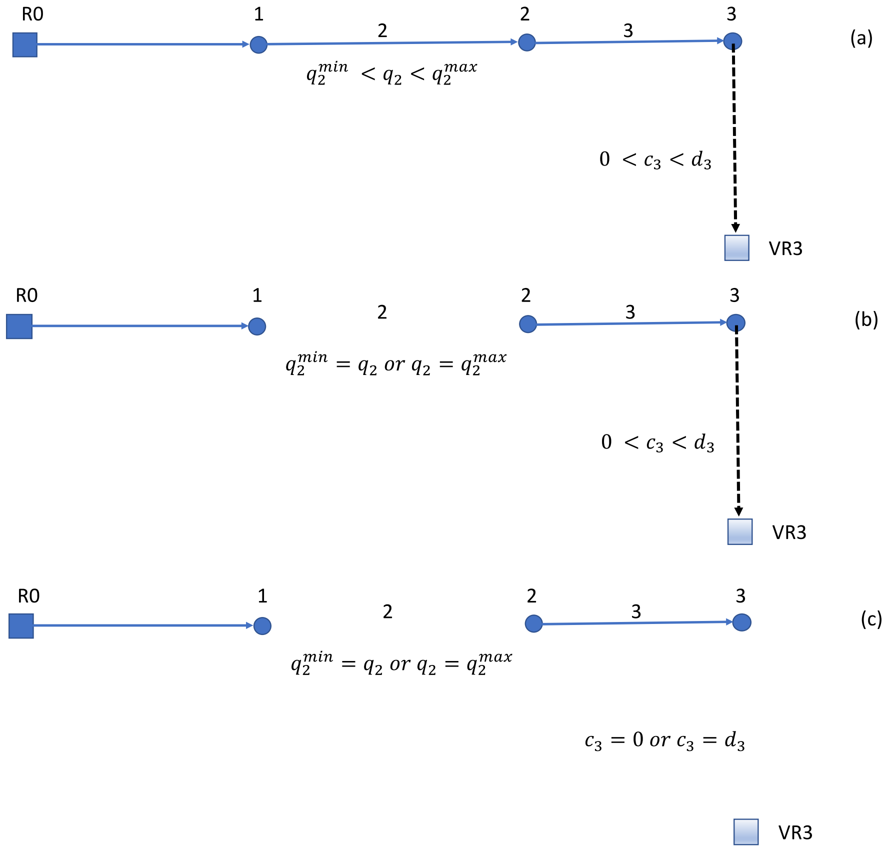

R0
1
2
2
3
3
(a)
$q_2^{min} < q_2 < q_2^{max}$
$0 < c_3 < d_3$
VR3
R0
1
2
2
3
3
(b)
$q_2^{min} = q_2$ or $q_2 = q_2^{max}$
$0 < c_3 < d_3$
VR3
R0
1
2
2
3
3
(c)
$q_2^{min} = q_2$ or $q_2 = q_2^{max}$
$c_3 = 0$ or $c_3 = d_3$
VR3

Figure 2. Graph representation of outflow constraints in PDM mode; at top (a), the linkflow constraint and outflow constraints are inactive; at middle (b), the linkflow constraint is saturated but nodes 2 and 3 are still connected to virtual reservoir VR3; at bottom (c), the linkflow and outflow constraints are saturated causing nodes 2 and 3 to be disconnected from source nodes R0 and VR3.

An unknown head loss (UHE) element is represented by a link whose head loss can be controlled to achieve a given set pressure at an assigned set pressure node (SPN). There is a one-to-one relationship between one UHE and its SPN. That means that there is exactly one SPN for each UHE. The UHE is assumed to be a unidirectional element that allows flow only in one direction ($q_j \geq 0$). In the very general case, the UHE headloss can have positive ($z_j \geq 0$) or negative sign (pumping, $z_j \leq 0$). The SPN can be upstream or downstream of the UHE; the SPN can also be at a distance from the UHE. UHE can be a pressure regulating device (PRV), or a pressure sustaining valve (PSV) or a variable speed pump (VSP).

Figure 3a shows the PRV V2 at equal distance between the two tank nodes R0 and R3. In Figure 3b, the PRV is open, the link V2 must be included in the network graph and in the conservation of energy ($h_1 = h_2$, neglecting the local head loss, or considering it $h_1 > h_2$); in Figure 3c, the PRV is active, node 2 is considered a fixed-head node, the head loss produced by the PRV explains the head difference: $z_2 = h_1 - h_2 > 0$, and the flow rate through the valve is not in one-to-one relationship with the head loss, so link 2 is removed from the graph; in Figure 3d, the PRV is closed, the heads at nodes 1 and 2 are separated.

It should be noted that having a demand at SPN 2 will not permit to satisfy the mass balance at this node, the outflow c_2 will not necessarily correspond to $q_2 - q_3$; this is because, the pressure head at node 2 is fixed by the pressure set (so the outflow would also be fixed) and does not reflect the characteristics of the system.

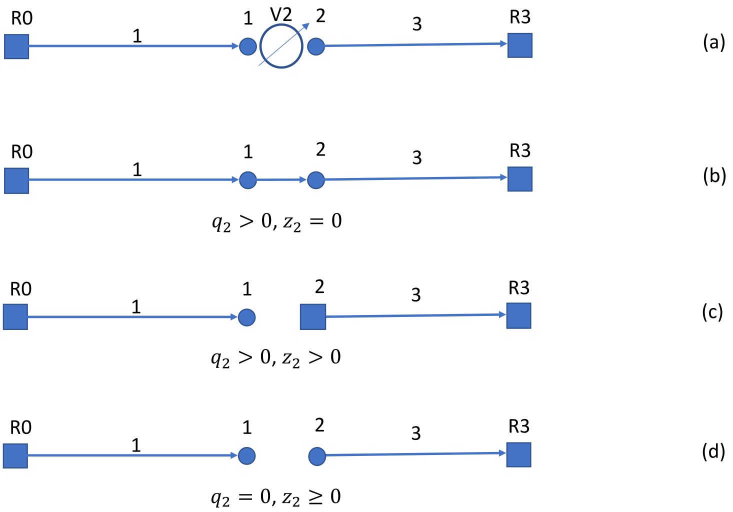


Figure 3. Graphical representations of a PRV depending on its different states; at top (a) configuration of the PRV V2 between the two tanks R0 and R3; at (b) V2 is open; at (c) V2 is active and its SPN is at the given set pressure; at (d), V2 is closed, and no water crosses the valve

The network graph may have small variations from one iteration to the next if the status of the output flow and linkflow constraints or the UHE pressure control devices change. If some junction nodes are disconnected from the source nodes, this causes a problem in the solution algorithm. We explain why below.

3 INCIDENCE MATRIX VS GRAPH REPRESENTATION

Incidence matrices are equivalent to the directed graph representations of the network. They are commonly used in the vectorised form of the steady state equations and are involved in the mass and energy equations.

Let **A** denote the $nl \times nj$, unknown-head arc-node incidence matrix (ANIM), with nl the number of links and nj the number of junction nodes. If some junction nodes are isolated from source/energy nodes then $rank(\mathbf{A}) < nj$. For the isolated junction nodes, the induced junction node incidence submatrix will therefore be rank-deficient (one of its columns is the sum of all others). As a consequence, the nodal head **h** is not defined uniquely [8]:

$$\text{if rank } (\mathbf{A}) < nj, \exists \mathbf{w} \neq \mathbf{0} / \mathbf{A}(\mathbf{h} + \mathbf{w}) = \mathbf{A}(\mathbf{h}) \quad (1)$$

Figure 4 shows the ANIMs for network variations in Figures 1 and 2. The rank of the ANIMs for network 1b and 2c are insufficient ($rank(\mathbf{A}) = 2 < 3$).

For the outflow and linkflow constraints and steady state equations, it is possible to formulate a convex minimization problem to solve [3,4]. The analysis of the existence and the search for an initial feasible solution can be achieved by solving a linear programming (LP) problem [9]; [10] used this LP problem to assess solvability in the presence of flow control devices. The verification of the full rank of the matrix, constructed from the Jacobian of the active inequality constraints and the Jacobian of the equality constraints, defines the so-called Linear Independence Constraint Qualification Condition (LICQ). Constraint qualifications such as LICQ, etc. are necessary to ensure that an optimal solution will satisfy the KKT conditions. If during the path to the solution, $rank(\mathbf{A}) < nj$, or some junction nodes are isolated, the LICQ conditions will not hold at the current iteration point, and there will be non-uniqueness of the KKT multipliers. Additionally, the Schur matrix (iteration matrix) for updating the heads has same structure/graph than $\mathbf{A}^T\mathbf{A}$, so there will be a problem solving this linear system.

BY NC SA

The pressure control problem cannot be expressed as a single optimisation problem. Some authors have proposed a bilevel optimisation [11], other Nash equilibrium [12]. The condition $\text{rank}(\mathbf{A}) = nj$, or checking if there are zero disconnected junction nodes from sources, is very robust and holds also for pressure control or UHE problem.

$$\mathbf{A}^{(\text{Fig 1a})} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow[\text{removing row 2}]{} \mathbf{A}^{(\text{Fig 1b})} = \left(\begin{array}{c:cc} -1 & 0 & 0 \\ \hdashline 0 & 1 & -1 \end{array}\right)$$

$$\mathbf{A}^{(\text{Fig 2a})} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[\text{removing row 2}]{} \mathbf{A}^{(\text{Fig 2b})} = \left(\begin{array}{c:cc} -1 & 0 & 0 \\ \hdashline 0 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right)$$

$$\mathbf{A}^{(\text{Fig 2a})} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[\text{removing rows 2 \& 4}]{} \mathbf{A}^{(\text{Fig 2c})} = \left(\begin{array}{c:cc} -1 & 0 & 0 \\ \hdashline 0 & 1 & -1 \end{array}\right)$$

Figure 4. unknown-head ANIM for network variations Figures 1 to 2; partition lines in the matrix indicates diagonal block matrices for different connected components.

4 NON-CONVERGENCE EXAMPLE

With Epanet 2.2, the simple configuration with a PSV and PRV in series cannot be handled correctly. The two UHE are between two tanks at 60 m (R0) and 30 m (R5). The set pressure for the PSV is 58 m, and the set pressure for the PRV is 35 m. Epanet has calculated in 7 iterations, with the PSV open but it cannot deliver the correct head (it should be $h_1 \geq 58\ m$).

It can be seen from Figure 5 that when the PSV and PRV are active, the two junction nodes in the middle are disconnected from the sources. This could explain Epanet's wrong results.

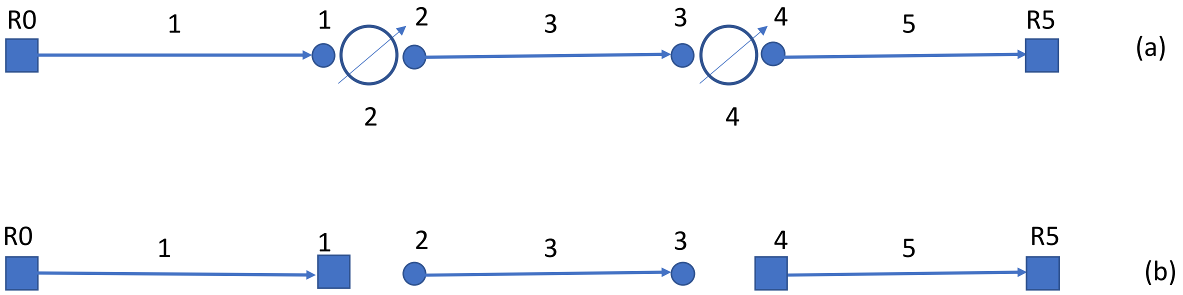


Figure 5. Graphical representation of a PSV and a PRV in series both active at same time; at top (a), configuration of the PSV/PRV between the two tanks R0 and R5; at bottom (b), the two UHEs are active causing junction nodes 2 and 3 to be disconnected from source nodes R0 and R5.

BY NC SA

5 CONCLUSIONS

In this paper graph representation of pressure control and flow constraints are made explicit.

In the context of the PDM simulation, the original graph is extended with virtual elements, one virtual link on which is imposed the outflow constraint and a virtual reservoir at the elevation of the connected demand node.

Depending on the saturation of the outflow and linkflow constraint or the status of the UHE, the corresponding link has to be removed from the graph, and some junction nodes are transformed in fixed head nodes. This process is repeated on the path to the solution, generating several variations of the extended graph. Analysis of each variation may reveal some junction node(s) disconnected from source nodes; this may cause some problem of convergence or convergence to a wrong solution. Moreover, the head is not uniquely defined in such a case.

We establish the link between the presence of disconnected junction nodes from source and rank deficiency for the junction node incidence matrix.

Future work will consist of finding rapid and robust algorithm for analysing the connection of junction nodes to the source nodes. One way is to consider only the impacts of graph variations during the iteration. Similarly, the rank of the incidence matrix of the junction nodes can be analyzed. Additionally, when dealing with large networks and a lot of constraints, finding efficient algorithms that can relax some constraints that might not be appropriate in a certain sense remains an open question.

6 REFERENCES

[1] E. Abraham, M. Blokker, and I. Stoianov, “Network Analysis, Control Valve Placement and Optimal Control of Flow Velocity for Self-Cleaning Water Distribution Systems,” PROENG, vol. 186, 2017, pp. 576-583, https://doi.org/10.1016/j.proeng.2017.03.272.

[2] A. Ulusoy, F. Pecci, and I. Stoianov, “An MINLP-Based Approach for the Design-for-Control of Resilient Water Supply Systems,” IEEE Systems Journal, 2020, 12 pp, https://doi.org/10.1109/JSYST.2019.2961104.

[3] O. Piller, S. Elhay, J. Deuerlein, and A. Simpson, “A Content-Based Active-Set Method for Pressure-Dependent Models of Water Distribution Systems with Flow Controls,” Journal of Water Resources Planning and Management, vol. 146, no. 4, 2020, 04020009, doi:10.1061/(ASCE)WR.1943-5452.0001160.

[4] S. Elhay, O. Piller, J. Deuerlein, and A. Simpson, An Interior Point Method Applied to Flow Constraints in a Pressure-Dependent Water Distribution System, Journal of Water Resources Planning and Management, vol. 148, no. 1, 2022, 04021090, doi: doi:10.1061/(ASCE)WR.1943-5452.0001484.

[5] I. Lippai and L. Wright 2014. “Demand constructs for risk analysis.” Procedia Eng. 89: 640–647, 2014. https://doi.org/10.1016/j.proeng.2014.11.489.

[6] O. Piller and J. van Zyl, “Modeling Control Valves in Water Distribution Systems Using a Continuous State Formulation,” Journal of Hydraulic Engineering, vol. 140, no. 11, p. 04014052, 2014, https://doi.org/10.1061/(ASCE)HY.1943-7900.0000920.

[7] J. W. Deuerlein, O. Piller, S. Elhay, and A. R. Simpson, “Content-Based Active-Set Method for the Pressure-Dependent Model of Water Distribution Systems,” Journal of Water Resources Planning and Management, vol. 145, no. 1, p. 04018082, 2019, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001003.

[8] O. Piller, “Water distribution system modelling and optimisation”, Final research Habilitation thesis, Dept. of Engineering Sciences, Ecole doctorale Sciences des métiers de l’ingénieur, 2019 (in French). Available: https://hal.inrae.fr/tel-02608624.

[9] S. P. Boyd, and L. Vandenberghe. “Convex optimization”, 7th ed. Cambridge, UK: Cambridge University Press, 2009, pp. 579-581.

[10] J. Deuerlein, A. Simpson, and I. Montalvo, “Preprocessing of Water Distribution Systems to Assess Connectivity and Solvability in the Presence of Flow Control Devices,” in World Environmental and Water

Resources Congress 2012, E. D. Loucks Ed. Albuquerque, New Mexico, United States, 2012, pp. 3237-3247, https://ascelibrary.org/doi/10.1061/9780784412312.325.

[11] O. Piller and J. van Zyl, "Modeling Control Valves in Water Distribution Systems Using a Continuous State Formulation," Journal of Hydraulic Engineering, vol. 140, no. 11, p. 04014052, 2014, https://doi.org/10.1061/(ASCE)HY.1943-7900.0000920.

[12] J. Deuerlein, R. Cembrowicz, and S. Dempe, "Hydraulic Simulation of Water Supply Networks Under Control," in Impacts of Global Climate Change, R. Walton Ed. Anchorage (AK), US: ASCE, 2005, pp. 1-12, https://ascelibrary.org/doi/abs/10.1061/40792%28173%2924.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14755

IMPORTANT FACTORS FOR WATER MAIN BREAK PREDICTION ACROSS 13 CANADIAN SYSTEMS

Sadaf Gharaati[1] and Rebecca Dziedzic[2]

[1,2] Concordia University, Department of Building Civil and Environmental Engineering, Montreal, Quebec, Canada

[1] *gharaati.sadaf@gmail.com*, [2] *rebecca.dziedzic@concordia.ca*

Abstract

Water main breaks can jeopardize the safe delivery of clean water and incur significant costs. To mitigate these risks, water main breaks have been predicted through physical and statistical approaches. The latter are less complex and can provide satisfactory results with less data. While many factors can contribute to breaks, the factors applied in previous studies depended on local data availability. Because other studies have focused on a few systems at a time, a broad comparison of factor importance has not been possible. This limits the understanding of the impact of different factors on water main deterioration.

The present study identifies the most important factors driving water main breaks across 13 Canadian water systems. Twenty-eight factors describing physical, historical, protection, environmental and operational attributes were compiled and cleaned. Availability of each attribute differed by system. To evaluate the importance of both numerical and categorical attributes together, two approaches were tested, categorical principal component analysis (CATPCA) and recursive feature elimination with cross-validation (RFECV). The target variable in both cases was set as yearly break status, either broken or non-broken. While CATPCA provides the contribution of each attribute to the target, RFECV provides a tuned predictive model with selected attributes. The RFECV approach was applied with Random Forest and XGBoost models, both types of machine learning models which have been shown to produce accurate results in water main break prediction.

Results from both approaches showed that physical and historical attributes are generally important across all systems. Other types of data, i.e. protection and operational are less available. When protection data is available it was shown to be even more important than physical and historical attributes. Specifically, with CATPCA, lining age and lining material were found to have a higher contribution to break status than pipe age and lining status. With RFECV lining age and lining material were also included in the best models, in particular for systems with greater percentage of lined pipes. These results indicate the choice and timing of lining are key in extending the service life of water mains. Furthermore, this data should be collected if protection practices are in place, to more accurately predict deterioration and future costs.

The results also point to an opportunity to collect more operational data. Among attributes collected by only one utility, pipe pressure, roughness, and dead-end, were found to be important in RFECV. Thus, pipe dissipation and water stagnation could lead to greater pipe deterioration. Further studies are required to quantify the impacts of different pressure ranges and network designs on deterioration.

Keywords
Water main breaks, dimensionality reduction, machine learning, physical, historical, protection, environmental, operational.

1 INTRODUCTION

Water main deterioration is a global challenge that can jeopardize water systems' ability to deliver clean water safely. The failure of water mains can affect individuals, businesses, industries, and institutions. Water main breaks can directly disrupt the service provided by pipes [1]. According to the Canadian Infrastructure Report Card [2], the cost of upgrade and replacement of water and wastewater network in Canada is estimated to be more than CAD$ 80 billion. Hence, it is essential for water utilities to seek cost-effective rehabilitation and renewal strategies[3].

The factors that contribute to water main failure are diverse and collecting all data can be cumbersome. That is why statistical models have been preferred over physical models. However, the factors applied by previous studies differ, based on the availability of data for their given case studies. Accordingly, the present study seeks to identify the most contributing factors to water main breaks across 13 Canadian cities.

2 LITERATURE REVIEW

Physical pipe attributes such as diameter, length and age are widely collected and applied in predictive models. Diameter, length, age, soil type, previous failures, and failure type were consistently applied in historical studies[4]–[6]. Pipe length is identified as an important factor by many authors. However, there isn't a consensus about whether breakage is positively[7] or negatively [8], [9] correlated with breaks [10]. Break rates also notably differ by material due to their structural resistance and vulnerability to corrosion[11], [12]. Unsurprisingly, protection of the pipe material can also extend service life[1], [12], [13]. While previous break prediction studies included data on lining status and material, the impact of lining age has not been explored.

Barton et.al. [13] note that operational and environmental factors such as sudden changes of temperature, pressure, and soil moisture level, can also increase probability of failure by increasing internal and external stress on the pipes. Martinez et.al [14] accounted for average pressure, in addition to diameter, install year and pipe depth. Snider and McBean [12] found varying break year pattern depended on various factors, most importantly weather. The impact of weather, specifically on soil movement, i.e. freeze-thaw cycles and ground swelling is confirmed by other studies[15]. It is also observed that, pipes are more likely to break once they have broken before [5], [16]–[18]. Previous breaks can be a proxy for local conditions such as soil type, weather conditions, traffic load, etc. However, the importance of these factors is not clear.

In order to identify the most important factors contributing to a target, dimensionality reduction approaches are commonly applied. However, they have not yet been applied to the analysis of water main break contributing factors. There are two general approaches for dimensionality reduction: feature elimination and feature extraction. The first reduces the number of variables by eliminating some, whereas the latter creates new independent variables from combinations of previous independent variables. A useful example of the first type is Recursive Feature Elimination with Cross-Validation (RFECV). RFECV finds the most important factors through a backward elimination process. This approach was initially introduced by Guyon et.al. [19] and is employed along with predictive models, either classification [20] or regression [21]. Previous studies have found a higher performance of RFECV with Random Forest [22], [23] and XGBoost [24]. One well-known feature extraction approach is Principal Component Analysis (PCA) [25]. This method however cannot handle categorical variables. Non-Linear PCA, also known as categorical PCA (CATPCA), is a dimensionality reduction method that, unlike PCA, can handle a non-linear relationship among variables. Categories of variables are replaced with numerical values through optimal scaling.

3 METHODOLOGY

The analysis of important factors driving water main breaks was divided into three key steps, data cleaning, data visualization and analysis, and dimensionality reduction. Each is explained in more detail in the following paragraphs.

3.1 Data preparation

Data from the utilities was provided as separate pipe inventory and historical break datasets. Thus, the first step for analyzing characteristics of broken and non-broken pipes was to merge the data. The datasets for each utility were merged based on unique IDs, identifying each pipe. Next, missing values were filled with three approaches, depending on data availability and type of attribute: 1. assumed value; 2. mirroring attribute; and 3. homogeneous groups. The first was applied for binary attributes with a clear common value. For example, anode status was only collected for pipes with anodes and all missing values were assumed to be related to pipes without protection. In the second method, missing values were replaced based on other attributes with equivalent and more detailed information. For instance, lining status (yes/no) was filled based on values of lining material. If lining material was “unlined”, the lining status was set to “no”, and “yes” for other actual lining materials. The third approach used clusters of similar pipes to replace missing values. For example, pipes with the same install year were assumed to generally be of the same material. After filling gaps, inconsistencies and outliers were detected and removed from the analysis. Lastly, categorical variables were converted to numerical through optimal scaling in R (optiscale package) for input to the correlation analysis and RFECV.

3.2 Data visualization and analysis

To better understand variations and correlations in the data, multiple graphs were generated and correlation analyses run. Because correlation reveals the relationship between numerical attributes, optimally scaled categorical variables were used. An initial analysis was performed between break status and common attributes across all cities (diameter, age, length and material). Then, a correlation analysis was run for all data for each city, and presented in a boxplot.

3.3 Dimensionality Reduction

Two dimensionality reduction methods were applied to identify the most important factors driving watermain failure, CATPCA and RFECV. The target variable in both cases was set as yearly break status (broken or non-broken). The CATPCA analysis was conducted in R (princals function, Gifi package). The number of PCs selected for each city was determined to account for around 78-85% of variance. Important factors were identified as those with a contribution above a cut-off, calculated as 100% divided by the number of attributes in each utility. The RFECV approach was conducted in python (Scikit-Learn library). Highly correlated predictors (correlation>0.8) were excluded from the analysis. Two types of estimators were employed, random forest and XGBoost. Hyperparameters were tuned for each estimator and each city. Overfitting was checked with 5-fold cross validation. Lastly, to evaluate the effectiveness of dimensionality reduction. the fit of the full data model and reduced data model were compared with F1 score and recall.

4 DATA DESCRIPTION

This study is part of the project “Best Practices for Predicting Water Main Breaks,” a collaboration between the Canadian National Water and Wastewater Benchmarking Initiative (NWWBI) and the Concordia University research group “UrbanLinks”. Thirteen utilities across Canada, in the provinces of Ontario, Nova Scotia, Newfoundland, Manitoba, Saskatchewan, and British Columbia, shared their water main inventories and historical records of main breaks as spreadsheets or GIS shapefiles. The inventory file contains information on the characteristics of existing pipes in the system, and the break file lists the failure records of broken pipes. The utilities are identified herein anonymously by the letters A through M. Overall, the data collected for this study can be categorized into five types of factors: physical, historical, protection, operational, and environmental. The attributes available in the datasets differ by utility, as shown in Table 1.

Table 1. Data available by utility (grey cells indicate available data, blank not available)

Attributes		**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
Physical	**Joint type**								▨					
	Diameter	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
	Material	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
	Length	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
	Restrained	▨												
	Roughness									▨				
	Dead-end		▨											
Historical	**Failure Month**	▨		▨	▨	▨	▨	▨		▨	▨	▨	▨	▨
	Install Month												▨	
	Status	▨		▨				▨	▨					▨
	Age	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
	Replaced Status								▨					
Protection	**Casing Material**	▨												
	Lining Material			▨	▨	▨		▨	▨		▨	▨	▨	
	Lining Status			▨	▨	▨	▨	▨	▨			▨	▨	
	Lining Age			▨		▨	▨	▨	▨				▨	
	Cathodic Protection Status	▨		▨			▨							
	Cathodic Protection Age			▨			▨							
	Coating Material										▨			▨
Operational	**Service type**	▨									▨		▨	
	Pressure											▨		

It is clear that certain physical and historical attributes are collected consistently by all utilities: diameter, material, length and age. These attributes are not only among the easiest to collect as they are generally recorded at the time of design and installation, but are also the most commonly applied in predictive modelling. The majority of utilities also record information on lining. On the other hand, only one utility recorded pipe pressure in their inventory.

Pipe materials have evolved throughout the years. The most common pipe material installed in the early to mid-1900s was cast iron, as shown in Figure 1. This market was taken over by ductile iron pipes in the 70s. Soon after, with the advent of plastic pipes, these became the most popular, in particular PVC.

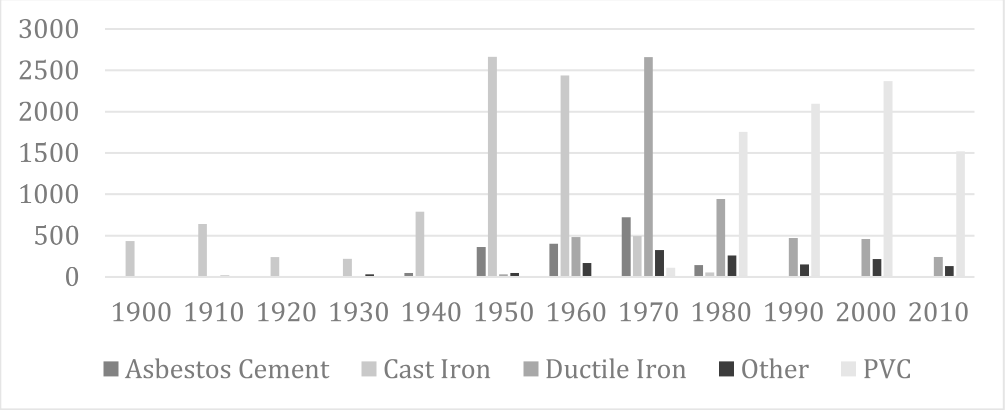


Figure 1. Total lengths of the pipes installed in each decade

Because material use trends changed over the years, the majority of pipes currently breaking are cast iron and ductile iron. This leads to a clear difference in material distribution for inventory and broken pipes, as illustrated in Figure 2. While almost 40% of pipes currently installed are PVC, more than 50%of the break records are for cast iron pipes.

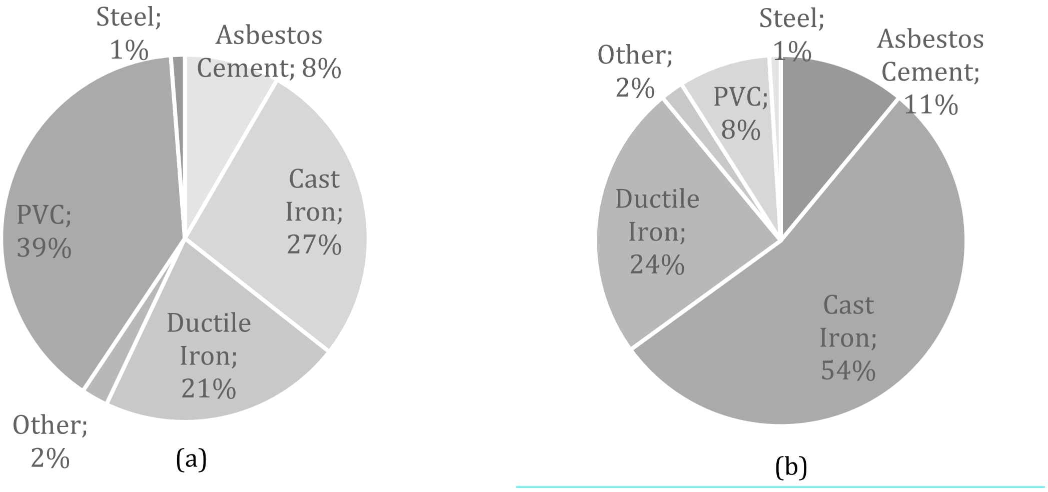


Figure 2. Breakdown of pipes by material for (a) all pipes and (b) broken pipes

It should be noted that he period of historical data collected by each utility differs. While the earliest data collecting utilities began in the 1950s (B, H and M), others only have the last two decades of data available (J and L). The size of the networks varies significantly as well. The largest is B with 6,811 km and the smallest is E with 12 km.

Table 2. General characteristics of pipes in each utility

Utility	Length (km)	Total breaks per KM	Break decades	% Cast Iron	% Ductile Iron	Average age	% Lined pipes	Average Lining age	% Protected pipes
A	897	15.1	1970-2010	16	25	27	-	-	4.3
B	6,811	15.1	1950-2010	21	23	30	-	-	-
C	3,183	13.3	1970-2010	17	20	31	13	32	13
D	2,710	14.9	1970-2010	44	48	39	44	-	-
E	12	956	1980-2010	24	35	34	0.3	018	-
F	1,501	15.7	1970-2010	9	14	23	11.2	33	27
G	392	10.2	1980-2010	10	32	32	25	26	-
H	1,363	27.4	1950-2010	19	0.2	34	0.9	5	-
I	694	15.4	1980-2010	43	44	40	-	-	-
J	1,577	25.5	1990-2010	43	54	43	47	-	-
K	351	9.3	1980-2010	49	37	56	5	-	-
L	481	13.4	1990-2010	31	15	36	11	14	-
M	4,862	33.2	1950-2010	25	1	38	-	-	-

5 RESULTS AND DISCUSSION

5.1 Correlation analysis

Analyzing all common utility attributes, material, diameter, length, age and their relation to break status does not reveal any high correlations. Figure 3 shows attributes are neither highly correlated with each other nor with the target break status. Among these common attributes, material and the length have the strongest association with the target.

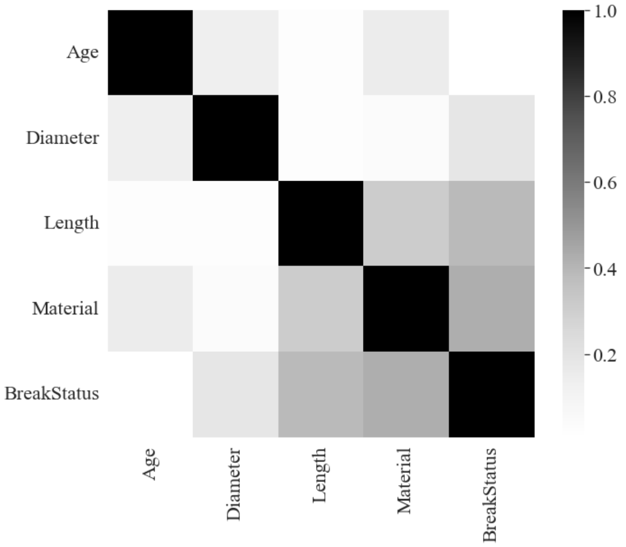


Figure 3. Correlation analysis between common attributes and break status

To further explore the correlations in each city and all attributes, separate correlation analyses were conducted. Results are summarized in Figure 4. The number of values in each box plot depends on how many utilities recorded that data. Attributes collected by only one utility are represented as a line. The most correlated attribute to break status is material. Nonetheless, results vary significantly by utility. This could be related to the variation of material within the utility. The lowest correlation (0.08) was found for utility J whose pipes are 97% either cast iron or ductile iron.

While previous studies commonly applied age in predicting watermain failure, results show cathodic protection age is more highly correlated with break status than age. This highlights the benefit of cathodic protection especially for largely metallic networks such as those analyzed herein.

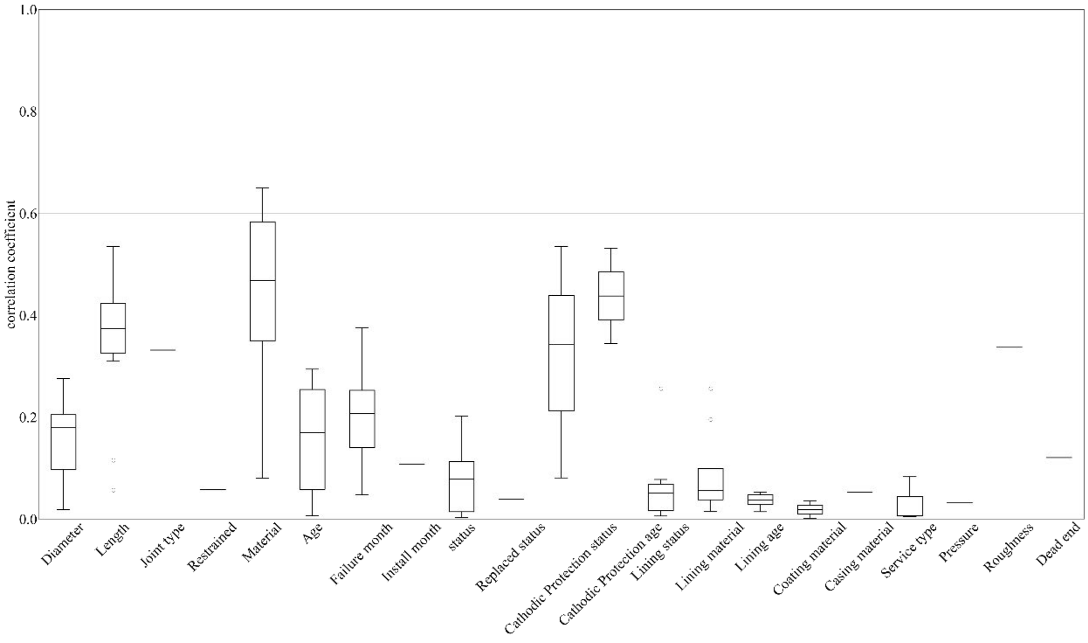


Figure 4 Correlation coefficients - Break status

5.2 Categorical Principal Component Analysis (CATPCA)

CATPCA results (Table 3) show the contribution of each attribute to break status. Overall results point to the importance of protection in general, i.e. lining, coating, and cathodic protection. In particular, the age of the protection, not only status is important. The type of lining material was often more important than the pipe material as well. Lining age was also found to be important for most utilities collecting this data, except utilities E and G, which have the lowest percentages of lined pipes, 0.3 and 0.9% respectively. Thus, collecting protection data can improve the selection and timing of protection activities, potentially reducing capital costs.

Table 3. CATPCA results - Break Status (Dark grey-important factors, light grey-available factors, blank cells-no available attribute)

Attributes		A	B	C	D	E	F	G	H	I	J	K	L	M
Physical	**Joint type**								8					
	Diameter	11	20	8	10	13	11	12	12	17	12	10	3	13
	Material	11	25	9	14	12	10	10	12	14	12	16	10	12
	Length	12	24	5	14	14	9	11	10	23	11	8	11	15
	Restrained	5												
	Roughness									12				
	Dead-end		10											
Historical	**Failure Month**	12		10	17	14	12	10		17	15	14	9	15
	Install Month												12	
	Status	11		9				11	4					16
	Age	9	21	8	14	11	9	11	12	16	11	10	10	15
	Replaced Status								10					
Protection	**Casing Material**	8												
	Lining Material			10	16	13		12	11		13	14	11	
	Lining Status			10	16	12	14	12	11			14	11	
	Lining Age			10		12	14	11	9				11	
	Cathodic Protection Status	12		10			11							
	Cathodic Protection Age			10			11							
	Coating Material										15			15
Operational	**Service type**	10									11		12	
	Pressure											14		
% Contribution cut-off level		**10**	**20**	**9**	**14**	**13**	**11**	**11**	**10**	**17**	**13**	**13**	**10**	**14**

BY NC SA

The results also consistently identified the most collected physical factors, i.e., material, diameter, and length as important. This data was shown to be particularly important when other attributes such as protection factors were not available, as for utilities A and B. Less commonly collected data, such as install month and pressure were also found to be important. However, the range of their contribution requires further investigation as only one or two utilities collect this data.

Failure month was found to be important for most utilities. Graphing the distribution of breaks over months as shown in Figure 5 elucidates the relationship between time of year and breaks. Breaks are more likely in colder months, i.e. January and February. An increase in breaks can also be seen at the height of summer in July when weather is dryer, confirming previous finding (Bruaset and Saegrov, 2018).

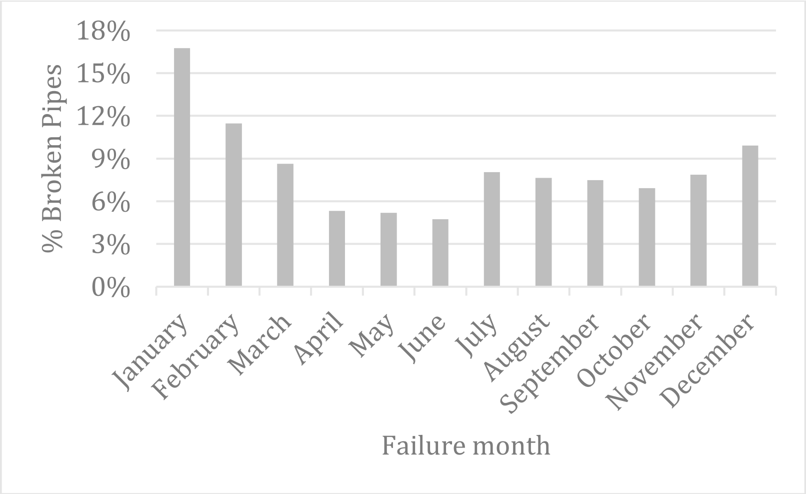


Figure 5. Total percentage of the failed pipes in each month

5.3 Random Forest Recursive Feature Elimination with Cross-Validation (RF-RFECV)

The RF-RFECV approach further reduces the number of selected features, compared with CATPCA. The reduced models, i.e. with fewer attributes, perform equally or slightly better than the full models, i.e. with all attributes, as shown in Table 4. Because the model predict the categorical target break status, they are evaluated according to F1 score and recall. Overall, the analysis rated physical and historical factors as the most important. Specifically length, age and material were consistently found to have the highest weights. Protection activities were rated less highly compared to CATPCA, but cathodic protection age and lining age were still found to be important. The results also selected joint type, pressure, roughness, and dead-end among the important features. This points to the opportunity to collect more operational data and explore the relation between operational decision and infrastructure service life. Pipe dissipation and water stagnation could lead to greater pipe deterioration.

To evaluate the performance and applicability of a model with even fewer attributes, models were developed with only common data (length, material, diameter and age). F1 and recall scores for these models are also provided in Table 4. The performance of these models is only slightly lower than the reduced models developed with RF-RFECV. Thus, for the purpose of predicting pipe deterioration for maintenance and capital planning, commonly available attributes should suffice. Nevertheless, in creating strategies for reducing maintenance and replacement costs, the relation between breaks and other adjustable factors such as pressure and protection should be explored.

Table 4. RF-RFECV weights and results - Break Status (Dark grey-important factors, light grey-available factors, blank cells-no available attribute)

	Attributes	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
Physical	**Joint type**								0.10					
	Diameter	0.07	0.07	0.08	0.06	0.06	░	0.10	0.06	0.06	0.05	0.07	0.04	0.03
	Material	0.35	0.34	0.32	0.33	0.21	0.40	0.08	0.19	0.11	0.03	0.09	0.14	0.40
	Length	0.22	0.26	0.19	0.34	0.38	0.17	0.35	0.28	0.27	0.43	0.31	0.34	0.30
	Restrained	░												
	Roughness									0.11				
	Dead-end		0.01											
Historical	**Failure month**	0.12		0.10	0.10	0.14	░	0.24		0.18	0.10	0.18	0.19	0.03
	Install month												░	
	Status	0.03		░				░	0.003					0.003
	Age	0.19	0.32	0.18	0.17	0.22	0.18	0.23	0.37	0.26	0.36	0.29	0.23	0.24
	Replaced status								░					
Protection	**Casing material**	0.02												
	Lining material			░	░	░		░	░		0.04	0.01	░	
	Lining status			░	░	░	░	░	░			░	░	
	Lining age			0.08		0.003	0.08	░	0.005				0.06	
	Cathodic Protection status	0.01		░			░							
	Cathodic Protection age			0.08			0.17							
	Coating material										░			
Operational	**Service type**	░									░		░	
	Pressure											0.04		
Full F1		**97.5**	**97.5**	**97.5**	**95.3**	**97.2**	**98.9**	**98.6**	**97.4**	**95.2**	**99.4**	**92.9**	**97.3**	**96.7**
Reduced F1		**97.5**	**97.5**	**97.5**	**95.3**	**97.2**	**98.9**	**98.6**	**97.4**	**95.2**	**99.4**	**92.9**	**97.2**	**96.7**
Common F1		**96.8**	**97.4**	**96.3**	**95.2**	**97.1**	**98.0**	**98.0**	**97.1**	**94.4**	**99.4**	**92.3**	**96.8**	**96.7**
Full Recall		**98.4**	**98.3**	**98.6**	**96.7**	**98.8**	**99.2**	**100**	**98.6**	**97.6**	**99.9**	**96.8**	**98.5**	**97.2**
Reduced Recall		**98.6**	**98.3**	**98.6**	**96.4**	**98.8**	**99.4**	**99.9**	**98.6**	**97.6**	**99.9**	**96.8**	**98.5**	**97.2**
Common Recall		**97.7**	**98.3**	**98**	**96.3**	**98.7**	**99.2**	**99.7**	**98.5**	**96.8**	**100**	**96.2**	**97.3**	**97.2**

5.4 XGBOOST Recursive Feature Elimination with Cross-Validation (XGB-RFECV)

The XGBOOST-RFECV approach yields slightly underperforming models, compared to RF-RFECV, as shown in Table 5. The features selected, however are similar in both RFECV approaches. Physical and historical attributes are rated as the most important, specifically material, diameter, length, age, and failure month. Similar to RF-RFECV, material is the most important factor. Lining age is consistently important for utilities collecting this data except E, which is only 0.3% lined. Lining material was found to be important for utilities D and J, with more than 40% of lined pipes. Thus, the contribution of certain factors depends on local practices and conditions.

Table 5. XGB-RFECV weights and results - Break Status (Dark grey-important factors, light grey-available factors, blank cells-no available attribute)

Attributes		A	B	C	D	E	F	G	H	I	J	K	L	M
Physical	**Joint type**								0.10					
	Diameter	0.06	0.05	0.02	0.02	0.10	0.02	0.20	0.07	0.06	0.10	0.12	0.07	0.02
	Material	0.53	0.66	0.69	0.85	0.59	0.78	0.13	0.52	0.41	0.36	0.34	0.41	0.77
	Length	0.04	0.08	0.04	0.06	0.12	0.04	0.17	0.07	0.09	0.10	0.14	0.12	0.05
	Restrained	0.05												
	Roughness									0.16				
	Dead-end		0.01											
Historical	**Failure month**	0.04		0.03	0.02	0.11	0.02	0.11		0.15	0.12	0.24	0.16	0.01
	Install month												0.02	
	Status	0.07		░				0.05	0.04					0.06
	Age	0.05	0.20	0.07	0.03	0.07	0.07	0.11	0.16	0.13	0.12	0.16	0.10	0.09
	Replaced status								░					
Protection	**Casing material**	0.05												
	Lining material			░	0.01	░		░	░		0.14	░	░	
	Lining status			░	░	░	░	░	░			░	░	
	Lining age			0.12		░	0.03	0.23	0.05				0.12	
	Cathodic Protection status	0.09		░			░							
	Cathodic Protection age			0.02			0.03							
	Coating material										0.05			░
Operational	**Service type**	0.04									░		░	
	Pressure											░		
Full F1		**86.5**	**91.8**	**89.7**	**90.9**	**79.6**	**90.3**	**61.3**	**88**	**77**	**35**	**69**	**74**	**86**
Reduced F1		**86.5**	**91.8**	**89.7**	**90.9**	**73**	**90.3**	**61.3**	**88**	**77**	**27**	**69**	**74**	**86**
Common F1		**85.9**	**85.9**	**88**	**91**	**78**	**90**	**58.8**	**86.2**	**72.5**	**24**	**68.5**	**66.5**	**85.6**
Full Recall		**84.8**	**89.3**	**85.9**	**89.1**	**78**	**86.6**	**60**	**84**	**71.5**	**27**	**62**	**68**	**83**
Reduced Recall		**84.8**	**89.3**	**85.6**	**89.1**	**72**	**86.6**	**60**	**84**	**71.5**	**24**	**62**	**69**	**84**
Common Recall		**84.8**	**84.8**	**84**	**89**	**71**	**87.5**	**52.6**	**81.8**	**66**	**17.5**	**61.6**	**62.6**	**83.5**

5.5 Discussion

Although all approaches applied in this study are of dimensionality reduction, they differ in nature. When data are numerical and linearly related to the target, correlation-based approaches are most appropriate. However, this is not the case for water main break prediction as a mix of numerical and categorical factors is available. CATPCA, also known as non-linear PCA, can handle a linear and non-linear relationship among variables and is recommended when mixed types of data in the analysis are not linearly related to the target.

RFECV can also handle different relations between predictors and targets through different estimators. The selection of an appropriate estimator based on the data structure is key to ensuring good results. In the present study XGBOOST and Random Forest estimators were selected and their hyperparameters tuned to maximize performance. In particular, the resulting random forest models were more accurate and, thus, provide more reliable feature selection results. However, the results are still largely dependent on the hyperparameter tuning.

The most important factors differed between approaches. While physical and historical factors were the most contributing factors in RFECV, CATPCA found protection activities to be the most important. CATPCA also selected a greater number of important factors compared to RFECV. Results also differed by utility, depending on the application of certain protection strategies and the variability of local practices, e.g. different types of installed materials. Thus, data collection strategies should be tailored to the factors impacting the most common materials and protection approaches in each utility.

6 CONCLUSION

Because data collection can be a time and cost intensive process, identifying the driving factors for water main breaks is a valuable endeavour. Based on the results of the dimensionality reduction approaches, a three-step data collection framework is proposed, summarized in Table 6. The first step represents the minimum level of data collection required to produce accurate water main break prediction models. Factors include material, diameter, length and age (calculated based on install date and failure date). This data is commonly collected across all utilities and was found to generate models with high F1 and recall scores, slightly below the optimal RFECV models.

Table 6. Three step data collection framework (1st step - dark grey, 2nd step- grey, 3rd step - light grey)

Physical	**Historical**	**Protection activities**	**Operational**
Material	Installation date	Cathodic Protection year	Pressure
Diameter	Failure date	Lining Material	Service Type
Length	Status	Lining Year	-
Join type	-	Coating Material	-
Roughness	-	Anode type	-
Dead-end	-	-	-
Restrained	-	-	-
Pipe Depth	-	-	-

The second step comprises factors that were found to be important when relevant, especially protection data. Collecting protection data can improve the selection and timing of protection activities, potentially reducing capital costs. More research is required on the extension of pipe service life caused by different types of protection and at different times. Lastly, the third step includes factors that were only collected by a few utilities and not identified as important in all approaches. Among attributes collected by only one utility, pipe pressure, roughness, and dead-end, were found to be important in RFECV. The results also point to an opportunity to collect more operational data and further research to quantify the impacts of different pressure ranges and network designs on deterioration.

7 REFERENCES

[1] I. American Water Works Service Co., “Deteriorating Buried Infrastructure Management Challenges and Strategies,” Environ. Prot. Agency, pp. 1–33, 2002.

[2] CIRC, “Informing the Future,” 2016. doi: 10.17226/11469.

[3] B. Rajani and Y. Kleiner, “Comprehensive review of structural deterioration of water mains: Statistical models,” Urban Water, vol. 3, no. 3, pp. 151–164, 2001, doi: 10.1016/S1462-0758(01)00032-2.

[4] U. Shamir and C. D. D. Howard, “An analytic approach to scheduling pipe replacement.,” Am. Water Work. Assoc. J., vol. 71, no. 5, May 1979, pp. 248–258, 1979, doi: 10.1002/j.1551-8833.1979.tb04345.x.

[5] T. M. Walski and A. Pelliccia, “Economic Analysis of Water Main Breaks.,” J. / Am. Water Work. Assoc., vol. 74, no. 3, pp. 140–147, 1982, doi: 10.1002/j.1551-8833.1982.tb04874.x.

[6] A. J. Kettler and I. C. Goulter, “Analysis of Pipe Breakage in Urban Water Distribution Networks.,” Can. J. Civ. Eng., vol. 12, no. 2, pp. 286–293, 1985, doi: 10.1139/l85-030.

[7] L. Berardi, O. Giustolisi, Z. Kapelan, and D. A. Savic, “Development of pipe deterioration models for water distribution systems using EPR,” J. Hydroinformatics, vol. 10, no. 2, pp. 113–126, 2008, doi: 10.2166/hydro.2008.012.

[8] Y. Wang, T. Zayed, M. Asce, O. Moselhi, and F. Asce, “Prediction Models for Annual Break Rates of Water Mains,” no. February, 2009.

[9] M. Nishiyama and Y. Filion, “Forecasting breaks in cast iron water mains in the city of Kingston with an artificial neural network model,” Can. J. Civ. Eng., vol. 41, no. 10, pp. 918–923, 2014, doi: 10.1139/cjce-2014-0114.

[10] Y. Wang, O. Moselhi, F. Asce, T. Zayed, and M. Asce, “Study of the Suitability of Existing Deterioration Models for Water Mains,” no. February, pp. 40–46, 2009.

[11] P. Rajeev, J. Kodikara, D. Robert, P. Zeman, and B. Rajani, “Factors contributing to large daimeter water pipe failure,” Water Asset Manag. Int., vol. 10, no. 3, pp. 9–14, 2014.

[12] B. Snider and E. A. McBean, “State of watermain infrastructure: A canadian case study using historic pipe break datasets,” Can. J. Civ. Eng., vol. 48, no. 10, pp. 1266–1273, 2021, doi: 10.1139/cjce-2020-0334.

[13] N. A. Barton, T. S. Farewell, S. H. Hallett, and T. F. Acland, “Improving pipe failure predictions: Factors effecting pipe failure in drinking water networks,” Water Res., vol. 164, p. 114926, 2019, doi: 10.1016/j.watres.2019.114926.

[14] Martínez-Codina, M. Castillo, D. González-Zeas, and L. Garrote, “Pressure as a predictor of occurrence of pipe breaks in water distribution networks,” Urban Water J., vol. 13, no. 7, pp. 676–686, 2016, doi: 10.1080/1573062X.2015.1024687.

[15] S. Bruaset and S. Sægrov, “An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions,” Water (Switzerland), vol. 10, no. 4, 2018, doi: 10.3390/w10040411.

[16] S. A. Andreou, D. H. Marks, and R. M. Clark, “A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory,” Adv. Water Resour., vol. 10, no. 1, pp. 2–10, 1987, doi: 10.1016/0309-1708(87)90002-9.

[17] R. Jafar, I. Shahrour, and I. Juran, “Application of Artificial Neural Networks (ANN) to model the failure of urban water mains,” Math. Comput. Model., vol. 51, no. 9–10, pp. 1170–1180, 2010, doi: 10.1016/j.mcm.2009.12.033.

[18] L. Scholten, A. Scheidegger, P. Reichert, M. Mauer, and J. Lienert, “Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis,” Water Res., vol. 49, pp. 124–143, 2014, doi: 10.1016/j.watres.2013.11.017.

[19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using DCA,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5139 LNAI, pp. 62–72, 2008, doi: 10.1007/978-3-540-88192-6_8.

[20] Q. Chen, Z. Meng, X. Liu, Q. Jin, and R. Su, “Decision variants for the automatic determination of optimal feature subset in RF-RFE,” Genes (Basel)., vol. 9, no. 6, 2018, doi: 10.3390/genes9060301.

[21] B. Butcher and B. J. Smith, “Feature Engineering and Selection: A Practical Approach for Predictive Models,” Am. Stat., vol. 74, no. 3, pp. 308–309, 2020, doi: 10.1080/00031305.2020.1790217.

[22] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, “Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products,” Chemom. Intell. Lab. Syst., vol. 83, no. 2, pp. 83–90, 2006, doi: 10.1016/j.chemolab.2006.01.007.

[23] S. Wang and S. Chen, “Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling,” J. Pet. Sci. Eng., vol. 174, no. November 2018, pp. 682–695, 2019, doi: 10.1016/j.petrol.2018.11.076.

[24] W. Chang, Y. Liu, Y. Xiao, X. Yuan, X. Xu, and S. Zhang, “A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data,” Diagnostics, vol. 9, no. 178, 2019.

[25] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14760

VALVE LOSS CURVE ESTIMATION USING REAL-TIME SCADA DATA

S.M. Masud Rana[1], Gal Perelman[2], Elad Salomons[3], James G. Uber[4]

[1,2,3,4] Xylem Inc., 1 International Drive, Rye Brook, NY 10573, United States

[1] *smmasud.rana@xylem.com*, [2] *gal.perelman@xylem.com*,
[3] *elad.salomons@xylem.com*, [4] *james.uber@xylem.com*

Abstract

Real-time operation of drinking water networks (DWN) through supervisory control and data acquisition (SCADA) systems can potentially reduce energy consumption. This can be achieved by controlling pumps and valves to reduce pumping costs and energy dissipated at valves. Moreover, other objectives can be improved, such as pressure levels and water quality throughout the DWN. With the integration of hydraulic models and real-time data sources, optimization of the real-time operation of DWNs has become more feasible in recent years. An accurate hydraulic model of a DWN is necessary to realize the benefits of real-time modelling and control, as the model is required for estimating the state of the network for different operating conditions. Pumps and valves curves estimation is a key step toward achieving an accurate model. Valves, which act as energy sinks, are vital elements in DWNs and stabilize pressures in certain parts of the network. Therefore, accurate flow-head loss relationships of the valves thorough a range of valve openings are essential for real-time modeling. For example, to estimate pressures downstream of the valves and optimize the operation of pumps and valves for forecasted demands to reduce the overall energy spent for DWN operation. Head loss across valves is conventionally modeled as proportional to the velocity head across the valve, with the proportionality factor called the minor-loss coefficient. The minor-loss coefficient, in turn, is a non-linear function of the valve opening degree, and valve manufacturers generally provide relationships between the minor-loss coefficient and the valve percent opening. Our experience is that using these original manufacturer curves can lead to significant prediction errors compared to operational data and should be tested and estimated. We show how to use measurements of valve opening, pressure differential, and flow through an individual valve or a valve facility for the estimation of valve curves. The approach can be used as a batch process as part of network calibration or may also be used in real-time to maintain accurate valve characteristics over time. A method is presented to estimate valve loss curves represented by a simple function (e.g., power-law or polynomial), using measured flow and pressure differential and measurement of valve opening degree. The method uses a gradient-based optimizer to solve a least-squares estimation (LSE) problem, where the residual error is the difference between predicted and measured flow through the valves. This method is applicable for valve facilities that contain more than one valve in parallel, where flow measurements are available for the valve facility or individual valves. The method developed was applied to estimate 77 valve curves for a large water transmission network in the USA as part of the development of a digital twin for the DWN. Results will summarize the statistical prediction errors for this set of control valves, before and after estimation, and will show the impact on network hydraulic predictions from using estimated valve headloss curves.

Keywords
Valve curve estimation, SCADA, real-time modelling, digital twin.

1 INTRODUCTION

Drinking water networks (DWN) consume about 4% of the total energy produced in the United States of America [1,2]. A large fraction of the consumed energy is used for pump operation. Pumps and valves are key control elements that are used by the network operators to drive the operation of a WDN to maintain adequate water quality and pressure throughout the system. While pumps are sources of energy (head) into the drinking water system, valves act as sinks of energy as they locally lose some of the energy. The use of valves has varied purposes, such as reducing pressure and maintaining water quality (e.g., enhancing chlorine residuals by altering hydraulic paths [3]). In most utilities across the USA, the pumps and valves are operated by experienced operators to achieve objectives such as maintaining minimum pressure at certain locations and tank recycling. Such manual operation objectives rarely consider energy consumption, and hence opportunities exist for energy consumption reduction by optimizing DWN operations.

Increasingly SCADA systems are being used by DWNs to collect a considerable amount of data (e.g., tank levels, pressures, water quality, and pump station flow measurements), which describes the real-time state of DWNs. Digital twins, which are real-time representations of DWNs, can predict the state of the network in space and time by coupling the collected SCADA data with hydraulic models and enable better operational decision-making [4,5]. The accuracy of digital twins is dependent on the accuracy of the underlying hydraulic model. For that, water demand is predictied using history of past states of the network from SCADA [6,7]. With the availability of accurate demand predictions and accurate hydraulic models, real-time operation optimization becomes feasible [8–11]. Therefore, an accurate hydraulic model is essential for generating a digital twin configuration. A calibrated hydraulic model can accurately predict the pressures and flows throughout the network. Two of the most important parameters that need to be predicted by the hydraulic models are the head gains across pumps and headloss through the valves. Hydraulic models, such as the EPANET [12], often use curves to predict the flow through pumps and headloss through valves. Therefore, the calibration of these pumps' and valves' curves using measured data is an important first step toward developing an accurate hydraulic model. Traditionally, these curves are estimated in field tests. However, field tests are expensive and can not be done often. Thus, using SCADA data for curve estimation is advantageous [13].

Headloss through valves is conventionally modeled as the product of the velocity head and a minor loss coefficient expressed as a function of valve opening. The valve curve relates the valve opening degree to a minor loss coefficient. While these curves are provided by the manufacturers of the valves, experience shows that the use of the manufacturer-provided curves most often leads to significant model prediction errors when compared to operational data [14]. Therefore, pump and valve curves need to be tested and calibrated before use for successful developments of digital twins and real-time operation optimization capabilities. The manufacturer valve curves can be use as base curves for the calibration process using real-time operational data to ensure the valve characteristics are accurately reflected by the curves.

Despite the importance of well-calibrated pumps and valves curves for many applications, only a few studies have addressed this issue. Araujo et al. [15] used a genetic algorithm to calibrate valves curves to minimize leakage pressure control technique. Campisano et al. [16] suggested real-time pressure control for the same goal of leakage reduction, they presented a calibration method based on a comprehensive dimensionless analysis of simplified hydraulic systems under real-time scenarios. The battle of the water calibration network [17] presents the challenge of overall network calibration. Within this challenge, calibration of pump curves and opening percentage of valves had to be addressed.

The objective of this paper is to present a novel method to calibrate valve curves using real-time SCADA measurements of pressures and flows along with valve opening data stream. The method can be used for real-time calibration of valve characteristics. The method is applicable for a single valve and for a group of valves in a valve facility that are arranged in parallel with individual or totalized flow measurements. The method can be used as a batch process as part of network calibration or may also be used in real-time to maintain accurate valve characteristics over time.

2 METHOD

2.1 Curve parameter estimation

The valve loss curve relating valve opening x with its minor loss coefficient k is conventionally modeled with a power-law curve of the form $k = ax^b$, with the parameters a and b to be estimated. To model a positive and monotonically decreasing minor loss coefficient with increasing valve opening, the parameter a needs to be positive, while the parameter b needs to be negative. For a given minor loss coefficient, headloss h_L across a valve is estimated as a multiple of the velocity head as

$$h_L = k\frac{V^2}{2g} \tag{1}$$

where $V = q/A$ is velocity, q is flow, A is cross-sectional area of the valve, and g is the acceleration due to gravity. Rearranging Eq. 1 in terms of valve flow and replacing k with the power law equation, the estimated flow q_e through a valve facility with N valves in parallel is given by Eq. 2.

$$q_e = \sum_i^N \left(2gh_L\left(a_i x_i^{b_i}\right)^{-1} A_i^2\right)^{\frac{1}{2}}. \tag{2}$$

Measured flow q_m through the valve facility can be used to develop a least-square estimation problem to estimate the parameters of the power-law curves for the N valves such that the mean of squared flow errors (MSE), given in Eq 3, for T discrete time steps is minimized.

$$\begin{gathered} \min_\theta \frac{1}{T}\sum_{t=1}^{T} \left(q_{e,t} - q_{m,t}\right)^2, \theta = \{a_1, b_1, \dots, a_N, b_N\} \\ a_i > 0, i = 1, \dots N \\ b_i \leq 0, i = 1, \dots, N \end{gathered} \tag{3}$$

Although a power-law curve was used in this research, any other suitable form can be used, and their parameter can be estimated similarly using Eq. 2. A parameter estimation program was written in the Python programming language, which can connect with SCADA through a digital twin of any network for which valve curves need to be estimated. The estimator program utilizes the Scipy library [18] to perform the least square optimization using the L-BFGS-B algorithm.

2.2 Data connections

The required data for valve curve calibration include the cross-sectional area of the valve, valve position data, the pressure upstream and downstream of the valve facility, accurate elevations of the pressure measurement locations, and flow measurements through the valve station. Except for the static information (i.e., valve diameter and elevations), the rest of the data are acquired from SCADA.

A facility description file describes the structure of the valve facilities in a network, i.e., the groupings of valves that forms a valve facility and the pressure and flow measurement locations for the facility in the network model. An in-house digital-twin system powered by EPANET-RTX

[19] contains information about the different available SCADA data streams and their mapping to the network model elements within a configuration file. The configuration file also contains different timeseries transformation instructions, e.g., resampling, outlier removal, moving average filters, etc., that is required to transform the raw SCADA data to data suitable for parameter estimation.

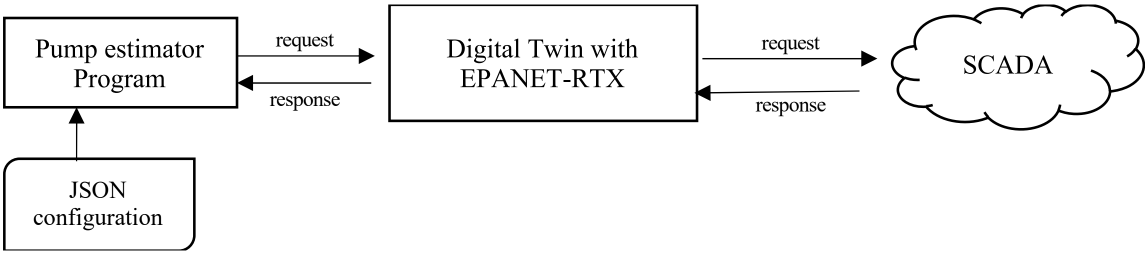


Figure 1. Data request and response structure from the estimator code to SCADA

Figure 1 shows how the data requests by the parameter estimator code are handled by the different parts of the digital twin and SCADA system. The curve estimator program uses the facility description file to request relevant data from a server that the digital twin program listens to using a representational state transfer (REST) protocol. After receiving the request, the digital twin connects to the SCADA and fetches the required data. The data is then processed by the digital twin and returned to the parameter estimator program (residing on a local machine or on a server). The parameter estimator uses the processed data for the optimization to estimate the parameters of the valve characteristics curves.

2.3 Case Study

The parameter estimation method was applied to estimate valve curves for 77 flow control valves for a large metropolitan water utility (peak production of around 800 MGD). The water utility operates a transmission network and supplies water to smaller regional distribution networks through these flow control valves. Figure 2 shows one such valve facility through which the transmission network supplies a distribution network tank. A digital twin for the transmission network for the utility was developed, which enable real-time network information monitoring and operation optimization for energy consumption reduction and water quality objectives. The utility collects high-density pressure measurements upstream and downstream of the valves, flow measurements across individual valves, and valve opening degree measurements. No manufacturer curves for the valves were available, and hence a power-law curve with the parameters $a = 35, b = -1.25$ was assumed to be the base (an initial guess) curve for all valves. The curves were estimated using measured operational data ranging from 2021-07-05 through 2021-11-15.

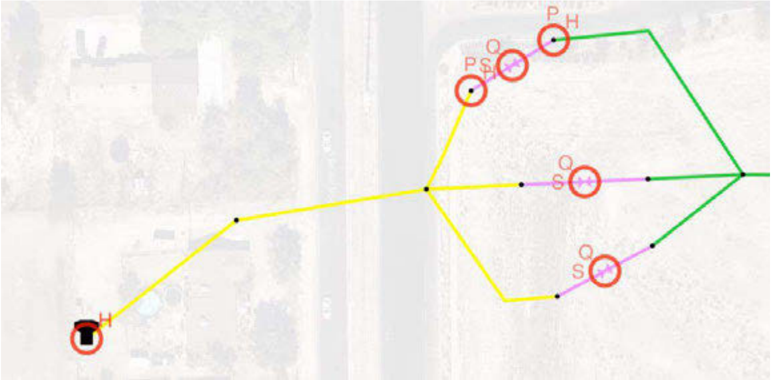


Figure 2. A valve facility with three valves supplying water to a tank in a regional distribution network. The red circles indicate the availability of measurement data at those model elements, with letters H, P, Q, and S indicating the availability of head, pressure, flow, and valve settings data streams, respectively.

Results for four valve facilities, VF1 through VF4, are presented to demonstrate the application of the estimator program. Flow measurements for all individual valves were available for this network, and therefore it was possible to estimate valve curves by considering each individual valve as one valve facility (i.e., $N = 1$ in Eq. 3). VF1 has three valves in parallel, and Figure 3 shows its valve percent opening timeseries for the three valves obtained from SCADA.

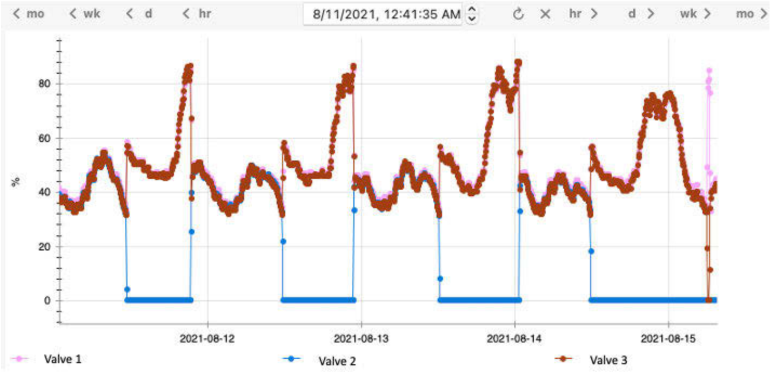


Figure 3. Valve opening degree in percentage for the three valves at VF1.

Since individual flow measurements were available for each of the valves in VF1, two scenarios were considered: Scenario 1, the valve curves were estimated by considering the sum of the individual flow measurement as the facility flow ($N = 3$ in Eq. 3). Scenario 2, each of the three valves curves were calibrated with their respective flow by considering each valve to be a valve facility ($N = 1$ in Eq. 3) for calibration purposes.

3 RESULTS

Starting from the initial guess for the parameters, the least square estimation method successfully converged to optimal parameter sets for the valve facilities. These parameter sets minimized the MSE between the estimated and measured flow timeseries of the facilities. For comparison of flow prediction error before and after calibration, the metric normalized root mean squared error $\mathrm{NRMSE} = \sqrt{MSE}/q_{avg}$ is used. Where q_{avg} is the averaged measured flow for T timesteps for the facility. Figure 4 shows the empirical cumulative frequency distribution function (CDF) of NRMSE for the initial and the calibrated curves for the 77 valve facilities (i.e., each valve is treated as a facility). The median value for the NRMSE for the base curve and the calibrated curve is 0.66 and 0.07, respectively, and about 80% of the valves have NRMSE less than 0.25 after calibration (Figure 4). Valves with post calibration NRMSE greater than 0.25 have all been found to have data issues such as bad valve status information or inconsistency of measured flow with valve opening.

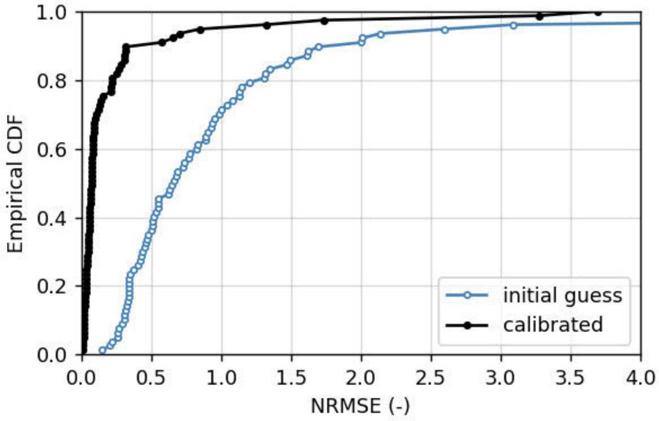


Figure 4. Empirical CDF of NRMSE for the base curve and the calibrated curve

Figure 5 shows the estimated vs. measured flow scatter plot for the initial guess and calibrated valve curves for VF1 for Scenario 1, and most of the estimated flow after calibration lies close to the 45-degree line of perfect correlation. The NRMSE for the initial guess and calibrated valve curves for Scenario 1 are 0.99 and 0.01, respectively. Figure 6 shows the calibration results for Scenario 2. Figures 6a through c show scatter plots similar to Figure 5 for the individual valves. Figure 6d through f shows scatter plots of calculated minor loss coefficients (rearranging Eq. 1 for k) with valve opening degree along with the initial and calibrated valve curves (a similar figure is not possible for Scenario 1). The estimated parameters for the two scenarios are shown in Table 1. The NRMSEs for Scenario 2 for the three valves for the base curve were 2.14, 0.27, and 1.00, and for the calibrated curve were 0.02, 0.02, and 0.01. Estimated flows for both scenarios matched well with measured flows and had low NRMSE (≤ 0.02). The difference in estimated parameters for each valve shown in Table 1 translates to very little difference in the estimated flows for VF1, and the minor loss curves for each valve for the two scenarios are essentially indistinguishable.

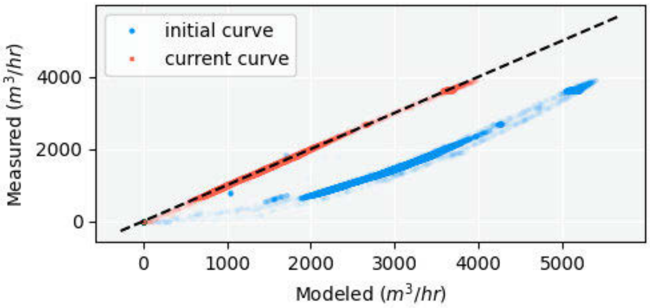
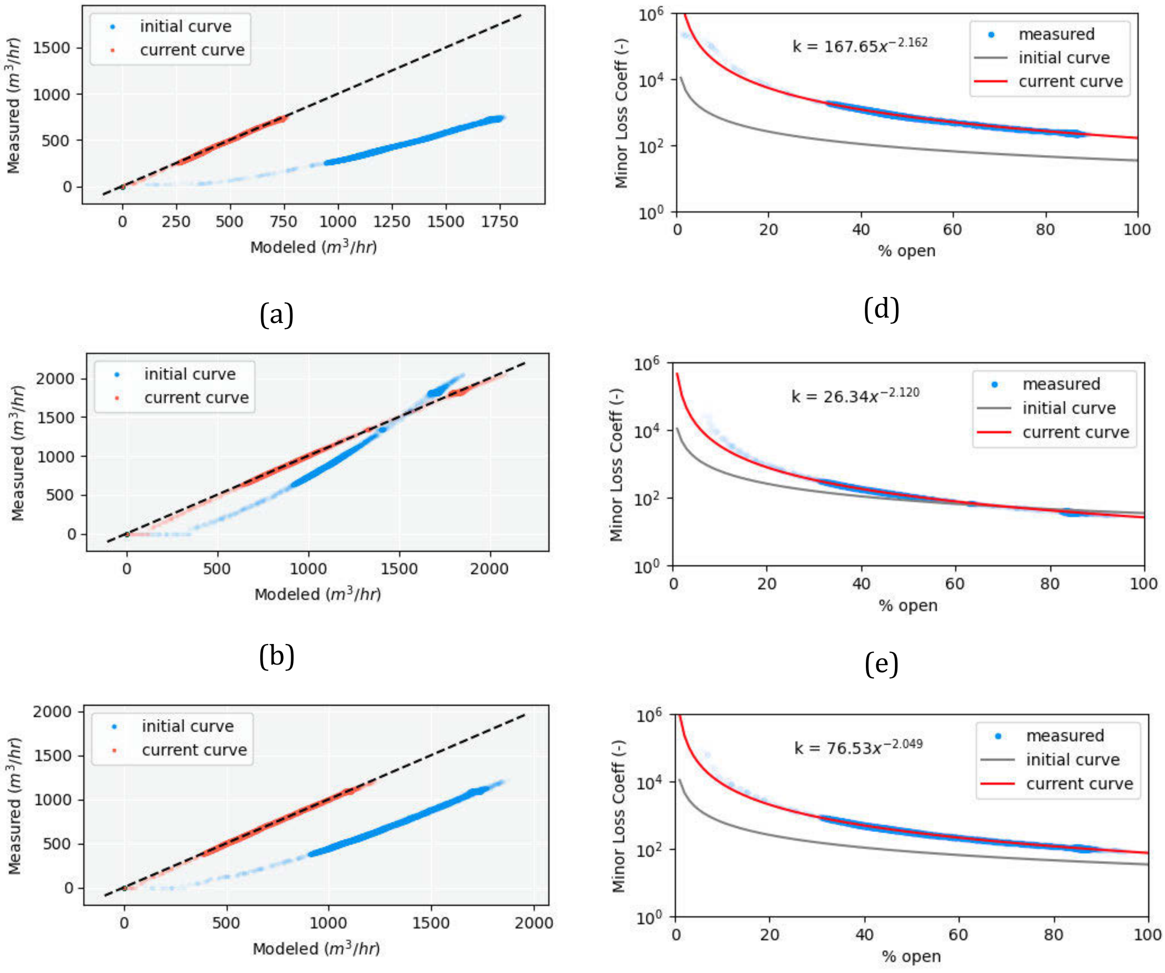


Figure 5. Scatter plot for measured vs. modeled totalized flow for the valve facility VF1. The initial curve refers to the initial guess of the parameters, and the current curve refers to the calibrated parameters of the power-law curve. The black dashed line represents the line of perfect correlation.

Figure 7 shows the calibration results for facility VF2 and VF3, both of which had NRMSE greater than 0.25, and demonstrates some of the issues in the measured data. The facility VF2 (Figure 7a) had a post-calibration NRMSE of 0.31 and is suspected to have inconsistent measurements as a significant part of the measured flow stays constant while the estimated flow varies (e.g., with varying valve opening or pressure). The facility VF3 (Figure 7b) had a post calibration NRMSE of 0.84, and the vertical scatter of measured flow data at zero modeled flow indicates that the valve opening degree timeseries is out of sync with the flow timeseries and therefore the estimated flow is zero when the measured flow is not. Additionally, VF3 flow data is relatively more noisy than flow data at other locations. The resolution of such data issues (e.g., fixing the misalignment of valve opening degree and measured flow) should improve the calibration results at valve facilities with poor NRMSE. Although the calibration process did not achieve the purpose of high accuracy prediction for these valves, it is still beneficial for exposing data streams issues that need to be resolved for the accuracy of the digital twins model and in general.

Table 1. Calibrated parameters for the three valves at VF1

	Scenario 1		Scenario 2	
	a	b	a	b
Valve 1	173.03	-2.221	167.65	-2.162
Valve 2	26.78	-2.113	26.34	-2.120
Valve 3	73.49	-2.031	76.53	-2.049

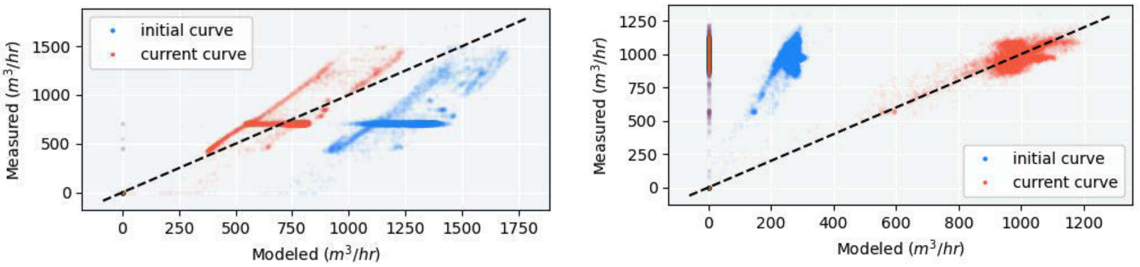


(a) (d)

(b) (e)

(c) (f)

Figure 6 (a) through (c) scatter plot for measured vs. modelled flow measurement through the three valves at VF1. The black dashed 45-degree line represents the line of perfect correlation. (d) through (f) calculated minor loss coefficients, initial and calibrated minor loss curves. The initial and current curve refers to the initial guess and the calibrated parameters of the power-law curve, respectively.



(a) (b)

Figure 7. scatter plot for measured vs. modeled flow measurement through the (a) VF2 and (b) VF3 facilities. The initial and current curve refers to the initial guess and the calibrated parameters of the power-law curve, respectively. The black dashed line represents the line of perfect correlation.

4 CONCLUSION

A framework for valve curve calibration using SCADA data was presented which was used to calibrate the valve curves of flow control valves in a transmission network of a large water utility in the USA. Valve curves were successfully calibrated and resulted in flow prediction with high accuracy obtained by low NRMSE. A power-law model was used to model the valve curves relating valve minor loss coefficients with valve opening measurements obtained from daily operations. While the power-law model worked well, other models can be easily incorporated into this optimization framework. Poor calibration for some of the valves helped uncover data issues, and the calibration can be improved by resolving those underlying data issues.

5 REFERENCES

[1] Copeland, C.; Carter, N.T. Energy - Water Nexus: The Water Sector's Energy Use R43200; Washington, D.C., 2017;

[2] Sanders, K.T.; Webber, M.E. Evaluating the energy consumed for water use in the United States. Environ. Res. Lett. 2012, 7.

[3] Marquez Calvo, O.O.; Quintiliani, C.; Alfonso, L.; Di Cristo, C.; Leopardi, A.; Solomatine, D.; de Marinis, G. Robust optimization of valve management to improve water quality in WDNs under demand uncertainty. Urban Water J. 2018, 15, 943–952.

[4] Conejos Fuertes, P.; Martínez Alzamora, F.; Hervás Carot, M.; Alonso Campos, J.C. Building and exploiting a Digital Twin for the management of drinking water distribution networks. Urban Water J. 2020, 17, 704–713.

[5] Ramos, H.M.; Morani, M.C.; Carravetta, A.; Fecarrotta, O.; Adeyeye, K.; López-Jiménez, P.A.; Pérez-Sánchez, M. New Challenges towards Smart Systems' Efficiency by Digital Twin in Water Distribution Networks. Water 2022, 14, 1304.

[6] Alvisi, S.; Franchini, M.; Marinelli, A. A short-term, pattern-based model for water-demand forecasting. J. Hydroinformatics 2007, 9, 39–50.

[7] Mu, L.; Zheng, F.; Tao, R.; Zhang, Q.; Kapelan, Z. Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model. J. Water Resour. Plan. Manag. 2020, 146, 05020017.

[8] Jamieson, D.G.; Shamir, U.; Martinez, F.; Franchini, M. Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks. J. Hydroinformatics 2007, 9, 3.

[9] Salomons, E.; Housh, M. A Practical Optimization Scheme for Real-Time Operation of Water Distribution Systems. J. Water Resour. Plan. Manag. 2020, 146, 04020016.

[10] Creaco, E.; Campisano, A.; Fontana, N.; Marini, G.; Page, P.R.; Walski, T. Real time control of water distribution networks: A state-of-the-art review. Water Res. 2019, 161, 517–530.

[11] Odan, F.K.; Ribeiro Reis, L.F.; Kapelan, Z. Real-time multiobjective optimization of operation of water supply systems. J. Water Resour. Plan. Manag. 2015, 141.

[12] Rossman, L.A. EPANET Manual. Soc. Stud. Sci. 2000.

[13] Salomons, E.; Shamir, U.; Housh, M. Optimization methodology for estimating pump curves using SCADA data. Water (Switzerland) 2021, 13, 586.

[14] Alvisi, S.; Franchini, M. Calibration and sensitivity analysis of the C-town pipe network model. In Proceedings of the Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference, WDSA 2010; American Society of Civil Engineers, 2012; pp. 1573–1584.

[15] Araujo, A.S.R.; Ribeiro, M.F.M.; Enzveiler, A.; Schenkel, P.; Fernandes, T.R.G.; Partata, W.A.; Irigoyen, M.C.; Llesuy, S.; Belló-Klein, A. Pressure Control for Leakage Minimisation in Water Distribution Systems Management. Water Resour. Manag. 2006 201 2006, 20, 133–149.

[16] Campisano, A.; Modica, C.; Vetrano, L. Calibration of Proportional Controllers for the RTC of Pressures to Reduce Leakage in Water Distribution Networks. J. Water Resour. Plan. Manag. 2012, 138, 377–384.

[17] Ostfeld, A.; Salomons, E.; Ormsbee, L.; Uber, J.G.J.G.; Bros, C.M.C.M.; Kalungi, P.; Burd, R.; Zazula-Coetzee, B.; Belrain, T.; Kang, D.; et al. Battle of the Water Calibration Networks. J. Water Resour. Plan. Manag. 2012, 138, 523–532.

[18] Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 2020 173 2020, 17, 261–272.

[19] Hatchett, S.; Uber, J.; Boccelli, D.; Haxton, T.; Janke, R.; Kramer, A.; Matracia, A.; Panguluri, S. Real-time distribution system modeling: Development, application, and insights. In Proceedings of the Urban Water Management: Challenges and Oppurtunities - 11th International Conference on Computing and Control for the Water Industry, CCWI 2011; 2011; Vol. 3.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14761

MULTI-OBJECTIVE INSIGHTS AND ANALYSIS ON DATA DRIVEN CLASSIFIERS FOR ANOMALY DETECTION IN WATER DISTRIBUTION SYSTEMS

Elizabeth Pauline Carreño-Alvarado[1], Mayra Hernández-Alba[2] and Gilberto Reynoso-Meza[3]

[1]Dpto. de Ingeniería Geomática e Hidráulica, Universidad de Guanajuato, Guanajuato Guanajuato, México.
[2] Universidad de Guanajuato, Guanajuato Guanajuato, México.
[3]Pontifícia Universidade Católica do Paraná (PUCPR) Programa de Pós-graduação em Engenharia de Produção e Sistemas (PPGEPS), Curitiba, Paraná, Brazil.

[1] ep.carreno@ugto.mx, [2] m.hernandezalba@ugto.mx, [3] g.reynosomeza@pucpr.br

Abstract

Machine learning techniques have shown to be a powerful tool for extracting and/or inferring complex patterns from data. In the case of the so-called supervised learning, a given learner representation could learn such patterns using labeled data. For example, a helpful approach is to adjust a learner to detect anomalies: historical data can be used, where those events are identified, to find a pattern to classify new data as an anomaly (true event) or not (false event). In this example, the learner's objective is to act as a binary classifier, where a balance between false negatives (predict a typical operation, when in fact an anomaly exists) and false positives (predict an anomaly, when there is not). This balance is attained via an optimization (learning phase), where the learner representation is adjusted. Multi-objective optimization techniques have a natural way of dealing with such problems. They perform a simultaneous optimization of conflicting objectives. As a result, a set of Pareto-optimal solutions, the Pareto front, is calculated. This idea could be used in the training process of binary classifiers.

Nevertheless, this requires an integral methodology, merging multi-objective optimization and multi-criteria decision making. While it is true that this idea is not new, methodologies and guidelines are still missing to conduct this process. In this work, we move toward the definition of an integrated methodology of multi-objective learning for binary classifiers for anomaly detection. An anomaly detection database for water distribution systems is used for such a purpose. Preliminary results show to be competitive regarding the F1-score to similar approaches.

Keywords
Machine learning, Logistic regression, Multi-objective optimisation, Water distribution systems.

1 INTRODUCTION

Machine learning techniques have shown to be a powerful tool for different kinds of applications [1,2,3,4,5]. Their data-driven approach makes them suitable for finding complex patterns and relationships with enough (and well processed) data. After a supervised learning process, a given machine learning representation can identify anomaly events from regular events [6]. In such cases, they are referred to as virtual or soft sensors: instead of having a physical device, information from other sources is mixed in order to infer, in this case, an anomaly event [7].

Such a learning process is usually performed via optimization with a single objective cost function [8]. Nevertheless, the trade-off between false positives and false negatives is evident for classification purposes. That means it is worthwhile to analyze the trade-off of a given classifier for anomaly detection between triggering a false alarm or letting pass risk situations. Even if both instances are considered equivalent misclassifications for practical purposes, they are not. On the

BY NC SA

one hand, false positives trigger an alarm that requires attention; an excess of false alarms could overwhelm the technical staff. On the other hand, false negatives are undetected situations that could be hazardous to the system's operation. Therefore, there is a trade-off that a learner must achieve through the learning process.

In such instances, where there is a clear trade-off between conflicting objectives, multi-objective optimization could be an interesting tool [9]. Multi-objective optimization deals with conflicting objectives simultaneously. Consequently, a Pareto front is approximated: a set of Pareto optimal solutions. In such a set, the only difference between two solutions is the trade-off that they exhibit among conflicting objectives. Therefore, it is possible to ponder benefits and drawbacks actively when the decision-maker favors one objective over another. Such an idea could bring compelling solutions for machine learning [10,11].

Nevertheless, the multi-objective nature demands a multi-criteria analysis to select a solution from the set to be implemented. While it is true that this idea is not new, methodologies and guidelines are still missing to conduct this process. Therefore, it is necessary to move toward the definition of an integrated methodology of multi-objective learning for binary classifiers.

This work deals with the binary classification problem using logistic regression and explores the advantages of using a multi-objective optimization approach for its training. Additionally, we point out some interesting facts and guidelines for the multi-criteria analysis. The remainder of this paper is as follows: In Section 2, a brief background on multi-objective optimization and supervised machine learning are given. In section 3, the proposal of this exploratory work is presented, whilst in Section 4, the study design is explained. In Section 5, results are commented on and discussed, and finally, conclusions and future works are presented.

2 DESCRIPTION

Next, fundamental ideas on machine learning and multi-objective optimization are given.

2.1 Machine learning, supervised learning, and binary classification

Machine learning refers to computer algorithms that improve themselves automatically through experience in the form of data. Such a learning process could be supervised (requiring inputs and targets), unsupervised (inputs required), or by reinforcement (via interaction with the surroundings) [12]. Supervised learning uses as inputs M instances (observations) with N features (explanatory variables) to train a given learner representation using reliable information of the targets T for each one of the instances. The main goal is to construct a relationship to provide an output (target prediction) for any new instance (with its features). Such training could be oriented for classification or regression. In both cases, several representations such as artificial neural networks [13], support vector machines [14], or decision trees [15] exist, among others.

A binary classification must predict if arriving data belong to class 0 or class 1. Such classes could be identified in the anomaly detection case by answering the following question: Is there an anomaly? A class 0 event is a situation where no anomaly exists; on the opposite, a class 1 event is an anomaly situation. A given learner will enter a training phase using a dataset $[MxN/T]$ to adjust its parameters β via an optimization phase, using some evaluation criteria or cost function.

Usually, this learning process's cost function for optimization is an aggregation function of correct classification and misclassifications. Furthermore, given a parameter vector β for a learner representation, it is usual to evaluate its final performance in the same way, merging true positives ($TP(\beta)$), false negatives ($FN(\beta)$), false positives ($FP(\beta)$) and true negatives ($TN(\beta)$). For example:

$$Error\ rate = \frac{FP + FN}{TP + FP + TN + FN} \tag{1}$$

$$F1\ score = \frac{2 * TPR * PPV}{TPR + PPV} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

Also, it is usual to depict such trade-offs with a confusion matrix (Figure 1). All performance index reveals that a trade-off between *FN* and *FP* is pondered. Both should be minimized, and normally, they are conflicting objectives. Therefore, multi-objective optimization could be an interesting tool to deal with such a situation.

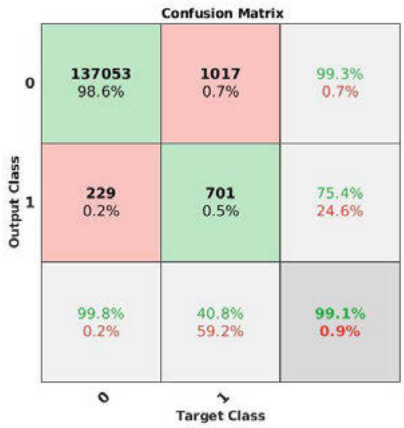


Figure 1. Confusion matrix. It is possible to visualise the performance of a given learner, identifying TN and TP (green boxes) and FP and FN (red boxes).

2.2 Multi-objective optimization

As commented in [16], a multi-objective problem (MOP) with m objectives can be stated as follows:

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = [J_1(\boldsymbol{\theta}), \dots, J_m(\boldsymbol{\theta})] \tag{6}$$

Subject to:

$$\boldsymbol{K}(\boldsymbol{\theta}) \leq \boldsymbol{0} \tag{7}$$

$$\boldsymbol{L}(\boldsymbol{\theta}) \leq \boldsymbol{0} \tag{8}$$

$$\underline{\theta_i} \leq \theta_i \leq \overline{\theta_i}, i = [1, \dots, n] \tag{9}$$

Where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]$ is defined as the decision vector with dim($\boldsymbol{\theta}$) = n; $\boldsymbol{J(\theta)}$ as the objective vector and $\boldsymbol{K(\theta)}$, $\boldsymbol{L(\theta)}$ as the inequality and equality constraint vectors respectively; ; $\underline{\theta_i}$, $\overline{\theta_i}$, are the lower and the upper bounds in the decision space.

It has been noticed that there is not a single solution in MOPs because there is not generally a better solution for all the objectives. Therefore, a set of solutions, the Pareto set $\boldsymbol{\Theta_P}$, is defined. Each solution in the Pareto set defines an objective vector in the Pareto front $\boldsymbol{J_P}$ (See Figure 2). It is important to notice that most of the time, we rely only on the Pareto front and set approximations $\boldsymbol{J_P^*}$, $\boldsymbol{\Theta_P^*}$. P. All the solutions in the Pareto front are a set of Pareto optimal and non-dominated solutions, where:

- Pareto optimality [16]: An objective vector $\boldsymbol{J}(\boldsymbol{\theta}^1)$ is Pareto optimal if there is not another objective vector $\boldsymbol{J}(\boldsymbol{\theta}^2)$ such that $J_i(\boldsymbol{\theta}^2) \leq J_i(\boldsymbol{\theta}^1)$ for all *i* ∈*[1,2,...,m]* and $J_j(\boldsymbol{\theta}^2) < J_j(\boldsymbol{\theta}^1)$ for at least one *j, j* ∈*[1,2,...,m].*
- Dominance [17]: An objective vector $\boldsymbol{J}(\boldsymbol{\theta}^1)$ is dominated by another objective vector $\boldsymbol{J}(\boldsymbol{\theta}^2)$ iff $J_i(\boldsymbol{\theta}^2) \leq J_i(\boldsymbol{\theta}^1)$ for all *i* ∈ *[1,2,...,m]* and $J_j(\boldsymbol{\theta}^1) < J_j(\boldsymbol{\theta}^1)$ for at least one *j, j* ∈ *[1,2,...,m]*. This is denoted as $\boldsymbol{J}(\boldsymbol{\theta}^2) \preccurlyeq \boldsymbol{J}(\boldsymbol{\theta}^1)$.

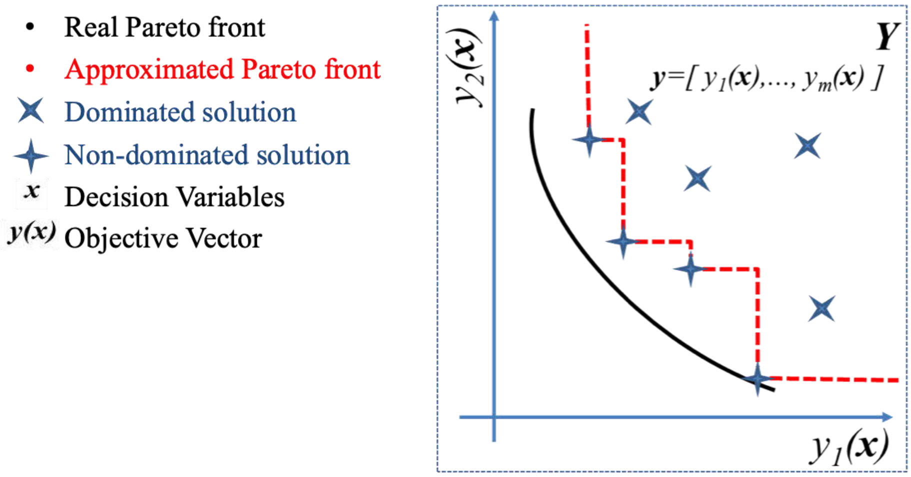


Figure 2. Pareto optimality and dominance concepts for a min-min MOP. Non-dominated solutions approximate (red dotted line) the unknown Pareto front (blsck solid line) in the objective space Y . Remainder solutions are dominated solutions.

The multi-objective optimization approach, from a practical point of view, requires three main steps:

Multi-objective optimization statement: this implies defining the design objectives to optimize, the decision variables, and the parametric model to establish an unequivocal correspondence. The multi-objective optimization (MOO) process corresponds to the optimization process itself. Requires to define an optimization algorithm (with its hyper-parameters), running platform, and hardware requirements/conditions. Multi-criteria Decision Making (MCDM) step: the final process where a solution from the Pareto front approximation should be selected, and its correspondent design alternative from the Pareto set implemented. As a multi-criteria analysis needs to be performed, visualization tools and multi-criteria methods are usually required. The procedure integrating those three steps is named the multi-objective optimization design (MOOD) procedure [18]. Next, it will be presented how to use such a procedure in the multi-objective learning process.

3 TOOLS AND METHODS

Here, the proposal is presented to adjust a binary classifier via a MOOD procedure. It covers the MOP statement, the MOO process, and the MCDM stage.

3.1 Multi-objective problem statement

The parametric model to be used is the logistic regression due to its simplicity and because it is the simplest binary classifier. Future work will focus on different representations or tribes [19]. The logistic regression uses the sigmoid function (equation 10) to compute the probability of a given observation to be 1.

$$h_\theta(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} \tag{10}$$

Where $x = [x_1,\dots,x_N]$ are the N explanatory variables or features of the learner; $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]$ are the regression coefficients adjusted given M observations or instances. Usually, given this set of M instances, the parameter β is adjusted using the loss function of Equation (11):

$$Loss(h_{\boldsymbol{\beta}}(\boldsymbol{x}), y) = \{ \begin{matrix} -\log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) & \text{if } y = 1 \\ -\log(1 - h_{\boldsymbol{\beta}}(\boldsymbol{x})) & \text{if } y = 0 \end{matrix} \tag{11}$$

and the cost function for optimisation of (12):

$$CE(\boldsymbol{\beta}) = -[y \log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) + (1 - y) \log(1 - h_{\boldsymbol{\beta}}(\boldsymbol{x}))]/M \tag{12}$$

Instead of using an aggregation function for FP and FN, such performance will be evaluated simultaneously via multi-objective optimization. Therefore, a multi-objective problem is considered as shown in Equation (13):

$$\min_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) = [FP + CE_m(\boldsymbol{\beta}), FN + CE_m(\boldsymbol{\beta})\ ] \tag{13}$$

With:

$$CE_m(\boldsymbol{\beta}) = -[y \log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) + (1 - y) \log(1 - h_{\boldsymbol{\beta}}(\boldsymbol{x}))]/(M \cdot \log(\epsilon)) \tag{14}$$

Subject to:

$$\underline{\beta_i} \leq \beta_i \leq \overline{\beta_i}, i = [1, \dots, n] \tag{15}$$

3.2 Multi-objective optimization process

For the experiments presented here, the spMODEx algorithm will be used. It is a multi-objective evolutionary algorithm based on Differential Evolution [20, 21], using as diversity mechanism a spherical pruning [22]. The following hyperparameters are used [23]:

Mutation: binomial;

Scaling factor: 0.5;

Crossover rate: 0.9;

Population: 20; Function evaluations: 1e4; Spherical arcs: 20.

Next, a proposal for the decision-making step is placed. It includes an analysis of the Pareto front and the Pareto set approximations.

3.3 Multi-criteria decision making

Next, a proposal for the decision-making step is placed. It includes an analysis of the Pareto front and the Pareto set approximations.

- Evaluate the design alternatives from the Pareto set with the data set for training (J_P^{t*}) and depict them using a 2-dimensional plot.
- Evaluate the design alternatives from the Pareto set with the data set for testing (J^{v*}) and plot it in the same 2-dimensional plot as J_P^{t*}.
- Perform a work scenario comparison [24]. Different from a design concept comparison [25], where different concepts are used to perform the multi-objective process, the work scenario comparison evaluates the performance of a given Pareto set approximation under different conditions.
- Determine the deformation from the J_P^{t*} towards J^{v*}. This evaluation will be helpful in measuring the internal coherence of the learners and their trade-offs. That is if the trade-off ordering among solutions is preserved.
- Define a region of interest (or pertinent region).
- Pick the most suitable solution and evaluate it with the test data.
- Plot Pareto set via parallel coordinates and boxplot. Perform a critical analysis of the most important features.

4 STUDY DESIGN

Providing clean and safe drinking water is crucial for any water supply company [26]. To guarantee such a supply, automatic anomaly detection plays a critical role in drinking water quality monitoring [27]. Recent anomaly detection techniques incorporate tools from the machine learning area [1]. This work uses a real-world data set generated in a research project on drinking water. The data set consists of data from *Thüringer Fernwasserversorgung*, a major German water supplier located in central Germany. This data set has been used for different competitions about anomaly detection for drinking water in major international conferences [26]. The data and additional documentation are available for download [28]. In Table 1, the features used are depicted. More details are provided in [6].

For this example, the following data science methods are implemented:

- Pre-processing: an imputing mechanism has been implemented for instances with missing or not interpretable values (repeat last value).
- Feature engineering: No additional features are included. This means that this problem is being treated as a static problem instead of a dynamic time series.
- Processing: A moving average of 1440 samples was used for detrending.

Table 1. Some Letters and Numbers [Caption, Cambria, 10pt, Italic, centred]

Parameter	**Unit**	**Description**
Time	*datetime*	Time Stamp
WT	°C	Water Temperature
$ClO2_1$, $ClO2_2$	*mg/l*	Chlorine Dioxide (2 values)
pH	*pH*	pH Value
Redox	*mV*	Redox Potential
EC	*µS/cm*	Conductivity
TURB	*NTU*	Turbidity
FR_1, FR_2	m^3/h	Water Flow Rate (2 values)
EVENT	*binary*	Anomaly Label

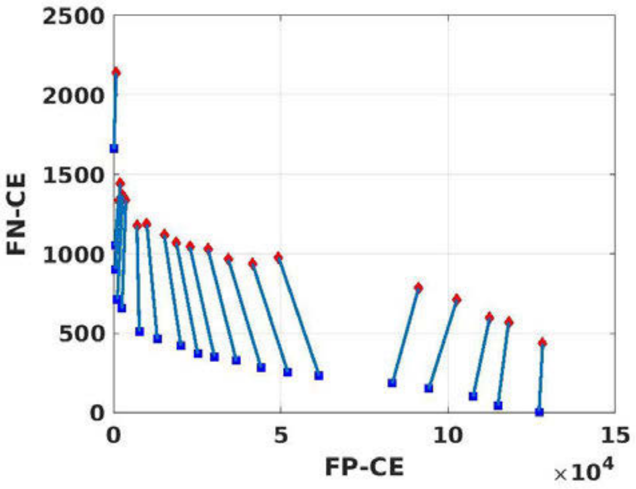
5 MULTI-OBJECTICE LEARNING RESULTS AND DISCUSSION

In Figure 3a, the Pareto front approximation is depicted. With blue squares, the Pareto front approximation J_P^{t*} while the red diamonds represent the deformation of the approximation J^{v*}. When evaluated with the testing data, the translation of a given design alternative β from J_P^{t*} is represented with blue lines. Such an analysis could reveal when an over-fitting occurs: it is expected to keep the trade-off coherence, as well as a similar trade-off or a reasonable mismatch between performance with the training set and the testing set (see [29]). In this case, as practically no blue lines are crossing, the learners in J_P^{t*} keep their internal coherence for decision making. Multiple crosses or gathering to a single region could reveal an over-fitting.

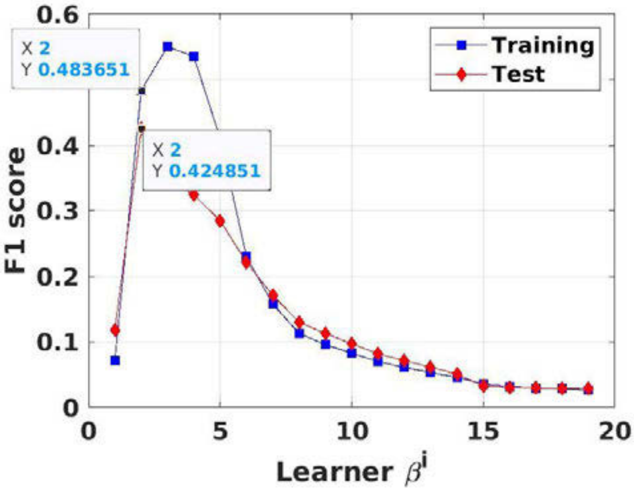
In Figure 3b F1-score is calculated for each of the trained learners. This index is a usual choice to evaluate overall performance whit imbalanced data. The closest to 1, the better the classifier according to this score. As it can be noticed, the learner with the highest F1 score in the training phase is not the better in the validation test. This reveals how important could be the decision-making stage, given that a good performance on the training set does not assure good generalization abilities. The best learner in the validation set is the second one, which F1-score value is relatively close to that achieved in the training set. Given that, this is the learner that is recommended for further implementation. Sixth to nineteenth learners exhibit practically the same performance.

In Figure 3c, the values of the parameter vector β of the approximated Pareto set are depicted using boxplots and parallel coordinates. With such an analysis, it is possible to appreciate the impact of a given feature on the prediction capabilities of the classifier. For example, zero values on β_8 and β_9 (features 8 and 9, FR1 and FR2) indicate that this measure could be potentially omitted. That could mean saving a couple of measures and sensors in a practical sense. On the opposite, β_1 and β_6 (features 1 and 6, WT and EC) seems to be the one that the most impact has in the classification; β_2 and β_7 (features 2 and 7, ClO21 and TURB) seems to be responsible on the trade-off exchange in the set.

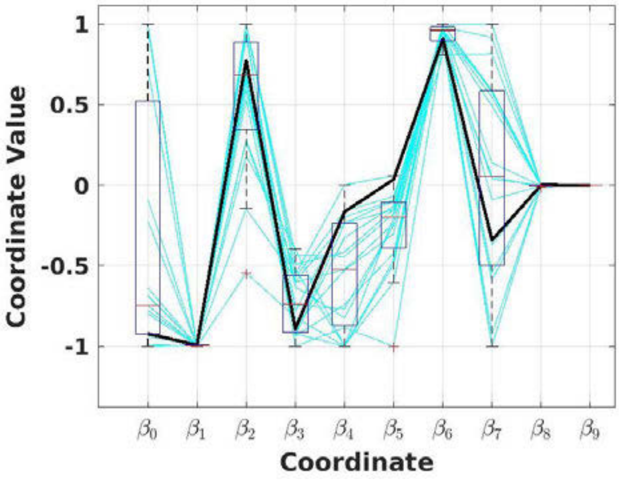
Finally, Figure 4 depicts the confusion matrix of the design alternative (learner) number two, the one selected in figure 3. The F1 score is competitive compared to other classifiers, with similar data science methods (pre-processing, feature engineering, processing, and splitting). For example, ensemble methods using support vector machines and decision trees (third and fourth place) had an F1 score of 0.39 and 0.45, respectively. Interestingly, no additional treatment for the imbalanced data was required by simultaneously considering both classes with the multi-objective approach.



(a) *Pareto front approximation*



(b) *F1-score*



(c) *β distribution in the Pareto set approximation*

Figure 3. Performance visualisation of the approximated Pareto front and set

BY NC SA

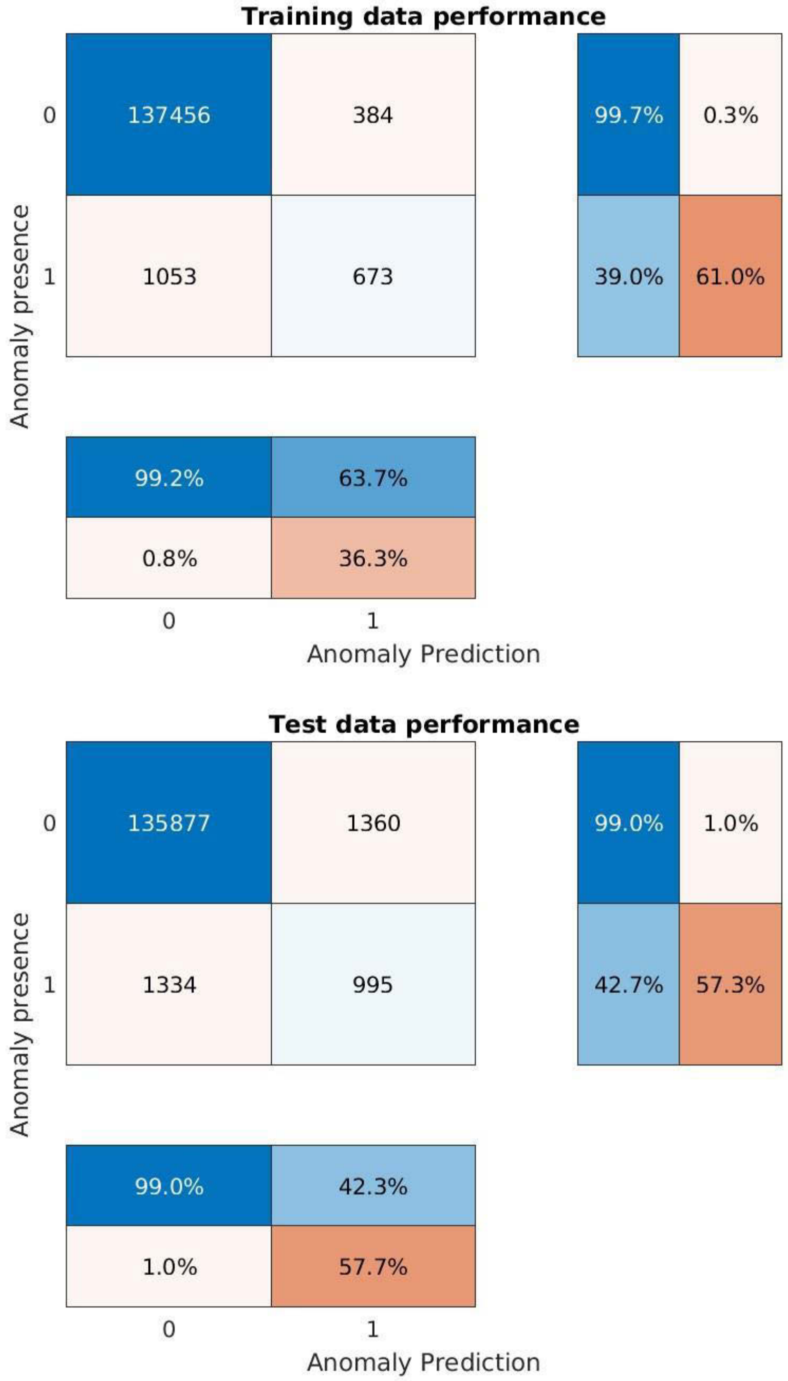


Figure 4. Confusion matrix of the learning vector selected using the F1 criteria.

6 CONCLUSIONS

In this work, it has been proposed a MOOD procedure for multi-objective training in binary classifiers. The multi-objective approach allows considering the *FP* and FN ratio simultaneously. That is, it is possible to train a set of learners with a different trade-off between false positives and false negatives. This could be interesting in the decision-making stage, given that it is possible to select a learner with an affordable cost regarding not-detected threats and/or triggering false

CC BY NC SA

alarms. Furthermore, a proposal for the decision-making stage has been presented to help in the selection of a suitable learner for implementation. An example using logistic regression was presented for anomaly detection in water distribution systems. It includes a Pareto front comparison to validate trade-off coherence with the data set for testing; this is important given that it also allows for verifying the generalization capabilities of the learner. An example of anomaly detection in a water distribution system was presented. Obtained results are competitive with other approaches under the same mechanisms for pre-processing, feature engineering, and processing of data. Future work will concentrate on the design concept comparison of different learners' representations and verifying different options for the MOP statement and indexes for the MCDM step.

7 REFERENCES

[1] Dogo, E.M., Nwulu, N.I., Twala, B., Aigbavboa, C.: A survey of machine learning methods applied to anomaly detection on drinking-water quality data. Urban Water Journal 16(3), 235–248 (2019).

[2] Weichert, D., Link, P., Stoll, A., Ru¨ping, S., Ihlenfeldt, S., Wrobel, S.: A review ofmachine learning for the optimization of production processes. The International Journal of Advanced Manufacturing Technology 104(5), 1889–1902 (2019)

[3] Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y., Ranganath,R.: A review of challenges and opportunities in machine learning for health. AMIA Summits on Translational Science Proceedings 2020, 191 (2020).

[4] Kadlec, P., Gabrys, B.: Soft sensors: where are we and what are the current andfuture challenges? IFAC Proceedings Volumes 42(19), 572–577 (2009).

[5] Alsrehin, N.O., Klaib, A.F., Magableh, A.: Intelligent transportation and controlsystems using data mining and machine learning techniques: A comprehensive study. IEEE Access 7, 49830–49857 (2019).

[6] Ribeiro, V.H.A., Moritz, S., Rehbach, F., Reynoso-Meza, G.: A novel dynamicmulti-criteria ensemble selection mechanism applied to drinking water quality anomaly detection. Science of The Total Environment 749, 142368 (2020).

[7] Souza, F.A., Arau´jo, R., Mendes, J.: Review of soft sensor methods for regressionapplications. Chemometrics and Intelligent Laboratory Systems 152, 69–79 (2016).

[8] Domingos, P.: A few useful things to know about machine learning. Communications of the ACM 55(10), 78–87 (2012).

[9] Meza, G.R., Ferragud, X.B., Saez, J.S., Dura´, J.M.H.: Controller tuning with evolutionary multiobjective optimization (2017).

[10] Ribeiro, V.H.A., Reynoso-Meza, G.: Ensemble learning by means of a multiobjective optimization design approach for dealing with imbalanced data sets. Expert Systems with Applications 147, 113232 (2020).

[12] Mitchell, T.M., et al.: Machine learning. McGraw-hill New York (1997).

[11] Ribeiro, V.H.A., Reynoso-Meza, G.: Feature selection and regularization of interpretable soft sensors using evolutionary multi-objective optimization design procedures. Chemometrics and Intelligent Laboratory Systems 212, 104278 (2021).

[13] Najah, A., El-Shafie, A., Karim, O.A., El-Shafie, A.H.: Application of artificialneural networks for water quality prediction. Neural Computing and Applications 22(1), 187–201 (2013).

[14] Deka, P.C., et al.: Support vector machine applications in the field of hydrology: a review. Applied soft computing 19, 372–386 (2014).

[15] Kotsiantis, S.B.: Decision trees: a recent overview. Artificial Intelligence Review39(4), 261–283 (2013).

[16] Miettinen, K.: Concepts. In: Nonlinear multiobjective optimization, pp. 5–36. Springer (1998).

[17] Coello, C.A.C., Lamont, G.B.: Applications of multi-objective evolutionary algorithms, vol. 1. World Scientific (2004).

[18] Reynoso Meza, G.: Controller tuning by means of evolutionary multiobjective optimization: a holistic multiobjective optimization design procedure. Ph.D. thesis, Editorial Universitat Polit`ecnica de Val`encia (2014).

[19] Domingos, P.: The master algorithm: How the quest for the ultimate learningmachine will remake our world. Basic Books (2015).

[20] Storn, R., Price, K.: Differential evolution–a simple and efficient heuristic for globaloptimization over continuous spaces. Journal of global optimization 11(4), 341–359 (1997).

[21] Pant, M., Zaheer, H., Garcia-Hernandez, L., Abraham, A., et al.: Differential evolution: a review of more than two decades of research. Engineering Applications of Artificial Intelligence 90, 103479 (2020).

[22] Reynoso-Meza, G., Sanchis, J., Blasco, X., Mart´ınez, M.: Design of continuouscontrollers using a multiobjective differential evolution algorithm with spherical pruning. In: European Conference on the Applications of Evolutionary Computation. pp. 532–541. Springer (2010).

[23] Reynoso-Meza, G., Sanchis, J., Blasco, X., Garcia-Nieto, S.: Physical programming for preference driven evolutionary multi-objective optimization. Applied Soft Computing 24, 341–362 (2014).

[24] Reynoso-Meza, G., Carrillo-Ahumada, J., Torralba, L.M.: Pareto front analysis ofcontrol structures in multimodel systems: The work scenario concept introduction.

[25] Mattson, C.A., Messac, A.: Pareto frontier based concept selection under uncertainty, with visualization. Optimization and Engineering 6(1), 85–115 (2005).

[26] Rehbach, F., Chandrasekaran, S., Rebolledo, M., Moritz, S., Bartz-Beielstein, T.: GECCO Challenge 2018: Online Anomaly Detection for Drinking Water Quality (2018).

[27] Banna, M.H., Imran, S., Francisque, A., Najjaran, H., Sadiq, R., Rodriguez, M.,Hoorfar, M.: Online drinking water quality monitoring: review on available and emerging technologies. Critical Reviews in Environmental Science and Technology 44(12), 1370–1421 (2014).

[28] Moritz, S., Rehbach, F., Chandrasekaran, S., Rebolledo, M., Bartz-Beielstein, T.: GECCO Industrial Challenge 2018 Dataset: A water quality dataset for the 'Internet of Things: Online Anomaly Detection for Drinking Water Quality' competition at the Genetic and Evolutionary Computation Conference 2018, Kyoto, Japan. (Feb 2018). https://doi.org/10.5281/zenodo.3884398.

[29] Reynoso-Meza, G., Vignoni, A., Boada, Y., Picu¨, J., Picu¨, E.: Model mismatch in multi-objective optimisation and preservation of trade-off order. IFACPapersOnLine 52(26), 249–254 (2019)

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14766

ANALYSIS OF PDA-BASED WATER DISTRIBUTION SYSTEM SUSPENSION RISK USING STATISTICAL AND MACHINE LEARNING METHOD

Yoojin Oh[1], Haekeum Park[2], Jinseok Hyung[3], Taehyeon Kim[4], Kibum Kim[5] and Jayong Koo[6]

[1,2,3,4,6] University of Seoul, 219 room, 2 Engineering, Seoulsiripdae-ro 163, Dongdaemun-gu, Seoul, South Korea,
[5]Purdue University, HamptonHall of Civil Engineering 1237, United States of America

[1] *cordial5670@uos.ac.kr*, [2] *hkpark@uos.ac.kr*, [3] *gudwlstjr@uos.ac.kr*, [4] *kimth0712@uos.ac.kr*,
[5] *kim3854@purdue.edu*, 6 *jykoo@uos.ac.kr*

Abstract

Recently, there have been frequent cases of water shortages caused by failure to old water pipes. As water is the most basic resource in life and is an indispensable resource in various fields such as industry and agriculture, the scale of the failure is significant in the event of accidents in the water supply pipe network, and in order to minimize the damage of accidents, it is important to prevent accidents through timely maintenance. At this time, the risk map of the water shortage of the water pipeline needs to be prepared for efficient maintenance, and it needs to be managed first from the high-risk area.

To this end, water shortage risk analysis due to pipe failure was performed in this study. Risk analysis is one of the ways in which water pipes are evaluated and decisions on investment plans, such as replacement or repair, can be supported. The risk is generally calculated by multiplying PoF (Probability of failure) with the resulting direct and indirect effects of CoF (Consequence of failure). In this study, PoF was derived as the failure of an individual water pipe was set as the probability of failure caused by corrosion, and in order for it to be predicted, MLP (Multi-layer perceptron) and XGBoot were developed as a data-based machine learning model. In addition, it was analyzed by setting the amount of water (supply shortage) that CoF could not be supplied due to failure, considering that the failure to the water pipe was directly linked to water shortage. In order to analyze the supply shortage at this time, the mathematical analysis of PDA (Pressure driven analysis) was performed.

Finally, the developed methodology was applied to the cities of the Republic of Korea, and the risks were analyzed by calculating the PoF and CoF of individual water pipes, and the GIS technique was used to create the risk map.

The results of this study can be more accurate in predicting the condition of water pipes, which can be helpful when water utilities establish maintenance plans.

Keywords
Water Distribution System, Risk analysis, Risk map, Logistic regression, XGBoost regression, Pressure Driven Analysis(PDA).

1 INTRODUCTION

Recently, cases of water shortage caused by failure to old water pipes are occurring frequently. As water is the most basic resource in life and is an indispensable resource in various fields such as industry and agriculture, the scale of the failure is significant in the event of accidents in the water supply pipe network, and in order to minimize the damage of accidents, it is important to prevent accidents through timely maintenance.

In order to predict the condition of water pipes, the study of corrosion depth prediction models and failure rate prediction models using physical and statistical techniques has continued, and with the recent application of machine learning techniques, more accurate condition predictions have been attempted[1],[2],[3]. In previous studies, however, it has been difficult to develop physical models such as corrosion prediction models due to the limitation of data collection, and there is a limit that the correlation between corrosion and failure was not considered.

Risk, on the other hand, can be represented by the multiplication of PoF (Probability of failure) and CoF (Consequence of failure) of the direct and indirect effects resulting from it. So far, the risk analysis has only analyzed the possibility of failure using data on the history of failure to the pipe and considering that the accident in the water pipe is directly linked to the water shortage of the consumer, it is reasonable for the extent of the failure to be considered along with the probability of failure depending on the condition of the pipe. At this time, the scale of the failure can be calculated as supply shortage, which is the amount of water that is not supplied due to failure, and pressure driven analysis based on water pressure should be performed so that errors in interpretation such as the occurrence of negative pressure do not occur.

In this study, in order to evaluate risk based on the condition of the pipe, a corrosion depth prediction model according to the laying environment was developed to predict the depth of corrosion of individual water pipes, and the failure probability prediction model was developed by utilizing the results of the prediction and the history of failure in the past. In addition, the risk of individual pipelines was analyzed by calculating the supply shortage in case the pie is damaged through PDA mathematical analysis. Finally, the water shortage risk map of the water pipeline in the area subject to the study was drawn up.

2 METHODS

2.1 Study area

In this study, the research on the actual condition of the old water pipe was used to secure data on the influence factors of corrosion, and B city, where the history of failure to the water pipe is managed by GIS data, was selected as the target area of the study. Most of the water pipes laid in B city are DICP (94.1%), which were laid between 1968 and 2020. In addition, the date on the research on the actual condition is data surveyed between 2020 and 2021, which records the corrosion depth of the water pipeline and the environment in which it was laid, and a record of failure that occurred between 1999 and 2020.

2.2 Multi-Layer Perceptron

The artificial neural network model is a methodology that mimics the human brain, divided into single-layer perceptron and multi-layer perceptron according to the number of layers in which the neural network is constructed. Multi-layer perceptron model consists of input layer, hidden layer, and output layer, and between the nodes that make up each layer, weight is learned, and weighted sum is calculated. The output value of the weighted sum is calculated by activation function, and the output value is received as input value again, and the output value is calculated repeatedly through activation function.

$$z_j^k = w_{j,0}^{k-1} x_0^{k-1} + w_{j,1}^{k-1} x_2^{k-1} + \cdots w_{j,n}^{k-1} x_n^{k-1} \tag{1}$$

where, z_j^k : j-th node value of the k-th layer

$w_{j,n}^{k-1}$: Weight at which the n-th node of the (k-1)-layer and the j-th node of the k-layer are connected

x_n^{k-1} : n-th node value of the (k-1)-layer

2.3 XGBoost Algorithm

XGBoost stands for Extreme Gradient Boosting, which is one of the ensemble methods where Weak learner is combined sequentially to improve errors. Decision tree is the most widely used as the basic learner of XGBoost and is considered to show high performance as an overfitting-proof model due to the use of loss function and regularization term. When XGBoost's objective function is mathematically expressed, it is as the following (2) formula.

$$obj(\theta) = l(\theta) + \Omega(\theta) \tag{2}$$

where, $obj(\theta)$: Objective function

$l(\theta)$: loss function (Mean Square Error, MSE in case of regression, and loss of logistic in case of classification)

$\Omega(\theta)$: regularization term

2.4 Pressure Driven Analysis

In order to analyze the scale of the failure caused by the failure to the water pipe, a simulation through the interpretation of pipe network should be carried out, and it is necessary to conduct a PDA (Pressure driven analysis), not a DDA (Demand driven analysis), in which unrealistic results such as the occurrence of negative pressure can be produced under certain conditions. In particular, the consumer's supply capacity is determined by HOR (Head-Outflow Relationship), and the water supply is impossible if the consumer's water pressure becomes less than the minimum water pressure, and if the head of the consumer's pressure falls between the minimum and marginal water pressure, the supply is supplied according to the HOR, providing only a fraction of the required demand supply.

$$\begin{cases} q_j^{avi} = q_j^{req}, \quad if\ H_j^{avi} \geq H_j^{des} \\ 0 < q_j^{avi} = q_j^{req} \left(\dfrac{H_j^{avi} - H_j^{min}}{H_j^{des} - H_j^{min}}\right) < q_j^{req}, \quad if\ H_j^{min} < H_j^{avi} < H_j^{des} \\ q_j^{avi} = 0, \quad if\ H_j^{avi} \leq H_j^{min} \end{cases} \tag{3}$$

3 RESULTS

3.1 Probability of Failure

The corrosion depth prediction model was developed to predict the probability of failure caused by corrosion, and the result was used as an input variable for the failure probability prediction model. Parameters used for the development of the corrosion depth prediction model and the failure probability prediction model are as follows in Table 1. The input variables of each model were selected for items of high correlation and items mainly used in prior studies as a result of the analysis of the correlation with the output variables. Hyperparameters were set through Bayesian Optimization techniques.

Table 1. Parameters of corrosion depth prediction model and failure probability prediction model

Classification	Input		Output
Corrosion depth prediction model	Pipe data	Pipe age, pipe material, pipe diameter	Corrosion depth
	Laying environment	Laying depth, sulphide concentration, soil pH, oxidation-reduction potential, moisture content	
	Operating conditions	Water quality corrosiveness (LI), maximum water pressure	
Failure probability prediction model	Pipe data	Pipe age, pipe material, pipe diameter, joint type, pipe thickness (predicted value)	Failure probability
	Laying environment	Laying depth, road shape	
	Operating conditions	Average flow rate, maximum water pressure	
	Failure history	The number of failure in the last 5 years, the number of failure in the last 10 years	

Training set and Test set are divided into 8:2, and the result of the development of the corrosion depth prediction model and the failure probability prediction model is the same as Table 2 and Table 3. In the evaluation of the model, the coefficient of determination (R^2), and RMSE were used.

Table 2. Results of the development of corrosion depth prediction model and failure probability prediction model (Regression)

Classification	Method	Training set		Test set	
		R^2	RMSE	R^2	RMSE
Corrosion depth prediction model	MLP	0.88	1.6	0.68	1.92
	XGBoost	0.90	1.45	0.84	1.41
Failure probability prediction model	MLP	0.97	2.23	0.42	8.64
	XGBoost	0.89	0.05	0.69	0.08

In the case of the corrosion depth prediction model, the Test set's R^2 was 0.84 and the RMSE was 1.41, which was confirmed to predict the depth of corrosion in the water pipe more accurately than the MLP model. In the failure probability prediction model, the XGBoost model more accurately predicted than the MLP model, with Test set's R^2 was 0.65 and RMSE was 0.19 for the probability of failure to the water pipe. The developed model was applied to B city to derive the probability of failure to individual water pipes.

3.2 Consequence of Failure

In this study, EPANET 2.2 toolkit was used to calculate the supply shortage caused by failure to water pipes, which simulated the failure of individual pipelines. First of all, the supply shortage was calculated in consideration of the time of recovery from the time the failure occurred based on the maximum supply of water in normal condition where no failure occurred. At this time, the results of the survey of the time of recovery by pipe diameter in the past were utilized as the time of recovery. The analysis showed that the supply shortage in case of failure was greater in the main drain than in the pipe-end area, and that the blocking made it possible to supply water despite failure to the pipe, leaving some areas where water shortage does not occur.

3.3 Risk analysis and risk map

Finally, the risk of individual pipelines multiplied by the impact (supply shortages) resulting from failure probability and failure was calculated. The result of visualizing the zone-specific risk of B city is as follows in Figure 1. The characteristics of pipe laid in the highest and lowest risk areas have been analyzed, and it has been confirmed that pipe network's blocking is constructed even if the probability of failure to pipe is high, resulting in a lower score if water shortage does not occur.

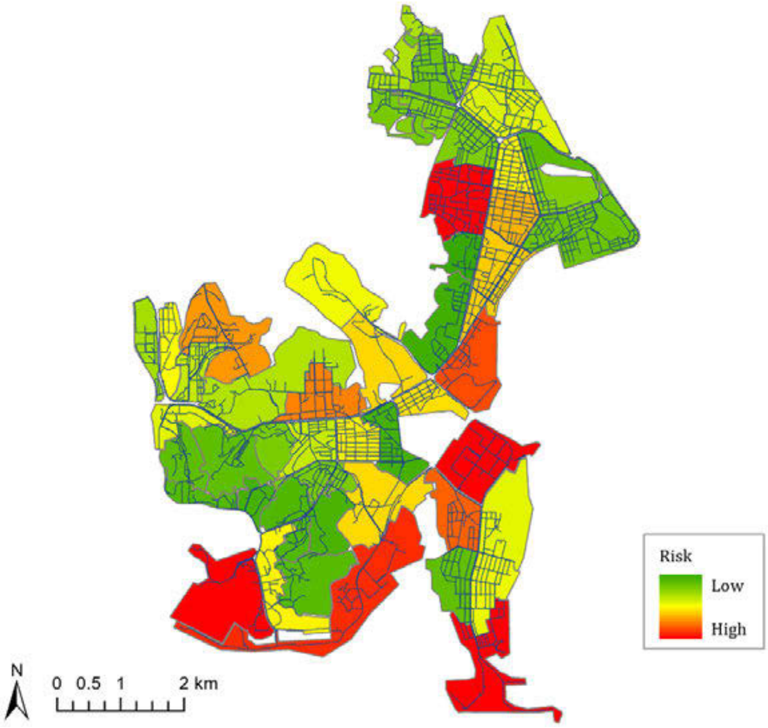


Figure 1. Risk map of B city

4 CONCLUSIONS

In this study, corrosion depth prediction model and failure probability prediction model were developed as data from research on the actual condition in water pipes and failure history data were used to analyze the risk of water shortage caused by corrosion of water pipes. As a result of the development of the model, the performance of the model using XGBoost was shown to be superior.

In addition, the PDA mathematical analysis technique was used to quantify the impact of failure on water pipes, resulting in a shortage of supply in the event of failure to individual pipelines. In addition, the PDA mathematical analysis technique was used to quantify the impact of the failure

CC BY NC SA

of the water pipeline, calculating the supply shortage in the event of failure to individual pipelines. Finally, as the probability of failure of individual pipelines was multiplied by the impact of failure (supply shortage), the risk of water shortage in the individual pipeline was calculated, and the zone-specific risk was visualized, and risk map was expressed.

This study can be helpful when water utilities identify the status of pipes and determine areas where maintenance is needed first.

5 ACKNOWLEDGEMENTS

The Korea Ministry of Environment supported this work titled as "Project for developing innovative drinking water and wastewater technologies (2020002700016)".

6 REFERENCES

[1] Uri Shamir and Charles D.D. Howard, "An Analytic Approach to Scheduling Pipe Replacement," Journal AWWA, vol. 71, no. 5, 1979, pp. 248-258.

[2] Robert M. Clark, "Water Distribution Systems : A Spatial and Cost Evaluation," Journal of Water Resources Planning and Management, vol. 108, no. 3, 1982, pp. 243-256.

[3] Xudong Fan, Xiaowei Wang, Xijin Zhang, P.E.F. ASCE Xiong (Bill) Yu, "Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors," Reliability Engineering and System Safety, vol. 219, no. 108185, 2022, pp. 1-15.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14768

EFFICIENCY ANALYSIS AND EVALUATION MODEL DEVELOPMENT OF WATER DISTRIBUTION SYSTEM REBUILDING PROJECT USING DEA METHOD

Haekeum Park[1], Kibum Kim[2], Jinseok Hyung[3], Taehyeon Kim[4] and Jayong Koo[5]

[1,3,4,5] University of Seoul, 219 room, 2 Engineering, Seoulsiripdae-ro 163, Dongdaemun-gu, Seoul, South Korea,
[2]Purdue University, HamptonHall of Civil Engineering 1237, United States of America

[1] *hkpark@uos.ac.kr*, [2] *kim3854@purdue.edu*, [3] *gudwlstjr@uos.ac.kr*, [4] *kimth0712@uos.ac.kr*,
[5] *jykoo@uos.ac.kr*

Abstract

The water supply facilities of Korea have achieved a rapid growth, along with the other social infrastructures consisting a city, due to the phenomenon of urbanization according to economic development. However, as an adverse effect of rapid growth, the quantity of aged water supply pipes are increasing rapidly, while pipe aging causes water suspension accidents and the scale of such accidents is getting large. Accordingly, the Ministry of Environment has been promoting the local water distribution system rebuilding project since 2017 to build a block system, pipe network maintenance work, leak management work, etc. Related projects are being promoted. In this study, analyzed the efficiency of the improvement of the revenue water ratio according to the DEA(Data Envelopment Analysis) method-based water distribution system rebuilding project, and develop an evaluation model for the water distribution system rebuilding project using the efficiency analysis results.

In this study, DEA analysis was performed by selecting 15 local governments that showed cost-effectiveness among 20 local governments that were carrying out the performance evaluation for the 2017 rebuilding project. In this study, DEA analysis was performed considering the number of DMUs secured three times or more than the sum of input and output factors suggested in Banker et al.(1984)'s previous study.

The input indicators of this study are the values derived by dividing the block system construction cost, pipe network maintenance cost, leak management cost invested for the local water distribution system rebuilding project from 2017 to 2021 by the lengths of the project target water distribution pipes. As the output indicators, the rate of increase in the revenue water ratio, which was changed according to the input project cost, was applied.

In addition, a project cost evaluation model was developed for the improvement of the revenue water ratio by using the optimal efficiency point derived through the efficiency analysis with the effect of improving the revenue water ratio as a calculation index. For model development, multiple regression analysis was performed according to the step-selection method that included only independent variables with high explanatory power that had an influence on the dependent variable in the regression model to develop the model.

As such, in this study, the efficiency analysis of local governments performing the local water distribution system rebuilding project was performed, and a project cost evaluation model was developed to improve the revenue water ratio. It is judged that the methodology and development model used in this study can be utilized in the analysis and evaluation of the project efficiency of the additional target area of the local water distribution system rebuilding project in the future.

Keywords
Water distribution system, Efficiency analysis, DEA, efficiency model.

1 INTRODUCTION

Since the 1970s, the water supply facilities in Korea have achieved rapid growth with rapid economic growth. However, due to the rapid economic growth, the deterioration of water supply facilities has been accelerating recently. It brings water suspension and water quality accidents and lowers the quality of water supply services. According to this current situation, the need to manage old water supply facilities that lower water supply services are continuously raised.

The Ministry of Environment in Korea has been carrying out a Water distribution system rebuilding project since 2017 to maintain water supply facilities. The performance guarantee of the local government where the project was carried out with priority is implemented. This study analyses the efficiency by calculating the optimal size of input and output by separately applying input-oriented and output-oriented models under BCC conditions based on DEA (Data Envelopment Analysis).

2 METHODS

2.1 Study area

In this study, the efficiency evaluation of the project implementation of the local governments performing the Water distribution system rebuilding project to be carried out was targeted at 15 county-level local governments that had conducted the Water distribution system rebuilding project from 2017 to 2018. The efficiency of project execution was evaluated using the cost of block system construction, pipe maintenance, leakage management as the annual increase rate of revenue water ratio and the composition of project costs.

2.2 Data Envelopment Analysis(DEA)

Data Envelopment Analysis(DEA) was proposed based on the study of Charnes, Cooper, and Rhodes in 1978. It is a model to evaluate the efficiency of a production organization that produces many output indicators, and a production effect by putting a plurality of input indicators. In DEA literature, the production organization is called the decision making unit (DMU). It is essential to secure the appropriate number of DMUs to perform efficiency analysis. Banker et al. (1984) verified that the number of DMUs to be evaluated should be three times greater than the sum of input and output indicators to be discriminatory. Therefore, in this study, efficiency analysis was performed by securing the number of 12 or more DMUs, which is three times the sum of three input indicators and one output indicator. Table 1 shows the input and output variables for the efficiency analysis based on the BCC model.

Table 1. Input and output variables for efficiency analysis

Input	Output	DMU
Block system building cost within 2017-2021(100 KRW/number of block)	Increase in revenue water ratio within 2017-2021 years (%)	15 water utilities in Republic of Korea
Pipe management cost within 2017-2021(100 KRW/km)		
Leakage management cost within 2017-2021(100 KRW/km)		

Unlike other existing efficiency measurement methods, data envelopment analysis is a nonparametric method, not presuming a specific function form and estimating parameters in advance. It uses data between empirical input elements and the output of the evaluation target based on the linear planning method. After deriving the efficiency frontier change, the efficiency is measured by how far away the evaluation targets are from the efficiency frontier.

To evaluate efficiency, there are various methods of synthesizing input elements to construct a measure of productivity in the production process. In DEA analysis, the CCR model of Charnes, Cooper, and Rhodes (1978) and the BCC model of Banker, Charnes, and Cooper(1984) are used at most. These two models are distinguished by input or output oriented depending on whether they focus on input or output elements. To evaluate the performance of a general business, the input-oriented model is selected as the output effect versus the input cost is considered as a major decision-making factor.

The efficiency using the DEA model is divided into three categories: technical efficiency using the CCR model(assuming a constant return to scale), pure technical efficiency using the BCC model(assuming variable return to scale), and scale efficiency(SE, scale efficiency). Scale efficiency is calculated as in the following equation (1).

$$SE = \frac{\theta^*_{CCR(CRS)}}{\theta^*_{BCC(VRS)}} \tag{1}$$

$\theta^*$is the efficiency value of the CCR model and the BCC model of each specific DMU

There are local governments that have not been optimally implemented in the best state due to various realistic constraints such as local and operational conditions of water supply operators. As the unit of the cost in the project increased, constant returns to scale assumes that the revenue water ratiowill increase steadily. Variable return to scale assumes that the revenue water ratiowill increase significantly or small compared to the cost unit, and it assumes a return on scale. In the case of water flow rate, it is reasonable to follow the scale profit variable model since the technology to improve from 80% to 90% is larger than the technology required to improve from 30% to 40%. Therefore, this research performs efficiency analysis by applying a BCC model that applies calculations on the premise of VRS according to size.

2.3 Multiple regression model

There are many ways to estimate the regression equation. Typical examples are simultaneous input and step input methods. The simultaneous input method analyzes all independent variables considered by the researcher at once. The step selection method includes only variables that influence dependent variables when other variables exist in the regression equation. This study aims to perform regression analysis by the step selection method.

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \tag{2}$$

The above equation is called the regressive equation, and $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ is called the regression parameter. The purpose of the regression analysis is to estimate the value of the regression parameter from the relationship between variables obtained from the sample and to verify the hypothesis.

3 RESULTS

3.1 Efficiency analysis

Figure 1 below indicates the results of the analysis of technology efficiency according to CRS(Constant Returns to Scale), net technology efficiency according to VRS(Variable Returns to Scale), and SE(Scale of Efficiency) according to the input orientation of the BCC model.

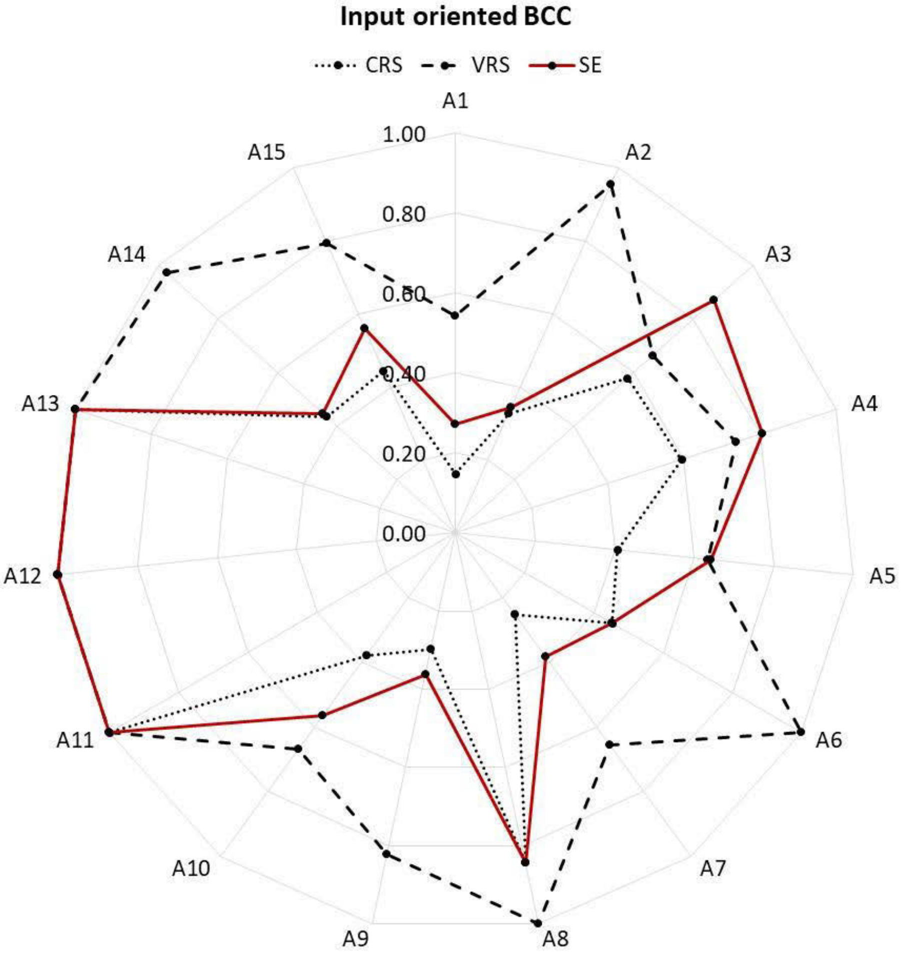


Figure 1. Result of DEA of water distribution system rebuilding project by Input oriented BCC

The efficiency of the input-oriented was analyzed for 15 local governments that implemented the water distribution system rebuilding project. The result shows that in the case of A13, A12, and A11 water suppliers, the technical efficiency under the insomnia of the scale and the net technical efficiency under the scale profit variable were derived as '1'. The efficiency of the scale is the maximum value of '1', which is the highest, and the performance of the project is efficiently achieved because of the effect of improving the revenue water ratio.

On the other hand, the efficiency of the project was low in the case of A9, A8, A2 and A14 water suppliers. Their effect of increasing the revenue water ratio was low compared to the cost invested and relative efficiency was higher than 0.95 under the VRS. However, the efficiency of the scale was small because the efficiency was low under the CRS. These results mean that if the current input cost is increased, it is difficult to see an increase in the output effect as much as the increased cost. Also, it indicates that the water service provider is not operating the optimal water supply facility maintenance.

Furthermore, in the case of A3 and A4 suppliers, the efficiency of each scale was relatively high, despite the low technical efficiency and net technical efficiency, which means that the result of the calculation is low compared to the cost. They may be considered that the cost of relatively input is excessive compared to other water suppliers.

The relative efficiency of local governments performing the water distribution system rebuilding project is shown in Table 2 based on the input-oriented and output-oriented analysis of the BCC model, which is a DEA analysis model.

Table 2. Result of efficiency analysis of water distribution system rebuilding project

Decision Making Units	Technical efficiency in CCR model	Pure technical efficiency in BCC model				Scale efficiency		Reference counts	
		Input-oriented		Output-oriented		Input	Output		
		Score	Score	Score	Score			Input	Output
A1	0.1477	0.5439	IRS	0.2666	DRS	0.2716	0.554	0	0
A2	0.3273	0.9534	IRS	0.3392	DRS	0.3433	0.9649	0	0
A3	0.578	0.6634	IRS	0.8595	DRS	0.8713	0.6725	0	0
A4	0.5939	0.7366	IRS	0.7748	DRS	0.8063	0.7665	0	0
A5	0.4078	0.6352	IRS	0.6122	DRS	0.642	0.6661	0	0
A6	0.4525	1	IRS	1	IRS	0.4525	0.4525	2	0
A7	0.2516	0.6552	IRS	0.3751	DRS	0.384	0.6708	0	0
A8	0.8435	1	IRS	1	IRS	0.8435	0.8435	5	0
A9	0.2967	0.8217	IRS	0.3529	DRS	0.3611	0.8407	0	0
A10	0.3781	0.6693	IRS	0.4829	DRS	0.5649	0.783	0	0
A11	1	1	CRS	1	CRS	1	1	0	10
A12	1	1	CRS	1	CRS	1	1	6	2
A13	1	1	CRS	1	CRS	1	1	10	10
A14	0.4343	0.9718	IRS	0.4458	DRS	0.4469	0.9742	0	0
A15	0.4437	0.7935	IRS	0.5492	DRS	0.5592	0.8079	0	0

This study divided and analyzed the input-oriented model and the output-oriented model of the BCC model, which are the efficiency measurement model based on the VRS. As a result of analyzing the efficiency measurement model, three of the 15 water suppliesrs(A11, A12, A13) had an efficiency of '1' in both the input-oriented model and the output-oriented model. The average efficiency of input orientation was 0.5437, and the average efficiency of output orientation was 0.6705.

Furthermore, the input-oriented model of the variable return to scale that the BCC model presupposes was analyzed as 12 IRSs, 0 DRSs, and 3 CRSs. In the output-oriented model, 2 IRSs, 10 DRSs, and 2 CRSs were analyzed. The DRS increased in the analysis of the output-oriented model. This means that the excess shortage of the revenue water ratio, the output factor, is greater than the excess input, and the output effect is not derived compared to the investment. In this case, advanced technology is required to improve the revenue water ratio in addition to the cost investment. The operation and maintenance of the water distribution system should be upgraded to increase the calculation even if the same cost is invested.

3.2 Appropriate project cost calculation model to improve the revenue water ratio

The appropriate project cost calculation model (Y_A) was developed using the optimal efficiency point derived through efficiency analysis using three input indicators and the effect of raising the revenue water ratio as a calculation index. Developing an appropriate project cost estimation model to improve the revenue water ratio, an optimization model was established by performing multiple regression analysis according to the step selection method including only the independent variables with high explanatory power that have an influence on the dependent variables in the regression equation. The developed model is shown in Equation (3).

$$y = 0.035 + 0.020x_1 \tag{3}$$

x_1 is leakage management cost(100KRW/km)

The analysis result of model variance showed that the significance probability of the developed regression model was 0.012, which is smaller than 0.05. It indicates that the developed regression model is statistically significant. The following Table 3 is the result of the significance analysis of the regression coefficient of the developed regression model.

The coefficient values of the model constant and independent variables were statistically significant, as below 0.05. The R-value of the model was 0.631. In the final regression model, the basic unit of leakage management cost was included as an independent variable.

Table 3. Result of the ANOVA of the cost evaluation model to improve the revenue water ratio

Division	B	Std.error	t	p
Constant	0.035	0.008	4.395	0.001
x_1	0.020	0.007	2.936	0.012

4 CONCLUSIONS

In this study, the efficiency of the input-oriented model and output-oriented model was compared and analyzed to evaluate the business efficiency of the water distribution system rebuilding project which has been implemented since 2017 through large-scale government support in Korea. The result of the analysis shows that the overload of the output element is larger than the overload of the project cost being invested in the current water supply facility scale. It suggests that it is necessary to apply more advanced water supply maintenance technology rather than facility expansion and improvement-oriented maintenance.

Through data enveloping analysis, the optimal efficiency points to improve the revenue water ratio of local governments performing water distribution system rebuilding project was derived. Based on the optimal efficiency point, this study developed a model for calculating the appropriate business cost to improve the revenue water ratio. As a result of testing the developed regression model, the correlation of the model was 0.631, indicating good explanatory power. If the number of related samples increases due to the increase of local governments in the performance guarantee stage of the water distribution system rebuilding project in the future, it has higher explanatory power than the model developed in this study. It leads to the development of a model for calculating the appropriate business expense considering various business details such as block system building cost and pipe management cost. It is expected that this study will be applicable in the implementation of the projects related to the improvement of the revenue water ratio through government support projects such as the water distribution system rebuilding project.

5 ACKNOWLEDGEMENTS

The Korea Ministry of Environment supported this work titled as “Project for developing innovative drinking water and wastewater technologies (2020002700016)”.

6 REFERENCES

[1] Banker, R. D., Charnes, A. and Cooper, W. W. (1984), Some models for estimating technical and scale efficiencies in data envelopment analysis, Management Science, Vol.30, pp.1078-1092.

[2] Charnes, A., Cooper, W. W. and Rhodes, W. (1978), Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through, Management Science, Vol.27, No.6, pp.668-697.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14770

MASS BALANCE CALIBRATION OF WATER DISTRIBUTION NETWORKS: APPLICATION TO A REAL CASE STUDY

Orazio Giustolisi[1], Francesco Ciliberti[2], Daniele Laucelli[3] and Luigi Berardi[4]

[1,3]Department of Civil, Environmental, Land, Building Engineering and Chemistry, Technical University of Bari, Bari (Italy)
[2,4] Department of Engineering and Geology, University "G. D'Annunzio" of Chieti-Pescara, Pescara (Italy)

[1] *orazio.giustolisi@poliba.it,* [2] *francesco.ciliberti@unich.it,* [3] *danielebiagio.laucelli@poliba.i,*
[4] *luigi.berardi@unich.it*

Abstract

Calibration of Water Distribution Networks (WDN) hydraulic models is mandatory to effectively support their analysis and management. As such, it should not be intended as the mere matching between model outputs and filed data but, rather, as a tool to understand WDN behavior under several different hydraulic functioning scenarios. The latest advanced hydraulic models encompass a phenomenological representation of the physical behavior of based on pressure-driven analysis. From such perspective, the problem variables include the parameters of the leakage model, the observable demand patterns and the average values changing day by day, as well as pipe hydraulic resistances, especially of the main water paths feeding the system. According to global mass-balance, the total water inflow recorded at source point (i.e., tanks, reservoirs, pumps) can be used to separate the stochastic component of water outflow, i.e., consumers' demands, from the deterministic component, i.e., pressure-dependent leakages at single pipe level.

This work demonstrates the innovative mass-balance paradigm applied for the calibration of few sub-networks of a large real WDN, which uses measurements collected in five characteristic days including summer, winter, holidays and working days. The resulting model allows a more robust prediction of the real system physical behavior and provide a reliable basis to support several design and management activities.

Keywords
WDN hydraulic model, calibration, leakages, demand patterns.

1 INTRODUCTION

Hydraulic models of Water Distribution Networks (WDN) are known to provide a methodological tool to support various asset planning and management tasks, as confirmed by the evolution of WDN hydraulic models of the last decades. Indeed, WDN models were initially conceived to support the design of new WDN, allowing to verify adequate pressure at delivery points, i.e., nodes of the model, under assigned demand and hydraulic capacity, i.e., pipe diameters of the sizing solution. Accordingly, the hydraulic modelling software (e.g., EPANET2, [1]) were built upon the assumption of fixed water outflows, i.e., demand-driven analysis (DDA). These models allowed WDN verification under both normal and abnormal water demand scenarios (e.g., firefighting).

Later, the increase in WDN deterioration and consequent leakage rates, motivated the introduction pressure-driven analysis (PDA) [2] and the conceptual basis through the definition of pressure-demand relationship for all types of indoor and outdoor components of demands, consistently with the Torricelli law [3]. The mostly adopted pressure-dependent function of customer demand resorted to Wagner's model [4], while pressure-dependent relationship for leakages was introduced in the eighties (e.g. [5]) and integrated into the PDA solving algorithm more than one decade ago [6] allowing the representation of leakages at pipe level depending on average pressure and deterioration.

Nowadays, WDN hydraulic models are used to support various asset management tasks including leakage reduction, pressure control, optimal design of District metering Areas (DMA), or sustainable pump operation. In order to support such tasks, the WDN hydraulic models have to provide a phenomenological representation of WDN, meaning that they should capture the emerging hydraulic behaviour of the network, which can be changed by proper management actions, e.g., installing pressure control devices (e.g. [7][8]), modifying water paths through district metering areas (e.g. [9] [10]) or implementing pipeline rehabilitation. This motivated the development of advanced WDN hydraulic models, like that used in this work, aimed at including many elements that are of direct relevance to support asset management and planning.

Such evolution in WDN modelling purposes also motivated some change in model calibration approach. Since the earliest contributions in this area, WDN model calibration was reported as the process of determining model parameters that will yield a reasonable match between measured and predicted pressures and flows in the network [11]. Accordingly, the calibration of DDA models took pipe hydraulic resistance as the main calibration variable aimed at maximizing the matching between simulated and measured pressure at some nodes (e.g. [12][13][14]). This framework emphasized the role of energy balance equations along pipes, while the mass balance, i.e., matching observed and simulated pipe flows over time, was used to adjust fixed nodal demands only, possibly including leakages patterns as additional fixed outflows. Recent works (e.g. [15][16][17]) included the calibration of demand patters, although they did not consider pressure-dependent background leakages.

The introduction of PDA motivated novel approaches to WDN model calibration since additional parameters of pressure-demand model components were introduced, beside pipe hydraulic resistance and water requests. In mode details, accounting for the leakage model in PDA requires the calibration of parameters representing leakage propensity. [18] reported that defining multiple groups of pipes that have different leakage propensity might provide accurate representation of leakages in WDN models. Nonetheless, as leakages depend on pressure and affect the distribution of water flows through the network, this asks for detailed monitoring, which is not always available, especially in those WDN which are in the process of installing meters and must use the hydraulic model to drive the designing of monitoring systems.

A novel approach to calibrate advanced WDN models was reported in [19] and inspired this contribution. It was based on the observation that time series of WDN water inflow equals the superimposition of two types of components of WDN water outflows, which can be referred to as *stochastic* and *deterministic*. *Stochastic* demand components refer to water requests from consumers, which changes day by day based on socio-economic or seasonal factors, i.e., working days, holidays, winter or summer. *Deterministic* demand components are directly related to asset conditions and pressure regime, i.e., pressure-dependent leakages. The variability of water requests of consumers over time (i.e., demand patterns) affect the WDN hydraulic status including pressures that, in turn, determine leakages. In PDA advanced models such variables should match energy and mass balance equations.

Using such paradigm [20] demonstrated that water flow measurements provide more effective information than pressure measurements for determining the distribution of leakages through the network. They also reported a strategy to assign different leakage model parameters to different pipes based on statistical model of failure propensity, consistently with the observation that the rate of pipe breaks increases with leakage rate (e.g., 21).

This contribution reports the application of WDN model calibration approach exploiting mass-balance, i.e. separating *stochastic* and *deterministic* demand components, on a real large WDN. The calibration of such WDN was part of a real procedure for planning district monitoring areas (DMA) integrated with leakage control actions, therefore it had to account for real constraints and uncertainties on available information and had to provide robust results to support next design activities. The implementation on such a real context allows to demonstrate the importance of some aspects which are usually neglected in classical WDN model calibration procedures: the existence of different WDN hydraulic status in different days; the change of leakage outflow over different days in consequence of change of pressure regimes; the actual observability of consumers' demand patters based on available measurements; the handling of unreducible uncertainties on actual status of some devices consistently with the intended model use.

2 REMARKS ON ENHANCED WDN HYDRAULIC MODELLING

This section briefly recalls the main features of enhanced WDN hydraulic model that is used herein to support the analysis and planning of works [22]. This is functional to identify the main variables to be calibrated and discuss some aspects of direct relevance in the real case applications.

It is worth to recall that the main assumption of WDN hydraulic modelling for planning purposes is the steady-state simulation conditions. This means that in each time interval ΔT, over an operating cycle (e.g., 24 hours), unsteady conditions are completely neglected and are not described in model equations. The choice of ΔT, e.g., 15, 30 or 60 minutes, depends on the peculiar modelling purpose (e.g. [23]), the variability of water demand patterns and the size of the network. During each ΔT water demands are assumed as stationary ([24][3]), i.e., with constant mean, and the filling/emptying process of water tanks is assumed as slow ([24]). On these premises, the mass balance and energy balance equations behind the enhanced hydraulic model can be written in matrix form (Giustolisi, 2020):

$$\begin{cases} \mathbf{A}_{pp}(t)\mathbf{Q}_p(t)+\mathbf{A}_{pn}\mathbf{H}_n(t) & = & -\mathbf{A}_{p0}\mathbf{H}_0(t) \\ \mathbf{A}_{np}\mathbf{Q}_p(t)-\dfrac{\mathbf{V}_n(t,\mathbf{H}_n(t))}{\Delta T} & = & \mathbf{0}_n \end{cases} \tag{1}$$

where p and n relate to the number of pipes and nodes (unknown heads), while the subscript "0" refers to the number of reservoirs (known heads). $\mathbf{Q}_p$, $\mathbf{H}_n$ and $\mathbf{H}_0$ are the column vectors of pipe flow rates, nodal heads and known nodal heads, respectively. $\mathbf{A}_{pn}$, $\mathbf{A}_{np}$ and $\mathbf{A}_{p0}$ are topological incidence sub-matrices of the general topological matrix, link-node, of the network. $\mathbf{A}_{pp}(t)\mathbf{Q}_p(t)$ is the column vector of pipe head losses containing terms related to internal head losses of pump systems, if any, minor head losses and evenly distributed head losses, i.e., depending on pipe hydraulic resistance parameters.

$\mathbf{V}_n$ is the column vector of volume outflows during a time interval ΔT lumped at nodes. The following equations reports the demand components that are included in $\mathbf{V}_n$, all dependent on pressure status ($\mathbf{H}_n$) and varying at each time step (t).

$$\begin{aligned} \mathbf{V}_n(t,\mathbf{H}_n(t)) = {} & \mathbf{V}_n^{cons}(t,\mathbf{H}_n(t))+\mathbf{V}_n^{priv-tank}(t,\mathbf{H}_n(t))+ \\ & +\mathbf{V}_n^{orif}(t,\mathbf{H}_n(t))+\mathbf{V}_n^{tank}(t,\mathbf{H}_n(t))+\mathbf{V}_n^{leak}(t,\mathbf{H}_n(t)) \end{aligned} \tag{2}$$

$\mathbf{V}_n^{cons}(t, \mathbf{H}_n(t))$ is the water demand supplied to consumers directly connected to the water network which are computed using Wagner's model in pressure-deficient conditions, otherwise it equals the *stochastic* water requests, changing day by day. $\mathbf{V}_n^{priv-tank}(t, \mathbf{H}_n(t))$ represents the volume of water feeding private storage tanks in ΔT, which are commonly installed in many areas worldwide, including that of the real WDN [26]. $\mathbf{V}_n^{orif}(t, \mathbf{H}_n(t))$ is the water volume from uncontrolled free orifices, e.g. hydrants.

$\mathbf{V}_n^{leak}(t, \mathbf{H}_n(t))$ represents leakage volume, which is computed at single pipe level. It is worth to remark that $\mathbf{V}_n^{leak}$ in a WDN hydraulic model used to support asset management refers to background leakages and outflows from undetected/unreported pipe bursts. Such leakages are

BY NC SA

known also as *volumetric leakages* since they entail the major volumes of lost water at annual scale. Although the used model is able to include model formulations according to the FAVAD concept (e.g. [27]) and Gemanopoulos [5], the latter is report herein:

$$\frac{V_k^{leak}(t)}{\Delta T}=\begin{cases} L_k \beta_k P_{k,avg}^{\alpha_k}(t) & if \;\; P_{k,avg}(t)>0 \\ 0 & if \;\; P_{k,avg}(t)\leq 0 \end{cases} \tag{2}$$

V_k represents distributed volumetric leakage volume in t along the kth pipe; L_k and $P_{k,avg}$ are the length and average pressure of that pipe; β_k is a pipe deterioration parameter and α_k is an exponent which can be assumed as 1, as reported in [23]. Half of V_k is then assigned to both nodes of the pipe in order to solve hydraulic equations in model (1). Assuming deterioration parameter at pipe level is useful for calibration purposes since it allows assigning values of β_k based on prior insight on leakage propensity (es in [20]).

3 MASS BALANCE APPROACH TO WDN MODEL CALIBRATION

The following figure depicts the separation of water demand components which is behind the mass-balance approach to WDN model calibration. Without impairing the general validity of the description, let's assume a WDN fed by one water "source" only (i.e., reservoir, tank or pump). The leftmost diagram shows in red the pattern of inlet water volumes recorded from the water "source", which can be thought as the superposition of *deterministic* volumetric leakages, depending on deterioration of pipes and pressure, and *stochastic* consumers' water demands.

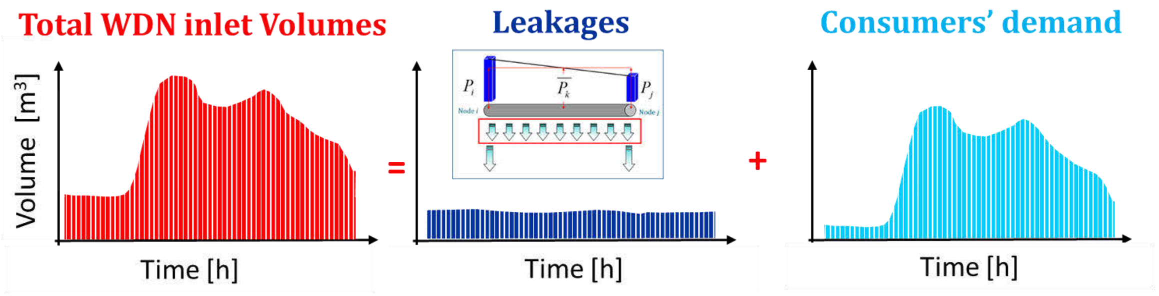


Figure 1. Separation of water demand components in mass-balance approach for WDN model calibration

The two components of water outflows should be in equilibrium with both global water inflow and the distribution of pressures in the network, which depends on flows circulating along pipes including both water supplied to consumers and leakages.

This approach gives rise to some remarks which are relevant while calibrating WDN hydraulic models to support asset management and planning actions.

The variables of the WDN model calibration problem are the deterioration parameters β_k, the hydraulic resistances of pipes, and the patterns of consumers' demand. The first two parameters deal with asset features, i.e., propensity to leak and deterioration; therefore, they can be assumed as invariant over multiple days.

About pipe deterioration parameters β_k, [20] used a methodology to assign different propensity to leak to different pipes. It was based on the using failure propensity models that can be borrowed from literature or developed *ad hoc* using data-modelling or statistical analyses on past pipe failure data. Such methodology was also adopted for the real case described herein.

The detailed calibration of hydraulic resistances at single pipe level requires flow and pressure measurements that usually are not available in real contexts. In addition, the vast majority of pipes in urban WDN represents the looped portion of the system. For these pipes multiple alternative water paths exists whose identification is not easy and make non unique the assessment of hydraulic resistance of these pipes base on limited number of flow/pressure measurements. Conversely, there are few pipes, e.g., those feeding the main distribution network from water sources. Such pipe can be longer than others (up do few kilometres) and carry larger discharges; therefore, assessing their hydraulic resistance values is of primary importance since they strongly affect the pressure regime in the entire system.

The calibration procedure adopted herein allows the assessment of hydraulic resistances of both types of pipes, starting from some prior values got from technical literature. The user can decide to calibrate the most relevant pipeline, groups of similar pipes or none of theme, assuming that prior values are reliable enough.

About the patterns of consumers' demand, it is known to change day by day and affect pressure regime, thus resulting into change of *volumetric leakages* over different days.

This confirms the findings of [12][28]: in order to avoid a mere error-compensation process and improve the robustness of model calibration, it is mandatory to consider a set of several independent steady-state observations of flow and pressure as well as the extended period simulation (EPS) of the network. Using sets of multiple operating cycles corresponding to working day and holidays in different seasons or even abnormal consumption days (e.g., New Year's Eve) allows more robust estimation of asset variables which are supposed to be invariant. This, in turn, increases the accuracy of identified consumers' demand pattern for each day.

The mass-balance approach also suggests that a unique demand pattern can be identified for each observable portion of the network (e.g., DMA), i.e. where al least inflow measurements are available at the boundaries of that network portion allowing mass balance estimate. Assuming demand patterns defined a priori (e.g., households, industrial, business, etc.) might introduce a strong bias for the identification of consistent demand patterns, deterioration parameters and hydraulic resistances. Prior information on demand patterns in some DMAs or in single consumers should be substantiated by detailed flow measurements.

It can be noted that the mass-balance calibration approach is quite flexible to exploit as much information as possible, while preserving the physical consistency of results to allow a robust phenomenological representation of WDN hydraulics. The main information required comes from inflow data and average consumption data (e.g., based on billing database). Information on pressure regime is useful to improve estimate of leakage model parameters. Nonetheless, it was demonstrated that changes in β_k, consistently with global mass balance, only change the distribution of leakages among pipes with negligible changes in pressure values. Pressure and flow information are both useful to calibrate the hydraulic resistance for the main feeding pipelines, while values borrowed from technical literature of pipe hydraulic resistances usually allow a reliable description of the distribution (looped) part of the system.

4 WDN MODEL CALIBRATION IN A REAL WDN

The mass-balance approach discussed above was applied on a real WDN in a large city in southern Italy. It was part of a larger procedure for asset management activities aiming at DMA design and pressure control for leakage reduction, with possible improvements of system hydraulic functioning.

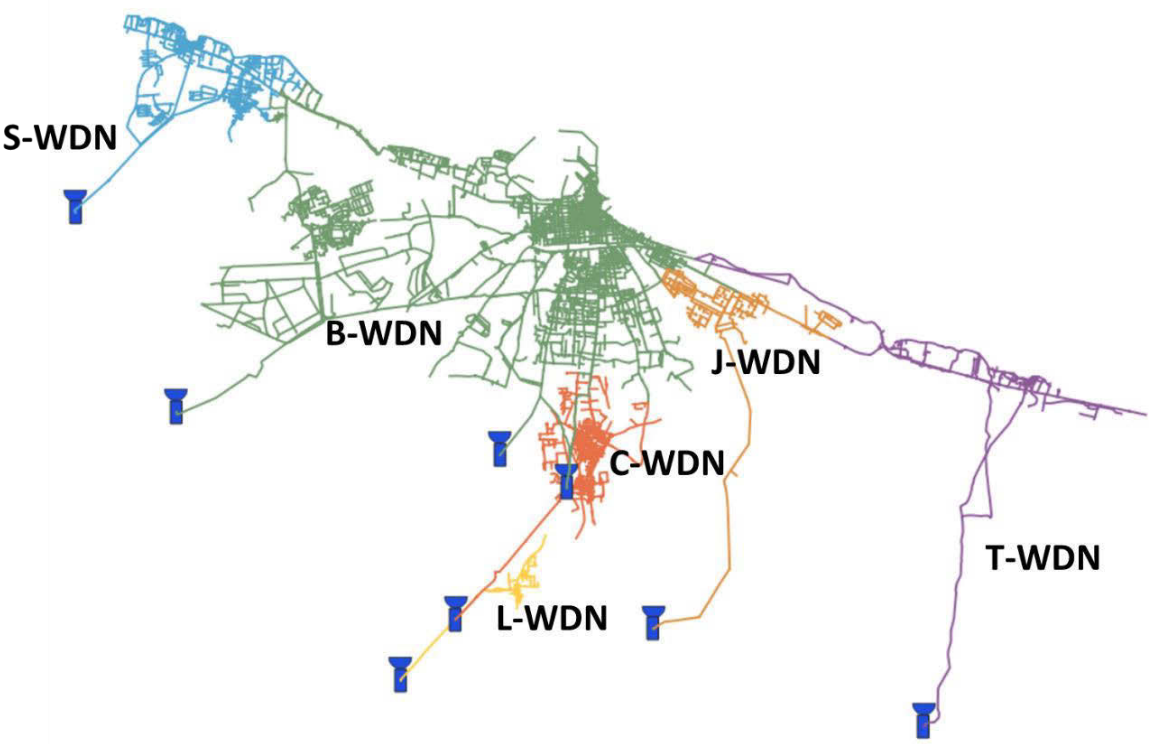


Figure 2. Plot of the real WDN and sub-networks

Table 1. Data of WD hydraulic model of each subnetwork

Sub-network	# Reservoirs	Length [km]	#Pipes	#Nodes	Control valves
T-WDN	1	52.71	350	325	1
J-WDN	1	43.27	358	332	2
L-WDN	1	12.87	169	144	-
S-WDN	1	86.59	1344	1181	-
C-WDN	1	75.82	1144	1007	1
B-WDN	3	424.33	5691	4783	3
Total	8	695.59	9056	7772	7

The WDN is composed of six subnetworks fed by eight reservoirs; the total length of pipeline is about 700km, with 7 pressure reduction valves. The system also includes about 80 partially closed valves (i.e., introducing minor losses) and about 150 closed gates. Figure 2 shows with different colours the six subnetworks and Table 1 summarizes key data for each WDN hydraulic model.

The mass-balance approach was used for calibrating the hydraulic model of each subsystem separately, using inflow data monitored from each reservoir and some pressure data in few monitoring points. It is worth noting to emphasize that mass balance paradigm requires synchronous data (i.e., data collected at all available meters in the same days), and average consumption of the same year. In this case, water consumption and monitored data at flow/pressure meters, along with information on settings of control valves referred to 2019.

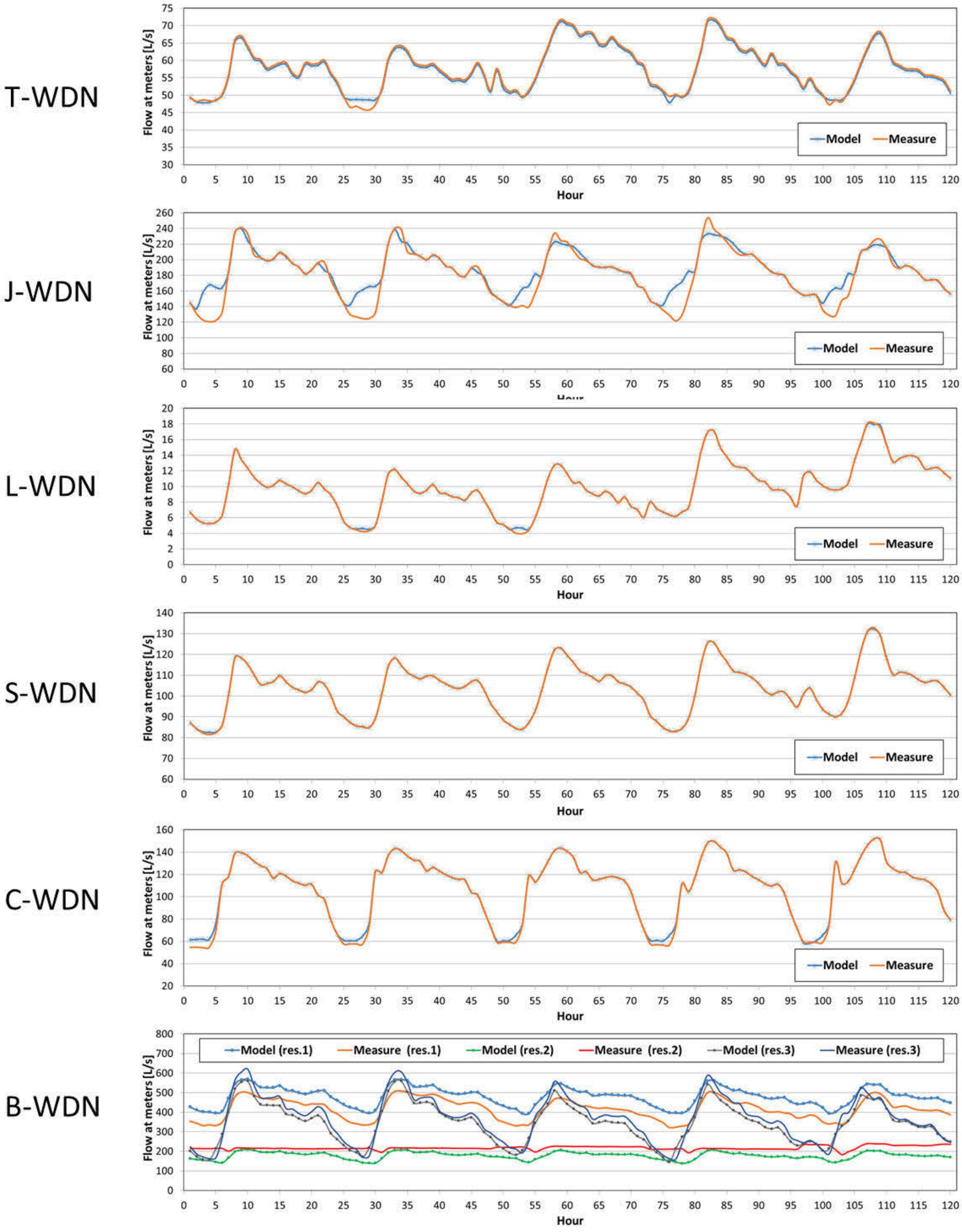


Figure 3. Calibration results: measured vs. model inflow data

For each subsystems the calibration was performed on five days (i.e., 120 timesteps of 1 hour each) representing week and weekend days during summer and winter, also including the 1st of January, representing an unusual pattern of consumptions. As each subnetwork was composed by one monitored "district", only one demand pattern can be observed and identified in each model.

Figure 3 reports the results of calibration in each model, comparing measured data and model output at flow meters. It can be noted that in all sublots data follows the same trend and in the majority of time step data almost overlap. This does not happen in few time intervals, mainly during night, because of some unknown/unreported controls of hydraulic devices or unpredictable water usage. In the case of the biggest B-WDN, which is fed by three reservoirs, there is a sort of compensation effect between underestimates and overestimates of model flows compared with observed flows. This mismatching is mainly related to unreducible uncertainties about the actual status of many valves not yes reported in the model as declared by the water utility. This is a typical scenario in real contexts where a detailed monitoring system in still at early stages. Nonetheless, the calibration procedure pursued the global mass-balance at every simulation time step.

Figure 4 shows the unique demand patterns identified in each WDN model, while Figure 5 shows the patterns of linear leakage indicator (m^3/km/day), M1a in Italian regulation [29]. It can be noted that identified demand patterns changes over different hours of the day in consequence of peculiar socio-economic factors. These patterns determine changes in pressure and leakages, as demonstrated by the variability of the linear leakage indicator M1a.

Also noteworthy is that some spikes are identified in demand patterns for J-WDN. This happens since leakages changes in consequence of pressure control (by time) performed by a pressure reduction valve ns mass-balance modify the demand pattern to match the total inflow recorded. The unexpected behaviour identified in J-WDN, in conjunction with the exceptionally high linear leakage indicator (i.e., about 138 m^3/km/day) unveiled that either all pipes of this small network are highly deteriorated or there are major unreported leaks in few pipes that need to be detected. From such perspective, the mass-balance calibration helped in *reverse engineering*, i.e., suggesting some feasible explanations to unexpected results which can improve the knowledge of the real system.

Figure 6 explicitly reports the separation of total inlet volume into leakage and consumers' demand components across the five days. It can be noted that leakage volume changes across different days, although its variation is lower than changes in consumers' demand. This happens because such systems are usually oversized, resulting into small changes in average daily pressure.

Finally, next figures show the results of model calibration in terms of average pressure at nodes (Figure 7) and linear leakage index M1a (Figure 8) at single piper level. Such plots confirms that pipes under the same average pressure might have different propensity to leak, due to many factors including age, number of connections to private properties and diameters, as represented by the calibrated parameter β_k. This information is of primary importance to drive leakage reduction actions like, for example, selecting pipes for rehabilitation works without using the information of pipe age only.

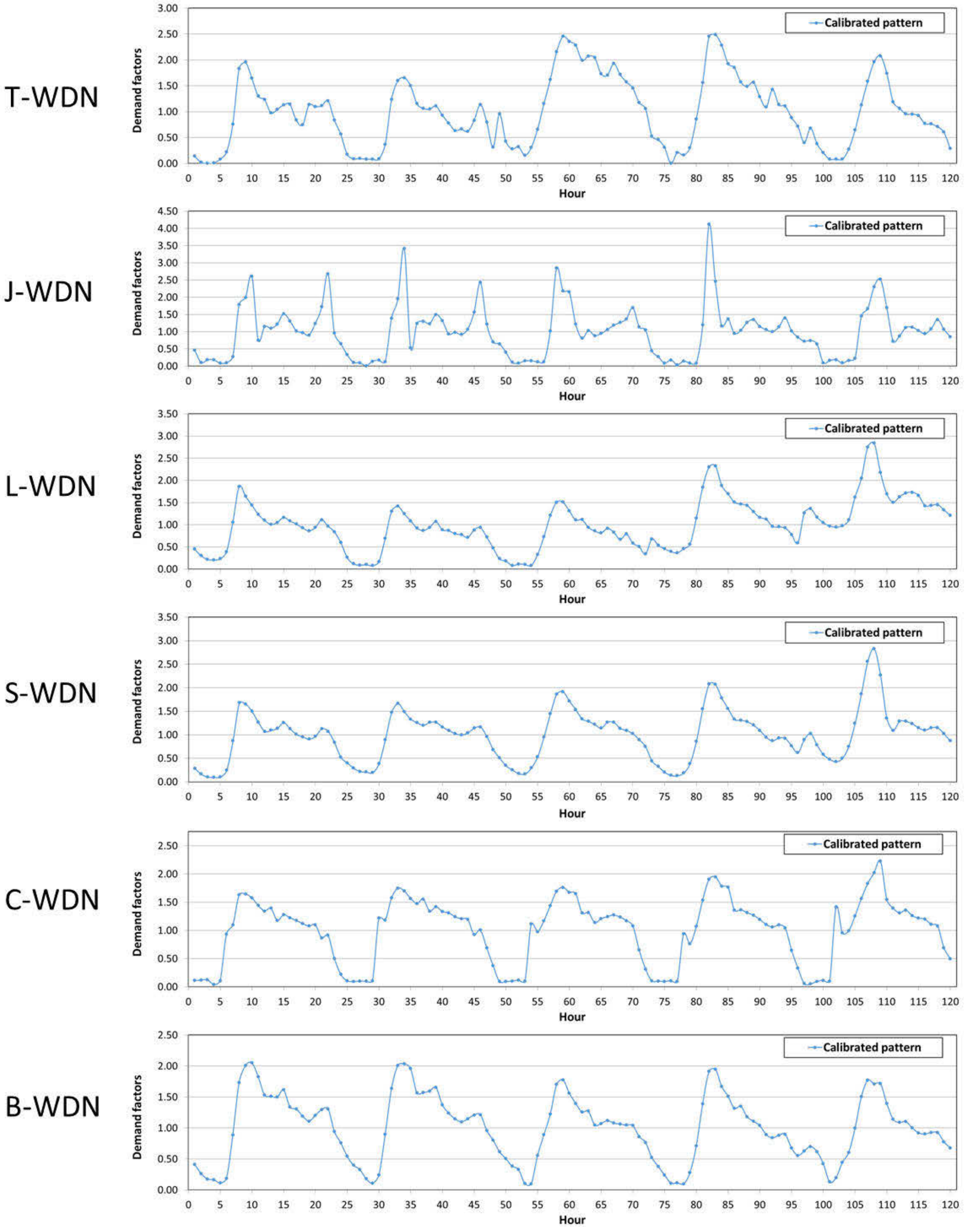


Figure 4. Calibration results: identified consumers' demand patterns

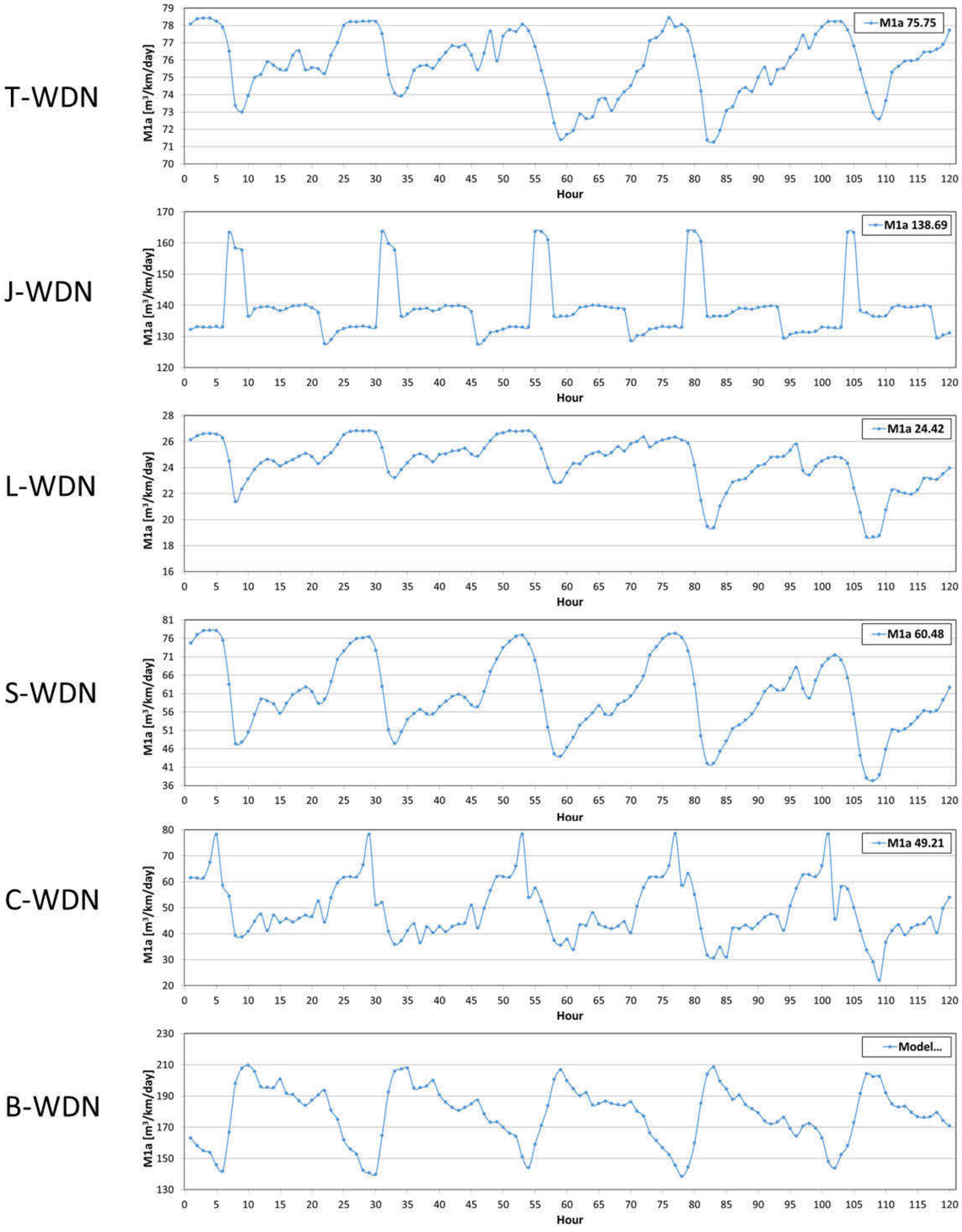


Figure 5. Calibration results: values of linear leakage indicator M1a [m³/km/day]

cc BY NC SA

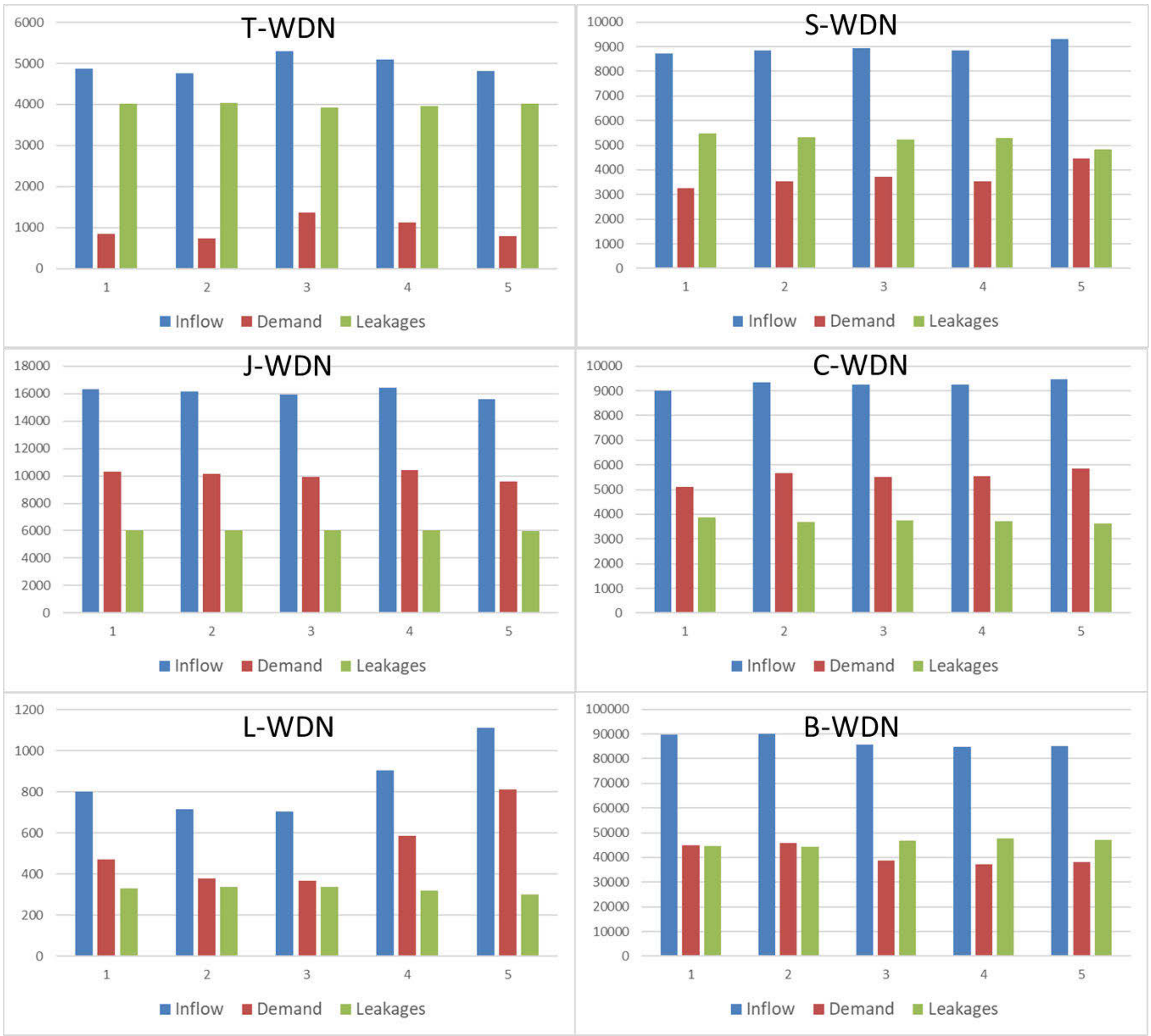


Figure 6. Calibration results: total inlet volume, volumetric leakages and consumers' demand over five operating cycles

BY NC SA

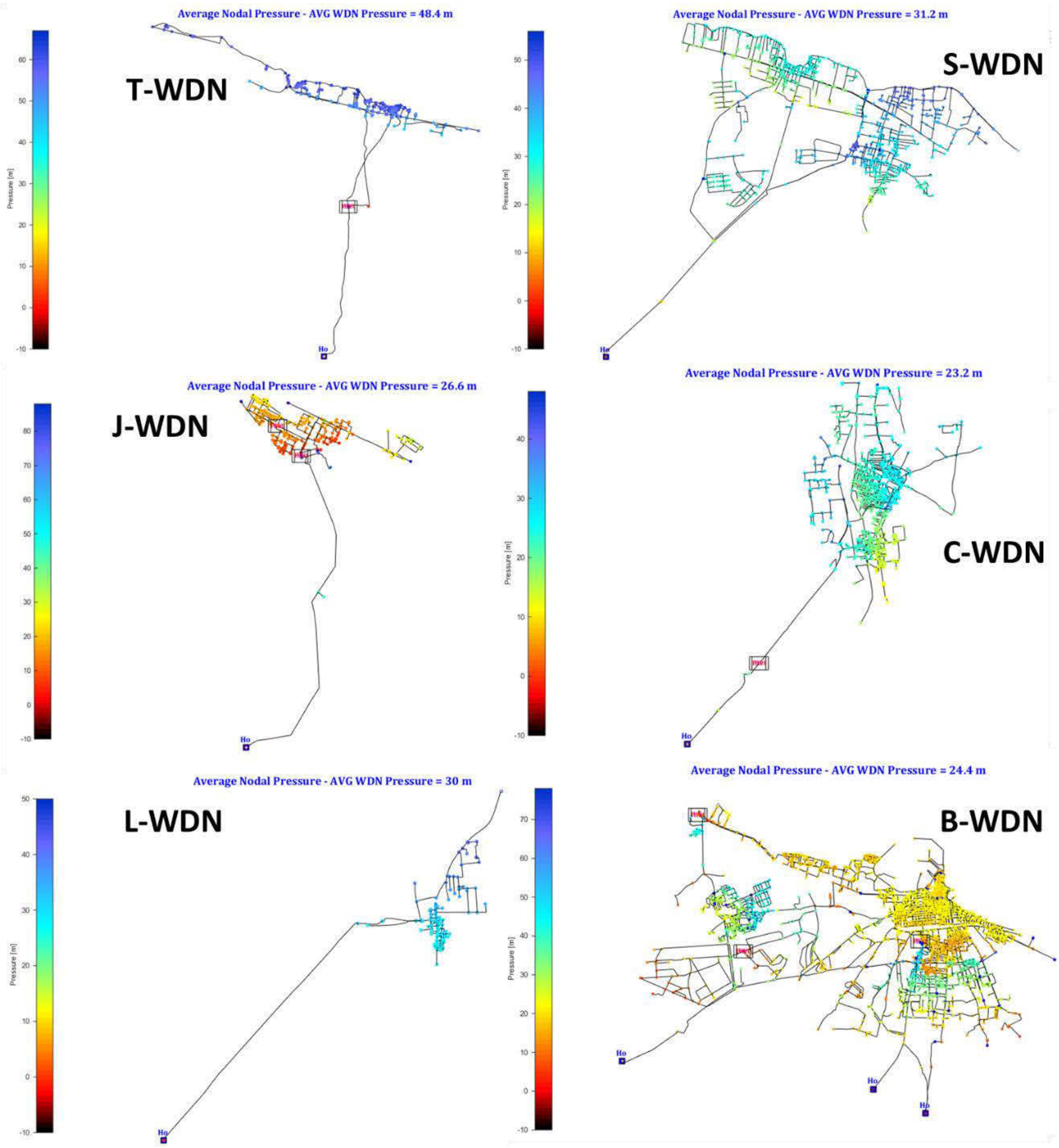

Average Nodal Pressure - AVG WDN Pressure = 48.4 m
T-WDN
Average Nodal Pressure - AVG WDN Pressure = 31.2 m
S-WDN
Average Nodal Pressure - AVG WDN Pressure = 26.6 m
J-WDN
Average Nodal Pressure - AVG WDN Pressure = 23.2 m
C-WDN
Average Nodal Pressure - AVG WDN Pressure = 30 m
L-WDN
Average Nodal Pressure - AVG WDN Pressure = 24.4 m
B-WDN
Pressure [m]
Ho

Figure 7. Calibration results: average pressure status at WDN model nodes

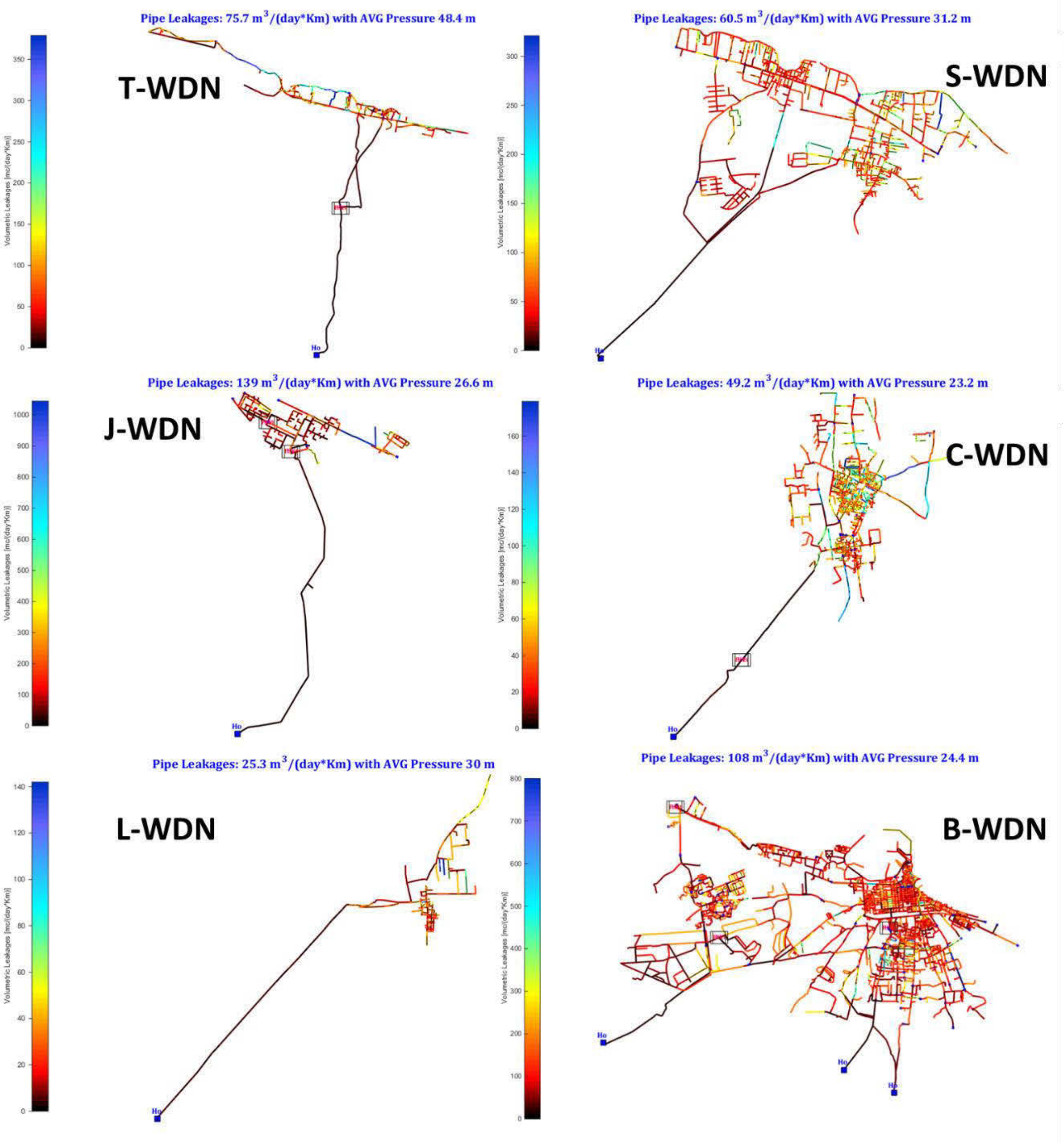

Pipe Leakages: 75.7 m³/(day*Km) with AVG Pressure 48.4 m
T-WDN
Pipe Leakages: 60.5 m³/(day*Km) with AVG Pressure 31.2 m
S-WDN
Pipe Leakages: 139 m³/(day*Km) with AVG Pressure 26.6 m
J-WDN
Pipe Leakages: 49.2 m³/(day*Km) with AVG Pressure 23.2 m
C-WDN
Pipe Leakages: 25.3 m³/(day*Km) with AVG Pressure 30 m
L-WDN
Pipe Leakages: 108 m³/(day*Km) with AVG Pressure 24.4 m
B-WDN
Volumetric Leakages [mc/(day*Km)]
Ho

Figure 8. Calibration results: linear leakage indicator at single pipes for each WDN

5 CONCLUSIONS

This contribution presents the application of mass-balance approach for calibrating WDN hydraulic model in a real large WDN, which is composed of six sub-systems hydraulically disconnected form each other. The adopted approach allows getting consistent results with the expected hydraulic behaviour of the networks in terms of mass balance accounting for leakages and consumers' demand. This is of primary importance to get a phenomenological representation of the system to support next design of asset management actions.

The advanced WDN hydraulic model, based on pressure dependent representation of all outflow components, is the main driver to accomplish such an approach. The discussion about the calibration variables and the role of flow and pressure measurements, substantiated by results on the real WDN, demonstrates the following key points.

- The calibration of a WDN model to support asset management cannot represent leakages using fixed demand patterns. In fact, pressure regime changes day by day in consequence of the stochastic change of water requests.
- The robust estimates of parameters representing asset features (i.e., pipe hydraulic resistances and deterioration parameters in the leakage model), which can be assumed as invariant during the calibration reference period (e.g., one year), need to be performed across multiple operating cycles representing characteristic system functioning.
- Flow measurements play a crucial role in mass-balance calibration, while pressure data are useful to get consistent identification of leakage outflow component and, consequently, leakage model parameters.

The analysis of results on the real WDN demonstrate that the mass-balance approach for calibration helps the identification of unexpected results due to possible erroneous information or missing data. This represents a powerful tool to identify possible corrections on available information to be verified on the field, which can help the understating of actual WDN behaviour. Such a *reverse engineering* process was proved also in other real WDNs analysed during the same procedure by the authors.

ACKNOWLEDGMENTS

This paper was partly financed by European Union - PON Ricerca e Innovazione 2014-2020 E FSC, call D.D. 13 July 2017 n. 1735, Project “Energidrica - Efficienza energetica nella gestione delle reti idriche” (code ARS01_00625). Data for figures were provided by IDEA-RT (www.idea-rt.com).

6 REFERENCES

[1] Rossman, L. A., Epanet2 Users Manual. EPA, Cincinnati, USA, 2000.

[2] Todini, E., “A more realistic approach to the ‘extended period simulation’ of water distribution networks”, In: Advances in Water Supply Management (C. Maksmovic, D. Butler & F. A. Memon, eds). A.A.Balkema Publishers, Lisse, The Netherlands, 2003, pp. 173–184.

[3] Giustolisi, O. and Walski, T., “Demand components in water distribution network analysis”, J. Water Res. Plan. Manage, vol. 138, no. 4, 2012, pp. 356–367.

[4] Wagner, J. M., Shamir, U. and Marks, D. H., “Water distribution reliability: simulation methods”, J. Water Res. Plan. Manage., vol. 114, no. 4, 1988, pp. 276–294.

[5] Germanopoulos, G., “A technical note on the inclusion of pressure-dependent demand and leakage terms in water supply network models”, Civil Engineering Systems, vol. 2, 1985, pp. 171–179.

[6] Giustolisi, O., Savic, D. A. and Kapelan, Z., “Pressure-driven demand and leakage simulation for water distribution networks”, J. Hydraulic Engineering, vol. 134, no. 5, 2008, pp. 626–635.

[7] Ulanicki, B., Bounds, P. L. M., Rance, J. P. and Reynolds, L., “Open and closed loop pressure control for leakage reduction”, Urban Water Journal, vol. 2, no. 2, 2000, pp. 105–114.

[8] Giustolisi, O., Ugarelli, R., Berardi, L., Laucelli, D. and Simone, A., “Strategies for the electric regulation of pressure control valves”, J. Hydroinformatics, vol. 19, no. 5, 2017, pp. 621–639.

[9] Ferrari, G., Savic, D. A., and Becciu, G. 2014, Graph-Theoretic Approach and Sound Engineering Principles for Design of District Metered Areas. J. Water Res. Plan. Manage. 140(12), 04014036.

[10] Laucelli, D., Simone, A., Berardi, L., and Giustolisi, O., “Optimal Design of District Metering Areas for the Reduction of Leakages”, J. Water Res. Plan. Manage., vol. 143, no. 6, 2017, pp. 04017017-1-12

[11] Shamir, U. and Howard, C.D.D, “Water distribution systems analysis”, J. Hydraul. Div., Am. Soc. Civ. Eng., vol. 94, no. 1, 1968, pp. 219–234.

[12] Walski, T. M., “Technique for calibrating network models”, J. Wat. Res. Planning and Management, vol. 109, no. 4, 1983, pp. 360–372.

[13] Kapelan, Z., Savic, D. A. and Walters, G. A., “Calibration of water distribution hydraulic models using a Bayesian-type procedure”, J. Hydraulic. Engineering, vol. 133, no. 8, 2007, pp. 927–936.

[14] Alvisi, S. and Franchini, M., “Pipe roughness calibration in water distribution systems using grey numbers”, J. Hydroinformatics, vol. 12, no. 4, 2010, pp. 424–445.

[15] Sanz G., and Pérez, R., “Demand pattern calibration in water distribution networks”, Procedia Engineering, vol. 70, 2014, pp. 1495-1504.

[16] Sanz, G., and Pérez, R., “Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition”, J. Water Resour. Plan. Manage., vol. 141, no. 1, 2015, pp. 04015020.

[17] Cheng, W., and He, Z., “Calibration of Nodal Demand in Water Distribution Systems”, J. Water Res. Plann. Manage., vol. 137, no. 1, 2011, pp. 31-40.

[18] Rokstad, M.M. and Ugarelli, R.M., “Investigation of the Ability to Accurately Estimate Background Leakage Parameters in WDS Network Simulation Models”, J. Water Res. Plan. Manage., vol. 143, no. 4, 2017, pp. 04017002.

[19] Berardi L., Simone A., Laucelli D. and Giustolisi O., “Feasibility of mass balance approach to Water Distribution Network model calibration”, in XVIII International Conference on Water Distribution Systems, WDSA2016, Procedia Engineering, vol. 186, 2017, pp. 551–558.

[20] Berardi, L. and Giustolisi, O. “Calibration of Design Models for Leakage Management of Water Distribution Networks”, Water Resources Management, vol. 35, 2022, pp. 2537–2551.

[21] Girard, M. and Stewart, R. A. “Implementation of pressure and leakage management strategies on the Gold Coast, Australia: case study”, J. Water Res. Plan. Manage., vol. 133, no. 3, 2007, pp. 210–218.

[22] WDNetXL, http://www.idea-rt.com/digital-water-latest-innovation, May 2022.

[23] Berardi L., Simone A., Laucelli D. B., Ugarelli R. M., and Giustolisi O., “Relevance of hydraulic modelling in planning and operating real-time pressure control: case of Oppegård municipality”, J. Hydroinformatics, vol. 20, no. 3, 2018, pp. 535–550.

[24] Todini, E., “Extending the global gradient algorithm to unsteady flow extended period simulations of water distribution systems”, J. Hydroinformatics, vol. 13, no. 1, 2011, pp. 167–180.

[25] Giustolisi, O., “Water Distribution Network Reliability Assessment and Isolation Valve System”, J. Water Res. Plan. Manage., vol. 146, no. 1, 2020, pp. 04019064.

[26] Giustolisi, O., Berardi, L., and Laucelli, D., “Modeling local water storages delivering customer-demands in WDN models”, J. of Hydraulic Engineering, vol. 140, no. 1, 2014, pp. 89-104.

[27] Van Zyl, J.E., and Cassa, A.M., “Modeling elastically deforming leaks in water distribution pipes”, Journal of Hydraulic Engineering, vol. 140, no. 2, 2014, pp. 182-189.

[28] Todini, E., “Using a Kalman Filter Approach for Looped Water Distribution Network Calibration”, in Proceeding of Water Industry Systems: Modelling and Optimisation Applications, Exeter, UK, D. A. Savic and G. A. Walters, eds., vol.1, 1999, pp.327-336.

[29] Autorità di Regolazione per Energia Reti e Ambiente (ARERA), Delibera 917/2017/R/IDR - Regolazione della qualità tecnica del servizio idrico integrato ovvero di ciascuno dei singoli servizi che lo compongono (in Italian), 2017.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14773

ENERGY DIAGNOSIS OF PRESSURIZED WATER SYSTEMS WITH THE ENERGOS TOOL

E. Gómez[1], R. del Teso[2], E Cabrera[3], E. Estruch[4], P. J. Maximino[5], M. Ortiz[6], G. Del Pozo[7] and C. Marco[8]

[1, 2, 3, 4] ITA. Department of Hydraulic Engineering and Environment. Universitat Politècnica de València. Valencia (Spain)
[5, 6, 7, 8]FACSA. SOCIEDAD FOMENTO AGRICOLA CASTELLONENSE, S.A.,Castellón (Spain)

[1] *elgosel@ita.upv.es*, [2]*rodete@ita.upv.es*, [3] *ecabrera@ita.upv.es*, *4maesjua1@ita.upv.es*,
[5]*pmaximino@facsa.com*, [6]*motiz@facsa.com*, [7]*guillermo.delpozo@facsa.com*, [8]*cmarco@facsa.com*

Abstract

Although the pressurized transport of fluids by pipeline is the most efficient one, it requires a high energy consumption, due to the volume transferred and the pressure required. This is the consequence of transporting large volumes of water (sometimes over considerable distances) and delivering it at the required pressure. In the current context of climate change, with both resources becoming increasingly scarce, the only way to minimize their impact and control energy consumption is to improve their efficiency, a process whose most relevant stages are the diagnosis, which identifies the starting point and the existing margin for savings, and the audit, which locates and quantifies inefficiencies.

This paper presents a simple tool, ENERGOS, which allows to perform the first stage: the energy diagnosis of a pressurized water system. The objective of this diagnosis is to know the current state of the system, and more importantly, the possible margin for improvement, if any, from the introduction of very few data. This is the first step to improve the efficiency of the system. The tool, and the energy indicators presented in it, have been designed under the premise that only the minimum information is required, which any manager should know about his system. That is, global volumes of water, billed and injected and energy consumed by the pumps (available in the electricity bills).

ENERGOS classifies the systems into three large groups, and performs the diagnosis according to the group. These are, firstly, the simple systems, defined as a pipeline, with one entry, one pump, generally, and one exit. Secondly, multi-scenario systems are systems with several inputs and outputs, and constant changes in their mode of operation, where each of these scenarios corresponds to a different layout. Finally, networks with one or more inputs and numerous outputs can operate differently, but without changing the layout.

In all cases, the tool has a schematic, simple and intuitive data entry part. From the data entered, it calculates the diagnosis, consisting of the comparison of the current energy intensity indicator (kWh/m^3) with the ideal energy intensity, the one that would imply the total absence of losses, both operational losses (friction, water losses, inefficiencies in the pumping stations) and structural losses (those due to topography). Since it is impossible to reach the ideal energy intensity value, an intermediate indicator, the target intensity, is defined. The calculation of it requires the establishment of targets for losses, reasonable reference values to be achieved with the current technology for the system under analysis. For example, the current efficiency of the pumping station is estimated, and a minimum acceptable level is calculated for it (establishing values for the efficiency of the motor, the variable frequency driver and the pump itself). The same is established for friction, water losses and excess pressure.

Keywords
Efficiency, Tool, Water and energy.

1 INTRODUCTION

In the current context, the need to improve the efficiency of all systems involving significant energy consumption is becoming increasingly evident. Proof of this concern is the upward revision of the efficiency target set by the European Commission (EC), established in 2011 at 20% [1], to 32.5% by 2030 [2]. It is therefore required to establish procedures and to have analysis tools for energy demanding sectors such as pressurized water transport.

The tool presented below is intended to ease the analysis of the state of water supply utilities systems from the point of view of energy consumption. The objective of this tool is to offer a series of indicators that make it easier to know the current status and, more importantly, to quantify the margin of improvement that these systems have. For this purpose, the energy diagnosis is carried out, depending on the type of system, classifying them as simple systems [3], networks [4] and multi-stage systems [5].

The first step to increase the efficiency of a system is to know its state. To this end, the first step proposed is to carry out an energy diagnosis. This will make it possible to establish an ideal (unachievable) efficiency limit which, compared with the current state, will show the system's margin for improvement. This ideal limit is established for the system under analysis. It does not make sense to establish the same limit for all systems, since each one has different characteristics, topography, demands and requirements. Similarly, since the ideal limit is unattainable, it is necessary to establish a target efficiency value. This value is established by reducing current operational energy losses to values that are reasonably achievable with current technology. Establishing these limits can be complex and, in any case, it will require a subsequent detailed cost-benefit study to determine whether they are economically and technically feasible.

This tool has been designed and conceived with the objective of being simple to use and mainly requiring little data from the system. Often the main problem when analyzing the systems lies in the difficulty of having the data, therefore, the diagnosis is made with the minimum data that is considered that a utility must know and control about their systems, such as topography, volumes and energy consumed. For the establishment of the objectives, ENERGOS has a series of assistants that guide the user and proposes a series of values based on experience, bibliography on the subject or regulations.

It should not be forgotten that this is an initial diagnosis and not an in-depth analysis. Once the diagnosis has been carried out and with the results obtained, an energy audit should be undertaken, if the diagnosis concludes so. The energy audit [6] requires a mathematical model of the network and, therefore, more time and detail both in the knowledge of the system and in the analysis.

2 CONCEPT OF ENERGY DIAGNOSIS AND METHODOLOGY

The proposed diagnosis, maintaining the same objective: to know the current state and compare it with the best possible one, presents some differences depending on the type of system. Therefore, the energy analysis advises to classify the water transport systems in three groups:

- Simple systems: One pipe, with one inlet, one outlet and, generally, one pump.
- Networks: With one or more inlets and numerous outlets. These may operate differently depending on the requirements of the outlets, but they do so without changing the layout.
- Multi-scenario systems: Systems with several inputs and outputs, and constant changes in their mode of operation (multi-scenario). A different layout corresponds to each one.

Simple systems, such as groundwater pumping, are an important part of the problem. Because of their simplicity, a single indicator (ratio of energy consumed to volume pumped) makes it possible

BY NC SA

to evaluate their energy efficiency. This enables to characterize energy intensity accurately. In networks, with multiple delivery points, as the operation of the system must meet the needs of the critical point [3], more energy than required is delivered to the other consumption nodes. These excesses, a function of the irregularity of the terrain, are related to topographic energy [7]. Their existence is, from the energy perspective, the great difference between simple and complex systems (networks).

The basic indicators proposed for the diagnosis of each of the types are presented below.

2.1 Simple systems

The metrics that characterize energy efficiency in simple systems are very straightforward. In particular, the normalized energy indicator, Ph5, of the IWA [8], quotient between the energy consumed during the analysis period and the product of the volume of water pumped times the pumping head in that time interval. A Ph5 value equal to or lower than 0.4 kWh/(m^3.100m) is customarily adopted as a boundary to qualify efficient management [9]. In simple systems it is a reference indicator [10]. However, being referred to pumping head (or manometric head), it encompasses the two most relevant inefficiencies (pumping and friction). They can be decoupled by adopting as reference, the geometric head.

Considering pump head to indicate efficiency instead of geometric head can lead to misdiagnosis, particularly when comparing systems with very different elevations and conveying distances. Also local losses (sometimes significant) and excessive piezometric heads [11] can lead to confusion, when interpreting the results. The basis of an energy audit is to disaggregate and quantify the three inefficiencies (pumping, friction and leakage), and it is only possible by normalizing the indicator with a geometric height independent of any friction losses.

To determine the contribution of each loss to the final energy consumption, the proposed diagnosis refers the energy intensity to the geometric head. Once the actual losses of each type are known, their comparison with those that would be reasonable allows us to evaluate both the margin for improvement and the strategy to be followed. This analysis can be extended to any simple system involving pumping.

If water transport requires overcoming a geometric difference in level $H_g = z_f - z_i > 0$, and there are no losses of any kind, the ideal energy intensity (I_{ei}) will be:

$$I_{ei} = 9810\, H_g \frac{N}{m^2} = 0.002725\, H_g \frac{kWh}{m^3} = 0.002725\,(z_f - z_i)\frac{kWh}{m^3} \tag{1}$$

The real energy intensity, I_{er}, is determined from the facility's operational data recorded during the analysis period. In the numerator, E_{sr} (actual energy supplied), easily known from electricity meters located at the pumping stations or from electricity bills, and in the denominator V_d, (volume delivered at the supply point), resulting in:

$$I_{er} = \frac{E_{sr}}{V_d} \tag{2}$$

The real energy intensity I_{er} can be expressed as a function of I_{ei} (the ideal energy intensity, equal to the energy required to lift, without losses, one cubic meter of water the geometric head H_g) and rh, the ratio of heads, the quotient between the geometric gradient and the pumping head. Thus Ier becomes:

$$I_{er} = I_{ei} \frac{1}{\eta_l \cdot \eta_p} \frac{1}{r_h} \tag{3}$$

Where, η_l, refers to the hydraulic efficiency (quotient between the volume delivered, V_d, and the volume pumped, V_i). The efficiency of the pumping station, η_p, considers the efficiency of all its elements, i.e. all electrical and mechanical losses. And finally, r_h, refers to the head ratio (quotient between the geometric head, Hg, and the head provided by the pump, H_p). All these inefficiencies can be grouped into the overall loss factor term, F_{Gl}, leaving:

$$I_{er} = F_{Gl} \quad I_{ei} \tag{4}$$

Where F_{Gl} is the global loss factor defined as follows:

$$F_{Gl} = \frac{1}{\eta_l \cdot \eta_p \cdot r_h} \tag{5}$$

The difference between both indicators, I_{ei} and I_{er}, establishes the margin of improvement that the current system presents. Since I_{ei} is unachievable, a reasonable intermediate limit must be established between them. For this purpose, the target energy intensity, I_{eo}, is defined. This is calculated from reasonably achievable inefficiency values, therefore:

$$I_{eo} = I_{ei} \frac{1}{\eta_{lo} \cdot \eta_{po}} \frac{1}{r_{ho}} \tag{6}$$

Consequently, it is appropriate to define guide values for each of these three parameters. The proposal is:

1. *With respect to the head ratio,* r_h. This is the quotient between the geometric head and the pumping head, the latter being dependent on friction losses. Therefore, target friction losses must be defined, and these are established from the concept of the most economical unit loss, J_o, [12], being L the length of the conduit that conforms the simple system. Therefore:

$$r_{ho} = \frac{H_g}{H_g + h_{fo}} = \frac{H_g}{H_g + L\, J_o} \tag{7}$$

2. *With respect to hydraulic losses,* η_l. In a simple pipeline, from an easily auditable hydraulic point of view, there should be no leakage, and, therefore, $\eta_{lo} = 1$ should be aimed for. In any case, for more complex systems there is a level of leakage above which it is not economically profitable to continue investing in tasks that try to reduce the leaked water, the so-called Economic Level of Leakage (ELL)

3. *Regarding pump inefficiency*. The European Directive “Establishing a framework for the setting of ecodesign requirements for energy-related products” [13], was completed shortly after with the corresponding for hydraulic pumps [14]. This last directive sets a minimum pump efficiency, dependent on its final quality established from the Minimum Efficency Index (MEI). The value adopted corresponds to that established by the directive for a given MEI.
   An efficiency limit is also established for the electrical elements, motor and variable frequency drive, which combined with the pump efficiency will result in the η_{po} value. The efficiency of the motor is established according to IEC/EN 60034-30-1. While the minimum efficiency of a motor is well regulated, this is not the case for the variable frequency drive whose efficiency is within the 96 - 98 % range [15]. Therefore, it is usually assigned an efficiency of 97 % [16] [17]. A new European standard, EN 50598, standardizes the efficiency of the entire assembly, the PDS (Power Drive System). That is, the whole assembly is regulated instead of addressing the elements separately [18]. However, in this work, the traditional, element-by-element approach is maintained, proposing the product of their efficiencies as the objective of the PDS.

2.2. Networks

Networks are defined as systems that may have one or more inputs, but they have multiple outputs (network users). The application of the continuity equation to this type of system over a given period of time is straightforward. The total injected volume of the pumping station V is equal (in the absence of water losses) to the sum of the demands of all the consumption nodes $V=\sum v_j$. Therefore, the ideal energy supplied to the system E_{si}, is that corresponding to the delivery of the entire water volume demanded by the consumption nodes at a head H_{hi}. This head is equal at all nodes (since there are no head losses) and it is the sum of the critical node elevation ($z_c=z_h$) and the service pressure (p_o).

$$E_{si} = \gamma V H_{hi} \tag{8}$$

This supplied energy corresponds to the sum of two energy terms. On the one hand, the one that really interests, which is the useful energy delivered to the users E_{uo}. This energy corresponds to the one to be delivered at all consumption nodes, at the level at which they are located (z_j), the volume demanded (v_j) at the minimum pressure or service pressure:

$$E_{uo} = \gamma \sum \ v_j \, [(z_j - z_l) + \frac{p_0}{\gamma}] \tag{9}$$

On the other hand, there is an energy term, which, although it is not useful, the system is obliged to deliver it due to the irregularities of the terrain. This is the topographic energy, E_{ti}. This term must be provided to supply adequately the critical node, while penalizing the rest of the consumption nodes. This is because in order to provide enough pressure to the highest node , the rest of the nodes of the system will have overpressure. In a completely flat network, E_{ti} (topographic energy) would be zero.

$$E_{ti} = \gamma \sum \ v_j \, (z_h - z_j) = \gamma \sum \ v_j \, \frac{p_{jt,i}}{\gamma} \tag{10}$$

This supplied energy can be of natural origin (gravitational) made available by the own elevation of the water source (h_{ni}) or shaft, i.e. supplied by the pump providing a head (h_{pi}). Therefore, in an ideal system, without friction and in the absence of losses, for any period of time, the piezometric head line is the sum of the natural head (h_{ni}) and the pump head (h_{pi}), this is constant and equal at all points of the system ($H_{hi}= H_{ji}=H_{li}$). This will not be the case in real systems.

There may be a third term in the energy supplied, and it is the excess energy (E_{ei}), which corresponds to the excess pressure, i.e., that which is supplied above that strictly necessary at the critical node and therefore at the rest of the nodes (p_{ei}). This excess energy is due to supplying more head than the minimum necessary (z_h+p_o), for example because the pump is operating at a point above the nominal operating point. This situation must be avoided either with operational measures (regulating the pump operation) or with structural measures (reconfiguring the system design).

In the case of excess pressure, the energy supplied to the system will be:

$$E_{si} = E_{uo} + E_{ti} + E_{ei} \tag{11}$$

From these concepts and energy balances the ideal energy efficiency can be stated as the ratio between the minimum energy required by the users and the energy supplied.

$$\eta_{ai} = \frac{E_{uo}}{E_{si}} \tag{12}$$

Which, like any other efficiency value, is the ratio between what is actually useful, E_{uo}, and what is delivered to the system, E_{si}. In this ideal case (in the absence of all types of losses), the inefficiencies of the system that cause E_{si} to be greater than E_{uo}, and therefore the value of the efficiency to be less than 1, can only be due to the topographic energy E_{ti}. This energy, although unavoidable (since it is due to the topography of the terrain), could be reduced if the configuration of the system is modified or even recovered (installation of PATs). Indeed, the topographic energy is an excess of energy that, due to the irregularity of the terrain, has to be over-delivered to the system. It is therefore, in a way, a loss to be avoided, as far as possible, at the design stage [3]. And if the system is operating, it has to be managed in the best possible way [7]. As they do not depend on the mode of operation of the system, these losses are called structura. These losses do not exist in simple systems, where it is sufficient to focus on operational losses (pumping station, those linked to leakage and, finally, pipe friction).

The topographical parameter, θ_{ti}, represents the fraction of the energy supplied that is intended to bridge the topographical unevenness of the system (assuming no energy recovery). This parameter will be lower the flatter the network is. Therefore, the ideal efficiency of a network, η_{ai}, represents the maximum value that the efficiency of the particular system under evaluation can reach. This value will be different for each network as it depends on the topographical characteristics of the system and the volume demanded by the system. This value determines the upper limit of efficiency, and it is unachievable since it assumes a total absence of losses, which is impossible in real systems. Real systems, as will be seen below, contain a series of losses that, although they can be minimized, they can never be completely removed, such as water losses, inefficiencies in the pressure groups or friction in the pipes.

On the other hand, the real energy supplied to the system can be calculated from the energy entering the system. This is, the real available natural energy $E_{sr,n}$ (proportional to the available head and the volume injected into the network, the sum of the demand plus the leakage) and the real energy supplied by the pump, $E_{sr,p}$, which can be obtained from the reading of electricity bills. Real systems share with the ideal systems, the numerator of the energy efficiency (E_{uo}), while the denominator changes to include the energy losses.

$$\eta_{ar} = \frac{E_{uo}}{E_{sr}} \tag{13}$$

The difference between the ideal performance, η_{ai}, and the actual performance, η_{ar}, (η_{ai} - η_{ar}), provides a first estimate of the system's margin for improvement. This difference provides an idea of where the system is and how far it could go (being this upper limit unreachable), if the difference between the two is significant it is because the energy losses in the system are considerable and therefore the margin for improvement is wide. If the difference is small, it means that the state of the network is practically the best possible scenario and it is not worth spending resources to deeply evaluate the system, since the situation in which the system is already is almost the best possible one.

The difference between the ideal performance, η_{ai}, and the actual performance, η_{ar}, provides a first estimate of the system's room for improvement. Once the difference between the two performances is known, the next step is to determine the potential margin for improvement. Or, in other words, how close the second term can come to the first term through the adaptation of measures that must subsequently be verified to maintain an acceptable cost-benefit ratio. A new reference value ($\eta_{ar,o}$) represents the achievable system efficiency or target efficiency, being $\eta ai > \eta_{ar,o} > \eta_{ar}$.

This new efficiency can be estimated from the Target Energy which will be the energy that would be supplied to the system if each of the losses were reduced to the minimum or target value. The more inefficiencies are removed from the system, the closer $E_{sr,o}$ will be to E_{si} (the closer the target will be to the ideal case). The target efficiency thus depends on the target level of energy losses. It is an intermediate value between the ideal (unachievable) and the actual (current) value. To calculate this value, the targets to be achieved for each of the losses must be set.

$$\eta_{Go} = \frac{E_{uo}}{E_{sr,o}} \tag{14}$$

For the establishment of the objectives, the same recommendations as those detailed for the simple systems are maintained.

2.3. Multi-scenario systems

Multi-scenario systems are those that operate in different modes, with constant layout changes. The diagnosis of a multi-scenario system without energy storage (no reservoirs inside the CV), compares the output powers with the input powers, regardless of what happens inside, an approach that invites to associate these systems (in the diagnostic phase) to a black box. Its definition [19] is: "a fiction representing a set of concrete systems into which stimuli S impinge and out of which reactions R emerge. The constitution and structure of the box are altogether irrelevant to the approach under consideration, which is purely external or phenomenological".

The existence of internal reservoirs (contributing or subtracting power) requires knowledge of the flow rate Qt, from additional inputs/outputs which, although internal to the control volume (such as the height of water in the reservoir z), are analogous to the external inputs/outputs of the system. Each scenario can be diagnosed. Subsequently, with the integration over time of the behaviour of each one, an overall assessment can be made. However, each scenario should be analysed to identify the most inefficient modes of operation and, if necessary, to correct them.

3 CONCEPT OF ENERGY DIAGNOSIS AND METHODOLOGY

ENERGOS is a simple tool for energy diagnosis. It is an MS Excel file that provides, from the introduction of the basic data of the system, the results of the energy diagnosis. These results will facilitate the decision-making process aimed at improving the energy efficiency of the drinking water or irrigation network.

The tool is divided into three tabs, one for simple systems, one for networks and one for multi-scenario systems. The first two have a very similar operation (Figure 1), first requiring the selection of the typology. Depending on the selected typology, a series of geometrical and operational data is required. Once the data for the selected analysis period has been entered, the result of the energy diagnosis is provided. Subsequently, the target values for operational losses must be established, for which the tool has a series of helping assistants, which ease the establishment of the targets.

3.1 Simple systems:

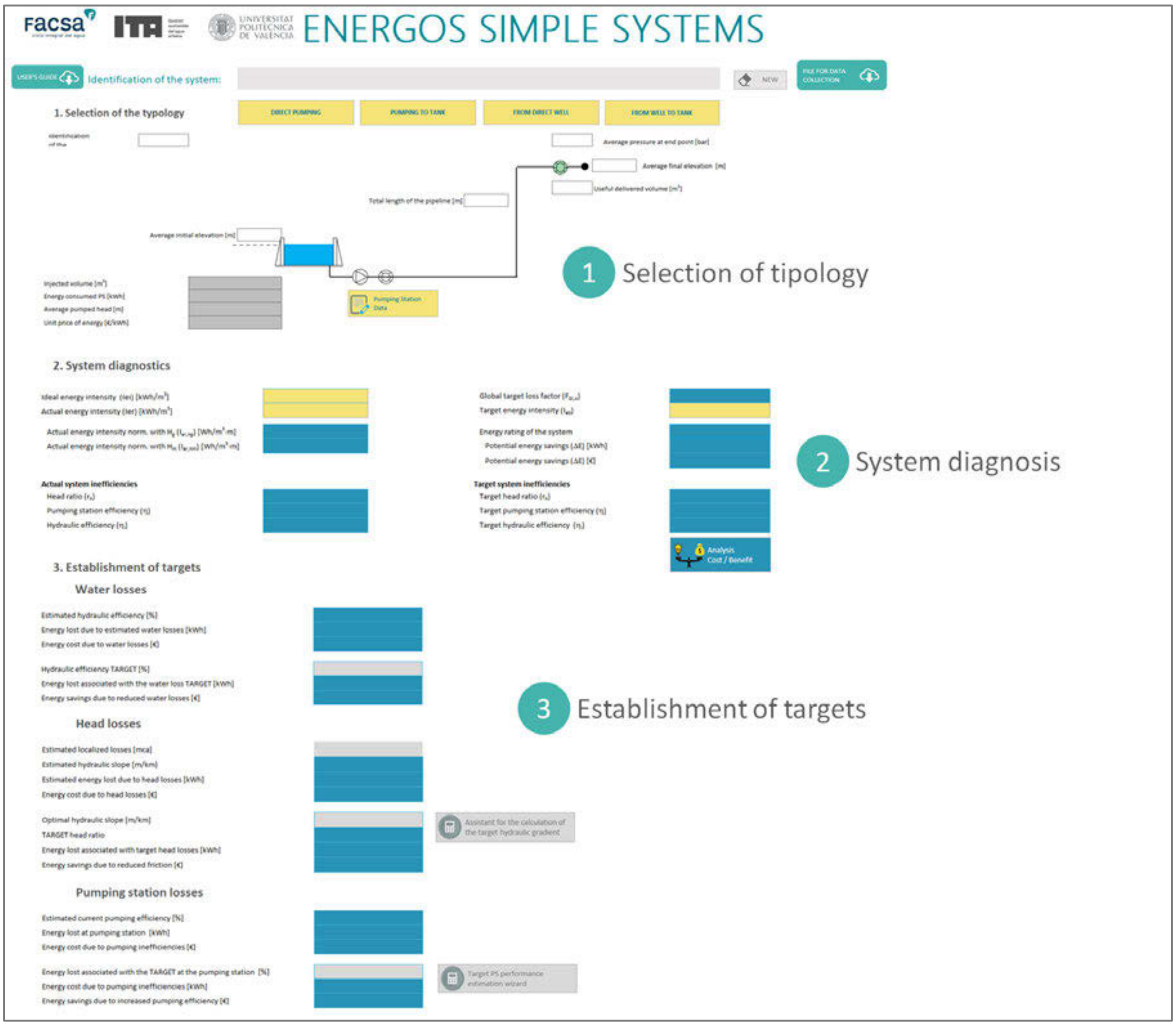
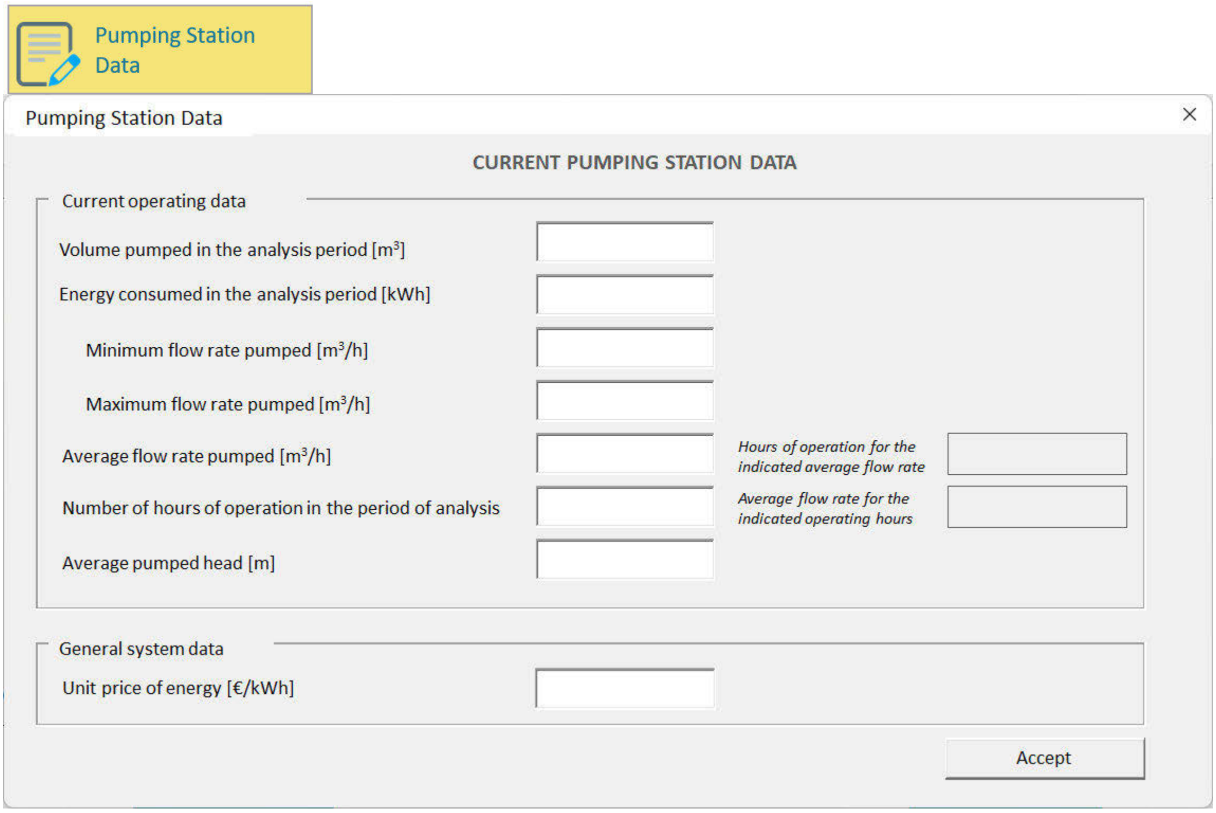

Facsa
ITA
UNIVERSITAT POLITÈCNICA DE VALÈNCIA
ENERGOS SIMPLE SYSTEMS
USER'S GUIDE
Identification of the system:
NEW
FILE FOR DATA COLLECTION
1. Selection of the typology
DIRECT PUMPING
PUMPING TO TANK
FROM DIRECT WELL
FROM WELL TO TANK
Identification of the
Average pressure at end point [bar]
Average final elevation [m]
Useful delivered volume [m³]
Total length of the pipeline [m]
Average initial elevation [m]
Injected volume [m³]
Energy consumed PS [kWh]
Average pumped head [m]
Unit price of energy [€/kWh]
Pumping Station Data
1 Selection of tipology
2. System diagnostics
Ideal energy intensity (Iei) [kWh/m³]
Actual energy intensity (Ier) [kWh/m³]
Actual energy intensity norm. with Hg [Wh/m³·m]
Actual energy intensity norm. with Hc [Wh/m³·m]
Actual system inefficiencies
Head ratio (rh)
Pumping station efficiency (ηp)
Hydraulic efficiency (ηh)
Global target loss factor
Target energy intensity
Energy rating of the system
Potential energy savings (ΔE) [kWh]
Potential energy savings (ΔE) [€]
Target system inefficiencies
Target head ratio (rh)
Target pumping station efficiency (ηp)
Target hydraulic efficiency (ηh)
Analysis Cost / Benefit
2 System diagnosis
3. Establishment of targets
Water losses
Estimated hydraulic efficiency [%]
Energy lost due to estimated water losses [kWh]
Energy cost due to water losses [€]
Hydraulic efficiency TARGET [%]
Energy lost associated with the water loss TARGET [kWh]
Energy savings due to reduced water losses [€]
Head losses
Estimated localized losses [mca]
Estimated hydraulic slope [m/km]
Estimated energy lost due to head losses [kWh]
Energy cost due to head losses [€]
Optimal hydraulic slope [m/km]
TARGET head ratio
Energy lost associated with target head losses [kWh]
Energy savings due to reduced friction [€]
Assistant for the calculation of the target hydraulic gradient
Pumping station losses
Estimated current pumping efficiency [%]
Energy lost at pumping station [kWh]
Energy cost due to pumping inefficiencies [€]
Energy lost associated with the TARGET at the pumping station [%]
Energy cost due to pumping inefficiencies [kWh]
Energy savings due to increased pumping efficiency [€]
Target PS performance estimation wizard
3 Establishment of targets

Figure 1. General diagram of the tool

The first step is to select the type of scheme that best suits the system under analysis, and enter the basic data that defines it. There are four different types of schemes (direct pumping, pumping to reservoir, from direct well and from well to reservoir). In fact, two different schemes are contemplated, either with direct pumping to the point of consumption or with pumping to the tank, either from the reservoir or from the well.

Once the scheme has been selected and its basic data entered (the one required for each scheme), the pumping station data must be defined. The requested data are the minimum data that are considered to be known by the managers of the utility, without too much complexity and, therefore, easy to know (Figure 2).

BY NC SA



Pumping Station Data

CURRENT PUMPING STATION DATA

Current operating data

Volume pumped in the analysis period [m^3]

Energy consumed in the analysis period [kWh]

Minimum flow rate pumped [m^3/h]

Maximum flow rate pumped [m^3/h]

Average flow rate pumped [m^3/h]

Hours of operation for the indicated average flow rate

Number of hours of operation in the period of analysis

Average flow rate for the indicated operating hours

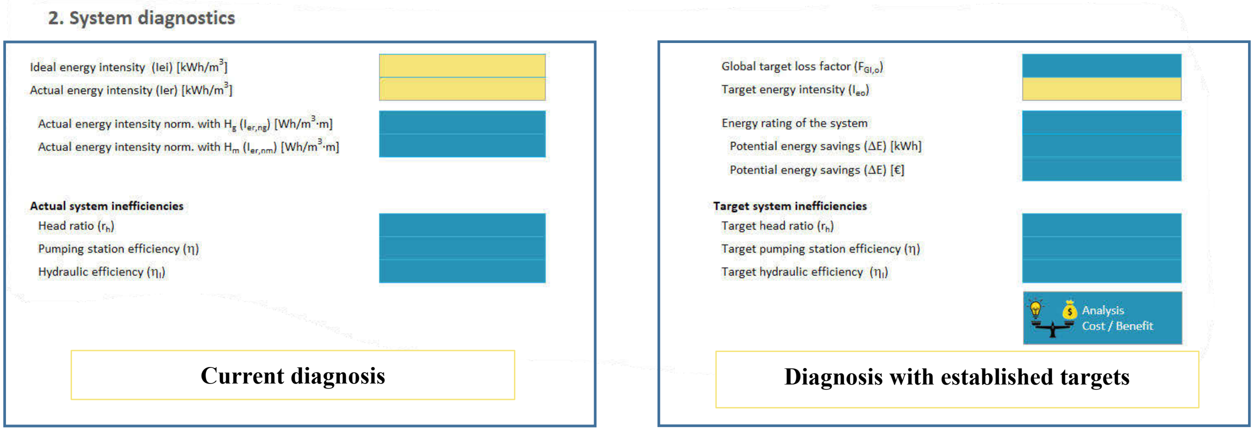
Average pumped head [m]

General system data

Unit price of energy [€/kWh]

Figure 2. Current data required from the Pumping Station

When the minimum parameters have been deined, the results of the current energy diagnostics will be displayed (Figure 3, left). To obtain the target diagnostic values (Figure 4, right), the targets must first be set.



2. System diagnostics

Ideal energy intensity (Iei) [kWh/m^3]

Actual energy intensity (Ier) [kWh/m^3]

Actual energy intensity norm. with H_g ($I_{er,ng}$) [$Wh/m^3 \cdot m$]

Actual energy intensity norm. with H_m ($I_{er,nm}$) [$Wh/m^3 \cdot m$]

Actual system inefficiencies

Head ratio (r_h)

Pumping station efficiency (η)

Hydraulic efficiency (η_l)

Global target loss factor ($F_{Gl,o}$)

Target energy intensity (I_{eo})

Energy rating of the system

Potential energy savings (ΔE) [kWh]

Potential energy savings (ΔE) [€]

Target system inefficiencies

Target head ratio (r_h)

Target pumping station efficiency (η)

Target hydraulic efficiency (η_l)

Figure 3. System diagnostics

Current inefficiencies

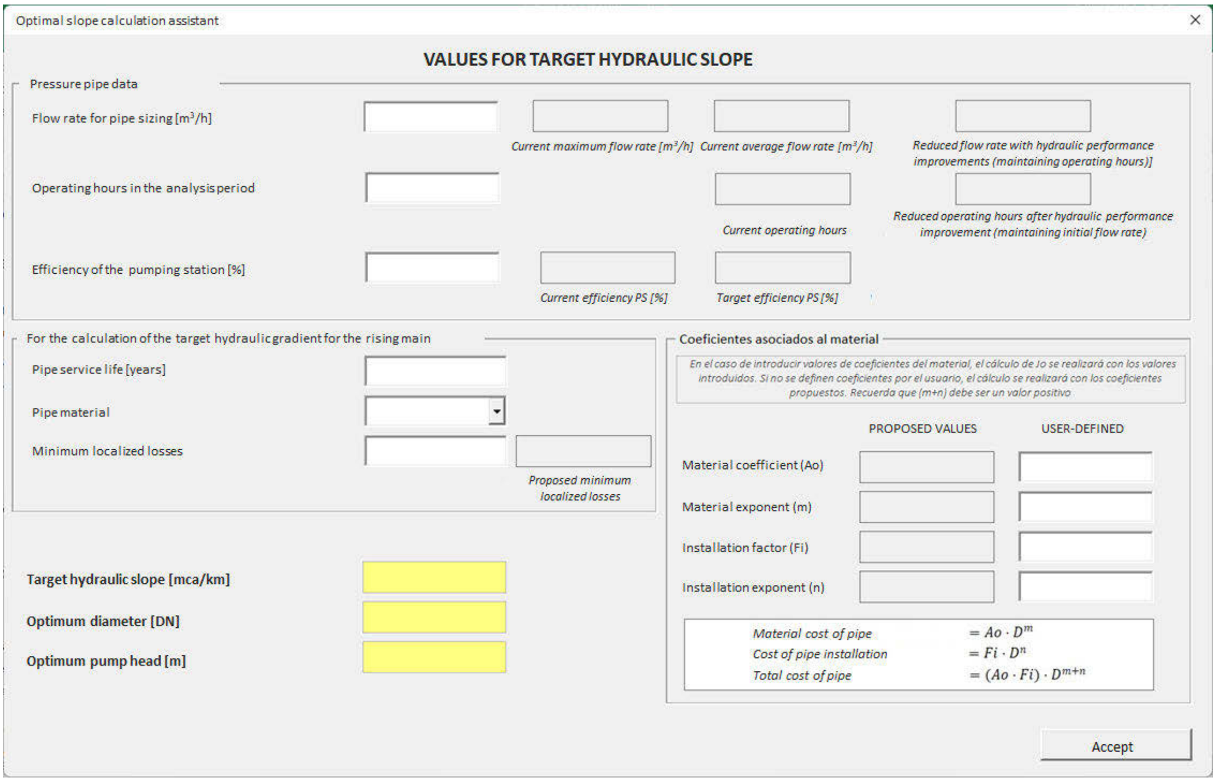
They are estimated from the data entered and the current performance of the facility obtained from the three inefficiency values of system: the head ratio (consequence of friction losses), the pumping station efficiency (consequences of energy losses in the pump and pumping station elements) and hydraulic efficiency (consequence of leaks or overflows).

BY NC SA

Establishment of objectives

Each of the possible operational losses that may occur in the system, "Water losses", "Head losses" and "Pumping station losses", are analysed below.

For each of the operational losses, ENERGOS has a helping assistant that allows to calculate the target to be reached (Figures 5 and 6).



Optimal slope calculation assistant

VALUES FOR TARGET HYDRAULIC SLOPE

Pressure pipe data
Flow rate for pipe sizing [m³/h]
Current maximum flow rate [m³/h]
Current average flow rate [m³/h]
Reduced flow rate with hydraulic performance improvements (maintaining operating hours)]
Operating hours in the analysis period
Current operating hours
Reduced operating hours after hydraulic performance improvement (maintaining initial flow rate)
Efficiency of the pumping station [%]
Current efficiency PS [%]
Target efficiency PS [%]

For the calculation of the target hydraulic gradient for the rising main
Pipe service life [years]
Pipe material
Minimum localized losses
Proposed minimum localized losses

Coeficientes asociados al material

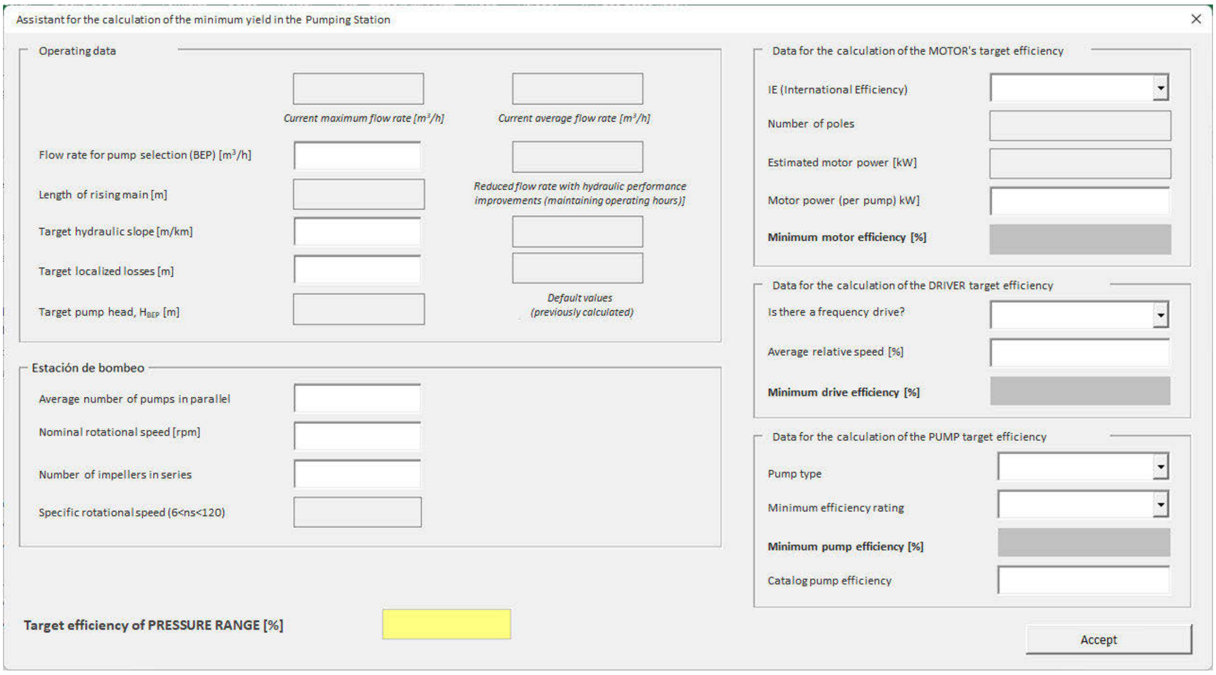
En el caso de introducir valores de coeficientes del material, el cálculo de Jo se realizará con los valores introducidos. Si no se definen coeficientes por el usuario, el cálculo se realizará con los coeficientes propuestos. Recuerda que (m+n) debe ser un valor positivo

PROPOSED VALUES
USER-DEFINED
Material coefficient (Ao)
Material exponent (m)
Installation factor (Fi)
Installation exponent (n)

Material cost of pipe $= Ao \cdot D^m$
Cost of pipe installation $= Fi \cdot D^n$
Total cost of pipe $= (Ao \cdot Fi) \cdot D^{m+n}$

Target hydraulic slope [mca/km]
Optimum diameter [DN]
Optimum pump head [m]

Figure 4. Helping assistant for the optimal hydraulic slope calculation



Assistant for the calculation of the minimum yield in the Pumping Station

Operating data
Current maximum flow rate [m³/h]
Current average flow rate [m³/h]
Flow rate for pump selection (BEP) [m³/h]
Reduced flow rate with hydraulic performance improvements (maintaining operating hours)]
Length of rising main [m]
Target hydraulic slope [m/km]
Target localized losses [m]
Target pump head, H_{BEP} [m]
Default values (previously calculated)

Estación de bombeo
Average number of pumps in parallel
Nominal rotational speed [rpm]
Number of impellers in series
Specific rotational speed (6<ns<120)

Data for the calculation of the MOTOR's target efficiency
IE (International Efficiency)
Number of poles
Estimated motor power [kW]
Motor power (per pump) kW]
Minimum motor efficiency [%]

Data for the calculation of the DRIVER target efficiency
Is there a frequency drive?
Average relative speed [%]
Minimum drive efficiency [%]

Data for the calculation of the PUMP target efficiency
Pump type
Minimum efficiency rating
Minimum pump efficiency [%]
Catalog pump efficiency

Target efficiency of PRESSURE RANGE [%]

Figure 5. Helping assistant for the calculation of the target performance in the pressure group

CC BY NC SA

Target system diagnosis

Once the reference values have been established (for friction, pumping station and leakage), both the overall target loss factor and the target energy intensity can be calculated. These indicators, although ambitious, are achievable, and should mark the level to be reached by the system.

Depending on the difference between the actual energy intensity (I_{er}) and the target energy intensity ($I_{er,o}$), the simple system is rated, similar to what is established by the European Union for refrigeration equipment, as shown in the following image:

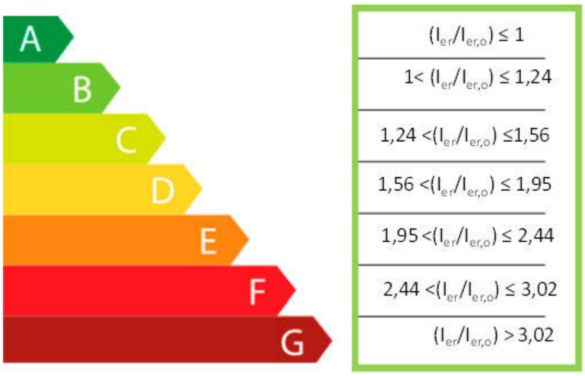

A
B
C
D
E
F
G
$(I_{er}/I_{er,o}) \leq 1$
$1 < (I_{er}/I_{er,o}) \leq 1{,}24$
$1{,}24 < (I_{er}/I_{er,o}) \leq 1{,}56$
$1{,}56 < (I_{er}/I_{er,o}) \leq 1{,}95$
$1{,}95 < (I_{er}/I_{er,o}) \leq 2{,}44$
$2{,}44 < (I_{er}/I_{er,o}) \leq 3{,}02$
$(I_{er}/I_{er,o}) > 3{,}02$

Figure 6. Energy qualification for simple systems

Finally, the tool displays a summary of all losses, as well as target values, and allows a preliminary cost/benefit analysis of each measure.

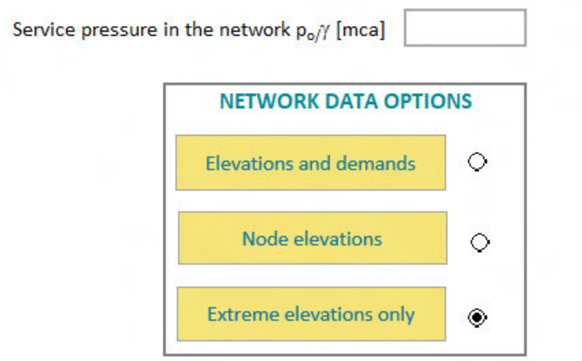
In order to help the prioritisation of measures, this sensitivity analysis is accompanied by the estimated margin for improvement of each of the inefficiencies and the potential savings that would result from improving each of them.

Finally, the cost of implementing the measure is estimated, i.e., the cost of modifying the current inefficiency value to the target value calculated or introduced. This cost is not calculated for water losses, but it does estimate the cost of installing a pipe of the optimum diameter and length corresponding to the one defined. Besides, it estimates the cost of a pump, providing the defined flow rate and head, according to the pump price database held by the tool. In any case, these costs are approximated, and the user is advised to define the cost of the measure. If so, this will be the cost considered and not the proposed cost.

Finally, the payback period is estimated based on the calculated potential savings and the cost of the measure.

3.2 NETWORKS

For networks the tool is similar to that for simple systems. First the typology is selected. Then, depending on the typology, the operating data are entered. The difference with respect to the simple systems tool is that, in this case, the characteristics of the subscribers or users of the system must be defined. These data are decoupled from the remaining input data. For this purpose, there are three levels of knowledge of the network, from more known data to less (figure 7).



Service pressure in the network p_o/γ [mca]

NETWORK DATA OPTIONS

Elevations and demands ○

Node elevations ○

Extreme elevations only ●

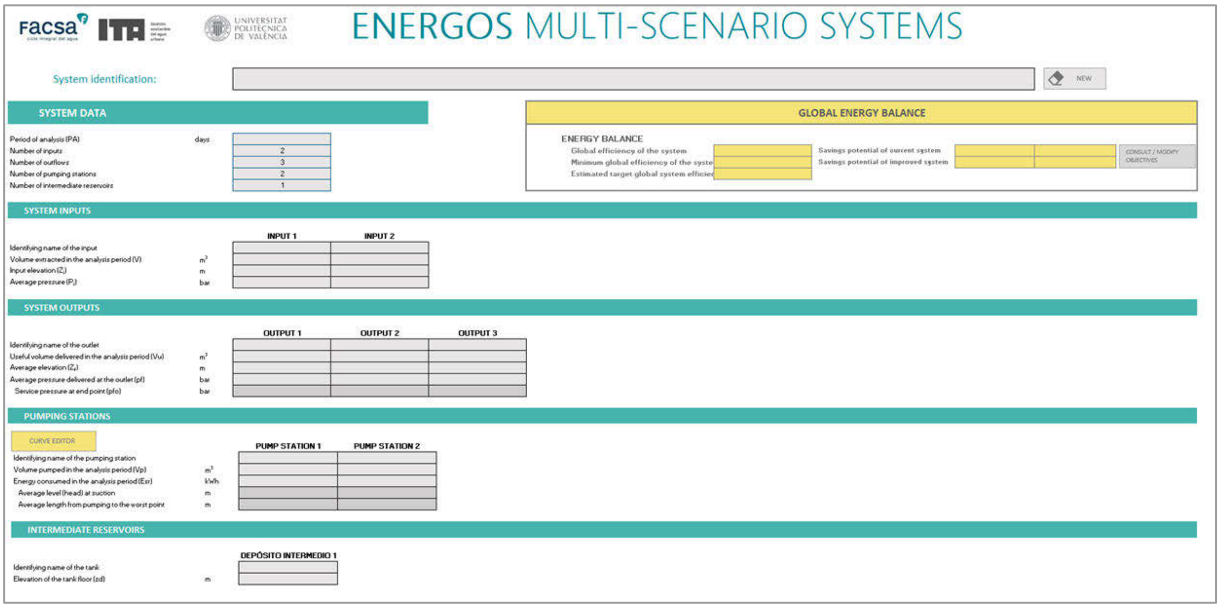
Figure 7. Selection of network knowledge

Each of these buttons leads to a different form. It is possible to fill in the data from the different forms, allowing comparisons between some results and others depending on the defined network detail.

Once all the data has been defined, the tool offers the system diagnosis, comparing the current state with the ideal state (unachievable). To define the target performance, target values must be specified for each of the operational losses in a similar way as in simple systems.

3.3 MULTI-SCENARIO SYSTEMS

This type of system is substantially different from the other two systems described. First, the basic data of the system are defined. That is, the number of inputs, outputs, pumping stations and intermediate reservoirs. Then, for each of these inputs and outputs, the average values of volume, pressures and heights that characterize them are defined. With the data entered, an overall energy balance is carried out to calculate the overall system efficiency, the minimum overall efficiency and the target efficiency.



ENERGOS MULTI-SCENARIO SYSTEMS

System identification:

SYSTEM DATA

Period of analysis (PA)	days	
Number of inputs		2
Number of outflows		3
Number of pumping stations		2
Number of intermediate reservoirs		1

GLOBAL ENERGY BALANCE

ENERGY BALANCE

Global efficiency of the system

Minimum global efficiency of the syste

Estimated target global system efficie

Savings potential of current system

Savings potential of improved system

CONSULT / MODIFY OBJECTIVES

SYSTEM INPUTS

		INPUT 1	INPUT 2
Identifying name of the input			
Volume extracted in the analysis period (V)	m^3		
Input elevation (Z_i)	m		
Average pressure (P_i)	bar		

SYSTEM OUTPUTS

		OUTPUT 1	OUTPUT 2	OUTPUT 3
Identifying name of the outlet				
Useful volume delivered in the analysis period (Vu)	m^3			
Average elevation (Z_f)	m			
Average pressure delivered at the outlet (pf)	bar			
Service pressure at end point (pfo)	bar			

PUMPING STATIONS

CURVE EDITOR

		PUMP STATION 1	PUMP STATION 2
Identifying name of the pumping station			
Volume pumped in the analysis period (Vp)	m^3		
Energy consumed in the analysis period (Esr)	kWh		
Average level (head) at suction	m		
Average length from pumping to the worst point	m		

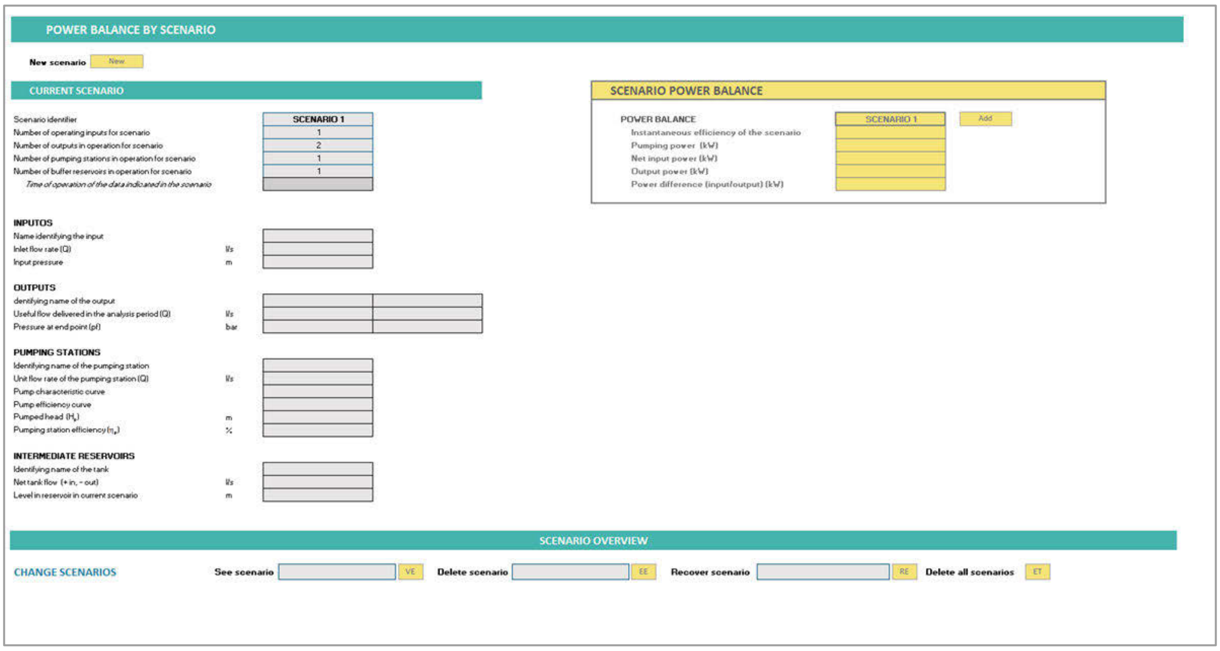
INTERMEDIATE RESERVOIRS

		DEPÓSITO INTERMEDIO 1
Identifying name of the tank		
Elevation of the tank floor (zd)	m	

Figure 8. Overall energy balance for multi-scenario systems

The overall energy balance (figure 8) considers that all inputs and outputs are in operation throughout the analysis period, for this reason average values are requested for the indicated period, since different operating modes are not contemplated in this analysis.

In order to be able to analyze the different modes of operation, which imply changes in the layout, the possibility of defining different scenarios is contemplated. For each of the scenarios, it is indicated which of all the elements defined at the beginning are involved, i.e. for each scenario which are the inlets, outlets, pumping stations and intermediate reservoirs. For each of the elements involved in the scenario, the specific values of the scenario are now requested, i.e., in the case of the inflows, for example, what is the value of the flow delivered for that scenario, similarly the data of the outflows or pumping stations are requested.



POWER BALANCE BY SCENARIO

New scenario New

CURRENT SCENARIO

Scenario identifier		SCENARIO 1
Number of operating inputs for scenario		1
Number of outputs in operation for scenario		2
Number of pumping stations in operation for scenario		1
Number of buffer reservoirs in operation for scenario		1
Time of operation of the data indicated in the scenario		
INPUTOS		
Name identifying the input		
Inlet flow rate (Q)	l/s	
Input pressure	m	
OUTPUTS		
dentifying name of the output		
Useful flow delivered in the analysis period (Q)	l/s	
Pressure at end point (pf)	bar	
PUMPING STATIONS		
Identifying name of the pumping station		
Unit flow rate of the pumping station (Q)	l/s	
Pump characteristic curve		
Pump efficiency curve		
Pumped head (H_b)	m	
Pumping station efficiency (η_b)	%	
INTERMEDIATE RESERVOIRS		
Identifying name of the tank		
Net tank flow (+ in, - out)	l/s	
Level in reservoir in current scenario	m	

SCENARIO POWER BALANCE

POWER BALANCE	SCENARIO 1	Add
Instantaneous efficiency of the scenario		
Pumping power (kW)		
Net input power (kW)		
Output power (kW)		
Power difference (input/output) (kW)		

SCENARIO OVERVIEW

CHANGE SCENARIOS

See scenario VE Delete scenario EE Recover scenario RE Delete all scenarios ET

Figure 9. Global power balance for multi-scenario systems

Once all the scenario data has been defined, a power balance is calculated (Figure 9) for the specific operation of the system in that scenario. The tool allows comparing the different scenarios so, with a simple glance, it is possible to identify which of the system's operating modes are the most efficient or the ones that consume more power

4 CONCLUSION

ENERGOS is a very simple tool for diagnosing pressurized water transport systems. The proposed diagnosis requires very little data and allows to establish the margin of improvement the system has, comparing it with the ideal state (without losses), with the current state and with an objective intermediate state that establishes reasonable loss limits.

The diagnosis is adapted to the different types of systems proposed, simple systems, water networks and multi-scenario systems, and although the procedure is generally the same, the indicators analyzed are adapted to each type of system.

5 REFERENCES

[1] EC (European Commission), 2011. Directive of the European parliament and of the council on energy efficiency and repealing directives 2004/8/EC and 2006/32. European Commission

[2] EC (European Commission), 2019. Fourth report on the state of the Energy Union. European Commission. Brussels, 9.4.2019 COM (2019) 175 end

[3] Cabrera, E.; del Teso, R.; Gómez, E.; Estruch-Juan, E.; Soriano, J. “Quick energy assessment of irrigation water transport systems”. Biosystems Engineering 2019, 188, 96–105.

[4] Cabrera E., Gómez E., Cabrera E. Jr, Soriano J., and Espert V. "Energy Assessment of Pressurized Water Systems", Journal of Water Resources Planning and Management, 2015. Vol. 141, No. 8.

[5] Cabrera E., Abreu J., Gómez E., Del Teso R., Estruch E., "Analysis of different approaches to improve the energy efficiency of water transport systems", Journal of Water Resources Planning and Management, (Revision)

[6] Cabrera, E., Pardo, M. E., Cobacho, R., Cabrera, E., "Energy Audit of Water Networks." Journal of Water Resources Planning and Management, 2010b, 136 (6): 669-677

[7] Del Teso R., Gómez E., Estruch M.E., Cabrera E., "Topographic energy management in water distribution systems." Water Resources Management. 2019, Vol 33, nº 12, pp 4385-4400

[8] Alegre H., Melo Baptista J., Cabrera E., Cubillo F., Duarte P., Hirner W., Merkel W., Parena R., Performance Indicators for Water Supply Services: Third Edition. 372 pages. IWA Publishing. London. UK, 2016

[9] ERSAR & LNEC. "Guide to the quality of service assessment for drinking water supply, urban wastewater management and municipal waste management services." 2nd generation of the assessment system (in Portuguese). 2nd ed. Lisbon, 2013

[10] Papa F., Rita Cavaleiro de Ferreira R. and Radul D. "Pumps: Energy Efficiency & Performance Indicators. Efficient" PI 2015 Joint Specialist IWA International Conference Cincinnati, USA.

[11] Pérez-Urrestarazu L. and Burt C. "Characterization of Pumps for Irrigation in Central California: Potential Energy Savings. Journal of irrigation and drainage engineering". ASCE. September 2012, pp 815-822.

[12] Cabrera E., Gómez E., Soriano J. and del Teso R." Eco-Layouts in Water Distribution Systems". Journal of Water Resources Planning and Management. DOI: 10.1061/(ASCE) WR.1943-5452.0001024. 2018b

[13] EC (European Commission), 2009. Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009, establishing a framework for the setting of ecodesign requirements for energy-related products. Official Journal of the European Union. 31.10.2009.

[14] EC (European Commission), 2012. Commission regulation (EU) No 547/2012 of 25 June 2012 implementing Directive 2009/125/EC of the European Parliament and of the Council with regard to ecodesign requirements for water pumps. Official Journal of the European Union 26.6.2012.

[15] ZVEI (German Electrical and Electronic Manufacturers' Association), 2015. Energy Efficiency with Electric Drive Systems. ZVEI. Electric Drive Systems Section. Frankfurt am Main, Germany.

[16] USDE (U.S. Department of Energy), 2012. Adjustable Speed Drive Part-Load Efficiency. Energy Efficiency & Renewable Energy. Washington. USA.

[17] NRCS (Natural Resources Conservation Service. United States Department of Agriculture), 2010. Variable Speed Drive (VSD) for Irrigation Pumping. Engineering Technical Note No. MT-14.

[18] Tsoumas P., Tischmacher H. and Köllensperger P., 2014. The European Standard EN 50598-2: Efficiency Classes of Converters and Drive Systems. International Conference on Electrical Machines (ICEM), pp 929 – 935..

[19] Bunge, Mario. "A General Black Box Theory." Philosophy of Science, vol. 30, no. 4, 1963, pp. 346–358.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14774

PYSIMDEUM – AN OPEN-SOURCE STOCHASTIC WATER DEMAND END-USE MODEL

D.B. Steffelbauer[1], B. Hillebrand[2] and E.J.M. Blokker[3]

[1]Hydroinformatics Group, KWB - Kompetenzzentrum Wasser Berlin , (Germany)
[2]KWR – Watercycle Research Institute, Nieuwegein, (the Netherlands)
[3]KWR – Watercycle Research Institute, Nieuwegein, (the Netherlands), Delft University of Technology, Delft, (the Netherlands)

[1] *david.steffelbauer@kompetenz-wasser.de,* [2] *bram.hillebrand@kwrwater.nl,*
[3] *mirjam.blokker@kwrwater.nl*

Abstract

Water demand is a crucial input parameter in water distribution system analysis because it can fluctuate over various temporal and spatial scales. In the past, researchers developed stochastic models that can provide realistic consumption patterns for simulations to account for those demand dynamics. Parameters for stochastic models are usually retrieved by fitting these models on smart water meter data. The stochastic demand model SIMDEUM uses an entirely different approach by generating highly realistic water demands based on (country-specific) statistical information only, without the need for measurements. While this approach makes SIMDEUM widely applicable in the water sector, its widespread usage within the community has been hindered due to its software implementation and availability.

We produced pySIMDEUM, an open-source and object-oriented implementation of the SIMDEUM software in the popular and freely available programming language Python. The pySIMDEUM software package is not only publicly available for usage within the water field — it is also intended to build the cornerstone of a widespread pySIMDEUM community of active developers. We want to use the WDSA/CCWI conference to address interested researchers or practitioners in the water sector and invite them to contribute to the software package as active part of the pySIMDEUM community. We will show SIMDEUM's history and past applications, the mathematical approach behind SIMDEUM and pySIMDEUM, where to download and install the pySIMDEUM package, the structure of the program, and a minimal example of how easily pySIMDEUM can be used to generate realistic stochastic water demand patterns from scratch. Furthermore, we will highlight possible future applications of the new pySIMDEUM tool. These applications include automatic parametrisation of pySIMDEUM parameters on smart meter data, coupling stochastic demands directly with hydraulic solvers, or how to enable city-scale stochastic demand simulations.

Keywords
Smart metering, water demand analysis, demand management, hydraulic simulations, water consumption, open-source software, Python.

1 INTRODUCTION

Water utilities use simulation models to design and operate distribution systems efficiently. In those models, water demand is a crucial input parameter that can fluctuate over various temporal and spatial scales. Historically, demand is modelled based on limited data through time-wise sparse consumption measurements (e.g., monthly, quarterly, yearly) and zonal meters with higher time resolution but are only available at a few locations. This approach resulted in inaccurate computer models, e.g., when simulating water quality. Researchers developed novel stochastic demand models that simulate water use with a high temporal and spatial resolution to overcome those inaccuracies, e.g., Poisson rectangular pulse or Neyman cluster process models [1]. Model

CC BY NC SA

parameters for those novel demand models are usually estimated through fits on consumer data, e.g., from Automatic Meter Readings (AMRs) data.

The stochastic demand model SIMDEUM [2] uses an entirely different approach by simulating end-use pulses based solely on statistical information of users and appliances, all without needing expensive measurement campaigns. This different approach makes SIMDEUM highly flexible in its application, especially for zones lacking AMRs. However, the applicability of SIMDEUM has some limitations, which are partly connected to its implementation in MATLAB. First, as proprietary software, MATLAB is accompanied by high license costs, hindering the parallel execution of SIMDEUM simulations in modern multi-core computing environments such as high-performance clusters. This limitation prevents the scalability of SIMDEUM from performing simulations for large networks in a reasonable time. Second, SIMDEUM consists of MATLAB structures and direct calls to its library. The SIMDEUM source code requires the user to be aware of all the different functions offered by SIMDEUM to implement custom-tailored simulations or if the user wants to enhance SIMDEUM's capabilities.

Consequently, the lack of an object-oriented software engineering approach hinders open-source development. To resolve all formerly mentioned limitations, we introduce pySIMDEUM, which simulates water demand using the same methodology as the MATLAB version of SIMDEUM, but is implemented in an object-oriented way in the freely available programming language Python. We chose Python because

1. it is nowadays one of the most popular programming languages, which replaces or already replaced MATLAB in the scientific community,
2. its simple syntax is easy to use (e.g., for Civil Engineering students not familiar with high-level programming),
3. it possesses a rich ecosystem of packages for data science, optimisation, and water distribution modelling (e.g., WNTR or OOPNET.

We will show where to download and install the pySIMDEUM package, the mathematical approach behind SIMDEUM and pySIMDEUM, how this translates naturally into an object-oriented program in Python, and a minimal example of how to use pySIMDEUM to generate realistic stochastic water demands from scratch. Additionally, we want to use the WDSA/CCWI presentation to show how users can contribute to the pySIMDEUM open-source initiative. Furthermore, we will highlight possible future applications of the new pySIMDEUM tool. These applications include automatic parametrisation of pySIMDEUM parameters on smart meter data, coupling stochastic demands directly with hydraulic solvers, or enabling city-scale stochastic demand simulations.

2 PURPOSE AND MATHEMATICS BEHIND SIMDEUM

SIMDEUM, abbreviated for SIMulation of water Demand and End-Use Model, is a stochastic drinking water demand model with a small spatial (end-use or device level) and temporal scale (up to one second). Initially intended to be a supporting tool for water quality modelling [2], for example, to calculate maximum demands within water distribution networks and the resulting maximum velocities to design self-cleaning networks related to discolouration, SIMDEUM has found many other applications, both in research and practice. The applications range from water quantity modelling (e.g., demand pattern generation for bottom-up and top-down allocation), water quality modelling (e.g., computing residence times in distribution networks, chlorine decay calculations, water temperature and bacterial growth simulations), design of distribution systems as well as drinking water installations, up to the prediction of future demands under changes in demographics or water usage behaviour. Additionally, novel use cases led to further developments and add-ons to the SIMDEUM software package (e.g., SIMDEUM hot and cold water use package or SIMDEUM's discharge pattern simulator for computing thermal energy and nutrients recovery potentials). A detailed overview of the many applications for SIMDEUM can be found in [3].

2.1 Mathematics behind the SIMDEUM approach

The mathematics behind SIMDEUM is relatively simple. SIMDEUM simulates rectangular water consumption pulses (B) similar to Buchberger's and Wells' Poisson Rectangular Pulse (PRP) model from 1996 [4]. Within the model, each pulse is defined by its height or intensity (I), its duration (D), and its arrival time over the day or start time (τ). The intensity, duration, and start time of each pulse are drawn from probability distributions that are different for each type of end-use (e.g., shower, toilet flush, …) as well as the type of user (defined through the user's age, gender, and occupation), respectively. In contrast to the PRP model, where the pulse parameters are retrieved through fitting the model on data, SIMDEUM generates the pulses from more high-level statistical information retrieved from country-specific statistics and surveys and is, hence, not dependent on available measurement data.

SIMDEUM generates a water demand pattern by summing up the randomly drawn end-use pulses per end-use (k), user (j) and frequency of use (i)

$$Q(t) = \sum_{k=1}^{M} \sum_{j=1}^{N} \sum_{i=1}^{F_{jk}} B(I_{ijk}, D_{ijk}, \tau_{ijk})$$

$$B(I_{ijk}, D_{ijk}, \tau_{ijk}) = \begin{cases} I_{ijk} \ldots \tau_{ijk} < t < \tau_{ijk} + D_{ijk} \\ 0 \ldots \quad else \end{cases}$$

M is the number of all different end-use devices, N is the number of users in the household, and F_{jk} is the frequency of a user using a specific end-use. All parameters in the formula above (I, D, τ) are described by probability density functions. Those functions are different for each end-use type and can differ for the kind of person or the household. The probability functions and statistics are taken from statistical information, in particular, from data taken from surveys and questionnaires in the Netherlands.

2.2 Households, water users and end-use devices within SIMDEUM

SIMDEUM consists of three different main types of components, which are *households*, water *users*, and end-use *devices*. Three main household types are implemented: (i) *one-person* households, *two-person* households representing double-income-no-kids apartments, (iii) and *family* households consisting of two or more people that can also include children.

The inhabitants of the houses are water users. Each water user is characterised by their age, gender, and occupation. These characteristics impact the daily schedules of users and hence, their water usage. Daily schedules are generated from probability distributions linked to statistics that depend on age, gender, and profession. For example, the profession defines whether a user has a job and leaves the house for work during the day. Consequently, this user's water usage at home during work hours is zero.

SIMDEUM takes eight different end-use devices into account, i.e., *bathtub, bathroom tap, dishwasher, kitchen tap, outside tap, shower, washing machine*, and *toilet*. The type of the house and the number of people living in it influence the end-use devices present in a home. For example, a dishwasher or a bathtub is more likely to be present in a house the more people live in it. Each of the end-uses devices can possess different end-use subtypes. The subtypes are distinct from one another by the average usage duration. For example, the bathroom tap can be used to wash, shave, and brush teeth. These different usage types possess different probability distributions and parameters connected to each end-use subtype. The frequency, duration, and intensity are drawn from those probability distributions.

The primary statistical information, i.e., specific probability distributions and their parameters (e.g., mean, standard deviation, etc.), are taken from surveys and questionnaires. For example, the frequency of use is taken from the residential water-use survey [5], while the statistics describing the daily schedules of the users (e.g., duration of sleep, time of leaving the house, …) are taken from the time-budget survey [6]. Household statistics (e.g., number of people per household, their gender, or their age distribution) are taken from statistics Netherlands [7]. A detailed description of the household statistics and end-use devices, end-use subtypes, and the specific probability distribution per end-use subtype can be found in [2].

3 THE PYSIMDEUM PYTHON PACKAGE

3.1 Shortcomings and limitations of SIMDEUM’s implementation

The variety of SIMDEUM’s practical applications demonstrated its widespread utility. However, its general use and development within the water community have been hindered by the software's availability, usability, and implementation.

First, SIMDEUM has been only available through directly communicating with its developers and signing a memorandum of understanding. After that procedure, the developers would send you the software. Direct contact with the developers might already be a barrier for users who simply want to try out suitable software for their use case. Additionally, SIMDEUM is written in the proprietary language MATLAB, which comes with high license costs.

Second, SIMDEUM was developed more than ten years ago when object-oriented programming was unavailable in MATLAB. That is why SIMDEUM is based on a workaround for a proper object-oriented approach, consisting of a mixture of MATLAB structures and cells, which is not ideal for the usability (e.g., the understanding of users of the software) nor enforcing collaboration between multiple developers as in an object-oriented software consisting of various sub-packages. Furthermore, the three main components of SIMDEUM (households, water users, and end-use devices) plus the statistical data are ideally implemented in software that uses an object-oriented structure.

3.2 Python implementation and overview of the object-oriented structure

To address the shortcomings and limitations of SIMDEUM’s MATLAB implementation, the decision was made to port SIMDEUM to Python and to refactor the software to enforce usability and collaboration through an object-oriented software design approach. Python is one of the most used programming languages worldwide, freely available, and includes a vast standard library that is topped up with a rich ecosystem of scientific software packages, all of them publicly available on *GitHub* (www.github.com) or the Python Package Index PyPI (*https://pypi.org*). Furthermore, Python is a multi-paradigm programming language that supports different programming approaches, including object-oriented programming. The usage of Python allows SIMDEUM to grow as an open-source software project and attract a big community of users, where each community member can contribute to the software’s source code.

The main components of SIMDEUM (households, users, devices, and end-uses) contain different information and statistics depending on their type. Each of them influences the other and translates naturally into an object-oriented structure. Slight differences or variations in similar objects can be implemented via inheritance. For example, a toilet with a water-saving flush option can be the child of a more general standard toilet. Additionally, pySIMDEUM contains a statistics object to store underlying (country-specific) statistical information that influences people’s water usage.

Figure 1 shows the overall structure of the pySIMDEUM package. The software flow is the following: First, a house is constructed based on underlying household statistics. Depending on the house type (one-person, two-person or family homes), the house is populated with water users.

Users are uniquely defined by age, gender, and job. The age defines the type of user. Four different age options exist, namely, (i) child, (ii) teen, (iii) adult, and (iv) senior. Each of them has different underlying statistics that define their daily schedules and water usage behaviours. For example, Dutch water use statistics show that teenagers shower more often than seniors. There exist two different options for the gender property (either male or female) and two different options for the job property, which have a strong influence on their daily water use schedule (either the user has a job where he leaves the house during the day [*True*], or the user stays at home [*False*]). The user's presence in the house and daily water use schedule are computed based on the user's properties and probability distributions defined in the (country-specific) statistics object.

Once the house is populated with users, the water use devices are generated by furnishing the house. This is again based on information contained in the statistics object. The users and devices present in a house define the possible end-uses with their statistical properties (i.e., underlying probability functions and parameters).

The house instance serves as a collector for storing different users and device objects (depicted as square brackets in Figure 1) and contains methods to simulate the water consumption of the entire household once users, devices and end-uses are defined.

The consumption is stored in a multi-dimensional labelled array, using the Python package *xarray* [10], with users, devices, samples and time as dimensions. The format of the output allows to access individual consumptions per user or device, compute statistics over all dimensions (e.g., mean, standard deviations, …), including moving averages over the time axis, algorithms to plot water consumption, and routines to write the results to different file formats (e.g., Excel, text files, or to netCDF). Additionally, wrappers exist to write consumption patterns to EPANET input files, allowing for directly connecting pySIMDEUM and hydraulic simulations.

It has to be noted that pySIMDEUM is highly flexible: All statistical information, household, user and device properties can be changed in runtime, either by directly changing the properties (of houses, users, devices, or end-uses), changing (default) values within the construction methods, or changing the values stored in the statistics object directly. Additionally, the user can introduce new objects (e.g., novel end-use devices) by inheriting from similar devices and adding the novel device simply to the method that furnishes the houses. This gives pySIMDEUM users total control over the software to tailor the software and statistics based on their needs or use-case.

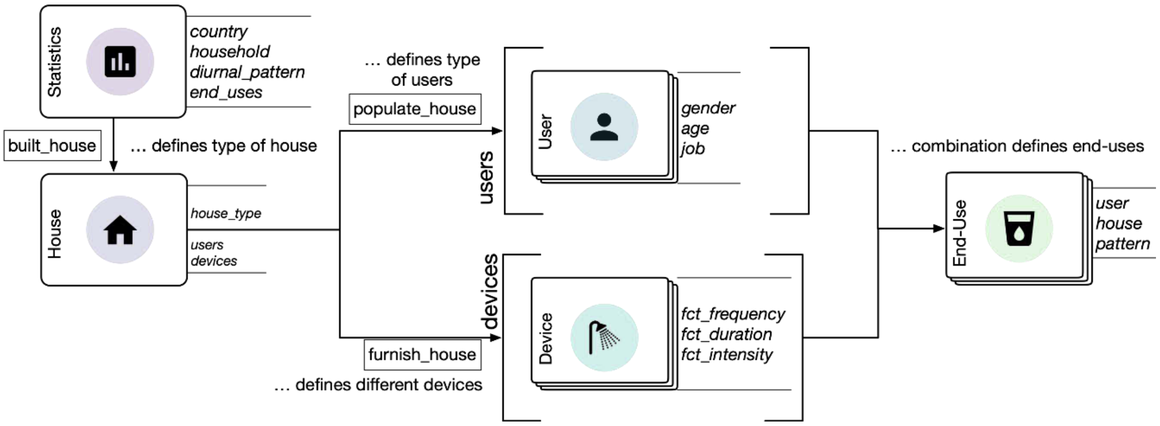


Figure 1: Object-oriented structure of pySIMDEUM

cc BY NC SA

3.3 Source code repository, installation and dependencies

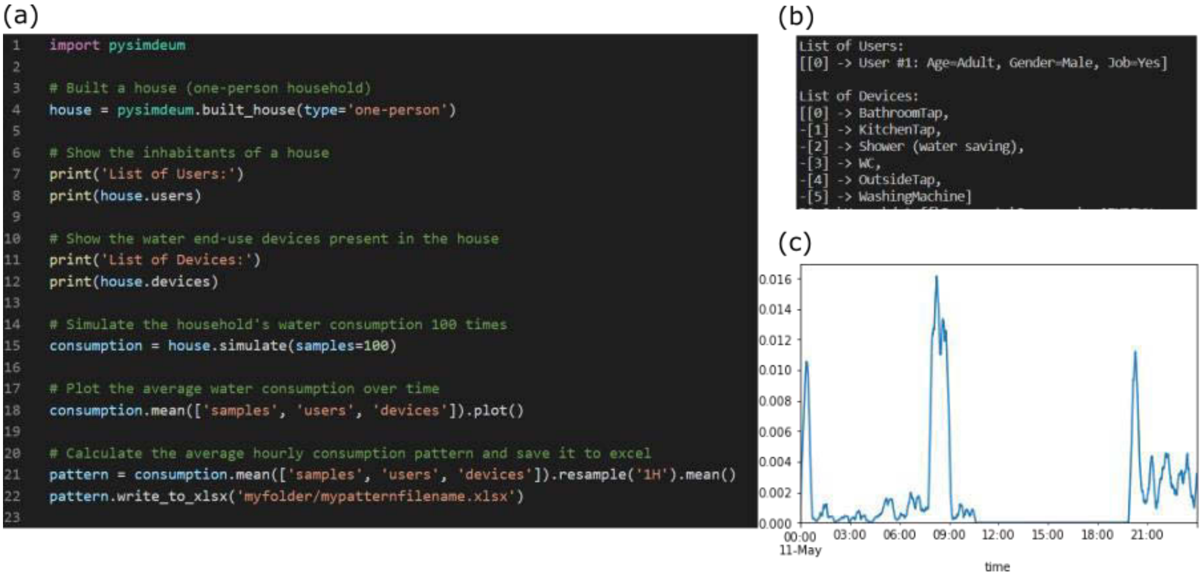
The pySIMDEUM software is available under the European Union Public License 1.2 on GitHub (from the conference's start date) under the following URL: https://github.com/KWR-Water/pySIMDEUM.

GitHub is a source-code hosting service that offers distributed version control and source code management based on the version control system Git. Additionally, it provides access control and advanced collaboration features, such as bug tracking, feature requests, task management, continuous integration, and wiki pages. The pySIMDEUM GitHub repository contains not only pySIMDEUM's source code but also a short README and a minimal usage example in the form of a Jupyter notebook, plus the complete documentation of the software. Furthermore, a link is available on the repository to the online documentation of pySIMDEUM on *Read the Docs* (https://readthedocs.org). Collaborators have to create a GitHub user account to contribute to pySIMDEUM, but anyone worldwide can download the repository.

Additionally, pySIMDEUM is hosted on PyPI. Hence, pySIMDEUM can also be installed via the command line by typing '*pip install pysimdeum*', which Python users are familiar with.

pySIMDEUM is available for Python version 3.7 or newer. The reason for this is that the objects in the software are implemented as Python data classes. This is a new feature for classes mainly containing data, which allows type hinting and saves a lot of boilerplate code (e.g., class initialisation and representation functions), making the source code more readable, understandable, and easier to maintain for pySIMDEUM developers. Further dependencies of the pySIMDEUM package include standard Python packages such as *numpy*, *scipy*, *matplotlib*, *pandas*, *xarray*, and *wntr* [8] for writing pySIMDEUM water usage patterns to EPANET input files.

3.4 Minimal usage example



(a)

```
1   import pysimdeum
2
3   # Built a house (one-person household)
4   house = pysimdeum.built_house(type='one-person')
5
6   # Show the inhabitants of a house
7   print('List of Users:')
8   print(house.users)
9
10  # Show the water end-use devices present in the house
11  print('List of Devices:')
12  print(house.devices)
13
14  # Simulate the household's water consumption 100 times
15  consumption = house.simulate(samples=100)
16
17  # Plot the average water consumption over time
18  consumption.mean(['samples', 'users', 'devices']).plot()
19
20  # Calculate the average hourly consumption pattern and save it to excel
21  pattern = consumption.mean(['samples', 'users', 'devices']).resample('1H').mean()
22  pattern.write_to_xlsx('myfolder/mypatternfilename.xlsx')
23
```

(b)

```
List of Users:
[[0] -> User #1: Age=Adult, Gender=Male, Job=Yes]

List of Devices:
[[0] -> BathroomTap,
-[1] -> KitchenTap,
-[2] -> Shower (water saving),
-[3] -> WC,
-[4] -> OutsideTap,
-[5] -> WashingMachine]
```

Figure 2: Minimal usage example for running a pySIMDEUM simulation: (a) code of the program, (b) Python console output, (c) graphical output of the simulation.

Figure 2 (a) shows the Python source code for simulating a water consumption pattern for a one-person household. First, the pysSIMDEUM package is imported into Python (Line 1), enabling the user access to pySIMDEUM's application programming interface (API). Then, a house is generated

BY NC SA

through the ‘built_house’ routine. The user can either specify a specific house type (i.e., a *one-person* household in this case) or if the type is not specified, pySIMDEUM will generate a random household (*one-person*, *two-person* or *family*) based on the underlying pySIMDEUM statistics. Furthermore, the number of inhabitants and the appliances within the house can be specified through the function’s keywords ‘*num_users*’ or ‘*devices*’, respectively. Additionally, instead of providing pySIMDEUM with the number of users, a list containing specific users can be pre-defined and given to the house constructor via the ‘*users*’ keyword. All information that is not particularly given to the function will be drawn from random distributions defined within pySIMDEUM’s underlying statistical information.

The house inhabitants can be depicted by printing the ‘*users*’ property of the generated ‘*house*’ object to the Python console (see Line 8 in the source code and the output in Figure 2 (b)). The printed information contains a list of users with the user’s index in the list (0 in this case), and the essential properties of the user, i.e., their age, gender, and profession. The generated user in the example is a male adult with a job so the user will leave the house during working days. The same print function can be used to depict the end-use devices present in the household by printing the ‘*devices’* property of the house (Line 12 in Figure 2 (a)). In this case, six water end-use devices are present in the house, a bathroom, a kitchen and an outside tap, a shower, a toilet, and a washing machine (see the output in Figure 2 (b)). Standard devices within pySIMDEUM are printed without additional information, while special devices are printed with other information in brackets. For example, the shower present in the one-person household is a water-saving shower, so the water usage intensity is less when a user takes a shower. Consequently, the water usage for shower events is lower than for houses having a standard shower.

In Line 15, the water consumption of the household is simulated. In particular, 100 samples are generated for the stochastic water usage of this specific house. The results are stored in a multi-dimensional labelled array using the *xarray* package. The array contains four dimensions: time, users, devices, and samples. Hence, each end-use device, user, sample, or point in time can be extracted by the user.

Furthermore, *xarray* allows building statistics over time over each dimension or combinations of dimensions. For example, the average consumption over time of the house can be retrieved by calculating the mean over the samples, users, and device dimensions (see Figure 2 (a) Line 18). Additionally, the statistics can be plotted (see Figure 2 (c)). It can be seen that the user’s water consumption shows high peaks, one in the morning and one in the evening when the user returns from work. During the day, the user’s consumption in the house is zero since he is not at home. Consumption during the night-time is low since the user is generally asleep. However, pySIMDEUM assumes that approximately 1.5 % of the water consumption happens during this time since users can get up during the night and use water, for example, when using the toilet. Of course, this parameter can be changed in pySIMDEUM as well. Finally, Line 21 in Figure 2 (a) shows how to generate water usage patterns by building averages over time intervals (i.e., one hour). Line 22 shows how to store the pattern subsequently in an MS Excel file with the ‘*write_to_xlsx*’ method.

4 CONCLUSION

SIMDEUM is a stochastic water demand end-use software that generates highly realistic drinking water demand patterns from statistical information without demand measurement data. While SIMDEUM has been used for many applications in the water sector, its widespread usage by researchers and practitioners in the water distribution system analysis field has been hindered mainly by its MATLAB implementation and availability. This paper presented pySIMDEUM, an open-source and object-oriented implementation of the SIMDEUM software in the popular and freely available programming language Python. The pySIMDEUM software package is publicly available for usage within the water field and is intended to build the cornerstone of a widespread pySIMDEUM community of active developers. That is why we hosted the software on GitHub to enforce the joint development of pySIMDEUM. Every interested researcher or practitioner in the water sector is invited to contribute to the software package actively.

Currently, pySIMDEUM produces water demand patterns in the same way as the original SIMDEUM software. But in future, we expect basic SIMDEUM functionality to be extended, allowing to simulate, for example, nutrient loads or realistic dry weather flows in sewers, test the effect of novel water end-use devices on the drinking water supply system, simulate the potential of peak demand load shifting and, minimise the energy consumption of drinking water systems, or in modelling premise plumbing systems, where water usage on small spatial and temporal scale plays a crucial role.

Furthermore, the implementation of the software in the non-proprietary programming language Python allows the parallel execution of SIMDEUM simulations on high-performance computers (HPCs). Due to the high licensing costs coming with MATLAB, this was impossible in the past. Consequently, massive pySIMDEUM simulations on HPCs can be used to calculate realistic water consumption on a high temporal and spatial resolution on the city scale. This will also allow to couple pySIMDEUM simulations with smart meter data: Instead of tediously retrieving country-specific statistics necessary for the SIMDEUM approach, pySIMDEUM parameters might be retrieved automatically from smart meter data and machine learning algorithms [9]. The underlying information and statistics of SIMDEUM might be learned to produce even more realistic water usage patterns tailored to specific case study areas.

5 REFERENCES

[1] E. Creaco, P. Kossieris, L. Vamvakeridou-Lyroudia, C. Makropoulos, Z. Kapelan, and D. Savic, "Parameterizing residential water demand pulse models through smart meter readings," Environmental Modelling & Software, vol. 80, pp. 33–40, 2016.

[2] E. J. M. Blokker, J. H. G. Vreeburg, and J. C. van Dijk, "Simulating Residential Water Demand with a Stochastic End-Use Model," Journal of Water Resources Planning and Management, vol. 136, no. 1, pp. 19–26, 2010.

[3] E. J. M. Blokker, C. Agudelo-Vera, A. Moerman, P. van Thienen, und I. Pieterse-Quirijns, „Review of applications for SIMDEUM, a stochastic drinking water demand model with a small temporal and spatial scale", Drinking Water Engineering and Science, vol. 10, Nr. 1, S. 1–12, 2017.

[4] S. G. Buchberger und G. J. Wells, „Intensity, Duration, and Frequency of Residential Water Demands", ASCE Journal of Water Resources Planning and Management, vol. 122, Nr. 1, S. 11–19, Jan. 1996.

[5] Foekema, H., and Engelsma, O. "A changing consumption pattern (water use at home 2001)", TNS NIPO, Amsterdam in Dutch, 2001.

[6] Van der Broek, A., and Breedveld, K. "Time-budget survey 1995"- TBO'95 computer file, Sociaal en Cultureel planbureau, Amsterdam in Dutch, 1995.

[7] CBS. 2007. http://statline.cbs.nl accessed on Dec. 1, 2007.

[8] Klise, K.A., Bynum, M., Moriarty, D., Murray, R. "A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study", Environmental Modelling and Software, 95, 420-431, 2017.

[9] Steffelbauer, D. B., Blokker, E. J. M., Buchberger, S. G., Knobbe, A., and Abraham, E. "Dynamic time warping clustering to discover socioeconomic characteristics in smart water meter data." ASCE Journal of Water Resources Planning and Management, 147(6), 04021026. 2021.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14775

DESIGNING ADVANCED ASSET MANAGEMENT FOR WATER DISTRIBUTION NETWORKS: APPLICATION TO A REAL CASE STUDY

Daniele B. Laucelli[1], Serena Spagnuolo[2], Antonio Rinaldi[3], Gianluca Perrone[4], Luigi Berardi[5], Orazio Giustolisi[6]

[1,6] Politecnico di Bari, Bari (Italy)
[2]Acquedotto Pugliese S.p.A., Bari (Italy)
[3]Studio AC3 Ingegneria S.r.l., Ferdinando di Puglia (Italy)
[4]IA.ING S.r.l., Lecce (Italy)
[5]University of Chieti-Pescara "G. D'Annunzio", Pescara (Italy)

[1] *danielebiagio.laucelli@poliba.it,* [2] *s.spagnuolo@aqp.it,* [3]*a.rinaldi@studioac3.com,* [4]*perrone@iaing.it,*
[5] *luigi.berardi@unich.it,* [6] *orazio.giustolisi@poliba.it.*

Abstract

In recent years, water utility companies are facing several challenges in managing water distribution networks caused by the combined effect of aging, network complexity, changing water needs and availability. In this sense, the main trend to be countered is the increase in the deterioration of pipelines, therefore of water losses and finally the increasing probability of pipe failures. Leaks represent a significant challenge for water services in terms of technical, economic and safety damage, as well as reliability of service to users. Therefore, it is necessary to support technicians and water utilities in the effective management of water distribution networks in the short and long term, allowing them to gain a better understanding of network behaviour and to implement a more efficient use of water. To this end, optimal design solutions can be developed for asset management based on an integrated approach that incorporates advanced hydraulic modelling, network monitoring and technical-engineering knowledge of the network.

This work presents a complete methodology for water distribution network asset management, integrating optimal district metering areas design and pressure control with optimal pipe replacement. The case study presented concerns the city of Bari, the capital of the Puglia region, in southern Italy, a municipality of over 300,000 inhabitants, as part of a project tender for the improvement of the asset managed by Acquedotto Pugliese, the largest water company in Italy. Each designed activity in the methodology has been evaluated in terms of leakage reductions performances, under given economic constraints, to fully support technicians and water utilities in addressing water distribution network management plans.

Keywords
Asset management, pressure control, DMA design, optimal pipe replacement, real scale application.

1 INTRODUCTION

In the early twentieth century, the aim of Water Distribution Networks (WDN) was to support economic-industrial development and provide fire protection. The objective of the modelling studies was the development of hydraulic verification criteria of the WDN projects with respect to water requirements of the various types of users (civil, commercial, industrial, fire prevention). The WDN hydraulic simulation models had to calculate the operating pressures at the network nodes, given the pipe roughness and nodal demands. The design, therefore, became an assessment of nodal pressures with respect to the minimum pressures for a correct service to users and the minimum residual flow rates and pressures for the correct hydraulic performance of hydrants. On these assumptions, with the advent of automatic calculation, Todini [1] introduced the Global Gradient Algorithm (GGA), which a few years later became the hydraulic engine of EPANET, developed by Rossman [2]. Most current software packages are based on such a hydraulic engine or similar strategies. In general, all classic hydraulic simulators are born for hydraulic verification, and not for management purposes since they are based on the assumption of fixed demands at WDN nodes. This is called *Demand-Driven Analysis* (DDA).

Over the years, WDNs have become increasingly large, complex and aged, implying management needs with respect to volumetric losses, reliability, water quality, energy optimization, rehabilitation, etc. Todini [3] noted the need for a hydraulic simulation modelling that was able to evaluate the "effective" demand that can be supplied to users in pressure deficit conditions (i.e., pressure lower than the minimum required for correct service), therefore suitable for the new management tasks to be faced for a water company. Such hydraulic simulations are now called *Pressure-Driven Analyses* (PDAs). Several studies proposed the model of Wagner et al. [4], to represent the link between pressure at the network nodes and water demand actually supplied to users, such as Giustolisi et al. [5]. This model has proved consistent with the real hydraulic operation of the network where users statistically control the flow rates at the taps if the pressures are sufficient, while drawing the maximum flow rate from them (i.e., the maximum volume in a certain time interval), allowed by the pressure available when the network is in deficit conditions.

PDA analysis can represent different demand components in addition to the human one. In the last decade, researchers have investigated the possibility of making hydraulic simulation consistent with support for management decisions. From this perspective, representing water losses having a volumetric effect in the network would be useful for several management activities such as, for example, pressure control or rehabilitation planning. Such kind of losses are widely spread in the network and cannot be repaired individually since they are small (i.e., dripping). They produce significant volumes of water lost over time. Giustolisi and Walski [6] framed the representation of leakages, and other demand components, in the PDA simulation scheme, by developing the representation of the volumetric leakage component (*background leakages*) as a function of pressure at the single pipe level. Background leakages are an important indicator for asset management as they can define the state of deterioration of the network, albeit maintaining the information at the level of a single pipe. In fact, they are linked to the average pressures acting on the pipe; the advanced hydraulic simulation model from [3] based on that by Germanopoulos [7], assigns to each pipe a parameter (β) which can be considered a global indicator of deterioration and, therefore, of significant decision-making utility for any leakage management activity. It is important to clarify that representing the volumetric losses with hydrants concentrated in the network nodes, in addition to making the hydraulic calculation inaccurate from different points of view, causes the loss of information on deterioration at the individual pipe level (since it is concentrated a priori in the nodes), which is important for decision-making in rehabilitation. Furthermore, the classic hydraulic simulators (e.g., EPANET) are *demand-driven* while the volumetric leakages cannot be fixed in advance since they are a function of the average pressures on the pipes. Therefore, the WDN simulation for management purposes always requires PDA, both for the presence of volumetric losses and hydraulic verification of the users' water demands satisfaction.

Background leakages refer to small-flow leaks widespread in the hydraulic system, which should not be confused with those of higher flowrates (in technical-scientific literature called *burst leakages*). The generation and development of burst leakages, as known in international literature in these last fifty years, is closely related to the level of background leakages of which they are the natural evolution due to internal or external factors of the WDN, including: fatigue conditions due to traffic loads; operating pressure; thermic stress; laying conditions, etc.. Background leakages have a global influence on the WDN but generally do not affect the quality of the service in terms of minimum pressures, burst leakages are events that can have a local or global influence on the service depending on their position in the network and flow rate and in relation to the water

CC BY NC SA

capacity of the system. Therefore, the rationalization of asset management operations must be designed considering the reduction of background leakages as a relevant indicator of management control, which consequently has a positive effect on the generation and development of burst leakage. In this context, the replacement of old and deteriorated pipes must be done by optimizing the investment cost, but also with the aim of reducing background leakages. This objective cannot be achieved with a local vision, i.e., evaluating it with respect to the replaced pipes only, but globally, i.e., evaluating the extended hydraulic consequences of these replacements on the whole network. This is to avoid the effect of increased background leakages caused by the significant greater hydraulic conductance of the new pipes, which increases the operating pressures on the downstream portions of the WDN [8].

The paper shows a full-scale application of the methodology developed by us for asset management based on the optimal planning of districtualization, pressure control, hydraulic monitoring [9] and rehabilitation which has as a core the advanced modelling of the hydraulic behaviour of WDNs, through a calibration of the project model with an innovative paradigm based on water balances, integrating in the hydraulic analyses the water volumetric losses as a function of pressures [10]. This modelling is an essential support for an integrated and efficient design, with the aim of achieving the quality of the investment by minimizing costs with respect to results, and effective, with respect to any technical-regulatory constraint, for example as outlined by ARERA [11] for Italy. The following sections illustrate in order: the methodological approach adopted; the case study and application of the methodology for an executive planning of asset management interventions for the city of Bari, the capital of Puglia, in southern Italy, a municipality of over 300,000 inhabitants, as part of a public tender for the improvement of the asset managed by Acquedotto Pugliese (AQP), the largest Italian water company.

2 METHODOLOGICAL APPROACH

The advanced phenomenological modelling of WDN behaviour plays the role of a methodological tool to support the various planning and design activities. In fact, the innovative hydraulic model overcomes the limits of classical modelling, of EPANET and of all commercial software packages, and allows to consider the volumetric water losses as a function of the average pressure at the individual pipes level. In general, it allows a more realistic simulation of the WDN functioning, for example, considering the different types of users (private tanks and pumping, real elevation of georeferenced users meters, etc.) or insufficient pressure conditions for a correct service [6].

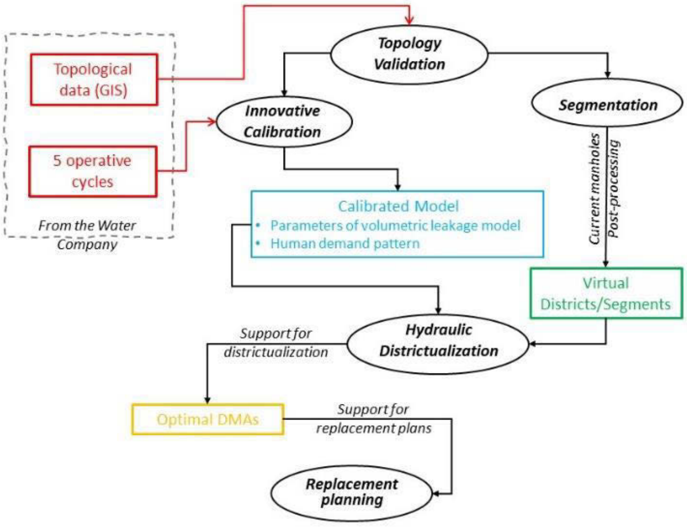

Topological data (GIS)
5 operative cycles
From the Water Company
Topology Validation
Innovative Calibration
Segmentation
Current manholes
Post-processing
Calibrated Model
• Parameters of volumetric leakage model
• Human demand pattern
Virtual Districts/Segments
Hydraulic Districtualization
Support for districtualization
Optimal DMAs
Support for replacement plans
Replacement planning

Figure 1. Flow-chart representing the adopted methodological approach.

BY NC SA

Based on advanced hydraulic analyses, multi-objective optimization procedures, integrated with technical constraints, allows to obtain design solutions that are efficient (in the cost/benefit ratio) and effective (linked to each WDN features), supporting from a managerial perspective both to the current and future design choices. In fact, the design methodology adopted offers design alternatives defined in a paradigm of flexibility, through a structured approach to asset management aimed also at future opportunities of the water company.

Figure 1 shows the methodological process used to support the executive planning of the asset management of the city of Bari WDN, detailed in the following sections. In the following, the individual phases, as indicated in Figure 1, are described in greater detail.

2.1 Topology validation, enhancement of SIT data and integration of consumption data

The topology validation consisted in checking isolated nodes and/or pipes, disconnected network portions, verification of elevations through DEM, coherence of diameters, etc., and was performed, during the import of GIS data provided by AQP as ESRI® shapefile format. The GIS information data was acquired from (i) geodatabase extracted from the company GIS, containing data relating to pipes, devices (valves) installed along the pipelines and georeferencing of user meters, (ii) indication of the recently replaced pipelines, (iii) consumption database for each user and (iv) data recorded at the remote-controlled meters with reference to the year 2019. These checks made it possible to evaluate the AQP source database both with respect to the network topology and the actual functioning of the system reproduced by the advanced hydraulic model, functional to carry out all subsequent design phases.

2.2 Calibration of the model based on the concept of mass balance

The calibration of the advanced hydraulic model, which includes the volumetric losses at single pipe level as a function of pressure, was performed using the innovative approach presented by Berardi and Giustolisi [10], based on the concept of mass balance. The strategy adopted aims to identify (i) the parameters of the volumetric water loss model and (ii) the demand curve of the different districts. The strategy also allows the simultaneous calibration of the hydraulic resistances of the most hydraulically important pipes when useful. To calibrate the system, the measures of the hydraulic quantities made available by AQP for the year 2019 were used, as well as the information regarding the state of the system, i.e., the topology of the network, the presence of closed or partialized valves, in particular referring to the sectioning between the subnets, and the setting of the pressure control valves (PCV). To obtain a robust model to support management and planning activities, the calibration was carried out on the basis of five real daily operating cycles representative of different states of the system, as a combination of working days/holiday, winter/summer conditions. The calibration made it possible to define the model of volumetric losses and the characteristic demand patterns of the different operating conditions. More details about the calibration methodology adopted in the case study are provided by [12] presented at this conference. It should be noted that, in this methodological framework, it would be useful to carry out measurements of hydraulic quantities before carrying out the works designed to allow for the evaluation of the system condition variation with respect to the hydraulic model calibrated with the 2019 conditions.

2.3 Districtualization methodology

The districtualization is an operation of dividing the network structure of a WDN into districts to improve its management with respect to the monitoring of water or mass balances. For this reason, the technical tradition speaks of District Metering Areas (DMAs). Normally, DMA planning is a trial-and-error operation with which, empirically, the technician divides the network into homogeneous segments from the point of view of a certain characteristic (e.g., elevation, material and pipes age, diameter, etc.), also considering technical constraints such as: vulnerability of the system in relation to the main water transfer pipes, operating pressures in relation to the height

BY NC SA

of buildings, types of users. However, classic districtualization is not an optimized design process and does not produce design choices that can be replicated and scaled to other networks. Furthermore, the classic DMA design does not evaluate the opportunities for controlling the pressures in the network, through the reconfiguration of water flows following the closure of sectioning valves, which can reduce pressures and therefore stresses on pipes, thus reducing both the failure of the system components (burst leakages), that volumetric losses. Finally, the classic districtualization does not integrate the design of pressure control with PCV setting and location design.

To overcome all this, adopting a rational, efficient and effective strategy, also guided by the need to reduce volumetric losses, an advanced districtualization procedure was used, divided into two phases: (1) topological segmentation; (2) hydraulic districtualization [9]. Topological segmentation is formulated as a two-goal optimization which, on the one hand, maximizes the modularity index which measures the efficiency of the topological division into segments/modules and, on the other hand, minimizes the number of conceptual cuts that separate the segments/modules [13]. The hydraulic districtualization, or rather the real DMA design, aims at the optimal selection, in each of the conceptual cuts of the segmentation, of (i) closed sectioning valves, to reduce volumetric losses in compliance with the requirements of a correct service to individual users; (ii) flow meters, in the minimum number possible, to create measurement districts; (iii) the set points (also variable) of the PCV. The motivation of the division into two phases is connected to the hydraulic functioning of the WDN conditioned in a dominant way by the topological connective structure of the network. Therefore, segmentation is a cost/benefit optimization process for the virtual division of the topology with conceptual cuts near the nodes to obtain segments or modules that respect given objectives and constraints. Hydraulic districtualization is a cost/benefit hydraulic optimization process for installing sectioning valves or flow meters in conceptual cuts (that can be seen as manholes). The technical purpose is to reconfigure the water flows inside the network by reducing the volumetric losses as an effect of pressure reduction, with the constraint of respecting the operating pressures required for the different users during the operating cycle. Finally, advanced modelling allows the planning of PCVs both with local control (immediately downstream the device) and with Remote Real-Time Control (RRTC) through a "sentinel" node.

2.4 Support for districtualization

The methodological scheme adopted supports the DMA design allowing the water company, once the sectioning valves are fixed, to choose the number of DMAs to monitor in an optimal way (costs/benefits), leaving the possibility of integrating the flow monitoring, activating the districts nested within the current DMAs. These districts are, in fact, already prepared in a "topologically" efficient way with respect to the sectioning valves determined with the hydraulic districtualization [9]. In particular, the strategy allows adopting various selection criteria that can be integrated with each other and are not exclusive, such as the “expert” choice, based on the judgment of the technicians based on the knowledge of the system as well as evaluations of the technical feasibility of the installations; or the "metrological" choice, based on the preliminary analysis of flow rates/speeds expected in the positions where the flow meters will be installed, with respect to their technical characteristics and the aim of reducing uncertainties on the district's water balances.

2.5 Support for the rehabilitation and replacement of old and deteriorated pipes

Rehabilitation is not a management option for a "massive" reduction of volumetric losses since replacement interventions affect, for economic reasons, a small percentage of the WDN. On the other hand, WDNs are commonly characterized by widespread deterioration in relation to the average age of the pipelines. In this context, the replacement of old and deteriorated pipes must be done by optimizing the investment cost, but also with the aim of reducing volumetric losses, as a management indicator of the overall effects of the rehabilitation on the network. This objective cannot be achieved with a local vision, evaluating it only with respect to the replaced pipes, but globally, evaluating the extensive hydraulic consequences of these replacements, on the whole network. This has led to the awareness, in the technical-scientific field, that rehabilitation based on a local methodological approach can become a negative element globally for the hydraulic system even if locally positive for the replaced pipelines.

In the proposed methodology, support for pipes replacement comes after the choice of DMA and setting of PCVs. In fact, the choice of pipes to be replaced with the best cost/benefit ratio, i.e., investments/reduction of volumetric losses, depends on the new hydraulic structure obtained following the reconfiguration of the flows. Therefore, once the hydraulic districtualization solution has been assigned, the rehabilitation support tool allows to identify the pipes which, for a given budget limit, maximize an efficiency index given by the ratio between the reduction in expected loss following the replacement and the intervention cost.

This decision support system for asset management, as well as being implemented in the WDNetXL-WDNetGIS platform, is provided by IDEA-RT through Digital Water services developed to allow the analysis and immediate comparison of design alternatives aimed at helping the water company understanding of solutions, and for the following needs to support the execution of works [14].

3 CASE STUDY: THE TOWN OF BARI – PUGLIA - ITALY

The methodological approach illustrated in the previous section was adopted as part of a public tender that AQP launched in relation to the new regulatory framework for the sector defined in Italy by ARERA, an agency that controls the water companies like OFWAT in the UK. In particular, the requested design operation is aimed at reducing losses and optimizing WDN management. Parameter M1a, introduced by ARERA, defined as the volume of water loss per km of network per day, is the indicator on which the reduction of losses is measured [11]. The results of the design described below, therefore, will refer to this indicator.

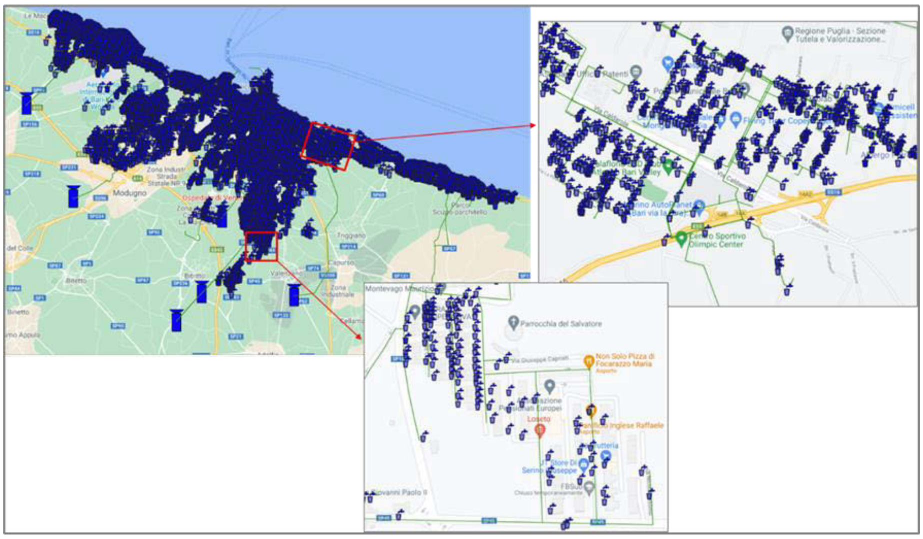


Figure 2. Model of the Bari network in QGis with georeferenced users.

The hydraulic model of the Bari network is made up of 9057 trunks and 7762 nodes. It was built starting from the acquisition of data provided by AQP in ESRI® shapefile format, with WDNetGIS [14]. The information on GIS data was acquired from (i) geodatabase extracted from the company GIS, containing data relating to pipes, devices (valves) installed along the pipelines and georeferencing of user meters, (ii) indication of the recently replaced pipelines, (iii) consumption

CC BY NC SA

database for each user and (iv) data recorded at the remotely controlled meters with reference to the year 2019. The model used considers the single georeferenced users as single points of consumption, and not aggregated in nodes as for the most of commercial software. This allows to better evaluate the effectiveness of management design with respect to the users. Figure 2 shows the Bari WDN in QGis with details on the representation of individual private users.

The Bari WDN is fed by eight reservoirs: (1) Palese-Santo-Spirito; (2) Bari-Modugno; (3) Nuovo di Bari-Bitritto; (4) Ceglie-Carbonara; (5) Bari-Ceglie del Campo; (6) Valenzano; (7) Torre a Mare; (8) Loseto. Figure 3 shows the altimetry of the Bari network with the aid of the 3D visualization of WDNetXL and indication of the flow meters at reservoirs.

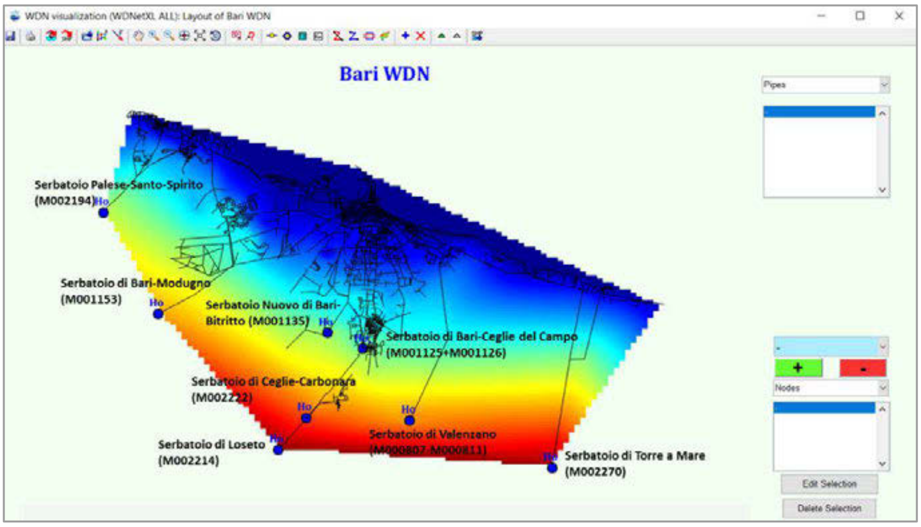

WDN visualization (WDNetXL ALL): Layout of Bari WDN
Bari WDN
Pipes
Serbatoio Palese-Santo-Spirito (M002194)
Ho
Serbatoio di Bari-Modugno (M001153)
Serbatoio Nuovo di Bari-Bitritto (M001135)
Serbatoio di Bari-Ceglie del Campo (M001125+M001126)
Serbatoio di Ceglie-Carbonara (M002222)
Serbatoio di Valenzano (M000807-M000811)
Serbatoio di Loseto (M002214)
Serbatoio di Torre a Mare (M002270)
+
-
Nodes
Edit Selection
Delete Selection

Figure 3. Altimetry of the Bari WDN by WDNetXL 3D visualization.

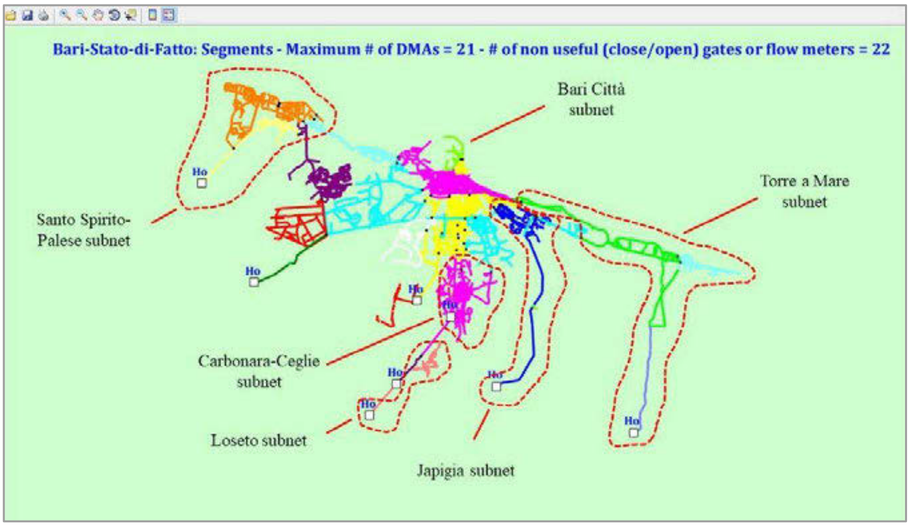

Bari-Stato-di-Fatto: Segments - Maximum # of DMAs = 21 - # of non useful (close/open) gates or flow meters = 22
Bari Città subnet
Torre a Mare subnet
Santo Spirito-Palese subnet
Carbonara-Ceglie subnet
Loseto subnet
Japigia subnet
Ho

Figure 4. Pseudo-districts and subnets of the Bari network.

Based on the data provided, it emerged that the Bari WDN is divided into six sub-networks, each hydraulically independent, called: Santo Spirito-Palese, Bari city, Carbonara-Ceglie, Loseto, Japigia and Torre a Mare. Furthermore, the presence of many closed sectioning valves in the network leads to the identification of pseudo-districts, so defined because, as there is no presence of flow meters, these districts cannot provide information on monitoring for management purposes. It follows that most of the closed valves do not identify actual DMAs but only attempts to control the pressures (by means of closed or partialized valves), which alter the water paths in the network

BY NC SA

on a heuristic basis. Figure 4 shows the existing pseudo-districtualization for the network of Bari (with the different colours) and the 6 subnets.

The modelling of the Bari WDN was performed by assuming the existing subdivision of the global network into the six subnets in Figure 4. Two possible optimal segmentation scenarios for the allocation of district wells were considered according to the complexity of each subnet: (i) based on the positions of existing devices, in order to reduce the number of new manholes; (ii) not assuming the aforementioned constraint but involving the selection of the largest number of manholes in the current positions on the basis of the engineering judgment. Typically, the second scenario resulted in fewer flow measurements and isolation valves at the edge of the districts.

For the sake of brevity, the following section shows the application of the procedure for the Bari-City subnet only, and a general picture of results for all 6 subnets.

4 RESULTS AND DISCUSSION

The Bari city subnet is the only one to be fed by three reservoirs (Bari-Modugno; Bari-Bitritto; Bari-Ceglie del Campo). The hydraulic model of the Bari city subnet has 148 closed/partialized valves. These devices do not create real DMAs but aim to produce a local pressure control on a heuristic basis. The closed valves shown in the model are generally isolation (gates) valves used improperly as sectioning valves used to modify the water flows in the network in order to locally reduce the pressures. The throttled valves, on the other hand, produce a greater effect of pressure reduction in the hours of maximum consumption and have a minimum impact during the night hours in which, however, the pressures and losses are higher, thus resulting ineffective for losses reduction. These valves represent an inherent element of uncertainty linked to their real shutter degree.

For the Bari city subnet, some throttled valves act on main feeding pipelines, conditioning the water flow distribution in the network in the current scheme, while others allow for local pressure reduction. The remaining throttled valves produce local pressure control effects, generally superimposed on the control carried out by the PCVs in operation, which in the Bari city subnet are 4. Figure 5 shows the Bari city subnet with closed valves (red X), throttled ("Km" black) and the PCV ("PV" purple).

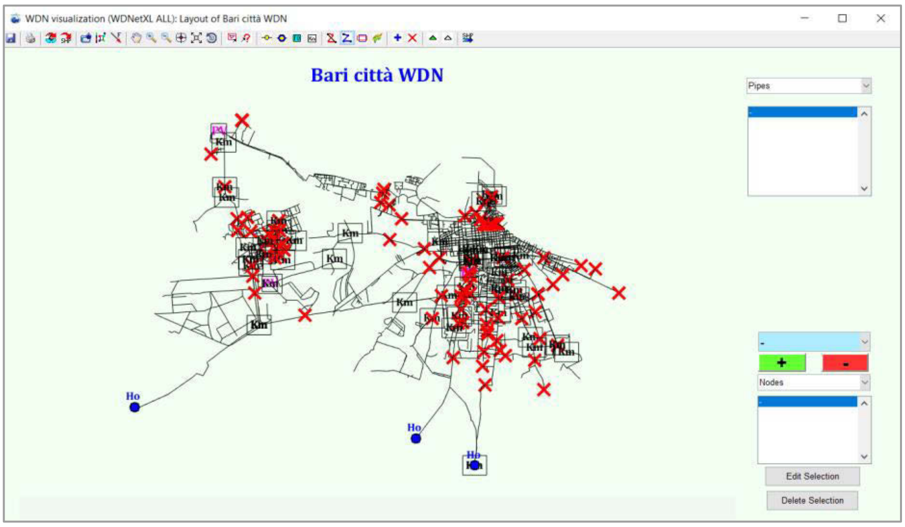


Figure 5. Bari city subnet with closed/throttled valves and PCV.

The subnet of Bari city has areas, in the neighbourhoods built between 1800 and 1900, with a widespread presence of doubling of pipes for final users' water distribution, determined by the stratification of different interventions over time. This configuration represents an anomaly in the

CC BY NC SA

topological structure and causes a greater risk of failure, or less reliability of the service, as well as producing greater volumetric losses due to the increase in the length of the pipelines and the number of joints. The study reported here, before proceeding with the DMA design, considered the removal of the double pipes (indicated in Figure 6 with the blue crosses) and the replacement of the remaining small diameters (40-80mm) to be replaced with a minimum diameter of 100mm.

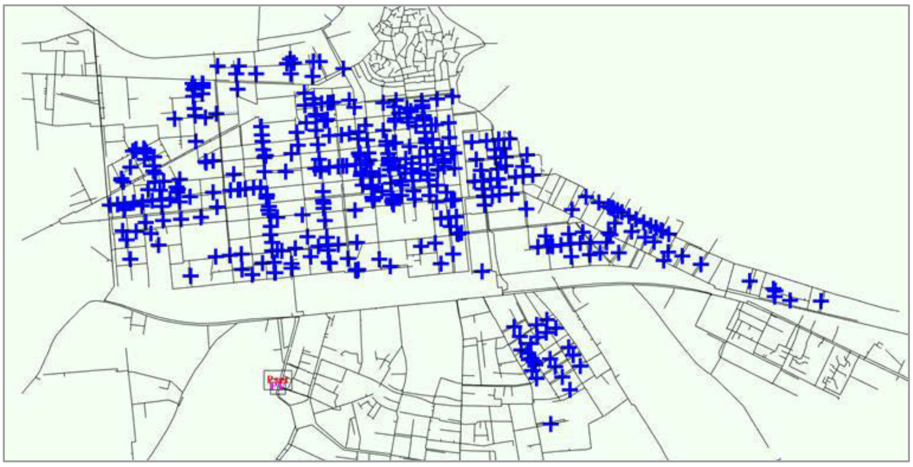


Figure 6. Double pipes removed in the Bari city subnet.

The calibration of the Bari city subnet model was carried out using the monitoring data provided by AQP for the year 2019, with reference to five days (operating cycles) representative of different operating conditions: weekdays and holidays in winter, weekdays and holidays in summer and New Year's Day (given the atypical functioning with respect to the remaining days of the year). Such a model calibration is more robust than any alternative based on data relating to a single specific day or representative of the "average" behaviour of the system, therefore more effective in supporting the DMA planning rehabilitation interventions aimed at reducing volumetric losses [10]. Figure 7 shows the hourly average flow rates measured at the three reservoirs of the subnet for the five days chosen (120 hours).

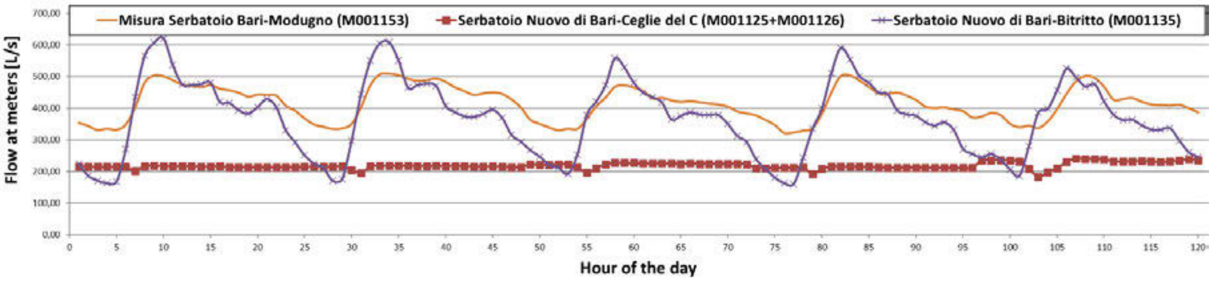


Figure 7. Hourly average flow rate measured at the three reservoir of Bari City subnet.

The calibration results return, for each calibration solution, the statistics relating to the average and maximum error at the flow meters, the average and maximum error of the mass balances at DMA level and the total real losses estimated as the ARERA M1a indicator.

Figure 8 shows the input, demand and leakages values for each of the five days considered in the calibration phase. It can be observed that day by day the calibration allows to identify a variable demand and the real losses as a function of pressure and deterioration. It is possible to note that the change in input is mainly associated with demand, while the losses remain almost constant, as expected for oversized networks such as the Bari WDN, further highlighting the robustness of the calibration strategy [10].

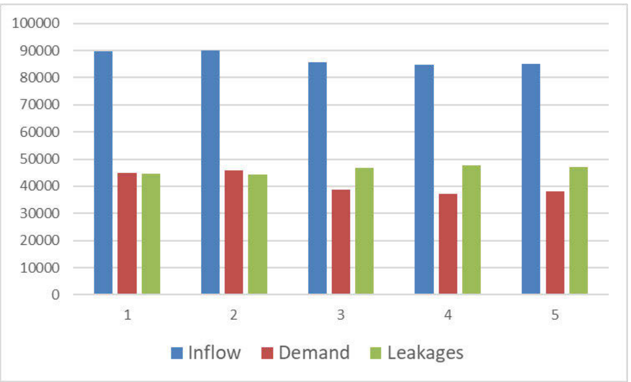


Figure 8. Input, demand and leakages for the five network operating conditions.

Figures 9-11 report the main results of the hydraulic analysis of the Bari City subnet following the model calibration procedure. Figure 9 shows the temporal trend of the input water volumes of the network, highlighting the trend of leakages as a function of the pressure and volumes feeding the private tanks, a particular element of the analysed network; Figure 10 shows the trend of the average pressures inside the network; and Figure 11 shows the spatial distribution of the linear leakages index (m³/ km/day) equivalent to the M1a index for each pipe of the network.

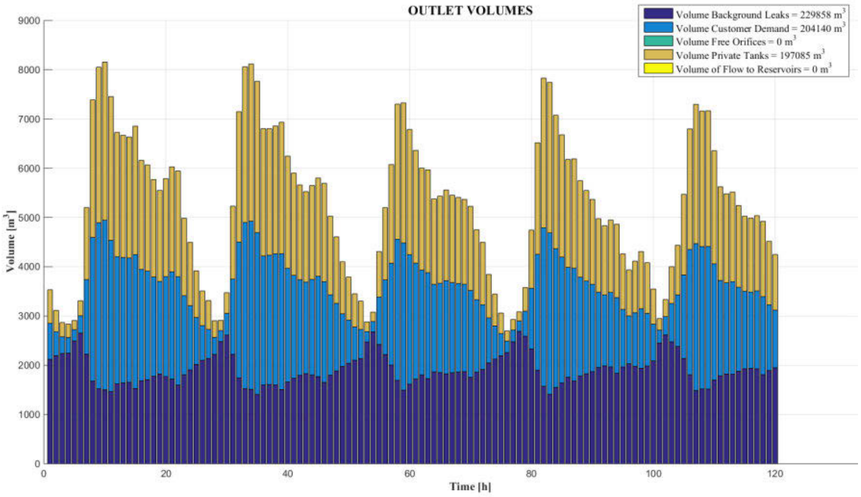


Figure 9. Volumes of demand components of the Bari city subnet during the five chosen operating cycles.

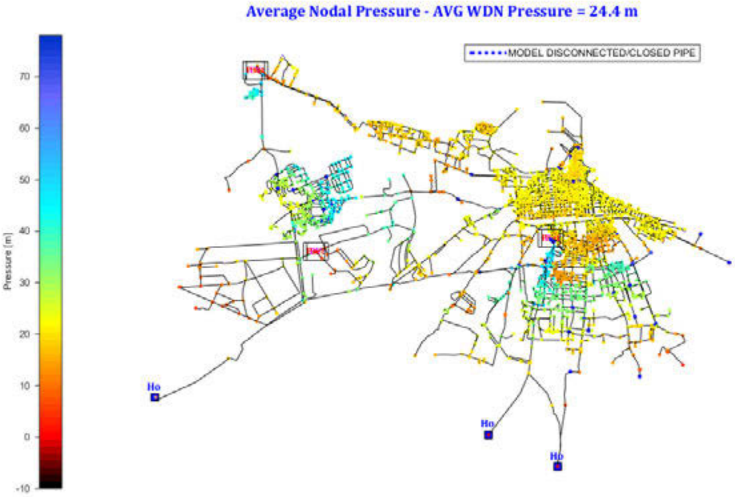

Average Nodal Pressure - AVG WDN Pressure = 24.4 m
MODEL DISCONNECTED/CLOSED PIPE
Pressure [m]
70
60
50
40
30
20
10
0
-10
Ho
Ho
Ho

Figure 10. Trend of average pressures within the subnet.

BY NC SA

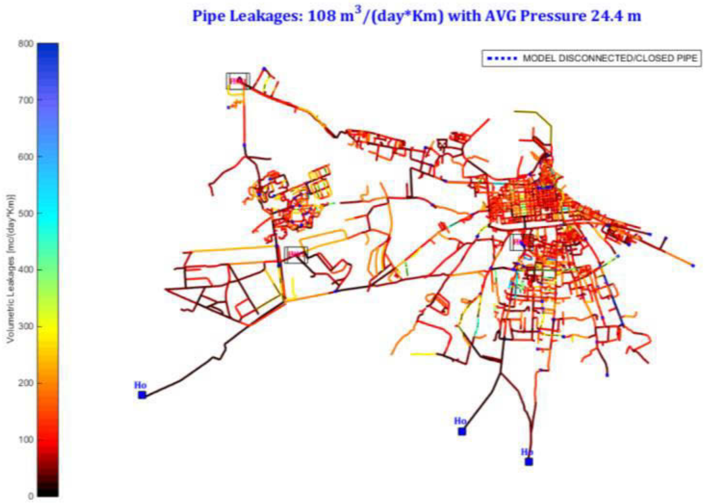

Pipe Leakages: 108 m3/(day*Km) with AVG Pressure 24.4 m
MODEL DISCONNECTED/CLOSED PIPE
Volumetric Leakages [mc/(day*Km)]
800
700
600
500
400
300
200
100
0
Ho
Ho
Ho

Figure 11. Linear losses for each pipe of the subnet that define the M1a indicator.

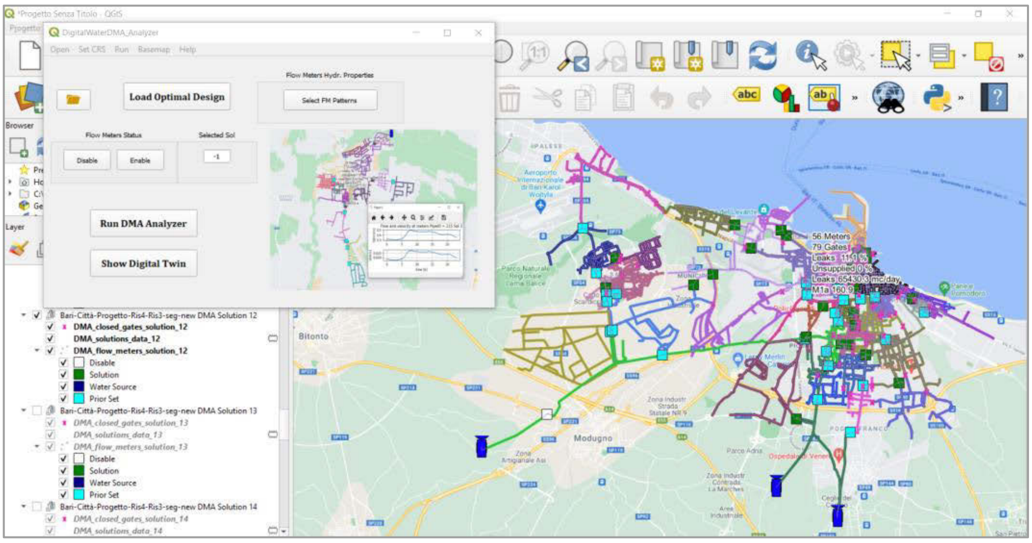
Based on the hydraulic analysis using the calibrated model, the DMA design had the main objective of rationalizing the subdivision of the network for future management purposes, pursuing the goal of containing the areas subject to pressure fluctuations and lowering of the pressure regime in the network. To support efficient design solutions, the design process took as reference two network conditions (i) the current condition, as per the calibrated model; (ii) the project condition following the removal of the double pipes described above. In addition, the design considered the need to preserve the operational continuity expressed by AQP, promptly evaluating the overlap between the existing hydraulic scheme and the optimal district configurations designed independently of the current scheme. As mentioned above, districtualization is a two-stage process: (i) the segmentation of the topological network structure of the WDN; (ii) and the real hydraulic districtualization in relation to the decision on the installation of sectioning valves and flow measurements.

For the Bari city subnet, this optimal districtualization process was not used to produce design solutions ex-novo, but to support engineering judgment and system knowledge in the allocation of sectioning valves or flow meters. The design process integrated information on the position of sectioning valves and optimized flow meters with evaluations on the engineering judgment also related to the position and setting of the PCV. The evaluation of each alternative was checked in detail in terms of impact on the hydraulic functioning of the entire subnet of Bari city. As a result, this objective has minimized the number of throttled valves currently present. This process reconfigured the water flows in a topologically and hydraulically optimal way while maintaining the current supply schemes and the existing PCVs. The hydraulic districtualization solution chosen (Figure 12) to support the subsequent engineering assessments represents the best compromise between the maximum leakage reduction and the minimum number of flow meters in the network.

It should be noted that the results of the districtualization procedure also allows metrological evaluations of the flow meters planned. For example, for each flow meter it is possible to carry out a survey on the expected speeds and flows, thus being able to select the most important meters that make the monitoring and water balance activity more reliable associated with the future management of the network, based on the planned districtualization (i.e., avoiding situations in which the operating conditions cause the flow in the meter to reverse). Figure 13 shows the general picture of the DMAs identified for the Bari city subnet, also considering the PCVs within the system. In fact, the DMA planning is integrated with the evaluation of the PCV set point that feed the system. Figure 14 shows the consequent trend of leakages and the average pressure of the network. The proposed solution allows to reduce the volumetric losses as a function of

BY NC SA

pressure, i.e., to reduce the values of the index of linear losses M1a for each pipe and for the entire network.


*Progetto Senza Titolo - QGIS
DigitalWaterDMA_Analyzer
Open
Set CRS
Run
Basemap
Help
Load Optimal Design
Flow Meters Hydr. Properties
Select FM Patterns
Flow Meters Status
Disable
Enable
Selected Sol
-1
Run DMA Analyzer
Show Digital Twin
Browser
Layer
Bitonto
Modugno

Bari-Città-Progetto-Ris4-Ris3-seg-new DMA Solution 12
- ☑ DMA_closed_gates_solution_12
- ☑ DMA_solutions_data_12
- ☑ DMA_flow_meters_solution_12
  - ☑ Disable
  - ☑ Solution
  - ☑ Water Source
  - ☑ Prior Set

Bari-Città-Progetto-Ris4-Ris3-seg-new DMA Solution 13
- ☑ DMA_closed_gates_solution_13
- ☑ DMA_solutions_data_13
- ☑ DMA_flow_meters_solution_13
  - ☑ Disable
  - ☑ Solution
  - ☑ Water Source
  - ☑ Prior Set

Bari-Città-Progetto-Ris4-Ris3-seg-new DMA Solution 14
- ☑ DMA_closed_gates_solution_14
- ☑ DMA_solutions_data_14

56 Meters
79 Gates
Leaks 11.1 %
Unsupplied 0 %
Leaks 65430.3 mc/day
M1a 160.9

Figure 12. Districtualization solution chosen for the Bari city subnet (solution 12).

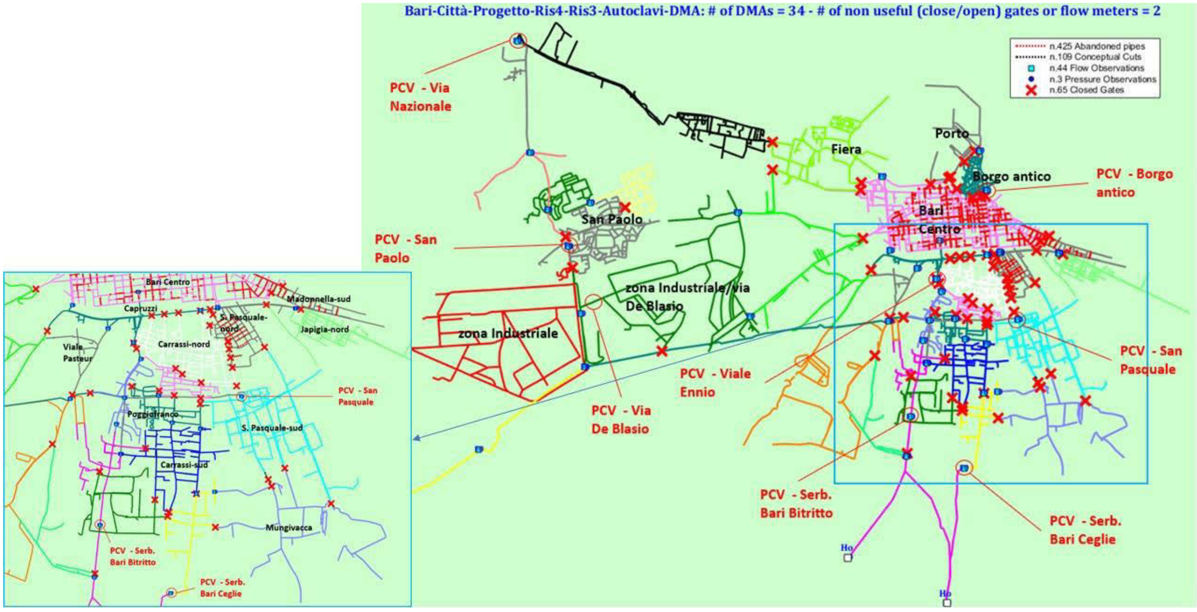

Bari-Città-Progetto-Ris4-Ris3-Autoclavi-DMA: # of DMAs = 34 - # of non useful (close/open) gates or flow meters = 2
n.425 Abandoned pipes
n.109 Conceptual Cuts
n.44 Flow Observations
n.3 Pressure Observations
n.65 Closed Gates
PCV - Via Nazionale
Fiera
Porto
Borgo antico
PCV - Borgo antico
Bari Centro
San Paolo
PCV - San Paolo
zona Industriale/via De Blasio
zona Industriale
PCV - Viale Ennio
PCV - Via De Blasio
PCV - San Pasquale
PCV - Serb. Bari Bitritto
PCV - Serb. Bari Ceglie
Bari Centro
Madonnella-sud
Capruzzi
S. Pasquale-nord
Japigia-nord
Viale Pasteur
Carrassi-nord
PCV - San Pasquale
Poggiofranco
S. Pasquale-sud
Carrassi-sud
Mungivacca
PCV - Serb. Bari Bitritto
PCV - Serb. Bari Ceglie

Figure 13. DMAs identified for the Bari City subnet, with indication of the project PCVs.

Table 1. Summary of devices and DMA after and before DMA design

	N. DMA	N. Closed valves	N. Partially closed valves	N. Flow meters	M1a index	Average pressure** [m]
Current condition	27*	95	56	3	108.65	24.40
Design condition	34	65	8	43	92.9	24.20

** Non-metered pseudo-districts; ** Assessed on 5 different operating cycles (120 hours)*

BY NC SA

Table 1 shows a summary/comparison between the devices and the districts present in the current status and in the project condition of the districtualization in the Bari city subnet, without any rehabilitation intervention.

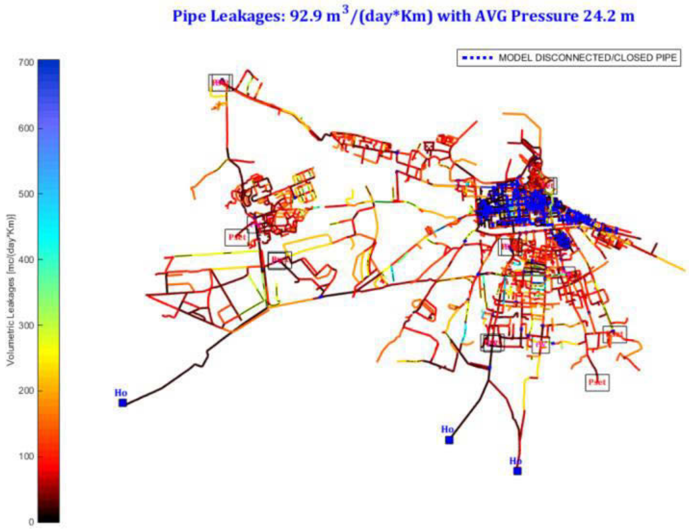

Pipe Leakages: 92.9 m3/(day*Km) with AVG Pressure 24.2 m
MODEL DISCONNECTED/CLOSED PIPE
Volumetric Leakages [mc/(day*Km)]
700
600
500
400
300
200
100
0
Ho
Ho
Ho
Pdst
Pdst

Figure 14. Distribution of volumetric leakages in the subnet of Bari city following the districtualization.

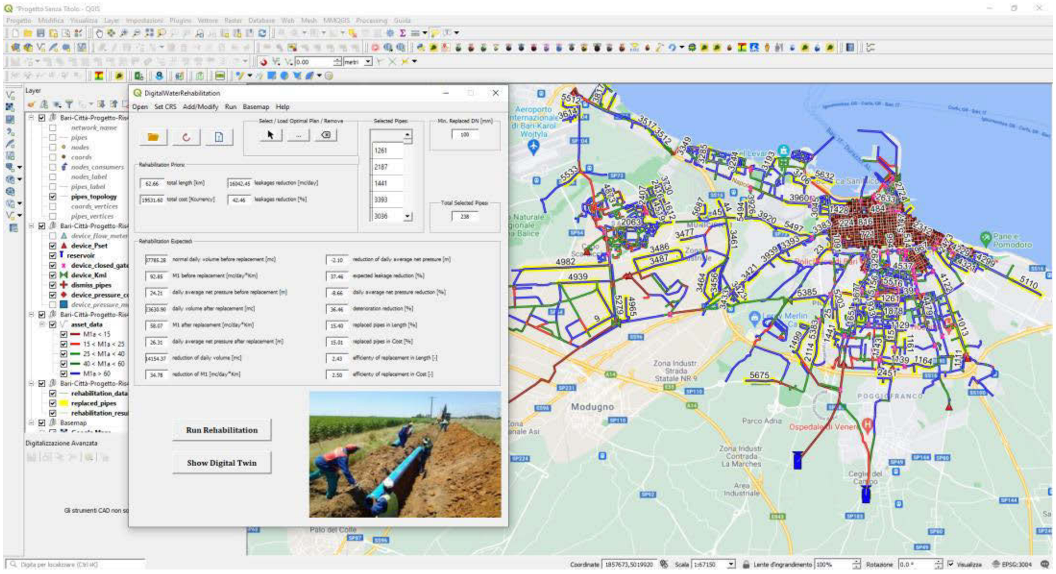
The rehabilitation support strategy was developed coherently with the integrated DMA design. Therefore, the rehabilitation interventions were planned with reference to the districtualization solution chosen (Figure 13).

For the Bari city subnet, different investment levels have been identified corresponding to a value up to 20% of the replacement cost of the entire system, considering the pipes with the maximum efficiency index, calculated as the ratio between the intervention cost and the expected leakage reduction. The rehabilitation support procedure determines the pipes to be replaced in a multi-objective scheme, looking at the best cost/benefit ratio, i.e., investments (cost of replacement compared to the cost of replacing the entire system)/reduction of volumetric losses (assessable through the advanced model starting from the new hydraulic structure obtained following the districtualization).

Table 2 – Rehabilitation plans for the subnet of Bari city as districtualized.

Cost [%]	Cost [€]	# Pipes	Replaced pipes length [m]	Volumetric leakages reduction [m^3/days]	Efficiency [-]
15.00%	19528731	238	62609	14161	2.50
15.98%	20801895	263	66968	14912	2.47
17.01%	22142303	286	70995	15633	2.43
18.03%	23472519	312	75344	16295	2.39
19.05%	24795509	334	79973	17051	2.37
20.00%	26037678	360	84145	17659	2.34

Table 2 shows 6 intervention plans among the 20 developed by the procedure. For each of the plans obtained, the number of pipes replaced, the total length of the same, the efficiency index and the expected reduction of losses on a daily basis are reported. Figures 15 and 16 show the rehabilitation plan corresponding to a cost of pipes replacing of 15% (displayed in the QGIS environment, indicating the ranges of the M1a value within the network and the position of the devices) and the relative results obtained in terms of volumes supplied in the 5 operating cycles considered.



Selected Pipes: 1261, 2187, 1441, 3393, 3036

Min. Replaced DN [mm]: 100

Total Selected Pipes: 238

Rehabilitation Priors

62.66 total length [km]
16042.45 leakages reduction [mc/day]
19531.60 total cost [Kcurrency]
42.46 leakages reduction [%]

Rehabilitation Expected

7785.28 normal daily volume before replacement [mc]
92.85 M1 before replacement [mc/day*Km]
24.21 daily average net pressure before replacement [m]
3630.90 daily volume after replacement [mc]
58.07 M1 after replacement [mc/day*Km]
26.31 daily average net pressure after replacement [m]
4154.37 reduction of daily volume [mc]
34.78 reduction of M1 [mc/day*Km]
-2.10 reduction of daily average net pressure [m]
37.46 expected leakage reduction [%]
-8.66 daily average net pressure reduction [%]
36.46 deterioration reduction [%]
15.40 replaced pipes in Length [%]
15.01 replaced pipes in Cost [%]
2.43 efficienty of replacement in Length [-]
2.50 efficienty of replacement in Cost [-]

Figure 15. Rehabilitation plan corresponding to a replacement cost of 15%.

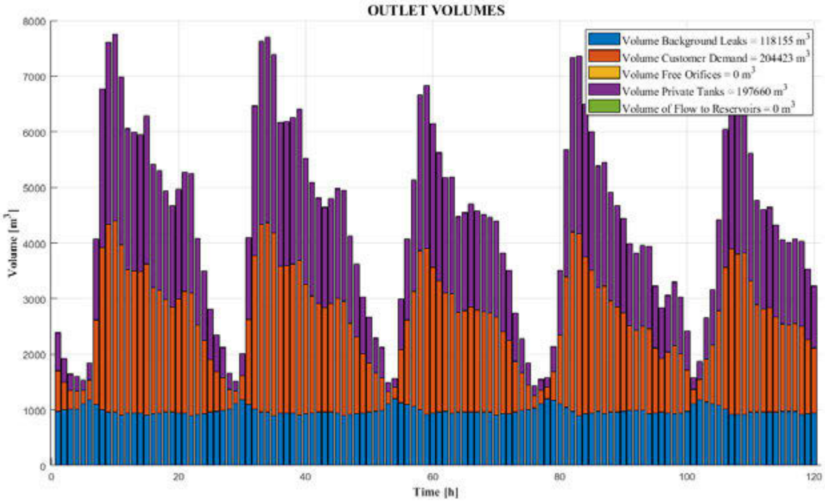


Figure 16. Volumes of the demand components of the Bari city subnet following rehabilitation.

In conclusion, the overall results should be reported. Table 3 shows the results relating to the districtualization phase. Figure 17 shows the complex of 68 DMAs identified for the entire city of Bari. Finally, Table 4 shows the expected reduction of annual leakages that can be achieved with a pipe replacement (optimal in a hydraulic/topological sense) approximately of 20% for each subnet.

5 CONCLUSIONS

The complexity of the Bari network, albeit divided into six separate subnets, suggested using districtualization as a strategic tool for all subsequent asset management activities. Therefore, with reference to each subnet, in addition to reducing water losses, the primary objective of the districtualization was the rationalization of the system in terms of enhancement of the network operating, with the possibility of its monitoring and control. This has allowed a drastic reduction in the use of empirical expedients aimed at local pressure control, such as gate valves currently partialized or closed, in favour of the reconfiguration of water flows in the network by means of sectioning valves with the minimum number of flow measurements at the edge of the designed DMAs and use of PCVs. The strategic value of the design support here reported is demonstrated in every aspect by the advanced hydraulic analysis and operational-managerial assessments.

Table 3 – Design solutions based on system conditions during the year 2019.

	Torre a Mare	Japigia	Bari Città	Carbonara-Ceglie	Loseto	Santo Spirito-Palese	Total
#DMA	5	5	34	12	1	11	68
#Closed Valves	3	4	65	18	-	10	123
#Flow meters	5	5	44	12	1	11	78
# PCV	2	2	8	2	-	2	16
Leakage Reduction [%]	36,31	11,31	18,14	17,29	-	25,36	19.95
M1a [m^3/km/day]	44,20	123	92,90	40,70	-	45,10	77.2
Leakage Reduction [m^3/year]	402.660	247.650	3.052.870	235.510	-	484.350	4.780.360

Table 4 – Reduction of volumetric losses following rehabilitation.

	Torre a Mare	Japigia	Bari Città	Carbonara Ceglie	Loseto	Santo Spirito-Palese	Total
Leakage Reduction [m^3/year]	434.815	1.269.567	6.445.684	438.202	-	636.631	9.224.899

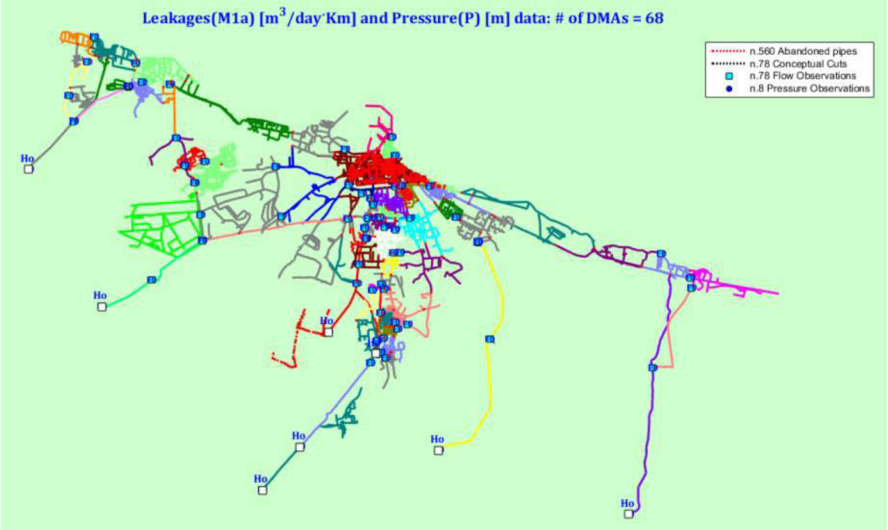

Leakages(M1a) [m3/day·Km] and Pressure(P) [m] data: # of DMAs = 68
n.560 Abandoned pipes
n.78 Conceptual Cuts
n.78 Flow Observations
n.8 Pressure Observations
Ho
Ho
Ho
Ho
Ho
Ho
Ho

Figure 17. DMA designed for the Bari network.

The decision support system for districtualization, using dedicated Digital Water services, allowed to select, among all the optimal positions for flow meters, the most effective with respect to management budget constraints (i.e., limiting the maximum number of meters to be installed), to the expert knowledge of the network about possible technical constraints, as well as metrological evaluations with reference to the expected flow rates at the measurement points. Therefore, the innovative scheme to support districtualization allows to choose the number of DMAs to be activated with flow meters, at the same time indicating perspective measurement stations nested in the chosen DMAs. These perspective meters are useful for planning future investments and preparing the system for the Internet of Things (IoT) framework of the near future. Furthermore, they can be used for mobile meters and/or to organize active leakage control activities even for long time, because they are not invasive for the system as they are based on sections already prepared in the project.

6 ACKNOWLEDGEMENTS

This paper was financed by European Union - PON Ricerca e Innovazione 2014-2020 E FSC, call D.D. 13 July 2017 n. 1735, Project "*Energidrica - Efficienza energetica nella gestione delle reti idriche*" (code ARS01_00625). Authors want to thank the academic spin-off company IDEA-RT s.r.l. (www.idea-rt.com) for providing WDNetXL software and digital services and related pictures.

7 REFERENCES

[1] Todini, E., and Pilati, S., "A gradient method for the solution of looped pipe networks", Computer Applications in Water Supply, vol.1, 1988, pp. 1-20.

[2] Rossman, L.A., EPANET 2.0 User Manual. Water Supply and Water Resources Division, National Risk management Laboratory, USEPA, Cincinnati, OH, 2000.

[3] Todini, E., "A more realistic approach to the extended period simulation of water distribution networks", in Proc. CCWI 2003 Conference, London, UK: Taylor&Francis, 2003, pp. 173-184.

[4] Wagner, J.M., Shamir, U., and Marks, D.H., "Water distribution reliability: simulation methods", Journal of Water Resources Planning & Management, vol. 114, no. 3, 1988, pp. 276-294.

[5] Giustolisi, O., Savic, D.A., and Kapelan, Z., "Pressure-driven demand and leakage simulation for water distribution networks", Journal of Hydraulic Engineering, vol. 134, no. 5, 2008, pp. 626-635.

[6] Giustolisi, O., and Walski, T.M., "A Demand Components in Water Distribution Network Analysis", Journal of Water Resources Planning & Management, vol. 138, no. 4, 2012, pp. 356 -367.

[7] Germanopoulos, G., "A technical note on the inclusion of pressure-dependent demand and leakage terms in water supply network models", Civil Engineering Systems, vol. 2, no. 3, 1985, pp. 171-179.

[8] Pelletier, G., Mailhot, A. and Villeneuve, J.P., "Modeling Water Pipe Breaks. Three Case Studies", Journal of Water Resources Planning & Management, vol. 129, no. 2, pp. 115-123.

[9] Laucelli, D., Simone, A., Berardi, L., and Giustolisi, O. "Optimal Design of District Metering Areas for the Reduction of Leakages", Journal of Water Resources Planning & Management, vol. 143, no. 6, 2017, pp. 04017017-1-12.

[10] Berardi, L., and Giustolisi, O. "Calibration of Design Models for Leakage Management of Water Distribution Networks", Water Resources Management, vol. 35, 2021, pp. 2537–2551.

[11] ARERA, Deliberazione 27 Dicembre 2017 917/2017/R/IDR - Regolazione della Qualità Tecnica del Servizio Idrico Integrato ovvero di ciascuno dei singoli servizi che lo compongono (RQTI), 2017 (in Italian).

[12] Giustolisi, O., Ciliberti, F., Laucelli, D., and Berardi, L., "Mass Balance Calibration of Water Distribution Networks: application to a real case study", in Proc. WDSA-CCWI 2022, Valencia, Spain 2022.

[13] Giustolisi, O., and Ridolfi, L. "A novel infrastructure modularity index for the segmentation of water distribution networks", Water Resources Research, vol. 50, no. 10, 2014, pp. 7648–766.

[14] Idea-RT, "WDNetXL-WDNetGIS", http://www.idea-rt.com/products-our-job/, June 2022.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14776

DOMAIN ANALYSIS OF WATER DISTRIBUTION NETWORKS

Francesco Ciliberti[1], Daniele Laucelli[2], Luigi Berardi[3] and Orazio Giustolisi[4]

[1,3] Department of Engineering and Geology, University "G. D'Annunzio" of Chieti-Pescara, Pescara (Italy)
[2,4] Department of Civil, Environmental, Land, Building Engineering and Chemistry, Technical University of Bari, Bari (Italy)

[1] *francesco.ciliberti@unich.it,* [2] *danielebiagio.laucelli@poliba.it,* [3] *luigi.berardi@unich.it,*
[4] *orazio.giustolisi@poliba.it*

Abstract

Water Distribution Networks (WDNs) represent spatially organized infrastructures, whose main function is to deliver water from hydraulic sources to meet the customers' demands and pressure requirements of the system through interconnected pipes. The time-varying water demands and the asset deterioration greatly contribute to the complex functioning of WDNs. Nonetheless, the hydraulics of these system in strongly determined by the topology of pipelines, as represented by the connectivity among pipes and nodes.

The Complex Network Theory (CNT) has been recognized to provide a wide range of metrics for studying WDNs , only based on the analysis of the topological domain. In addition, some latest advancements on tailoring CNT metrics for WDNs have been introduced to understand hydraulic behaviour of such systems, even before performing classical hydraulic modelling.

This contribution deals with the application of some tailored centrality metrics, such as betweenness and edge betweenness for capturing the topological domain of WDNs. The analysis has been extended to each subnetwork of a large real WDN, to provide helpful elements for various tasks including the validation of available hydraulic models, and the identification of main water paths to support model calibration and planning of maintenance/retrain works.

Keywords
Water Distribution Networks, Complex Network Theory, Centrality Metrics.

1 INTRODUCTION

Water Distribution Networks (WDNs) in urban areas stands as the main fundamental infrastructures for different activities of life. In the last years, water utilities are facing several technical issues, particularly related to the increase of water demand, asset degradation and water leakages. All such issues concur to downgrade the performance of the whole system in terms of the reliability of the water supply service.

In order to ensure an adequate level of service, water companies are interested in procedures and products to improve the understanding of such systems and hence make the WDN management more effective and efficient.

The complexity of WDNs is related to many aspects, including their topological configuration and the hydraulic behaviour, which follows water requests from consumers and leakages due to asset deterioration and pressure. Indeed, the hydraulic phenomena that determine the level of water supply service in WDNs take place in networked pipelines. Therefore, WDNs topological domain strongly affects the hydraulic status throughout the system in terms of flows along pipes and pressure regime.

Assessing WDNs topology is known to provide effective understanding of WDNs hydraulic behavior [1][2][3], besides classical hydraulic modelling approach. This means that information on WDNs topology can provide useful insights from various planning and management tasks, ranging from WDNs model calibration up to possible improvements of service conditions.

In recent years, Complex Network Theory (CNT) has been introduced in WDNs analysis for an in-depth understanding of the network from a topological point of view. CNT has been previously developed and applied in various technical and research areas, even far from water systems; they include social networks [4], biological networks [5], informative networks [6], and spatial networks [7]. In the case of WDNs, it has been reported that CNT is suitable for detecting the emergent hydraulic behavior [8] and it has been applied for a wide range of operational tasks, including WDNs vulnerability and structure assessment [9] [10], WDNs solutions optimization [11], optimal district design [12] and water quality analysis assessment [13].

The basic idea of CNT leads to the conceptualization of the network as a graph, an ordered set of vertices interconnected by a set of edges, and the evaluation of some peculiar properties of the network topology through metrics, that were introduces to quantify the properties of network structure [14].

Among such metrics, the centrality metrics, firstly introduced by [15], represent the candidate measures for better understanding the features of the network by ranking the importance of the nodes [14]. The most adopted are: degree [15, 14], eigenvector [16], closeness [17], Katz centrality [18], and PageRank [19].

WDNs are spatial networks whose topology is bounded by the constraints of urban layout, which influence the nodal degree distribution [20].

As pointed out by [7], in case of spatial networks, like WDNs, the centrality metrics are more relevant to capture the flux of information. This happens because centrality metrics are built upon the concept of shortest path, representing the optimal path between two nodes using the minimum number of links. In the case of WDNs, the flux is represented by water flows along pipes, reaching nodes of the graph (where consumers demand is allocated in classical WDNs representation) from water sources (reservoirs, tanks, pumps).

In addition, differently from other networked systems, nodes and pipes in WDNs have different relevance in terms of hydraulic status. This fact, motivated to define some tailored metrics, such as [21], to take into account that those pipes/edges are physical components of the WDNs; some nodes (reservoirs and tanks) play a completely different hydraulic role from the majority of nodes (demand and connection nodes); and pipes have different characteristics (length, diameter, hydraulic resistance, etc.) that can be included in CNT metrics tailored for WDNs. Later on [22] provided a first application of the edge betweenness centrality metrics developer in [22] for analysing the spatial domain of the networks.

Recently, [23] pointed out that classical centrality metrics cannot exhibit the information associated to vertices and edges, because the relevance of the network elements, identified as the role among the community, is not considered. Therefore, [23] relaxed the assumption of identical relevance of edges and vertices and proposed a weighted version of these metrics for understanding the networks' behavior. This way, a relevance function has been introduced, which takes different formulations according to the type of the network, for capturing the interplay between the network topological structure and the intrinsic relevance of the elements of the network.

In this paper, two metrics tailored by [23] have been applied for the first time to analyse the topological domain of a large real WDN, composed by few subnetworks. The topological analysis has been compared to the hydraulic simulation results, proving the consistency of the proposed approach. Thus, the preliminary understanding of the WDN behavior based on topological domain analysis is demonstrated to provide a useful support for several tasks of WDN design, including asset management and planning, to be used in conjunction with classical hydraulic modelling.

2 METHODOLOGY

[23] draws the attention about the importance of considering the information about the intrinsic relevance of each node, as exogeneous data independent from the network topology layout, into the CNT metrics. The intrinsic relevance for each node, defined as R_n ($n=1, \ldots, V$), is so assigned in different ways to nodes having different roles in WDN functioning. It is equal to the water demand supplied to each "demand node" and to the sum of supplied water demands for each "source node". In this way, all nodes assume a different relevance during the analysis, and the intrinsic-relevance metrics can be introduced using various formulations of the function $f(R_s,R_t)$ [23], where R_s and R_t represents the relevance of vertices s and t, respectively.

Some most adopted centrality metrics in WDNs analysis [24] have been applied in the relevance-embedded versions [23], however nodal and betweenness centrality metrics have demonstrated to detect well the WDN topological behavior [22].

The standard version of the betweenness centrality [15] assigns the number of shortest paths traversing a vertex v for all couples of vertices s and t of the network. Similarly, the edge betweenness, $EB(e)$, is the sum of the fractions for all the couples of vertices s and t of the network traversing an edge e. The classic formulations of these metrics, for a vertex v and edge e respectively, are reported below:

$$B(v) = \sum_{s \neq v \neq t \in N} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

$$EB(e) = \sum_{s \neq t \in N} \frac{\sigma_{st}(e)}{\sigma_{st}} \tag{2}$$

The relevance-embedded versions of such metrics are carried out by weighting them with the relevance function $f(R_s, R_t)$, which formulations are reported in [23], calculated as follows:

$$B(v) = \sum_{s \neq v \neq t \in N} f(R_s, R_t) \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

$$EB(e) = \sum_{s \neq t \in N} f(R_s, R_t) \frac{\sigma_{st}(e)}{\sigma_{st}} \tag{4}$$

where the intrinsic relevance terms of $f(R_s, R_t)$ function are assigned as in [23], i.e. equal to the customer water demand for each "demand node" of the network and the sum of supplied water demands for the "source node". Among the formulations of the relevance function, the $f(R_s, R_t)= R_s \cdot R_t$ one fits well the relevance domain of the network [23], and it is herein applied for the analysis of the domain characteristics of the network.

3 CASE STUDIES

The domain analysis reported above has been applied to a real-life WDN of a large city in Southern Italy. It was part of a workflow for asset management activities aiming at DMA design and pressure control for leakage reduction, with possible improvements of system hydraulic functioning.

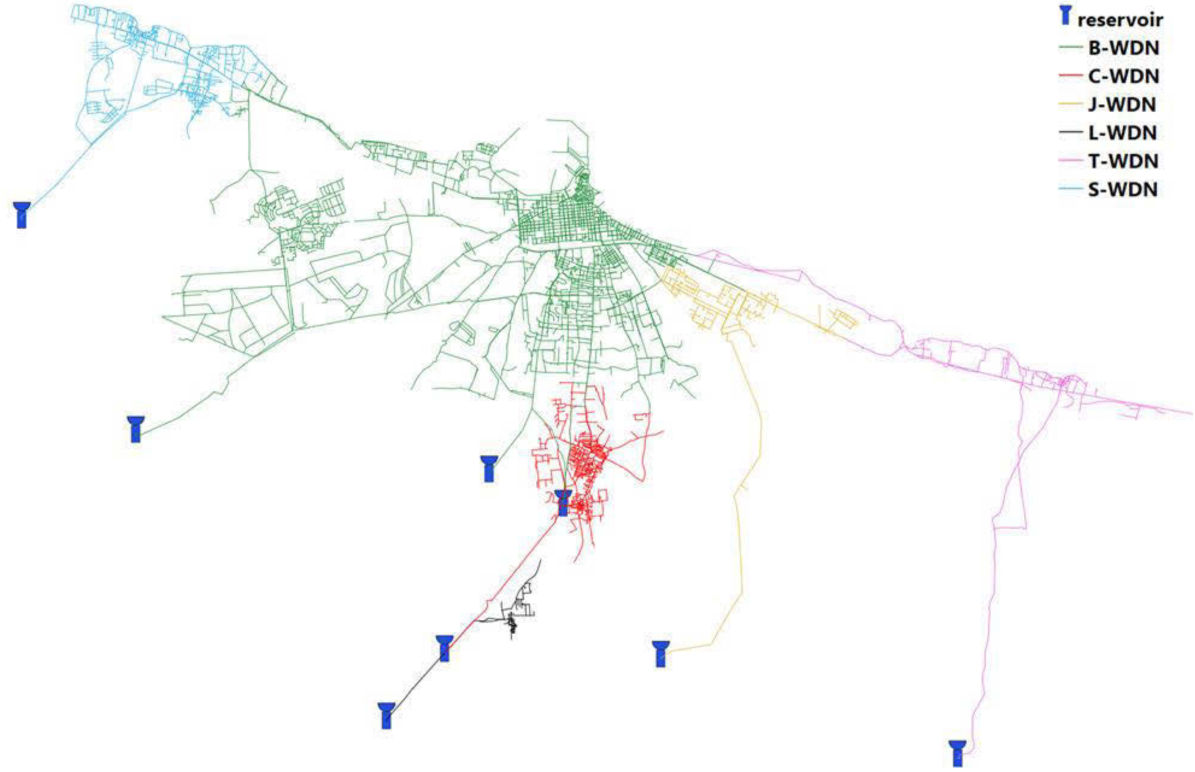


Figure 1. Visualization of the real WDN layout and its sub-networks.

The WDN is composed of six subnetworks fed by eight reservoirs; the total length of the whole network is about 700km, with 7 pressure reduction valves. The system also includes about 80 partially closed valves and 150 closed gates. Figure 1 shows the six subnetworks and Table 1 summarizes key data for each corresponding WDN hydraulic model.

Previous metrics has been applied to the whole set of subnetworks singularly, with respect to each one hydraulic layout, which comprises its own sources and pipelines. Indeed, relevance-embedded centrality metrics are weighted with the relevance terms R_n of nodes. The results of the domain analysis have been compared with the hydraulic simulation of each network.

The hydraulic analysis has been performed using the WDNetXL platform [25][26], which allows a phenomenological representation of the hydraulic behavior of the network by representing the volumetric leakages as a function of the average pressure and pipe deterioration, at single pipe level, and including different types of users' connection (i.e. direct connection to WDN, free orifices, private tanks) or insufficient pressure conditions for correct service.

BY NC SA

Table 1. Data of WD hydraulic model of each subnetworks

Subnetwork	Reservoirs	Length [km]	pipes	nodes	Control valves
T-WDN	1	52.71	350	325	1
J-WDN	1	43.27	358	332	2
L-WDN	1	12.87	169	144	-
S-WDN	1	86.59	1344	1181	-
C-WDN	1	75.82	1144	1007	1
B-WDN	3	424.33	5691	4783	3
Total	8	695.59	9056	7772	7

Figures 2 and 3 show the comparison between the representations of the relevance-embedded edge betweenness (3) and the hydraulic simulation outputs of the main pipe flows for each subnetwork. It is worth noting that higher values of edge betweenness, ranging from 65-100%, are related to the higher supplying paths. This is emphasized in the smaller WDNs cases, J-WDN, L-WDN, C-WDN and T-WDN, where the paths with edge betweenness values, about 75-100%, fits with pipes supplying the main amount of demand of each network.

In case of the B-WDN, which is the biggest one fed by three source nodes, the main feeding lines show different relevance values, which are consistent with the lines traversed by the highest average flow rates. This, in turn, confirm the main structure of the system behaviour as built, consistently with the information reported by the personnel of the water utility.

In the case of J-WDN, it has to remark that the main feeding line from the reservoir shows two different values of the relevance (i.e. tailored edge betweenness) for the upstream and the downstream half of that line. In more details, the upstream half shows a lower relevance than the downstream half. This happens because at about half of that pipeline there is a big demand node representing the feeding of another small city. Therefore, such node has high nodal relevance which affects all the shortest paths traversing the downstream half of the feeding pipeline only, according to formulation (4).

This circumstance confirms that the domain analysis, although largely consistent with the hydraulic behaviour of the system, cannot replace the physically-based modelling of WDN functioning, which unveils also flow directions.

Figures 4 and 5 show the comparison between the representations of the relevance-embedded nodal betweenness (4) and the hydraulic simulation outputs of the main pipe flows for each subnetwork, which confirms the observations, mentioned above in the case of the edge betweenness, where main flow paths of each network pass through the nodes with higher betweenness values.

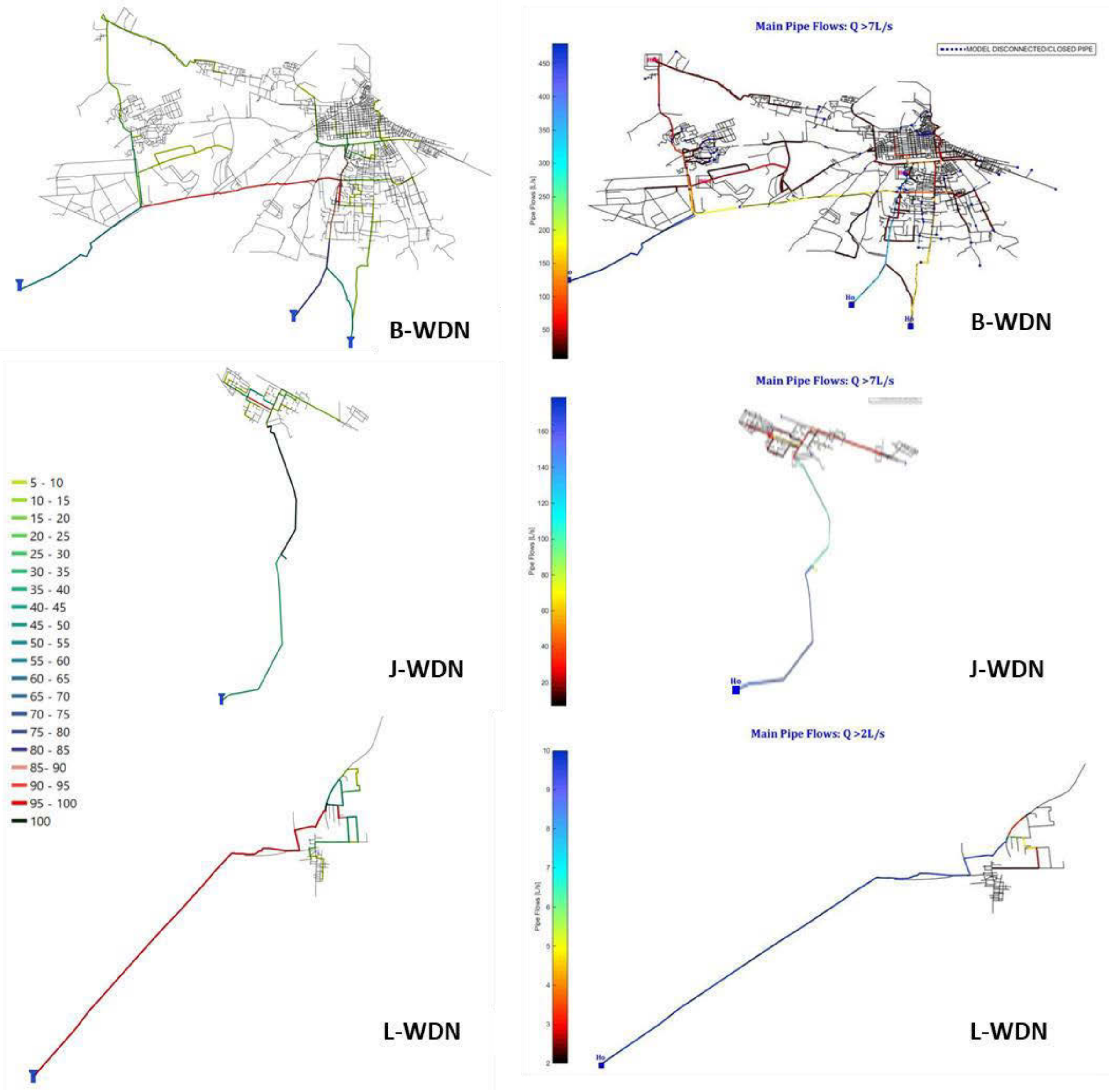

B-WDN
Main Pipe Flows: Q >7L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
450
400
350
300
250
200
150
100
50
Ho
B-WDN
5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40- 45
45 - 50
50 - 55
55 - 60
60 - 65
65 - 70
70 - 75
75 - 80
80 - 85
85- 90
90 - 95
95 - 100
100
J-WDN
Main Pipe Flows: Q >7L/s
Pipe Flows [L/s]
160
140
120
100
80
60
40
20
Ho
J-WDN
L-WDN
Main Pipe Flows: Q >2L/s
Pipe Flows [L/s]
10
9
8
7
6
5
4
3
2
Ho
L-WDN

Figure 2. Comparisons between the relevance-embedded edge betweenness and main pipe flows simulation for B-WDN, J-WDN and L-WDN.

cc BY NC SA

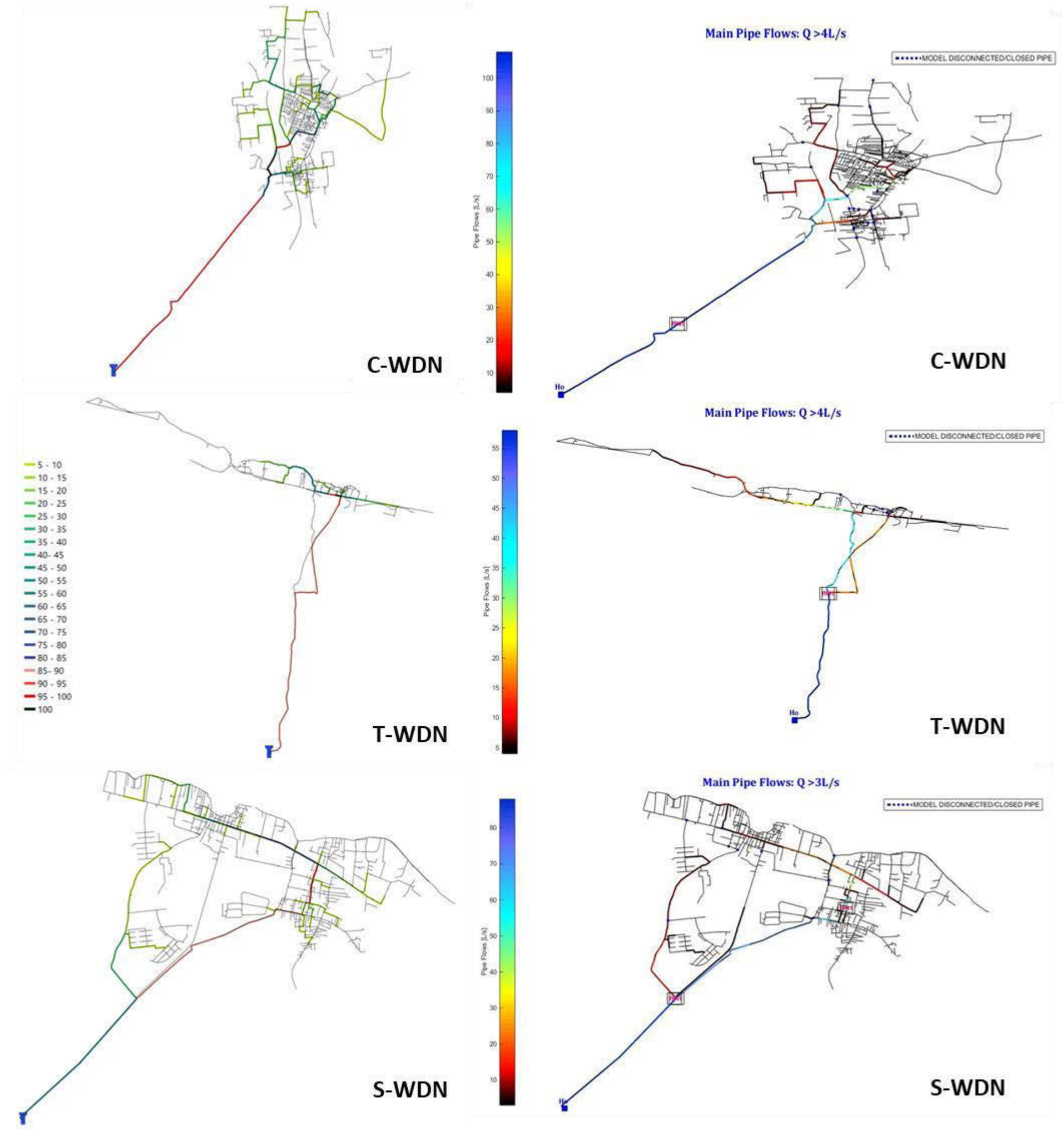

C-WDN
Main Pipe Flows: Q >4L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
100
90
80
70
60
50
40
30
20
10
Ho
C-WDN
5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40- 45
45 - 50
50 - 55
55 - 60
60 - 65
65 - 70
70 - 75
75 - 80
80 - 85
85- 90
90 - 95
95 - 100
100
T-WDN
Main Pipe Flows: Q >4L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
55
50
45
40
35
30
25
20
15
10
5
Ho
T-WDN
S-WDN
Main Pipe Flows: Q >3L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
80
70
60
50
40
30
20
10
Ho
S-WDN

Figure 3. Comparisons between the relevance-embedded edge betweenness and main pipe flows simulation for C-WDN, T-WDN and S-WDN.

cc BY NC SA

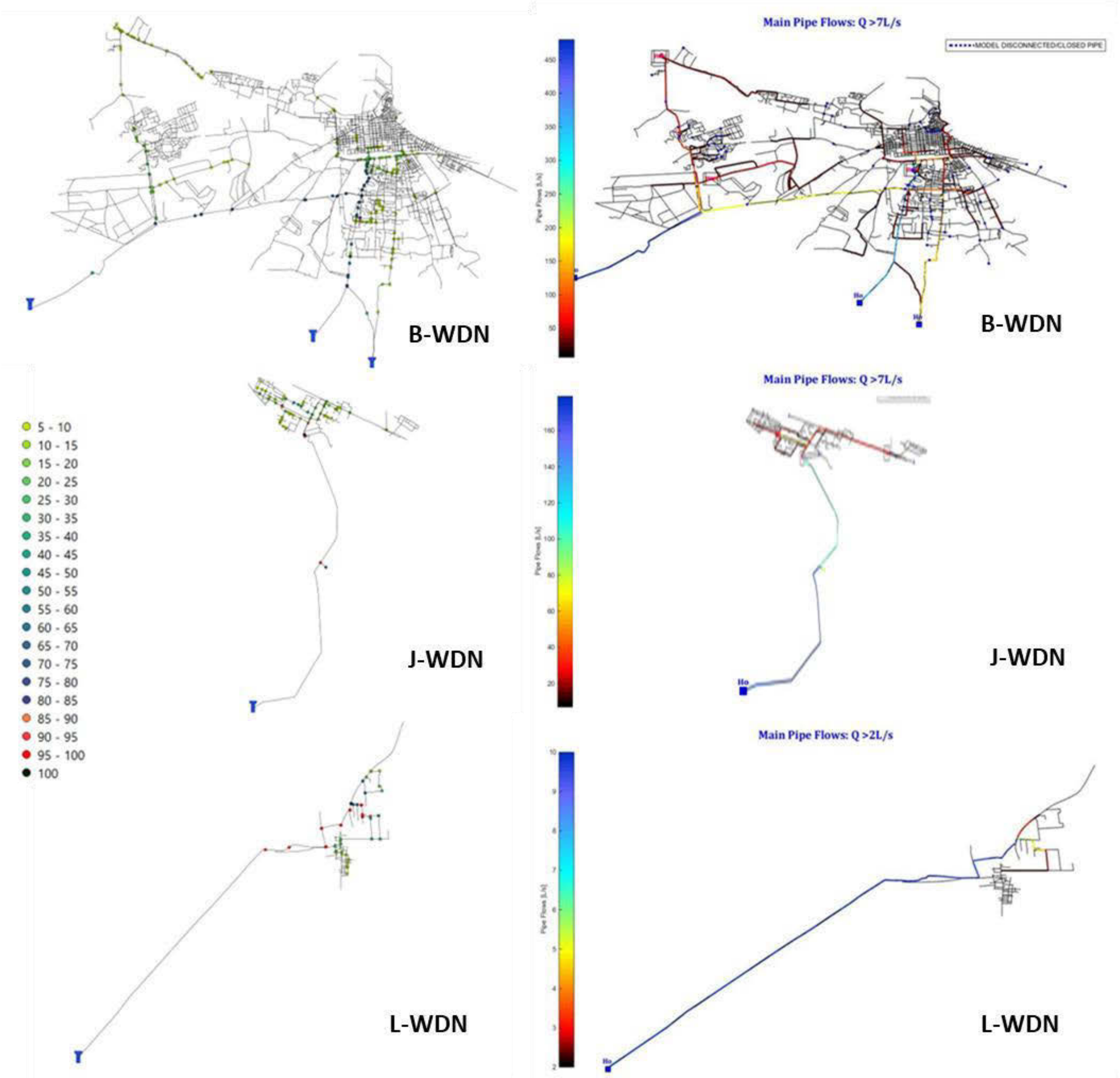

B-WDN
Main Pipe Flows: Q >7L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
450
400
350
300
250
200
150
100
50
Ho
B-WDN
5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40 - 45
45 - 50
50 - 55
55 - 60
60 - 65
65 - 70
70 - 75
75 - 80
80 - 85
85 - 90
90 - 95
95 - 100
100
J-WDN
Main Pipe Flows: Q >7L/s
Pipe Flows [L/s]
160
140
120
100
80
60
40
20
Ho
J-WDN
L-WDN
Main Pipe Flows: Q >2L/s
Pipe Flows [L/s]
10
9
8
7
6
5
4
3
2
Ho
L-WDN

Figure 4. Comparisons between the relevance-embedded betweenness and main pipe flows simulation for B-WDN, J-WDN and L-WDN.

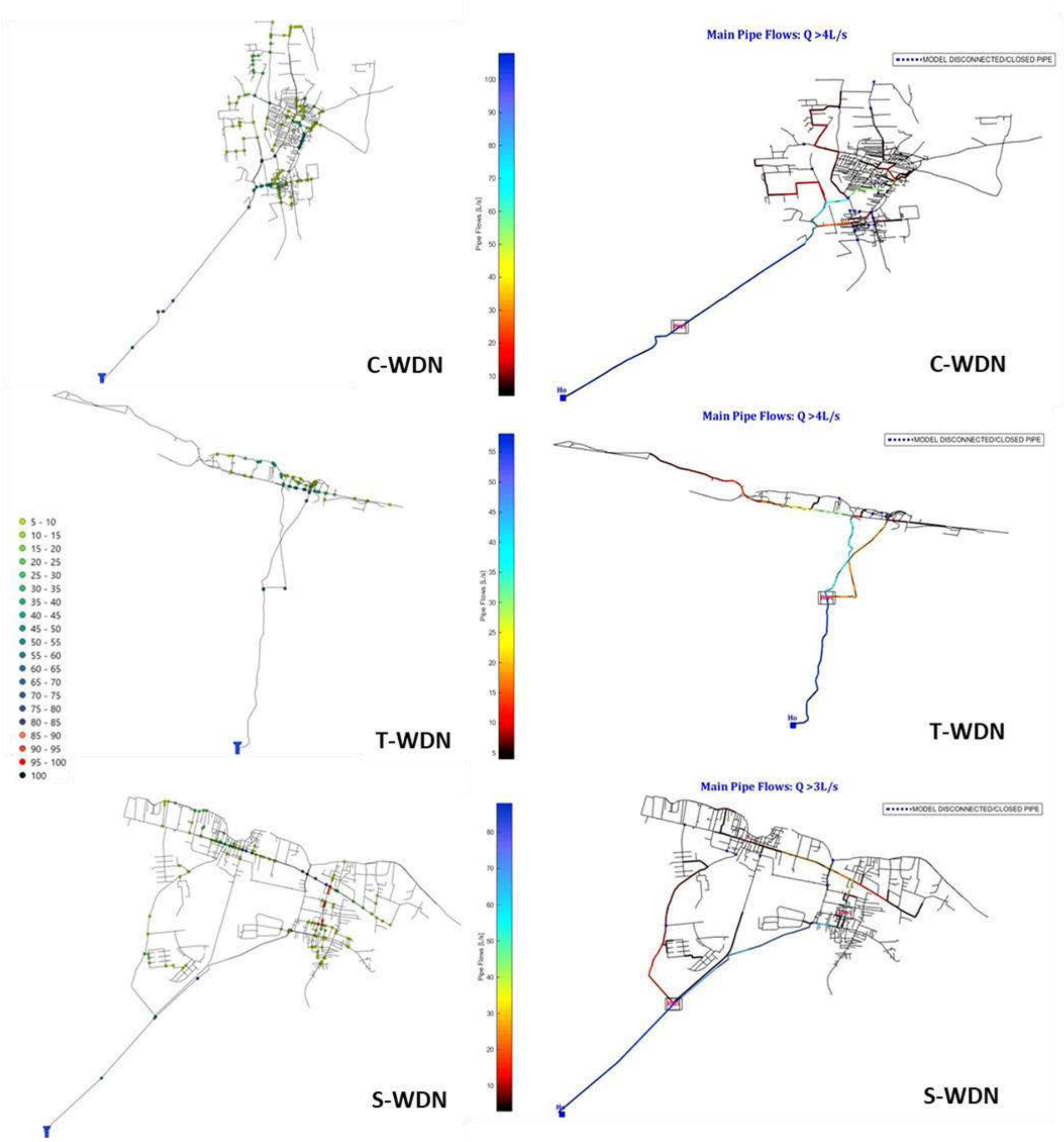

C-WDN
Main Pipe Flows: Q >4L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
C-WDN
5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40 - 45
45 - 50
50 - 55
55 - 60
60 - 65
65 - 70
70 - 75
75 - 80
80 - 85
85 - 90
90 - 95
95 - 100
100
T-WDN
Main Pipe Flows: Q >4L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
T-WDN
S-WDN
Main Pipe Flows: Q >3L/s
MODEL DISCONNECTED/CLOSED PIPE
Pipe Flows [L/s]
S-WDN

Figure 5. Comparisons between the relevance-embedded edge betweenness and main pipe flows simulation for C-WDN, T-WDN and S-WDN.

Finally, Figures 6 and 7, provides a comprehensive view of the original large WDN, showing the results discussed above. Such analyses, provided some preliminary information to drive for several operational and design tasks in each sub-system.

In case of model calibration, the most relevant pipes unveiled by domain analysis are those deserving accurate assessment of hydraulic resistance. In fact, they are traversed by highest flows and usually are quite long pipelines. Therefore, the head-losses along such lines strongly affect pressure regime in the main distribution part of the system and, consequently, the calibration of leakage model parameters [25]. *Viceversa*, pipes into the main looped distribution part of the WDN are assigned with lower relevance because there are many alternative paths connecting nodes with similar low relevance. For these reasons, from calibration perspective, the accurate assessment of their hydraulic resistance values has not a relevant impact on WDN hydraulic behaviour.

The visual identification of the main flow paths circulation provides also useful information about possible improvements in WDN supply service performances. For instance, in case of T-WDN, the rightmost part of the distribution network, which was reported to suffer from pressure deficit conditions, has a low relevance from domain analysis. This allowed to explore a possible change in global WDN configuration, where that part of the T-WDN is fed by the J-WDN. In fact, from domain analysis perspective, this change is going to not affect the main flow paths in T-WDN, while generating increase in pressure at demand nodes because of higher pressure in J-WDN.

In the case of DMA design, pipes with higher relevance represents pipes where closed sectioning valves should be avoided in order to not modify abruptly the existing hydraulic scheme. For the same reason, i.e. highest flows circulating along the most relevant pipes, they can be candidate for installing affective flow meters at DMA boundaries, since they are not affected by flow inversion thus minimizing uncertainties in DMA bass balance evaluation.

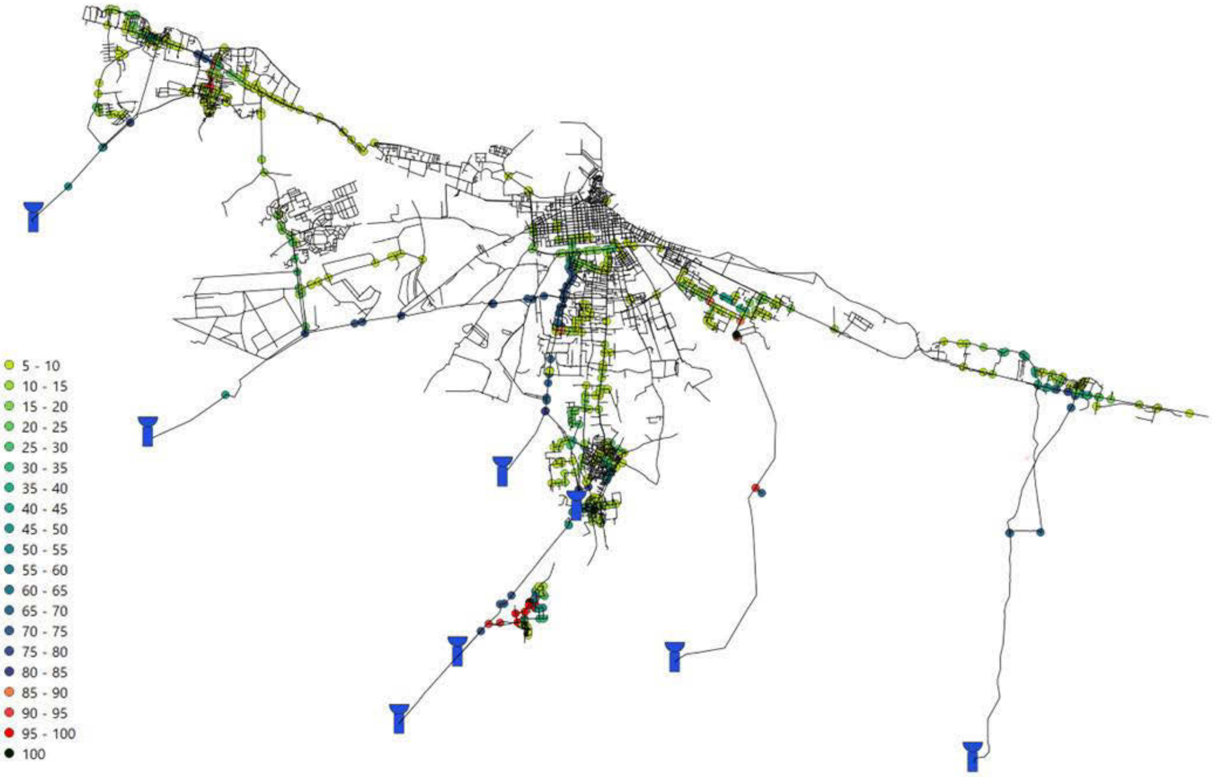


Figure 6. Visualization of the relevance-embedded betweenness centrality for the WDN case study.

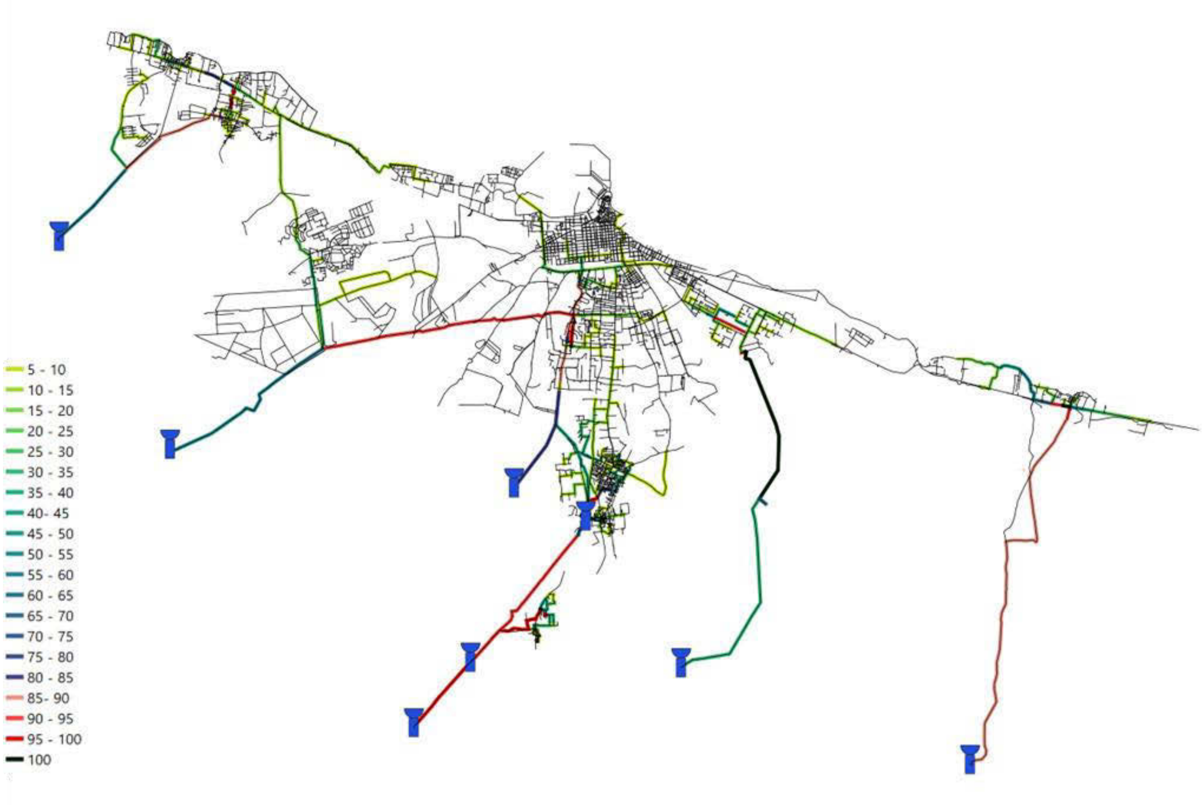


Figure 7. Visualization of the relevance-embedded edge betweenness centrality for the WDN case study.

4 CONCLUSIONS

In this work is presented the application of the domain analysis in a real large WDN, which is composed of six sub-systems hydraulically disconnected from each other. That analysis encompasses the evaluation of two relevance-embedded centrality metrics, weighted according to the formulation proposed by [23], for analysing the emergent hydraulic behavior of WDNs without performing any hydraulic simulation. Results show that it allows getting consistent outcomes through the comparison with the expected hydraulic behaviour, accounting for the simulated pipe flows along the networks. In particular, it has been reported that higher values of centrality metrics for pipes and nodes fit with the main simulated flow paths and demand nodes. The ability to capture the hydraulic behavior, only using the topological analysis, can be useful as a preliminary step for supporting WDNs operational and planning tasks.

Using a large real-life system, also revealed that such analysis can support various planning and management tasks, while it cannot replace the phenomenological representation provided by the hydraulic modelling, which is mandatory to validate each phase of the design workflow.

ACKNOWLEDGMENTS

This paper was partly financed by European Union - PON Ricerca e Innovazione 2014-2020 E FSC, call D.D. 13 July 2017 n. 1735, Project “Energidrica - Efficienza energetica nella gestione delle reti idriche”(code ARS01_00625). Data for figures were provided by IDEA-RT (www.idea-rt.com).

5 REFERENCES

[1] D. Hamberg, and U. Shamir, "Schematic models for distribution systems design. I: Combination concept," Journal of Water Resources Planning and Management, vol. 114, no. 2, 1988, pp. 129-140.

[2] L. Perelman, and A. Ostfeld, "Topological clustering for water distribution systems analysis," Environmental Modelling & Software, vol. 26, no. 7, 2011, pp. 969-972.

[3] O. Giustolisi, Z. Kapelan and D. Savic, "Algorithm for automatic detection of topological changes in water distribution networks," Journal of Hydraulic Engineering, vol. 134, no. 4, 2008, pp. 435-446.

[4] P.J. Carrington, J. Scott and S. Wasserman, Models and methods in social network analysis, 1st ed., Cambridge: Cambridge university press, 2005.

[5] M. Girvan, and M. Newman, "Community structure in social and biological networks," Proceedings of the national academy of sciences, vol. 99, no. 12, 2002, pp. 7821-7826.

[6] P. Pons, and M. Latapy, "Computing communities in large networks using random walks," in International symposium on computer and information sciences, Berlin, Germany: Springer, 2005, pp. 284-293.

[7] M. Barthélemy, "Spatial networks," Physics reports, vol. 499, no. 1-3, 2011, pp. 1-101.

[8] A. Simone, L. Ridolfi, D. Laucelli, and O. Giustolisi, "Complex network theory for water distribution networks analysis," EPiC Series in Engineering, vol. 3, 2018, pp. 1971-1978.

[9] A. Yazdani, and P. Jeffrey, "Robustness and vulnerability analysis of water distribution networks using graph theoretic and complex network principles," in Water Distribution Systems Analysis 2010, Tucson, Arizona, United States: ASCE, 2010, pp. 933-945.

[10] A. Yazdani, and P. Jeffrey, "Complex network analysis of water distribution systems," Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 21, no. 1, 2011, 016111.

[11] R. Sitzenfrei, Q. Wang, Z. Kapelan, and D. Savić, "Using complex network analysis for optimization of water distribution networks," Water resources research, vol. 56, no. 8, 2020, e2020WR027929.

[12] D.B. Laucelli, A. Simone, and O. Giustolisi, "Optimal design of district metering areas for the reduction of leakages," Journal of Water Resources Planning and Management, vol. 143, no. 6, 2017, 04017017.

[13] R. Sitzenfrei, "Using complex network analysis for water quality assessment in large water distribution systems," Water Research, vol. 143, no. 6, 2021, 117359.

[14] M. Newman, J. Scott and S. Wasserman, Networks, 1st ed., Oxford: Oxford university press, 2018.

[15] L.C.Freeman, "A set of measures of centrality based on betweenness," Sociometry , 1977, pp 35-41.

[16] P. Bonacich, "Power and centrality: A family of measures," American journal of sociology, vol. 92, no. 5, 2017, pp. 1170-1182.

[17] A. Bavelas, "Communication Patterns in Task-Oriented Groups," The Journal of the Acoustical Society of America, vol. 22, no. 6, 1950, pp. 725-730.

[18] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol. 18, no. 1, 1953, pp. 39-43.

[19] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," in Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia: Elsevier Science, 1998, pp. 161-172.

[20] O. Giustolisi, A. Simone, and L. Ridolfi, "Network structure classification and features of water distribution systems," Water Resources Research, vol. 53, no. 4, 2017, pp. 3407-3423.

[21] O. Giustolisi, L. Ridolfi, and A. Simone, "Tailoring centrality metrics for water distribution networks," Water Resources Research, vol. 55, no. 3, 2017, pp. 2348-2369.

[22] A. Simone, F.G. Ciliberti, D.B. Laucelli, L. Berardi, and O. Giustolisi, "Edge betweenness for water distribution networks domain analysis," Journal of Hydroinformatics, vol. 22, no. 1, 2020, pp. 121-131.

[23] O. Giustolisi, L. Ridolfi, and A. Simone, "Embedding the intrinsic relevance of vertices in network analysis: the case of centrality metrics," Scientific reports, vol. 10, no. 1, 2020, pp. 1-11.

[24] A. Simone, L. Ridolfi, D. Laucelli, B. Berardi, and O.Giustolisi, "Centrality metrics for water distribution networks," EPiC Series in Engineering, vol. 3, no. 1, 2018, pp. 1979-1988.

[25] WDNetXL 2022 http://www.idea-rt.com/digital-water-latest-innovation.

[26] L., Berardi, O. Giustolisi, "Calibration of Design Models for Leakage Management of Water Distribution Networks". Water Resources Management 35, 2021, 2537–2551

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14777

DEVELOPMENT OF PREDICTION MODEL OF OZONE DOSAGE AND RESIDUAL OZONE CONCENTRATION USING MACHINE LEARNING METHODS IN OZONE PROCESS OF DRINKING WATER TREATMENT PROCESS

J.Y.Kwon[1], J.S.Hyung[2], T.H.Kim[3], H.K.Park[4] and J.Y.Koo[5]

[1,2,3,4,5]Department of Environmental Engineering, University of Seoul
Seoul, Dongdaemun-gu, (Republic of Korea)

[1] *mattis94@naver.com*, [2] *gudwlstjr@uos.ac.kr*, [3] *kimth0712@uos.ac.kr*, [4] *phk940904@uos.ac.kr*,
[5] *jykoo@uos.ac.kr*

Abstract

The ozone process, which is the latter process of the water purification process, injects ozone to remove taste odor substances from tap water. Still, it is difficult to work the ozone process due to recent changes in water quality, such as taste and odor substances due to climate change. Therefore, this study developed an ozone injection rate determination model and a residual ozone concentration prediction model to properly remove flavor odor substances from raw water and proposed an operational diagnosis and optimal decision-making method for the ozone process in water purification. An ozone injection rate determination model and a residual ozone concentration prediction model were developed using data on water quality, flow rate, and operating conditions measured at Seoul's Y water purification plant. Two models were developed: the random forest and the MLP models. The performance difference between the two was verified by comparing the correlation coefficient and error index. Bayesian optimization, a global search method within a given composition space, was used to determine hyperparameters for each model. RMSE was selected as an objective function to determine the optimal hyperparameter through cross-validation. If the above model is applied to the ozone process, it is expected that an immediate response to changes in raw water quality and human error prevention will be possible.

Keywords
Machine Learning, Random Forest, MLP, Bayesian Optimization, Ozone Injection Rate, Residual Ozone Concentration.

1 INTRODUCTION

The water treatment process produces tap water essential for universal activities such as daily human life, economic activities, and industrial activities. Because of this, it should always produce safe tap water. However, due to abnormal phenomena such as torrential rain and algae caused by climate change, the water intake facility's water quality change has intensified. This has increased the difficulty in operating the water treatment process.

Due to an improper ozone injection following the inflow of taste odor substances, it may be possible for some taste odor substances to remain in tap water with the ozone process located at the rear of the water purification process. Therefore, this study developed an ozone injection rate determination model and a residual ozone concentration prediction model to properly remove flavor odor substances from raw water and proposed an operational diagnosis and optimal decision-making method for the ozone process in water purification.

2 MATERIAL AND METHOD

2.1 Current status of the subject

As shown in Figure 1, the Y water purification plant, which takes water from the P water intake plant and produces and supplies purified water, was selected as the study target area. In this plant, ozone is continuously injected at a concentration of 0.5 mg/L or less, and when the concentration of Geosmin and 2-MIB increases in such high temperatures in summer or low water temperatures in winter, ozone is injected in proportion to the concentration of odor substances.

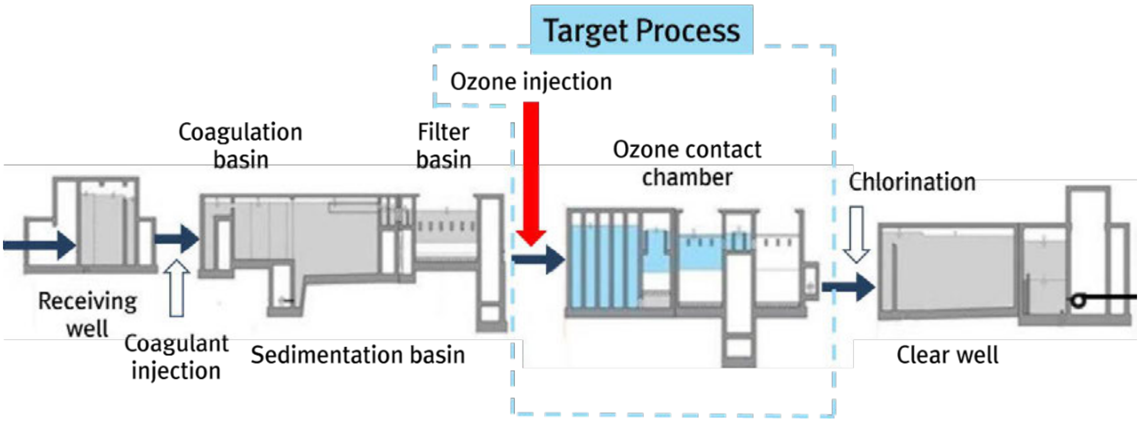


Figure 1. Target water treatment processes in Y water treatment plan

2.2 Configuration of collected data

The Y water purification plant data used in the study, which contains the data of ozone injection, residual ozone concentration, hydrogen peroxide injection, Geosmin injection, and 2-MIB injection, were collected from January 1, 2014, to December 31, 2015, and from January 1, 2017, to November 30, 2018. In the raw water of the study's target water purification plant, 2-MIB concentration reached 269.3 ng/L from 2014 to 2018, and Geosmin concentration reached 80 ng/L. The resulting ozone injection rate was operated up to 2.0 mg/L.

As shown in Figure 2, the change in ozone injection rate for the flavor odor substance concentration increases by injecting more ozone as the 2-MIB concentration and the Geosmin concentration increase. When the 2-MIB concentration increases, more ozone is injected than the Geosmin concentration. The decomposition rate of ozone varies depending on the water temperature, pH, and concentration of dissolved organic matter in water. This makes it difficult to control the proper injection rate and concentration. The ozone injection rate is determined by the concentration of the raw water's flavor odor substance. Still, excessive ozone may be injected to ensure safe water quality, thereby generating residual ozone. As residual ozone causes facility corrosion and impairment of the working environment, it must be controlled.

BY NC SA

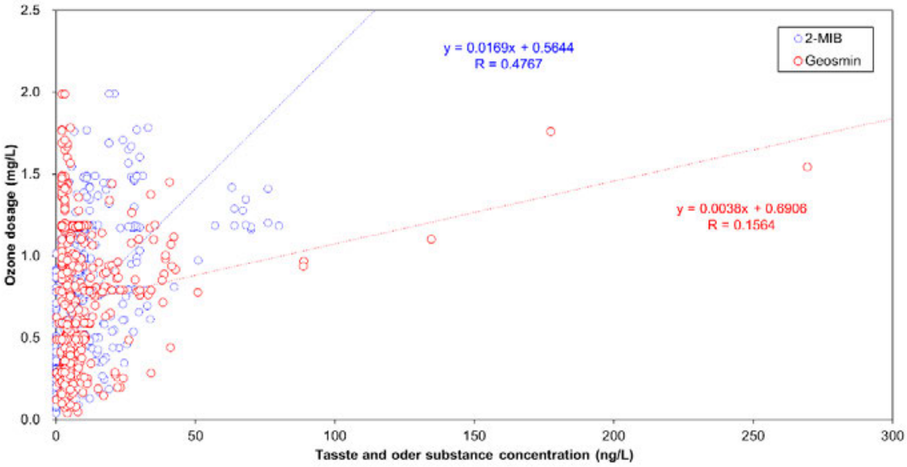


Figure 2. Trend analysis between taste and odor substansces

2.3 Establishment of ozone injection rate determination model and residual ozone concentration prediction model data

As input variables for developing the ozone injection rate model, the concentration of raw water flavored substances such as 2-MIB concentration and Geosmin concentration, raw water temperature, raw water TOC, filtered water turbidity, filtered water pH and ozone contact time, 2-MIB and Geosmin removal rate in the ozone process were considered as input variables. A model for determining ozone injection rate was developed using 32 data of 2-MIB concentration and Geosmin concentration collected during 2014, 2017, and 2018 for raw water, filtered water, ozone-treated water, and activated carbon filtered water. When constructing the ozone injection rate determination model data, learning data and verification data were sorted according to the size of the concentration of taste odor substances generated per day for the final prediction. The learning and verification data were learned with 32 datasets, and predictions were conducted with 6 datasets.

In the case of constructing residual ozone concentration prediction model data, 32 data were used to classify learning and verification data, as in the case of creating an ozone injection rate determination model.

Before developing an artificial intelligence model, a data preprocessing process, which removes and supplements outliers and missing values after data construction, was performed. Then, using the model developed in this study, a prediction period was selected to determine the rate of flavor odor removal by ozone injection, and the data for that period were utilized as prediction data. The rest of the period was used as data for learning and verifying the model.

2.4 Hyperparameter Optimization

As input variables for developing the ozone injection rate model, the concentration of raw water flavored substances In this study, Bayesian optimization was performed to select the optimal hyperparameters of the machine learning model. Bayesian optimization starts with any combination of hyperparameter composition spaces. Then, if the performance improves, change the hyperparameter combination to maintain the combination and, if not, revert to the previous combination. The most significant advantage of Bayesian optimization is that it determines hyperparameters by reflecting prior knowledge, and it generally outperforms random exploration (Bergstra et al., 2011; Bergstra et al., 2013; Falkner et al., 2018; Li et al., 2017).

BY NC SA

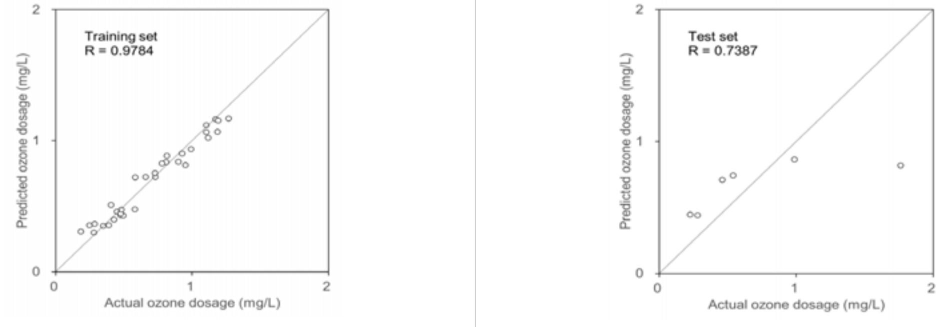
3 RESULTS AND DISCUSSION

3.1 Development of a Model for Determining Ozone Injection Rate

The ozone injection rate determination model was developed by applying optimal hyperparameters and dividing them into learning and verification data for 32 datasets. Based on 6 datasets, the developed model was compared with the actual ozone injection rate. The learning and verification results of the model are presented in Table 1, and the results of visualizing the correlation between the real value and the predicted value are provided in Figure 3. The correlation coefficient was shown as 0.9424 between learning data and model estimation data. In addition, 6 datasets not used for learning were used to predict the ozone injection rate and showed that the OZ-MLP Model had a higher number than the OZ-RF Model. Accordingly, it was found that the OZ-MLP Model is the optimal ozone injection rate determination model with the best performance.

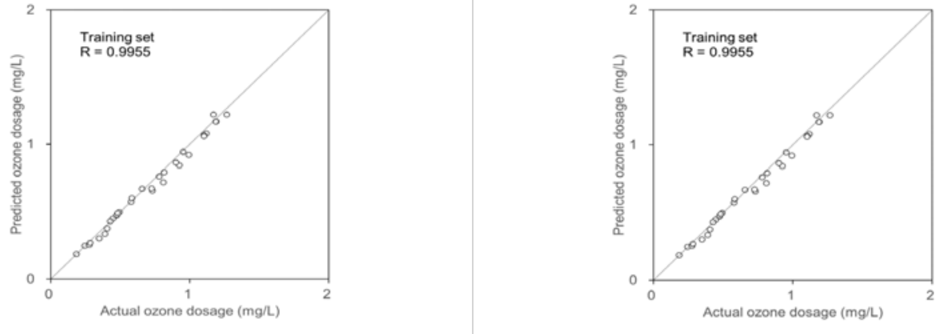
Table 1. Performance evaluation results of ozone dosage determination model using RF, MLP

Model	PI	Model performance	
		Training (n=32)	Test (n=6)
OZ-RF	R	0.9784	0.7387
	MAE	0.06 mg/L	0.32 mg/L
OZ-MLP	R	0.9955	0.9424
	MAE	0.03 mg/L	0.15 mg/L



(a) RF Training result

(b) RF Test result



(c) MLP Training result

(d) MLP Test result

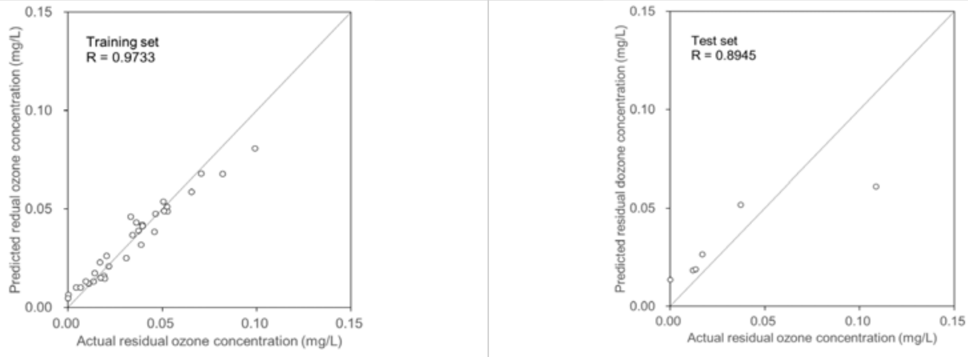
Figure 3. *Training and test results of ozone dosage determination model using RF and MLP*

3.2 Development of Residual Ozone Concentration Prediction Model

Like the ozone injection rate determination model, the residual ozone concentration prediction model was developed by dividing 32 datasets into learning and verification data by using optimal hyperparameters. Then, the developed model was used on 6 datasets and compared with the actual ozone injection rate. The learning and verification results of the model are shown in Table 2 below, and the results visualizing the correlation between the real and predicted values are shown in Figure 4. The correlation coefficient between the learning data and the estimated data was 0.9007. In addition, using 6 datasets not used for learning, the prediction of the ozone injection rate was made. It showed that the RO-MLP model had a higher number than the RO-RF model. Accordingly, it was derived that the RO-MLP model is the optimal residual ozone concentration prediction model with the highest performance.

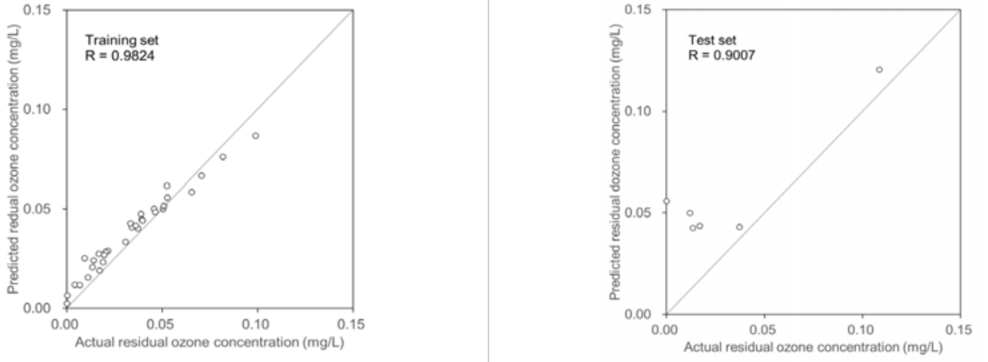
Table 2. Performance evaluation results of ozone concentration prediction model using RF, MLP

Model	PI	Model performance	
		Training (n=32)	Test (n=6)
RO-RF	R	0.9733	0.8945
	MAE	0.005 mg/L	0.016 mg/L
RO-MLP	R	0.9824	0.9007
	MAE	0.006 mg/L	0.028 mg/L



(a) RF Training result

(b) RF Test result



(c) MLP Training result

(d) MLP Test result

Figure 4. *Training and test results of residual ozone concentration prediction model using RF and MLP*

3.3 Comparison with previous studies

As for the ozone process, empirical modeling by experimental and statistical methods was conducted, as shown in Table 3. Still, there were limitations, such as not using data sets on raw water quality or considering Geosmin concentration among flavor-smelling substances.

Table 3. Ozone dosage and residual concentration modeling studies

Study	Input	Output	Model	Best performance
Cromphout et al. (2013)	CT	Ozone dosage	Empirical model	R2=0.86
Hyung et al. (2017)	Residual chlorine, temperature, conductivity, alkalinity, Geosmin concentration	Ozone dosage	Linear regression	(ordinary) R:0.858 (taste and odor substances) R:0.830
		residual ozone concentration		(ordinary) R:0.858 (taste and odor substances) R:0.856

Therefore, in this study, 2-MIB data, a taste-smelling substance other than Geosmin, was also considered, and raw water quality data according to seasonal changes were used to supplement the limitations. In addition, machine learning was applied instead of the regression model, which is a traditional model technique. The R values in previous studies were 0.83~0.858, in contrast to the R values that are 0.9007~0.9424. This increased the predictive power of this study.

4 CONCLUSION

This study developed a machine learning-based model that can perform the optimal operation and decision-making in the ozone process during water purification. The development model developed an ozone injection rate determination model with 2-MIB concentration, Geosmin concentration, water temperature, TOC, filtered water turbidity, filtered water pH, ozone contact time, 2-MIB and Geosmin removal rate in the ozone process as input variables, and a residual ozone concentration prediction model. The optimal ozone injection rate determination model is OZ-MLP Model whose correlation coefficient is 0.9424 based on the predicted data, MAE 0.028 mg/L. And, The optimal residual ozone concentration prediction model is RO-MLP Model which correlation coefficient is 0.9007 based on the predicted data, MAE 0.15 mg/L. For the high quality of tap water, the ozone injection rate for 100% removal of flavor odor substances was determined by the model through treatment up to the ozone process. Based on the model application, it was found that additional ozone should be injected at a minimum of 0.0971 mg/L and at a maximum of 0.2681 mg/L than the actual rate of ozone injection every day. Using the model and methodology developed in this research, it is judged that it is possible to determine the injection rate of hydrogen peroxide by predicting the appropriate ozone injection rate for removing flavor odor substances and the residual ozone concentration. When the above-developed model is applied within the actual process, it can be used as the basis for smart ozone process operation, such as supporting water supply operators' decision-making and responding to rapid changes in raw water. In addition, it is expected that the model's performance can be improved if additional filtered water data or additional taste odor substance measurement data after undergoing a water purification process on the ozone process leaflet are collected in the future.

5 ACKNOWLEDGEMENTS

This subject is supported by Korean Ministry of Environment as Global Top Project (2016002120006).

6 REFERENCES

[1] Bergstra, J., Bardenet, R., Bengio, Y. & Kegl, B, “Algorithms for hyper-parameter optimization,” In Advances in neural information processing systems 24, 2011, pp. 2546-2554.

[2] Cromphout, J., Goethals, S. & Verdickt, L, “, Optimization of the ozone dosage at the drinking water treatment plant of Kluizen,” Water Science & Technology : Water Supply, Vol.13, 2013, pp. 1569-1575.

[3] Hyung, J., Kim, K., Kim, M., Lee, I. & Koo, J, “A study on prediction method for ozone dosage and residual ozone concentration in advanced ozone water tratment,” Water Practice & Technology, Vol.12, 2017, No.1 pp. 87-96.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14779

WHY AREN'T SURROGATE RELIABILITY INDICES SO RELIABLE? CAN THEY BE IMPROVED?

Joaquim José de Oliveira Sousa[1], João Muranho[2], Marco Amos Bonora[3], Mario Maiolo[4]

[1]Polytechnic Institute of Coimbra, Coimbra, Portugal
[2]University of Beira Interior, Covilhã, Portugal
[3]Virtual Laboratory of Smart Water Management, NTT DATA Italia S.p.A., Cosenza, Italy
[4]Virtual Laboratory of Smart Water Management, University of Calabria, Cosenza, Italy

[1] *jjoseng@isec.pt,* [2] *jmuranho@ubi.pt,* [3] *marcoamos.bonora@unical.it,* [4] *Mario.maiolo@unical.it,*

Abstract

Water distribution networks are known to be costly infrastructures. A few decades ago, the research efforts concerning water distribution network design were focused on economic aspects and the goal was to obtain the least-cost solutions. Beyond economics, these infrastructures must mostly be reliable since they provide an essential service to society. But reliability assessment is a complex task and involves various aspects: mechanical, hydraulic, water quality, and water safety, among others. This paper focuses on hydraulic reliability.

As hydraulic reliability is computationally hard to measure directly, researchers came up with surrogate measures, like entropy and the resilience index. But these surrogate measures had some flaws and researchers quickly started suggesting new ones trying to avoid those known flaws, like the diameter-sensitive flow entropy or the modified resilience index. But even these new approaches are still not so reliable to be used in the design of water distribution networks.

This paper presents a performance analysis of the resilience index and a modified version as reliability surrogate measures, supported by illustrative examples. A new version of the resilience index is also proposed, introducing additional coefficients in the attempt to overcome some of the flaws of the previous versions. Some results are presented to compare the performance of the new index with those from the previous versions.

Keywords
Water distribution network design, Reliability, Entropy, Resilience index.

1 INTRODUCTION

The increase of water resources demand due to the development of inhabited areas and production processes, and, concomitantly, the progressive reduction in the availability of water resources due to climate change, impose the necessity for optimization and improvement of the existing infrastructure.

To assess the functioning regime performances, the network managers and the scientific community use synthetic indicators. These indicators numerically describe one or more intrinsic characteristics of a distribution network.

Very often water distribution networks (WDNs) designers refer to the concept of Reliability.

Kaufmann *et al.* [1] define reliability as the probability that the system will perform its specified tasks under specified conditions and during a specified time. Cullinane *et al*. [2] and Goulter [3] define WDN reliability as the ability of the system to meet the demands. In several studies, it is defined as the weighted time-averaged value of the ratio of the flow delivered to the flow required by the users [5,6].

Ciaponi *et al*. [7] define reliability as the ability of a WDN to satisfy users in all possible operating conditions. Reliability is assessed using the ratio between the volume of water delivered to users and the one requested in a given period.

Todini [8] considers the concept of reliability as not completely defined and influenced by numerous factors that are difficult to define. The author introduces the concept of resilience as the system's ability to overcome failures. The formulation allows estimating resilience without the need to analyse all types and combinations of possible failures. An increase in resilience leads to an increase in reliability.

Prasad *et al*. [9] consider the network resilience to be representative of reliability. Network resilience is based on the resilience index proposed in [8]. The resilience index has been modified to reward the presence of loops with similar pipe diameters.

Di Nardo *et al*. [10] consider reliability as an indicator difficult to define, due to the uncertainties affecting the WDN operating conditions knowledge. The authors consider the robustness a better metric. Robustness is defined as the ability of the system to maintain a certain level of service in the presence of unfavourable operating conditions.

Muranho *et al*. [11] define reliability as the ability to satisfy the water demand with sufficient pressure, even in the case of critical operating scenarios. The authors present a comparison between surrogate measures of reliability: resilience [8], network resilience [9], and flow entropy [12,13,14]. A new reliability surrogate measure is presented, the WNG Index (WaterNetGen Index). The index represents the ability to satisfy the water demand in the presence of a pipe failure. The authors present a design methodology that identifies the optimal solution with the maximum reliability subjected to a certain budget. The study shows no strong correlations between the analysed indices (resilience, network resilience and flow entropy) and reliability, but entropy maximization shows an improvement in reliability.

This paper presents a new formulation for the resilience index that keeps the simplicity of the original one but, at the same time, takes into account some network characteristics. Many of the commonly used resilience indices are based on an energy balance, considering the surplus of energy to be dissipated in the event of failure as an indicator of robustness. However, these formulations do not take into account the network topology. In some scenarios, this approach tends to overestimate the network resilience schemes, particularly in tree-shaped networks with a good pressure regime.

The new index presented here is based on the formulation of the resilience indices presented in [8] and [15-18]. Three weight coefficients are integrated within the classic formulation. These coefficients take into account the topology of the network, the importance of the nodes and the uniformity of the diameters of the pipes supplying each node.

2 MATERIALS AND METHODS

2.1 Todini resilience index

[8] introduces the concept of resilience as a surrogate measure of robustness. Robustness is a measure of the system's ability to overcome failures. The author defines the resilience index and the failure index as design metrics. The optimal design scheme is achieved by maximizing resilience and limiting the cost. [8] builds the resilience index from an energy balance.

The Todini resilience index, like many other known ones, refers to the "requested" or "design" conditions. These conditions are water demand and piezometric head values that must be reached to ensure proper network functioning. The requested conditions and variables are indicated by an asterisk apex.

The formulation is based on the concept that a network that has a pressure surplus is more robust in case of pipe breaks or anomalous hydraulic events. The resilience index, I_r - equation (1), is structured as the complement to the ratios between the power/energy dissipated in the WDN (P_D*) and the maximum dissipable value (P_{Dmax}) to meet the target/design values.

$$I_r = 1 - \frac{P_D^*}{P_{D\,max}} \tag{1}$$

The total available power (P_{tot} - equation (2)) at the entrance of a WDN is:

$$P_{tot} = \gamma \sum_{k=1}^{r} Q_k H_k + \sum_{j=1}^{p} P_j \tag{2}$$

in which:

- γ water specific weight;
- r number of sources (tanks, reservoirs);
- Q_k source discharge (flow entering the network);
- H_k source piezometric head;
- p number of pumps;
- P_j pump power.

The global minimum output power (P_{Emin} - equation (3)) is the global sum of the power that must be delivered at each demand node to satisfy the design piezometric head and demand:

$$P_{E\,min} = \sum_{i=1}^{n} p_i^* = \gamma \sum_{i=1}^{n} q_i^* h_i^* \tag{3}$$

in which:

- n number of network nodes;
- p_i^* design power of the ith node;
- q_i^* design water demand of the ith node ;
- h_i^* design piezometric head of the ith node.

The maximum dissipable power (P_{Dmax} - equation (4)) is the highest power that can be used without compromising the accomplishment of the design values:

$$P_{D\,max} = P_{tot} - P_{E\,min} = \left(\gamma \sum_{k=1}^{r} Q_k H_k + \sum_{j=1}^{p} P_j\right) - \gamma \sum_{i=1}^{n} q_i^* h_i^* \tag{4}$$

The total amount of actual power delivered to the demand nodes (P_E - equation (5)) is:

$$P_E = \gamma \sum_{i=1}^{n} q_i h_i \tag{5}$$

in which:

- q_i water delivered to the i^{th} node;
- h_i piezometric head of the i^{th} node.

The total amount of power dissipated in the network ($P_D{}^*$ - equation (6)) to satisfy the total demand is:

$$P_D^* = P_{tot} - \gamma \sum_{i=1}^{n} q_i^* h_i \tag{6}$$

The resilience index (I_r - equation (7)) can be written as:

$$I_r = \frac{\sum_{i=1}^{n} q_i^* (h_i - h_i^*)}{\sum_{k=1}^{r} Q_k H_k + \sum_{j=1}^{p} \frac{P_j}{\gamma} - \sum_{i=1}^{n} q_i^* h_i^*} \tag{7}$$

2.2 Di Nardo *et al.* alternative resilience index

Di Nardo *et al.* [15-18] present an alternative formulation for the resilience index, equation (8).

$$I_R = 1 - \frac{P_D}{P_{D\,max}} \tag{8}$$

The index uses the total dissipated power (P_D - equation (9)) instead of the amount of power dissipated in the network. The total dissipated power is obtained as:

$$P_D = \gamma \sum_{j=1}^{m} q_j \Delta h_j \tag{9}$$

in which:

- m total number of pipes in the network;
- Δh_j piezometric head dissipated along the j^{th} pipe;
- q_j flow along the j^{th} pipe.

The total dissipated power (P_D - equation (10)) can be obtained as:

$$P_D = P_{tot} - P_E \tag{10}$$

The resilience index (I_r - equation (11)) is:

$$I_R = 1 - \frac{P_D}{P_{D\,max}} = \frac{\sum_{i=1}^{n}(q_i h_i - q_i^* h_i^*)}{\sum_{k=1}^{r} Q_k H_k + \sum_{j=1}^{p} \frac{P_j}{\gamma} - \sum_{i=1}^{n} q_i^* h_i^*} \tag{11}$$

The use of resilience indices in WDN design rewards networks with an energy surplus that can be dissipated in the event of a failure or an increase in user demand. The limitation of the resilience indices is that their assessment does not take into account the network topology or the necessary connectivity and pipe diameter balance, assumes that nodes without demand do not contribute to the reliability and every node with demand has the same importance. In some cases, a tree-like network topology, obviously not very resilient to failure, with a high enough piezometric head surplus can obtain high resilience values.

2.3 Network resilience index

[9] present an alternative formulation of the Todini resilience index (I_r) which reward the presence of loops in the network, penalizing sudden changes in diameter. To take into account the variability of the diameter, the authors define a uniformity coefficient (C - equation (12)):

$$C_i = \frac{\sum_{j=1}^{npi} D_j}{npi * \max(D_j)} \tag{12}$$

in which:

- C_i uniformity coefficient for the i^{th} node;
- npi number of pipes connected to node i;
- D_j diameter of pipes connected to node i.

The coefficient gets a value $C = 1$ if pipes connected to a node have the same diameter and $C < 1$ if pipes connected to a node have different diameters. The weighted surplus power combines the effect of surplus power and nodal diameter uniformity. The weighted surplus power for node i (X_i - equation (13)) is:

$$X_i = C_i p_i = C_i \gamma q_i (h_i - h_i^*) \tag{13}$$

where p_i is the surplus power of the i^{th} node.

The total weighted surplus power (X - equation (14)) is:

$$X = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} C_i \gamma q_i (h_i - h_i^*) \tag{14}$$

The network resilience (I_r - equation (15)) is:

$$I_r = 1 - \frac{X}{X_{max}} = \frac{\sum_{i=1}^{n} C_i q_i (h_i - h_i^*)}{\sum_{k=1}^{r} Q_k H_k + \sum_{j=1}^{p} \frac{P_j}{\gamma} - \sum_{i=1}^{n} q_i h_i^*} \tag{15}$$

where X_{max} is the maximum surplus power. It is assumed that the nodal design demand is fully satisfied ($q_i^* = q_i$). This version of the reliability index intends to take into account the pipe diameter balance. However, the uniformity coefficient is computed from all the pipes connected to a node (in and out) and not only from the pipes supplying the node (in). A node that is supplied by one single pipe (like in a tree-like network) may present a good uniformity coefficient and consequently a good resilience index.

3 WEIGHTED RESILIENCE INDEX

This paper presents a set of three coefficients that modify the weight of each WDN junction in the resilience index assessment. The three coefficients represent characteristics of the network which are not commonly taken into account by the classical formulation of the resilience indices.

The *Topological coefficient* takes into account the network topology. This coefficient aims to penalize the junctions that have a low number of connections in the resilience assessment. In order not to add complexity to the calculations using graph theory algorithms, this coefficient is calculated knowing only the number of pipes supplying each node.

The *Importance coefficient* defines a hierarchy of importance of the pipes. The break of main pipes carrying larger volumes of water has a greater negative impact on the WDN hydraulic behaviour.

The *Uniformity coefficient* takes into account the pipes' diameter uniformity. A network for which there are several pipes of similar diameter connected to each junction reacts better in case of the failure of one of them.

3.1 Topological coefficient

The *topological coefficient* reduces the contribution of the junctions for which there is a single entering pipe. It is assumed that junctions with a single entering pipe contribute less to the network resilience.

The *topological coefficient* (K^T - equation (16)) is a multiplicative coefficient that can assume values between 0.5 and 1.5 and can be estimated as:

$$K_j^T = 0.5 + \frac{N_{in}^j - 1}{N_{in}^j} \tag{16}$$

where:

- N_{in}^j: Total number of pipes entering junction j.

This coefficient aims to reduce the flaw in the resilience assessment for tree-like networks. This type of network is notoriously non-resilient (the break of a pipe completely stops the supply of water for all the downstream pipes). The calculation of Todini's resilience index for a tree-like network can lead to misleading results (high resilience) in the presence of a fairly high pressure surplus.

The presence of a coefficient that tends to reduce the importance of the junctions for which there is a low connection redundancy should considerably reduce the resilience of the tree networks, reflecting more the reality.

As defined, this coefficient assigns a reduction value (0.5) to the configurations in which the network junctions are supplied by a single pipe. For configurations in which two pipes are supplying a junction, the coefficient is unitary. The increase in the number of incoming pipes has a progressively smaller effect on the improvement of resilience. Due to its mathematical structure, the coefficient has an upper limit of 1.5 (Figure 1).

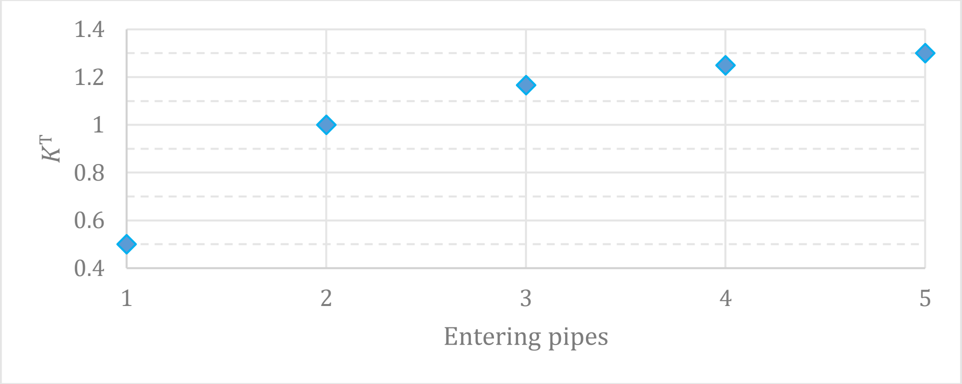


Figure 1. The topological coefficient variation.

3.2 Importance coefficient

The *importance coefficient* is a multiplicative coefficient that assumes values lower than 1. In drinking WDNs, not all junctions have the same importance. This coefficient is based on the assumption that the junctions through which a greater flow passes are more important for the network functioning, and therefore their resilience (or lack of it) is more impactful overall.

The *importance coefficient* (K^I - equation (17)) can be estimated as:

$$K_j^I = \frac{\sum_{i=1}^{N_{in}^j} Q_{in_i}^j}{Q_{in}^{MAX}} \tag{17}$$

where:

- N_{in}^j: Total number of pipes entering junction j;
- $Q_{in_i}^j$: Flow of the i^{th} pipe that enters junction j;

The denominator of the formula refers to the maximum value among all the flows entering the network junctions (Q_{in}^{MAX} - equation (18)):

$$Q_{in}^{MAX} = \max\,(Q_{in}^1, \dots, Q_{in}^j, \dots, Q_{in}^n) \tag{18}$$

where:

- Q_{in}^i: Flow entering junction i.

The presence of a coefficient that weights more on the junctions through which more flow passes, allows to better take into account the areas near the tanks or main pipelines in the resilience assessment. A junction located in the peripheral area of the network has a marginal impact compared to the water mains near a tank.

3.3 Uniformity coefficient

In a WDN, the connection redundancy does not ensure resilience by itself. The *uniformity coefficient* is based on the assumption that the pipes converging into a junction are effectively redundant, and therefore resilient, the more their diameters are similar.

The *uniformity coefficient* is a multiplicative coefficient that rewards the uniformity of the diameters and penalizes situations in which the diameters of the incoming pipes are very different because they are not very resilient. In general, it will assume values between 0 and 1, but in specific situations it can surpass 1.

The *uniformity coefficient* (K^U - equation (19)) can be assessed as:

$$K_j^U = \frac{\sum_{i=1}^{N_{in}^j} \left(D_{in_i}^j\right)^2}{MIN(N_{in}^j, 2) * \left(MAX(D_{in_i}^j)\right)^2} \tag{19}$$

where:

- N_{in}^j: Total number of pipes entering junction j;
- $D_{in_i}^j$: Diameter of the i^{th} pipe that enters junction j.

The structure of the *uniformity coefficient* is similar to that presented by [19] but here it takes into consideration only the pipes entering the junctions and not the diameter but its square because the pipe section is proportional to the square of the diameter. As it is formulated, the coefficient assumes unitary values when there is one single pipe entering the junction (in this case the resilience index is reduced by the topological coefficient) or when the two pipes entering a junction have the same diameter, and it can present lower or higher values in other situations.

3.4 Weighted resilience indices

The *Topological*, *Importance* and *Uniformity coefficients* are three dimensionless multiplicative coefficients that can be integrated into different formulas for the assessment of resilience indices. The coefficients are calculated for each junction and multiply the numerator of the formula. In general, these coefficients reduce the numerator, but in specific situations may increase it. As they are formulated, a resilience index that integrates these coefficients should be lower than the original one.

Equations (20) and (21) respectively show the suggested new formulations for the resilience index of [8] and [15-18].

$$I_r = \frac{\sum_{i=1}^{n} (K_i^I K_i^T K_i^U)\, q_i^* (h_i - h_i^*)}{\sum_{k=1}^{r} Q_k H_k - \sum_{i=1}^{n} q_i^* h_i^*} \tag{20}$$

$$I_R = \frac{\sum_{i=1}^{n} (K_i^I K_i^T K_i^U)(q_i h_i - q_i^* h_i^*)}{\sum_{k=1}^{r} Q_k H_k - \sum_{i=1}^{n} q_i^* h_i^*} \tag{21}$$

4 CASE STUDIES

The network used in the simulations serves the zone of Villa Rosa. This area is an isolated portion of the network serving the area called Northwest System, in the city of Tampa (Florida, USA). The served area covers approximately 2 square kilometres and is a residential area. The topography of the area is very regular. The junctions of the network have an average elevation of 17.6 m (and are between 16.4 and 19.62 m). The network is supplied by a reservoir which represents the connection of the subnet to the water main of the network. The reservoir has a total head of 54.75 m and supplies 162.95 l/s to the network.

4.1 The three scenario networks

Three WDNs (depicted in Figures 1, 2 and 3) called Normal, Looped and Treelike were used to test the performance of the proposed resilience index and compare it with the original version.

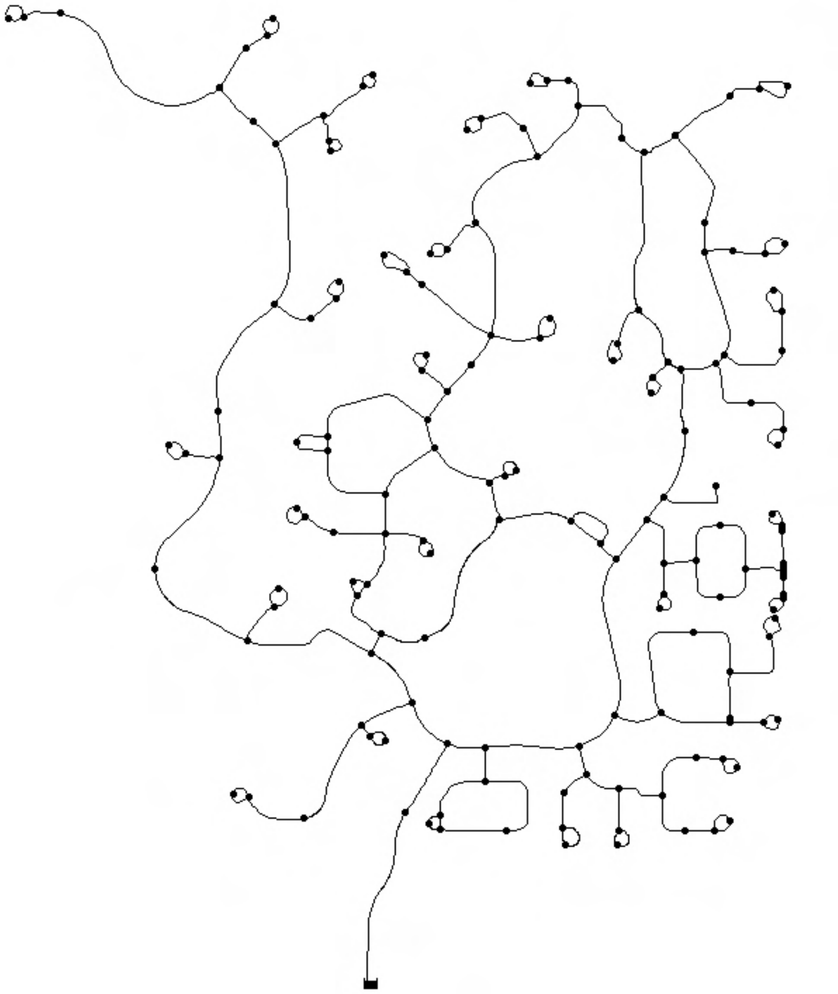


Figure 2. Model of the Villa Rosa neighbourhood WDN in the city of Tampa, Florida - Normal.

The Normal variant is the network that currently supplies Villa Rosa. The other two networks are based on the Normal one but have been modified to test the effect of the weight coefficients on the resilience index.

Table 1. The number of junctions and pipes in the three variants of the Villa Rosa network.

	Normal	Looped	Treelike
Number of Junctions	163	180	163
Number of Pipes	208	254	163

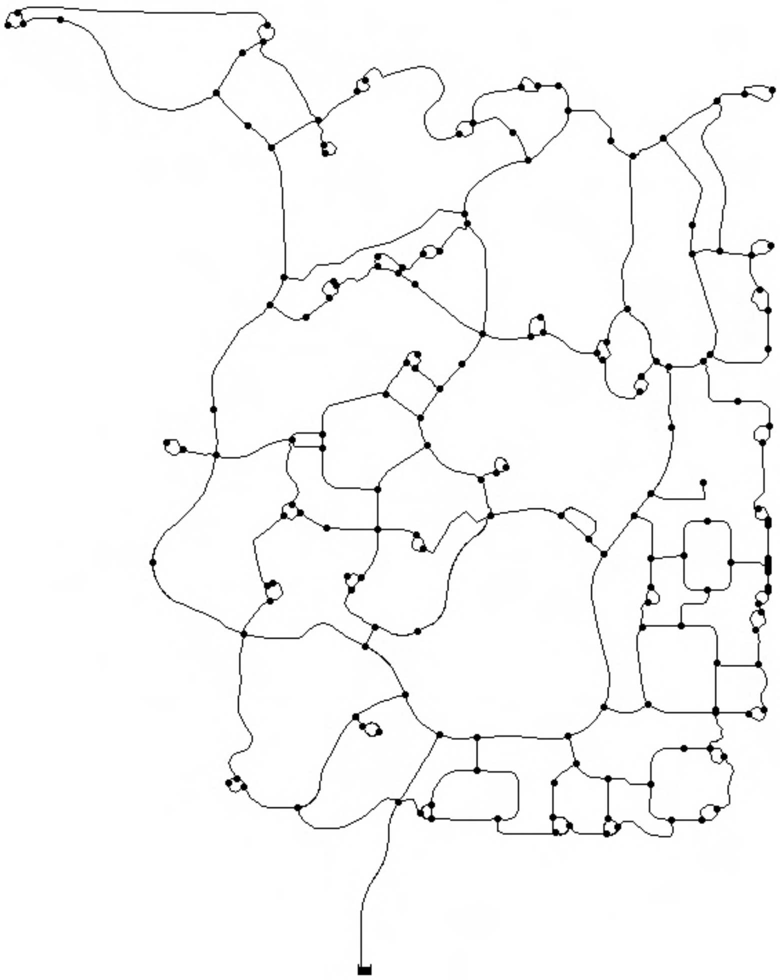


Figure 3. Model of the Looped variant of the Villa Rosa WDN.

To build the Looped network, 17 junctions and 46 pipes were added to the network currently in operation. The pipes added to the network were chosen to maximize the number of loops within the network. The diameter of the added pipes was chosen consistently with the diameters currently present. The 17 junctions added to the network have the function of simplifying the connection of the new pipes and do not modify the distribution or the value of the water demand.

To build the Treelike network, 45 pipes from the network currently in operation were deleted. Pipe loops and some connections between different areas of the network were removed. This operation made it possible to build a completely treelike network (each junction is supplied by a single upstream pipe). Also, in this case, the changes to the network did not change the distribution or the value of the water demand.

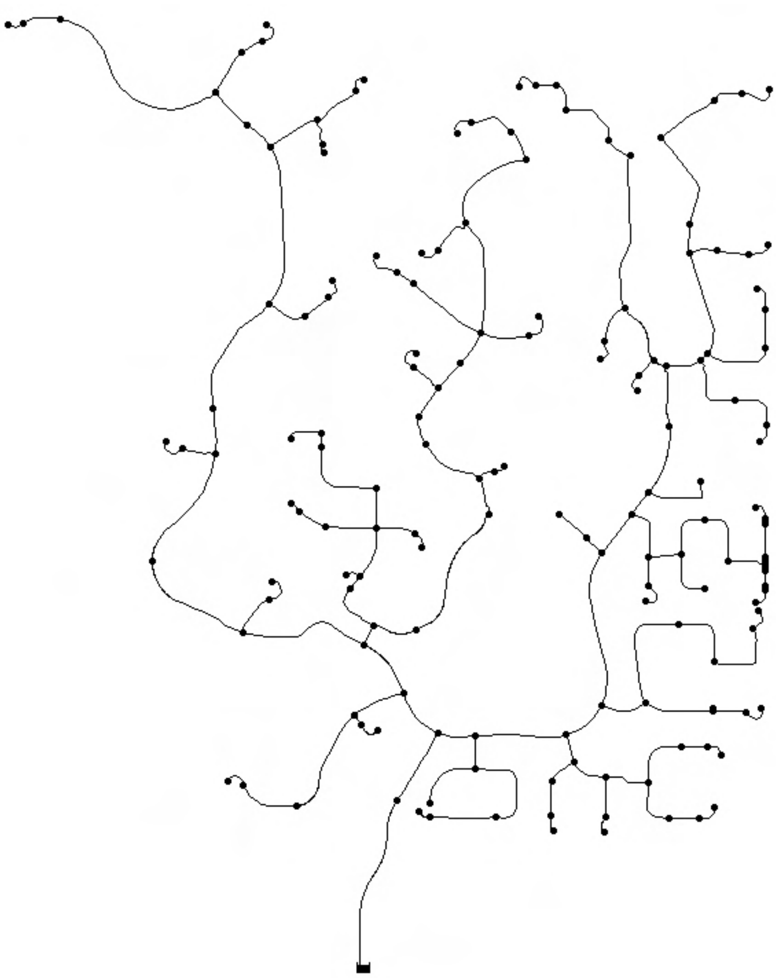


Figure 4. Model of the Treelike variant of the Villa Rosa WDN.

4.2 Resilience indices assessment

The assessment of the resilience indices requires knowledge of the WDN operating regime and a set of information that describes the design or required conditions. Design conditions are used in many of the resilience indices found in the literature [8-9,15-18]. The literature in most cases does not investigate the methodology or the criterion used to assess these required values.

Authors have investigated on several occasions the necessary hypotheses and simplifications useful for estimating the conditions required in different application cases [20-22].

Taking into account the purpose of this work, authors used a simplified approach, defining a single requested pressure value. This simplifying hypothesis does not overly trivialize the results thanks to the homogeneity of the served area. The Villa Rosa neighbourhood is almost entirely residential. The supplied dwellings are small, single-storey houses.

Many of the literature indices use a simplified approach by defining a single required pressure value (P^*) for the entire network [9,15-18]. In the cited bibliography the requested demand value is not defined, assuming that the supply always corresponds to the demand. This hypothesis was also used here. Under this hypothesis, the index defined by di Di Nardo *et al.* and the one defined by Todini are the same.

The resilience indices were assessed for the three networks (Table 2). The Villa Rosa network is characterized by a good pressure regime. The absence of pressure deficits makes it possible to achieve a fairly high resilience index value. The Treelike network has a slightly lower I_r value (-

BY NC SA

3.09%) while the Looped network has a slightly higher value (+1.79%). The variations are due to changes in the hydraulic operating regime of the WDN.

Table 2. Value of the resilience indices assessed for the three variants of the Villa Rosa network.

$P^* = 20m$	Treelike	Normal	Looped
Resilience index:	0.784	0.809	0.823
Weighted Resilience index:	0.0239	0.0282	0.0295

The weighted resilience index is characterized by lower resilience values. The multiplicative coefficients, which are usually less than unity, reduce the overall resilience value.

An interesting aspect of this index is its sensitivity to topological changes in the network. As known, tree networks are not resilient. In the case in question, the variant of the Villa Rosa network, modified to be a tree network, sees a reduction in the resilience index of 15.25% (the original resilience index was only reduced by 3.09%). Even the Looped network, characterized by a more redundant structure, undergoes an increase of 4.61% in the weighted resilience index (the original resilience index only increased by 1.79%). Tables 3 to 5 present some statistics of the results obtained for the three networks.

Table 3. Statistics on weight coefficients estimated for the Treelike network. The average value of the coefficients and percentage of the coefficients that are lower, higher and equal to 1.

Treelike				
	Average	<1	=1	>1
K_t	0.500	100.00%	0.00%	0.00%
K_u	1.000	0.00%	100.00%	0.00%
K_i	0.065	99.39%	0.61%	0.00%

Table 4. Statistics on weight coefficients estimated for the Villa Rosa (Normal) network. The average value of the coefficients and percentage of the coefficients that are lower, higher and equal to 1.

Normal				
	Average	<1	=1	>1
K_t	0.638	72.39%	27.61%	0.00%
K_u	0.994	1.84%	98.16%	0.00%
K_i	0.067	99.39%	0.61%	0.00%

Table 5. Statistics on weight coefficients estimated for the Looped network. The average value of the coefficients and percentage of the coefficients that are lower, higher and equal to 1.

Looped				
	Average	<1	=1	>1
K_t	0.702	60.00%	38.89%	1.11%
K_u	0.964	13.33%	85.56%	1.11%
K_i	0.063	99.44%	0.56%	0.00%

As from Tables 3-5, a tree network has the most reductive set of coefficients. The topological coefficient has the greatest reduction effect, while the uniformity coefficient is higher since each junction of the network is supplied by a single pipe. Values greater than 1 of the multiplicative coefficients are very rare and only arise in the parts of the network with greater redundancy.

5 CONCLUSIONS

This paper presents a variant of the resilience index proposed in [8] and in [15-18]. Commonly, the indices aim to assess the resilience of a WDN taking into account an energy balance (pressure surplus) and referring to the required conditions. Such an approach is often not very effective in representing resilience in specific cases (i.e. tree-shaped networks) since it does not take into account the topological characteristics.

The variant proposed here allows taking into account directly some aspects that are ignored or indirectly taken into account by the analysed indices. The use of three coefficients allowed to give greater weight to the redundancy of the connections, especially if using pipes with similar diameters, and to differentiate the impact on the resilience of junctions on water mains and the ones connected to smaller pipes.

Three case studies were analysed. The topological changes made to the Villa Rosa network correspond to a decrease (*Treelike*) and an increase (*Looped*) in resilience.

Using the weighted resilience index, as can be seen from the results, a change in the topology of the analysed network produces a greater variation in the resilience index compared to the classic formulations.

The proposed variant was found to be more effective in representing the various aspects that contribute to the resilience of the network. Although authors are convinced that using these coefficients is a step forward in improving the reliability assessment using the resilience index, there are still some issues to solve and those will be addressed in future works.

6 REFERENCES

[1] A. Kaufmann, R. Cruon, and D. Grouchko, “Mathematical Models for the Study of the Reliability of Systems,” Elsevier, 1977.

[2] M. J. Cullinane, K. E. Lansey, and L. W. Mays, “Optimization-availability-based design of water-distribution networks,” in Journal of Hydraulic Engineering, 1992, 118(3), 420-441.

[3] I. Goulter, “Analytical and simulation models for reliability analysis in water distribution systems”, in Improving efficiency and reliability in water distribution systems, Springer, Dordrecht, 1995, pp. 235-266.

[4] Y. Setiadi, T. T. Tanyimboh and A. B. Templeman, “Modelling errors, entropy and the hydraulic reliability of water distribution systems,” in Advances in Engineering Software, 2005, 36(11-12), 780-788.

[5] T. T. Tanyimboh, “An entropy-based approach to the optimum design of reliable water distribution networks”, Doctoral dissertation, University of Liverpool, 1993.

[6] T. T. Tanyimboh. And C. Sheahan, “A maximum entropy based approach to the layout optimization of water distribution systems,” in Civil Engineering and Environmental Systems, 2002, 19(3), 223-253.

[7] C. Ciaponi, L. Franchioli, and S. Papiri, “Simplified procedure for water distribution networks reliability assessment,” in Journal of Water Resources Planning and Management, 2011, 138(4), 368-376.

[8] E. Todini, “Looped water distribution networks design using a resilience index based heuristic approach,” in Urban water, 2000, 2(2), 115-122.

[9] T. D. Prasad and N. S. Park, “Multiobjective genetic algorithms for design of water distribution networks,” in Journal of Water Resources Planning and Management, 2004, 130(1), 73-82.

[10] A. Di Nardo, R. Greco, M. Di Natale, and G.F. Santonastaso, “Resilienza ed entropia come indici di robustezza delle reti di distribuzione idrica”, in Quinto seminario su “La diagnosi e la gestione dei sistemi idrici”, 2012, (pp. 225-232).

[11] J. Muranho, J. Sousa, A. S. Marques, and R. Gomes, “Water distribution network reliability: are surrogate measures reliable?,” in 13th International Conference on Hydroinformatics (HIC 2018), EPiC Series in Engineering, 2018, Vol. 3, pp. 1470-1461..

[12] T. T. Tanyimboh, and A. B. Templeman, “Calculating maximum entropy flows in networks,” in The Journal of the Operational Research Society, 1993n 44(4), 383-396.

[13] T. T. Tanyimboh, & A. B. Templeman, “Optimum design of flexible water distribution networks,” in Civil Engineering Systems, 1993, 10(3), 243-258.

[14] H. Liu, D. Savic, Z. Kapelan, M. Zhao, Y. Yuan, H. Zhao, “A diameter-sensitive flow entropy method for reliability consideration in water distribution system design,” in Water Resources Research, 2014, 50, pp. 5597–5610.

[15] A. Di Nardo and M. Di Natale, “A design support methodology for district metering of water supply networks,” in Water Distribution Systems Analysis, 2010, pp. 870-887.

[16] A. Di Nardo and M. Di Natale, “A heuristic design support methodology based on graph theory for district metering of water supply networks,”in Engineering Optimization, 2011, 43(2), 193-211.

[17] A. Di Nardo, M. Di Natale and G.F. Santonastaso, “A comparison between different techniques for water network sectorization,” in Water Science and Technology: Water Supply, 2014, 14(6), 961-970.

[18] A. Di Nardo, M. Di Natale, G.F. Santonastaso, V. G. Tzatchkov and V. H. Alcocer-Yamanaka, “Performance indices for water network partitioning and sectorization,” in Water Science and Technology: Water Supply, 2015, 15(3), 499-509.

[19] T.D. Prasad, H. Sung-Hoon and P. Namsik, “Reliability based design of water distribution networks using multiobjective genetic algorithms,”in KSCE J. of Civil Engineering, 2003, 7(3), pp. 351-361.

[20] M. A. Bonora, F. Caldarola, and M. Maiolo, “A new set of local indices applied to a water network through Demand and Pressure Driven Analysis (DDA and PDA),” in Water, 2020, 12(8), 2210.

[21] M. A. Bonora, F. Caldarola, M. Maiolo, J. Muranho and J. Sousa, “The New Set Up of Local Performance Indices into WaterNetGen and Application to Santarém’s Network,” in Environmental Sciences Proceedings, Multidisciplinary Digital Publishing Institute, 2020, Vol. 2, No. 1, p. 18..

[22] M. A. Bonora, F. Caldarola, J. Muranho J. Sousa, and M. Maiolo, “Numerical experimentations for a new set of local indices of a water network,” in International Conference on Numerical Computations: Theory and Algorithms, Springer, Cham,2019, pp. 495-505.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14781

ENERGY EQUATIONS TO ANALYZE PRESSURIZED WATER TRANSPORT SYSTEMS

Roberto del Teso[1], Elena Gómez[2], Elvira Estruch-Juan[3] and Enrique Cabrera[4]

[1,2,3,4] ITA. Department of Hydraulic Engineering and Environment. Universitat Politècnica de València.
Valencia, Valencia, (Spain)

[1] *rodete@ita.upv.es*, [2] *elgosel@ita.upv.es*, [3] *maesjua1@ita.upv.es*, [4] *ecabrera@ita.upv.es*

Abstract

Pressurized water systems are high energy demanders. They must deliver the demanded volume of water at the minimum service pressure. There is currently a marked change in production and energy tariffs, which is causing prices for water services to rise. In addition to the economic aspect, the energy demand of water systems has an environmental implication linked to gas emissions. With increasing demands, the climate change situation and the shift in energy tariffs, it is necessary to improve the efficiency of pressurized water transport systems. The lower the kWh required to supply, the more energy efficient they will be.

The energy analysis of the systems will allow us to know the current situation of the systems, and to know if it is necessary to undertake improvement measures. There are different processes to know the energy status of the networks. The simplest one is to carry out a diagnosis, with little data, which will give an initial idea of the energy status of the networks. This diagnosis can be carry through by applying Bernoulli's equation. Applied from the supply points to the most unfavorable point of the system.

If it is desired to know more precisely the energetic status of the networks, the use of the integral energy equation seems more reasonable. This equation makes it possible to audit the system and to know in detail the energy use for a given control volume. It allows the energy supplied to be broken down into: useful energy, structural energy losses (linked to topography) and operational energy losses (friction losses, pumping losses, leakage losses and excess energy). In order to be able to apply this method, it is mandatory to have the mathematical model of the system.

This paper discusses the advantages and disadvantages of carrying out the energy analysis of a system using the Bernoulli equation (diagnosis) or the energy integral equation (audit), and when it is convenient to apply one or the other. On one hand, the Bernoulli equation makes it possible to estimate the energy level of the network with very little data, without knowing exactly in which processes the energy introduced into the system is invested. The audit based on the integral energy equation, on the other hand, requires precise data collection and mathematical modelling, but it will provide a detailed understanding of the energy breakdown of the network.

Depending on the objective of the energy analysis, it seems reasonable to apply one process or another. For a first estimation and as a start of the energy analysis, it will be sufficient to carry out a diagnosis as quickly as possible, which will allow to know if it is necessary to continue with a more in-depth research of the system status such as an energy audit.

Keywords
Energy audit, water-energy nexus, energy equation, water distribution systems, energy efficiency.

1 INTRODUCTION

Energy consumption in water transport is large. In Europe, pumps are responsible for 10 % of the electrical energy [1], while in California, pressurised water transport accounts for 6 % of the energy demand [2]. In Spain, energy consumption due to irrigation accounts for 3% of the country's total electrical energy [3], while in Israel pumping accounts for more than 8% of the total electricity consumed in the country [4]. Climate change calls for increased efficiencies, both water and energy.

In order to improve the efficiency of pressurised water systems, it is necessary to analyse their energy status and find out the type of energy losses in order to propose improvements and changes to optimise their efficiency. Water transport systems can be differentiated into simple and complex systems. Simple systems are made up of a pipe and a pump that transfers water between end nodes. Complex systems are networks with multiple delivery points, where the operating conditions are imposed by the most unfavourable or critical point [5].

In complex systems, by guaranteeing the necessary energy at the critical node, the rest of the nodes are given more energy than necessary. The sum of all these excesses, the greater the more irregular the terrain, is the topographic energy [6]. From an energetic perspective, topographic energy is the big difference between simple and complex systems. This energy, as far as possible, should be minimised at the design stage [7], and if the system is already operating, it should be managed as efficiently as possible to minimise its losses [8]. Since these losses do not depend on the system's operating mode, they are called structural losses, which do not exist in simple systems. In these systems, it is sufficient to minimise operational losses (related to the pumping station, those linked to leaks and pipe friction).

The energy analysis in simple and complex systems can be carried out in different ways, depending on the type of system to be analysed and the depth of the analysis, it will be more convenient to choose one method or another. The energy analysis to be carried out will depend on the quantity and quality of the installation's data. A first diagnosis allows detecting whether there is a need for a more specific analysis detailing the energy consumed by the system. This diagnosis can be carried out with little data. If the result of the diagnosis indicates that there is considerable room for improvement, it will be necessary to carry out a network audit, for which it is essential to have the corresponding mathematical model, and therefore much more precise data.

Two main equations can be distinguished for the energy analysis of pressurised water transport. On the one hand, the Bernoulli equation, on the other hand, the energy integral equation.

Contrary to what happens in thermal fluid mechanics, which is always governed by the integral energy equation, the resolution of hydraulic energy problems, in which the thermal term is neglected, is often based on Bernoulli's equation. This equation does not allow heat balances to be carried out, so when these effects are negligible, the fluid is incompressible and certain conditions are given that are typical of the transport of water under pressure, its use is accepted. In this case, both equations are valid, both the general energy equation and the Bernoulli equation, although with their nuances (such as the one-dimensionality of the Bernoulli equation as opposed to the spatiality of the integral energy equation). This work indicates which equation should be applied depending on the type of water transport system and the depth of the study, highlighting similarities and differences between the two equations.

2 BERNOULLI EQUATION

The well-known Bernoulli equation, expressed in energy per unit weight (m), is usually formulated as follows:

$$\frac{p_1}{\gamma}+\frac{1}{2g}V_1^2+z_1=\frac{p_2}{\gamma}+\frac{1}{2g}V_2^2+z_2=C \tag{1}$$

With p the pressure, γ the specific weight of the fluid (9810 N/m^3), g the acceleration of gravity, v the velocity and z the elevation.

Most hydraulic flows are assumed to be one-dimensional, which means that Bernoulli's equation is often used to solve this type of problem. It is an expression that relates the different ways of storing energy in a fluid, with the exception of thermal energy, which is irrelevant in most cases. It poses an energy balance between two points on the same streamline, distinguishing three different summands that allow each type of energy to be recognised: energy in the form of pressure, kinetic energy and potential energy.

For the use of Bernoulli's equation in the energy analysis of pressurised water transport, it must be applied to a pipe, associating its axis to the reference streamline. Furthermore, it must be generalised to include every existing energy input or loss. As an input, the shaft work contributed by a pump, h_p, must be added (if there is shaft work subtracted by a turbine, this would also be included). As energy losses, the friction between the water and the walls of the pipe is included (Darcy Weisbach equation). In this way we arrive at the generalised Bernoulli equation [9], equal to:

$$\frac{p_1}{\gamma}+\frac{1}{2g}V_1^2+z_1+h_p=\frac{p_2}{\gamma}+\frac{1}{2g}V_2^2+z_2+f\frac{L}{g}\frac{V^2}{2g} \quad (2)$$

where f is the friction factor (dimensionless) and D is the pipe diameter.

2.1 Recommendations of Bernoulli's equation

The Bernoulli equation is suitable for diagnosing the energy efficiency of simple systems because, when applied between two points on the same power line, it is perfectly suited to the condition of simple systems. In complex systems, with well identified end nodes, it provides important but incomplete information.

Bernoulli's equation is well suited to diagnose the energy efficiency of simple systems. From the results of Bernoulli's equation and knowing that the units of the energy intensity indicator (energy per unit volume) coincide with those of the pressure. A quick energy diagnosis can be established, where the energy intensity can also be expressed in metres of water column (p = γ H). Their ratio 1 m = 0.002725 kWh/m^3. Under these conditions, the energy intensity required by the system, I_{ee}, can be estimated:

$$I_{ee} = 0.002725\, H_e \text{ kWh/m}^3 \quad (3)$$

He is the estimated head needed to get from the supply source to the critical node at the required pressure. This He would be obtained by applying the Bernoulli equation:

$$H_e = \frac{1}{\eta_{pe}\eta_{le}}[(z_c-z_s)+(z_s-z_l)\eta_{pe}+h_{fe}+\frac{p_o}{\gamma}]\, m \quad (4)$$

Where η_{le} is the estimated water yield, η_{pe} the pumping efficiency. The remaining terms in equation 1 are, 0.002725, a unit change factor (metres to kWh/m^3), z_c, the most energy demanding, or critical, node elevation. On the other hand, z_s is that of the supply source, and z_l is the minimum level. The remaining variables are h_{fe} (head loss from the source to the critical node), po the service pressure and γ the specific weight of the water (9810 N/m^3).

Once the estimated energy intensity for the installation is known, it can be compared with the actual energy intensity value (total energy consumed divided by volume of water billed), in which case pump inefficiencies and system leakage must be included.

In previous work [5] [10], energy efficiency has been analysed in both simple and complex systems using the Bernoulli equation. For this purpose, I_{er}, the real energy intensity (energy consumed divided by the final measured volume), is compared with I_{ee}, the intensity estimated from the application of the Bernoulli equation, a comparison that, in order to be carried out correctly, requires disaggregating the energy supplied into natural and shaft energy.

2.2 Limitations of diagnosing systems with the Bernoulli equation

As it is a static equation, it provides information for a specific instant. Therefore, its accuracy depends on the representativeness of the selected instant (there are no problems if the flow is stationary) and the critical node adopted. The same happens if the head losses between the source and the critical node are estimated with a peak flow rate that does not correspond to the average state of the system.

Bernoulli requires a careful selection of the end nodes, usually the supply source and the most energy demanding point (or critical point). This result extrapolates to the rest of the nodes [10]. However, by ignoring the topography (elevation) of the rest of the nodes, it does not quantify the individual pressure exceedances, and the information obtained is incomplete. To obtain it, the Bernoulli equation has to be applied repeatedly between the source and each node, losing the appeal of simplicity. This is information that the integral energy equation, which does require the coordinates of the nodes, provides [6].

Another case that conditions the application of the Bernoulli equation is networks with several supply sources. In these types of complex systems, as there are two or more energy sources, the selection of the points between which to apply the Bernoulli equation is no longer immediate. One of them is the critical one (the one that requires the most energy) and the other the energy source. However, if there are different energy sources, one must be chosen and, at the same time, the energy contribution of the whole must be weighted [10].

Finally, in networks with backflows, the energy balance of the nodes of the system is broken, a balance that is maintained within the subsystems corresponding to each backflow. This makes it necessary to apply Bernoulli's equation to each subsystem, where the nodes do maintain an energy balance.

The integral energy equation, the energy balance, extended in space and time, is not conditioned by these factors that make energy analysis using Bernoulli difficult.

3 ENERGY INTEGRAL EQUATION

The energy integral equation is more general than Bernoulli's equation, but also more demanding, as it requires more data. In its initial approach, the energy integral equation is the result of applying Reynolds' Drag Theorem to the first law of thermodynamics [9]. The total energy per unit mass is the sum of internal energy (u), kinetic energy ($v^2/2$) and gravitational energy (gz). Thus, in its most general form, the equation results:

$$\frac{dQ}{dt}+\frac{dW_{shaft}}{dt} = \frac{\partial}{\partial t}\int\int\int_{VC}\left(\frac{v^2}{2}+u+gz\right)\rho dV + \int\int_{SC}\left(\frac{v^2}{2}+u+gz+\frac{P}{\rho}\right)\rho(\vec{v}\cdot d\vec{A}) \tag{3}$$

It validates after disaggregating the total work into flow work (between the system and the external medium through the control surface, SC) and shaft work, $(dW_{shaft})/dt$. The remaining variables are, the heat exchange between the external medium and the control system, SC, v, the velocity, u, the internal energy per unit mass, z, the geometric coordinate, the fluid density, dV, the volume differential of the control volume (VC), p, pressure and $\vec{v}\cdot d\vec{A}$ the flow differential across the SC. This general formulation, with few limitations, is widely used in Thermal Fluid Mechanics and very rarely in Mechanical (or Hydraulic) Fluid Mechanics, which usually resorts to the much simpler Bernoulli equation.

The three terms of this equation have a clear physical meaning. The first is the shaft work, which can be contributed (with pumps), subtracted (with turbines) or simply be zero. It is the hydraulic useful work (given to the turbine or delivered by the pump), as the energy changes of the fluid (the system) are evaluated. The inefficiencies (turbomachines and engine/generator) are added afterwards. The first integral of the second member represents the energy change inside the VC, which will be non-zero if there are reservoirs inside the VC (they gain or lose energy depending on whether they are filled or emptied). Without them, the fluid being incompressible, it is zero. The second integral, the flow term, is the balance of powers between the outgoing flow (uses and leaks) and the incoming flow in the VC. Both powers include three summands, the kinetic (negligible in network analysis), the piezometric ($gz + p/\rho$) and the internal energy, whose variation is the power dissipated by friction.

As a consequence of its initial approaches, Bernoulli's equation formulates an energy balance between two points on the same streamline, while the energy equation formulates it for a control volume [9]. With hardly any limitations, the energy integral equation is valid for any regime, static or dynamic, compressible or incompressible, with or without heat transfer. Thus, the main difference with Bernoulli is the application framework (a current tube versus a VC) and the possibility of studying transient flows and including internal compensation reservoirs in the study [3].

4 CASE STUDY

The above is applied to a case study that has a supply source with a height of 84 m, from which the pumping station in charge of supplying the sector draws water. The physical data for the sector are as follows:

- Total length of the sector = 45.1 km

- Height of the critical junction (coincides with the highest point), z_c = 120,66 m

- pump suction elevation, z_S = 84.00 m

- Height of the lowest node, z_l = 35,64 m.

- Distance between the source and the most demanding junction = 4,5 km.

- Minimum operating pressure p_o/γ = 20 m.

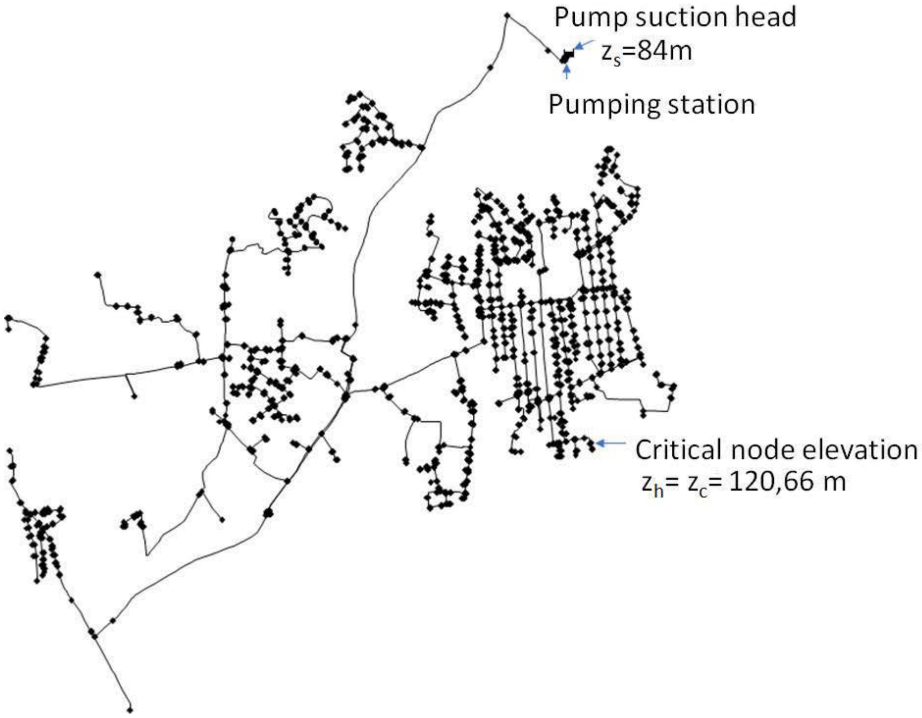


Figure 1. Case study network

Sector operating data:

- Volume injected = 15386 m^3/month
- Volume registered = 8994 m^3/month
- Real water efficiency (including real and apparent losses) η_{lr} = 0,58
- Energy consumed by the pumping unit E_p = 3902 kWh/month
- Average efficiency of the pumping unit, η_{pr} = 0,70
- As far as the head loss is concerned, a value of 1.4 m/km is assumed (Cabrera et al., 2018), resulting in an estimated friction height, h_{fe}, equal to 6.3 m.

With this information we obtain:

$$I_{ee} = \frac{0.002725}{\eta_{pe}\eta_{le}}[(z_c - z_s) + (z_s - z_l)\eta_{pe} + h_{fe} + \frac{p_o}{\gamma}]$$
$$= \frac{0.002725}{0,7 \cdot 0,58}[(120,66 - 84) + (84 - 35,64)0.7 + 6,3 + 20]$$
$$= 0,64\ kWh/m^3$$

After analysing the energy efficiency of the studied sector, it is desired to improve it. To this end, the energy intensity that could be achieved by improving the volumetric efficiency of the network and the efficiency of the pumping stations will be estimated. To do this, target values are set that are achievable in an urban supply network of these characteristics (η_{pe} = 0.75 and η_{le} = 0.80) and the volume demanded by the subscribers, V_r, is maintained. It can be seen how a large margin of improvement in water efficiency is imposed, from 0.58 to 0.80. This results in the following:

$$I_{ee} = \frac{0.002725}{\eta_{pe}\eta_{le}}[(z_c - z_s) + (z_s - z_l)\eta_{pe} + h_{fe} + \frac{p_o}{\gamma}] = 0{,}45\ kWh/m^3$$

There is a large margin for improvement between the stipulated target efficiencies and the actual efficiency of the system.

The above analysis does not include the spatial distribution of network demands. Consequently, nothing is known about the weight of structural losses, as well as the specific breakdown of other operational losses. For this it would be necessary to carry out an energy audit.

The following table shows the breakdown of the energy audit carried out on the basis of the data from the mathematical model:

Table 1. Energy audit.

Total energy input	**3059.13**	**kWh/5days**
Energy supplied by the pumps	1522.573	kWh/5days
Natural energy	1536.557	kWh/5days
Total energy consumed	**3059.064**	**kWh/5days**
Energy delivered to users	2170.936	kWh/5days
Minimum energy required	1528.718	kWh/5days
Energy excess of minimum pressure	643.045	kWh/5days
Energy dissipated by friction	41.672	kWh/5days
Energy dissipated in valves	70.647	kWh/5days
Energy lost through leakage	395.166	kWh/5days
Energy lost in pumps	380.643	kWh/5days

As can be seen, the information extracted from the energy audit made from the integral energy equation is much more extensive than that obtained with the diagnosis made using the Bernoulli equation.

5 CONCLUSIONS

In order to analyse the efficiency of a pressurised water conveyance, it is necessary to define the physical framework to which the analysis is to be made and to apply the energy equation of choice correctly. Otherwise, the results obtained will be inconsistent.

The Bernoulli equation is simple to use and works well for simple systems, in complex systems its use is not recommended due to its limitations. Bernoulli is useful for a first diagnosis of the energy state of the network. In complex systems it is better to use the Bernoulli integral equation, whose main limitation is the need to have a series of concrete data extracted from the mathematical model. However, it allows a detailed energy audit to be carried out.

6 REFERENCES

[1] Grundfos, 2014. “High efficiency motor technology that reduces energy waste in pump applications”. Grundfos. Denmark.

[2] WW (Water in the West), 2013 Water and Energy Nexus: A Literature Review. Stanford University. USA.

[3] Cabrera E, Pardo MA, Cobacho R, Cabrera E. Jr. 2010. Energy audit of water networks. J Water Resour Plan. Manag Vol. 136, pp 669 - 677.

[4] Book review: Groundwater around the World: A Geographic Synopsis, by Jean Margat and Jac van der Gun. 2013, CRC Press.

[5] Cabrera E., Gómez E., Soriano J., del Teso R., 2019. Towards eco-layouts in water distribution systems. J. Water Resour Plan Manag 145(1):04018088.

[6] Cabrera E, Gómez E, Cabrera E, Soriano J, Espert V., 2015. Energy assessment of pressurized water systems. J Water Resour Plan Manag 141(8):04014095.

[7] Cabrera E., Gómez E., Cabrera E., Jr., Soriano J., 2018. Calculating the economic level of friction in pressurized water systems. Water, 9, 763. DOI: 103390/w10060763.

[8] Del Teso R., Gómez E., Estruch M.E., Cabrera E., 2019. Topographic energy management in water distribution systems. Water Resources Management. Vol 33, nº 12, pp 4385-4400

[9] White F. M., 1979. Fluid mechanics. New York, USA: McGraw-Hill, Inc. ISBN 10: 0070696675.

[10] Cabrera, E.; del Teso, R.; Gómez, E.; Cabrera, E., Jr.; Estruch-Juan, E., 2021. Deterministic Model to Estimate the Energy Requirements of Pressurized Water Transport Systems. Water 2021, 13, 345. https://doi.org/10.3390/w13030345

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14782

UNSTEADY FRICTION MODELING TECHNIQUE FOR LAGRANGIAN APPROACHES IN TRANSIENT SIMULATIONS

M. Zeidan[1] and A. Ostfeld[2]

[1,2] Faculty of Civil and Environmental Engineering, Technion- Israel Institute of Technology, Haifa 32000, Israel.

[1] smjz@campus.technion.ac.il, [2]ostfeld@technion.ac.il

Abstract

This study tackles the problem of simulating the head damping effect in transient flows when modeled in the Lagrangian approach rather than Eulerian. The Lagrangian approach normally requires orders of magnitude fewer calculations, which allows very large systems to be solved in an expeditious manner, and it has the additional advantage of using a simple physical model as the basis for its development. Moreover, since it is continuous in both time and space, the method is less sensitive to the structure of the network and the length of the simulation process, resulting in improved computational efficiency. Nevertheless, most recent studies used an Eulerian approach when simulating the systems transient response (e.g., method of characteristics), thus focused on developing and improving different computational routines for modeling and simulating unsteady friction models that are better fixated for Eulerian methods and are not suited for Lagrangian ones. One-dimensional methods of representing unsteady contributions to skin friction based on instantaneous acceleration have a long track record (e.g., Daily et al. 1956). And it is still the most popular method in software used for practical simulations, despite it cannot accurately depict the system's transient responses without additional calibrations and tunings. However, the more accurate models (e.g., Brunone 2000) are not suited for the Lagrangian approach. The lack of a mesh structure in the Lagrangian approach makes it challenging to consider the convective acceleration terms in addition to the local acceleration. Therefore, there is a need for a more accurate friction model that is suited for Lagrangian methods without compromising their performance. Unfortunately, such a model is yet to be published in the literature. This study presents a new friction modeling technique that compensates for both the local and convective acceleration terms for the Lagrangian transient modeling approach, without compromising the computational time. Additionally, fixating only on the Eulerian approach for transient modeling and undermining the Lagrangian based models is concerning since it can provide different perspectives for developing novel solutions and tools that take advantage of transient events.

Keywords
Transient, WCM, Water hammer, Unsteady friction.

1 INTRODUCTION

Water hammer models are becoming more present in the designing and analysing complex pipeline system. In addition, they are more frequently used for the identification of system leakage, closed or partially closed valves, and the assessment of water quality problems. Turbulence models have been developed and used to perform numerical experiments in turbulent water hammer flows for a multitude of research purposes, such as the computation of instantaneous velocity profiles and shear stress fields, the calibration and verification of water hammer models, the evaluation of the parameters of unsteady friction models, and the comparison of various unsteady friction models.

Understanding the governing equations that are in use in water hammer research and practice and their limitations is essential for interpreting the results of the numerical models that are based on these equations, for judging the reliability of the data obtained from these models, and for minimizing misuse of water hammer models. While the most common approaches used to describe transient events in literature are inherently Eulerian, they require a dense mesh to mimic real life transient events and guarantee accurate results. subsequently, significantly increasing the computational capacity, time and resources required. Consequently, deeming these models impractical to use in advanced optimizing algorithms. For instance, stochastic algorithms (e.g., genetic algorithms) are inconvenient to work with when dealing with time extensive simulation, since their efficiency relies on the speed of the simulation parallelly performed.

This work presents a framework for simulating the head damping effect in transient flows when modelled in the Lagrangian approach. The Lagrangian approach normally requires orders of magnitude fewer calculations, which allows very large systems to be solved in an expeditious manner, and it has the additional advantage of using a simple physical model as the basis for its development. Moreover, since it is continuous in both time and space, the method is less sensitive to the structure of the network and the length of the simulation process, resulting in improved computational efficiency. The suggested model relies on the wave characteristics method (WCM) and describes an unsteady wave celerity in addition to unsteady friction factor. The celerity is modelled in such that it decreases while the waves propagate through the pipelines due to energy dissipation.

While advanced sophisticated methods tris to mathematically describe and model the shear stresses and velocity profiles to better mimic transient events in a lab, it can lead to misuse and inaccurate estimation in real life settings. Therefore, the suggested method attempts to “catch” the transient behaviour by introducing a calibrated factor that can accommodate for the unknown uncounted for parameters that exist in real life networks.

2 MODELS

The equations governing unsteady turbulent flows in pipes form a system of hyperbolic-parabolic partial differential equations which cannot, in general, be solved analytically. As a result, numerical solutions are used to approximate them. Extensive research has been devoted to developing theories and models to better describe and manage hydraulic transients in pipeline systems. Previous studies have employed one-dimensional models to analyze the efficiency of systems' transient responses (Wylie et al. 1993; Duan et al. 2010b; Chaudhry 2014). More recent studies have suggested and explored more complex two-dimensional and quasi-two-dimensional models that can contain various factors (e.g., Brunone 1999; Lee et al. 2013; Meniconi et al. 2013; Gong et al. 2014, 2016, 2018; Duan and Lee 2016; Kim 2016, 2020; Wang and Ghidaoui 2018; Che et al. 2018, 2019; Wang et al. 2019; Keramat et al. 2019; Zhao and Ghidaoui 2003; Zouari et al. 2020).

3 MODEL FORMULATION

The model formulation described below is composed of a description of the WDS mapping and modeling process, the transient model simulations, followed by a comparison with the TSNET package (Xing and Sela, 2020).

3.1 Mapping the water distribution system

The network is imported from a database file (INP) and prepared for the transient simulation. The water distribution system is mapped onto an undirected graph G= (V, E), in which the vertices V represent the consumers, sources, and valves, and the edges E represent the connecting pipes. The different types of vertices are defined as different discontinuities with different resistance coefficients. All the transient calculations and simulations were carried out via MATLAB codes and the TSNET Package in Python.

3.2 The wave characteristic method model

The wave characteristic method (WCM) introduced by Wood (1965) is based on the notion that transient pipe flow is caused by the propagating of pressure waves which occurs when a disturbance is introduced into the system. These pressure waves are described as a rapid pressure change that travel at pressure wave speed in the liquid-pipe medium. In pressurized water pipes, the pressure waves travel about 1000 m/s in metallic pipes and around 400 m/s in polymeric pipes (Duan et al. 2020, Wan and Mao 2016). When these waves encounter discontinuities, they are partially reflected and transmitted through the pipe system.

3.3 Pressure magnitude

The magnitude of a pressure wave is calculated using the Joukowsky equation (Eq.1.) while assuming an immediate change in the valve opening. The Joukowsky equation does not only describe the correlation between pressure change ΔP and flow change ΔQ, but also forms the basis for the WCM mathematical model. In order to work with head-pressure units, the ΔP is replaced by the term ρ g ΔH as shown in Eq.1.

$$\Delta P = \rho c \frac{\Delta Q}{A} \xrightarrow{\Delta P=\rho g \Delta H} \Delta H = c\frac{\Delta Q}{gA} \tag{1}$$

Where ΔP represents the pressure, ΔH is the change in head pressure, ΔQ is the change in flow rate, c is the pressure wave's celerity; A represents the flow section area, ρ is the fluid viscosity, and g is the gravitational acceleration constant.

3.4 Unsteady friction attenuation

Previous works have captured the influence of line friction (viscous resistance) on the propagation of pressure waves using the "orifice analogy" (e.g., Jung et al 2009) where (n) imaginary friction orifices are added to the pipe, thus dividing it into (n+1) pipes that are similar in length. The friction orifices are modeled as a square-law orifice with a properly chosen orifice coefficient. This takes the form of a quadratic correlation between the pressure head ΔH and the flow rate Q, as follows:

$$\Delta \mathrm{H} = \mathrm{A(t)} + \mathrm{B(t)}|\mathrm{Q}| + \mathrm{C(t)}\mathrm{Q}|\mathrm{Q}| \tag{2}$$

The terms A, B, and C represent the coefficients for a general representation of the characteristic equation. The coefficient may be time-dependent but known at every time step. The absolute values of Q are employed to make the resistance term dependent on the flow direction. The new modification suggests a simple linear damping factor to the wave's celerity (Cel) as follow:

$$Cel = 1000 - \alpha \cdot t \tag{3}$$

where α is a calibrated coefficient and t represent the lifespan of the wave.

The model presented by Brunone et al. 2000 is the most commonly used in conventional transient analyses, However, it does not fit the Lagrangian approach for transients modeling. Consequently, in addition to the Darcy-Weisbach equation, the Daily (1955) empirical correction to the wall shear stress model was adopted here to compute the head-loss and shear stresses along the pipelines. The wall shear stress is expressed as:

$$\tau_{wall} = \frac{\rho f V^2}{8} + \frac{k_u \rho D}{4}\frac{\partial V}{\partial t} \tag{4}$$

where τ_{wall} is the combined steady and unsteady wall shear stresses, ρ is the water density, f is the Darcy-Weisbach coefficient, V is the flow velocity, D is the pipe diameter, $\partial V/\partial t$ is the local instantaneous acceleration, and K_u is Brunone's friction coefficient.

$$K_u = 0.16 + \frac{t}{\beta} \tag{5}$$

where β is a calibrated coefficient and t represent the lifespan of the wave.

The result of including the unsteady shear stress model into the square- law orifice analogy is described by the following equation:

$$\Delta H = \underbrace{[\frac{k}{2gA}\frac{\partial Q}{\partial t}]}_{A(t)} + \underbrace{[\frac{-fL}{2gDA^2}]}_{C(t)} Q|Q| \tag{6}$$

4 COMPUTATIONAL RESULTS

4.1 Case study 01:

The first network resembles a simple pipeline system with a valve at the downstream. In this case study the new suggested parameters (α, β) are calibrated and to be further used in the second case study. The desire is to get constant parameters that will fit in other layouts and would be considered constants coefficients in the future.

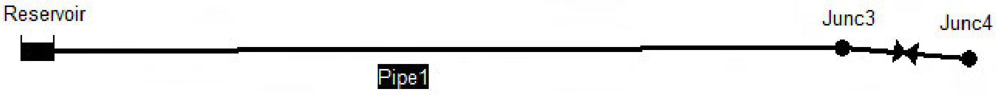


Figure 1. The layout of the first case study, describing a simple pipeline system with a valve downstream

In this section, the performance of the refined WCM model and the other models available at the TSNET package are compared. The main purpose is to find the main differences and refine the WCM accordingly. As seen in figure 2, the WCM, the steady friction model and the quasi-steady model almost align with each other. While the unsteady friction model differs dramatically. The main reason is the Celerity attenuation in the unsteady friction model. hence the main focus in this work will be introducing celerity attenuation into the WCM.

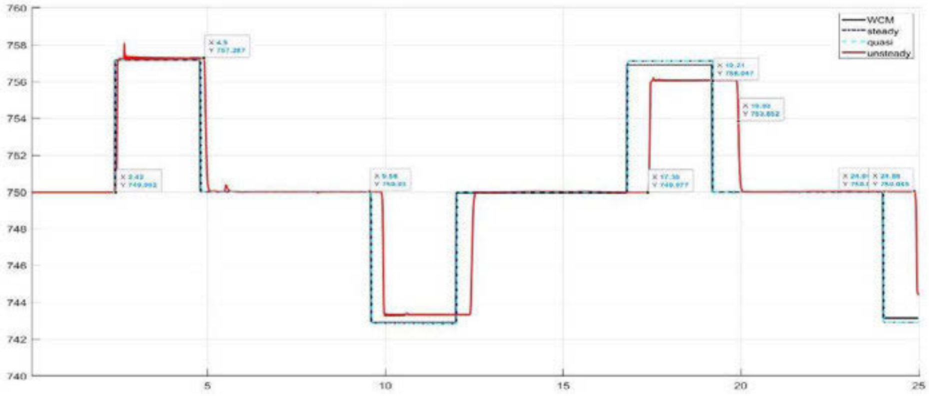


Figure 2. Comparison between the WCM model and the three available models in the TSNET package.

The next step is to introduce the celerity and friction modification to the WCM model and observe the influence of each parameter over the transient response, at node 3 in this case. The transient response presented at Figure 2.a will serve as the baseline for these modifications, since it holds no celerity or friction attenuations. The WCM pulses arrive much faster than their counterparts in the unsteady model, mostly due to different wave celebrities. Therefore, by adjusting the wave celerity to be 960 m/s and the Ku to 0.16, the phase difference is eliminated (Figure 2.b), however, the amplitudes are still different. As shown in Figure 2.c, by introducing the damping effect for the friction factor, the transient response received from both models are similar to a great extent. Lastly, Figure 2.d shows the potential of adjusting both the celerity and friction factor so that the transient responses match to a great extent. The Parameters α and β were calibrated to be 0.5 and 0.15, respectively.

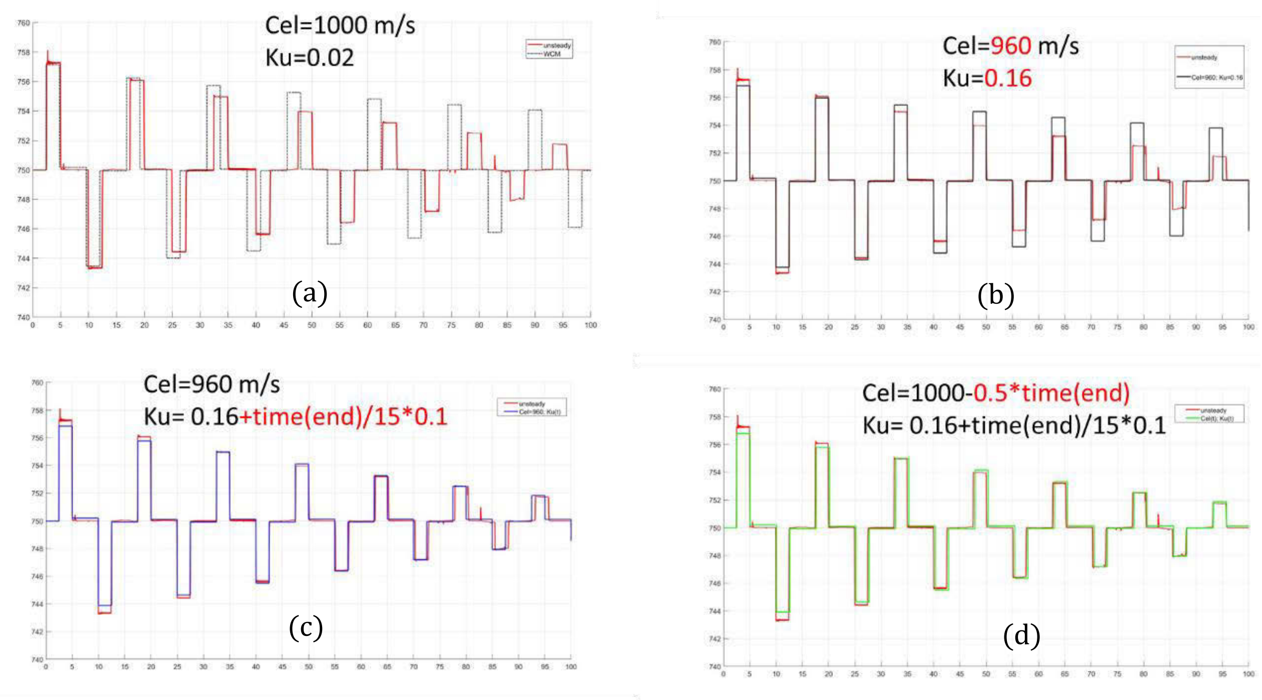
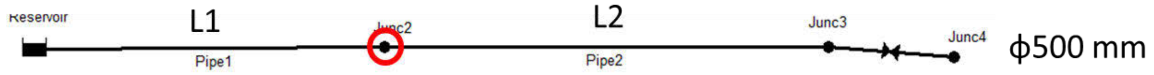


Figure 3: different adjustment to the WCM in comparison to the unsteady model from TSNET.

BY NC SA



Case study 02: different adjustment to the WCM in comparison to the unsteady model from TSNET.

Another pipe was added to the first case study layout to better test the calibrated parameters [α=0.5, β=0.15]. the transient respond at Node 2 is observed and compared to the unsteady friction model from the TSNET package. The pipe lengths L1 and L2 are changed to avoid over fitting scenarios. In this first case L1 is 1.5 Km while L2 is 2.0 Km. the transient response of the refined WCM, and the unsteady friction model are presented in Figure 4.

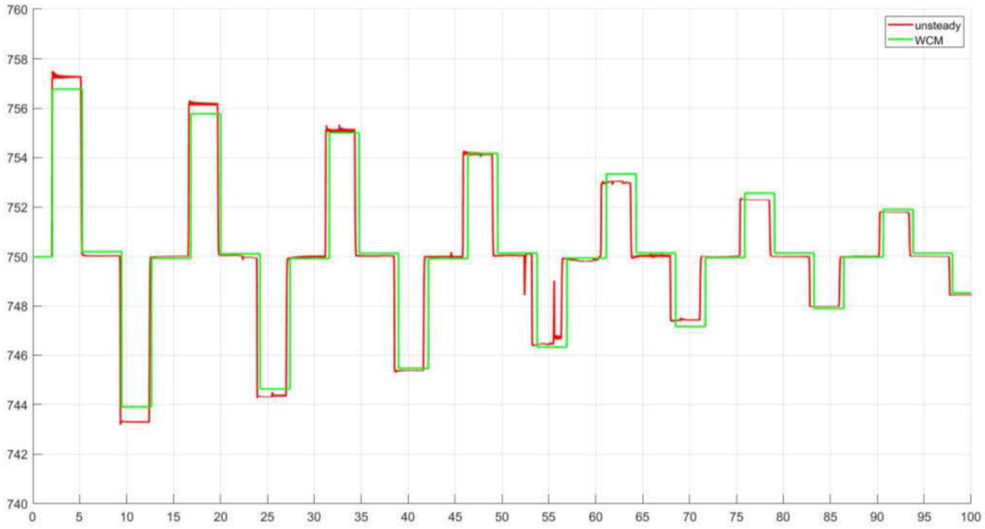


Figure 4: the transient response for the refined WCM and the unsteady TSNET model for second case study.

As seen in Figure 4, the Transient response for the two models are quiet similar, with few anomalies. However, the improvement from the regular WCM is clear in this case. Moreover, these differences are negligible from an engineering standpoint when dealing with unpredicted noise and disturbances. To further check the parameters, the lengths of L1 and L2 are now 3.0 Km each. The results of the transient simulation for both models are presented at Figure 5.

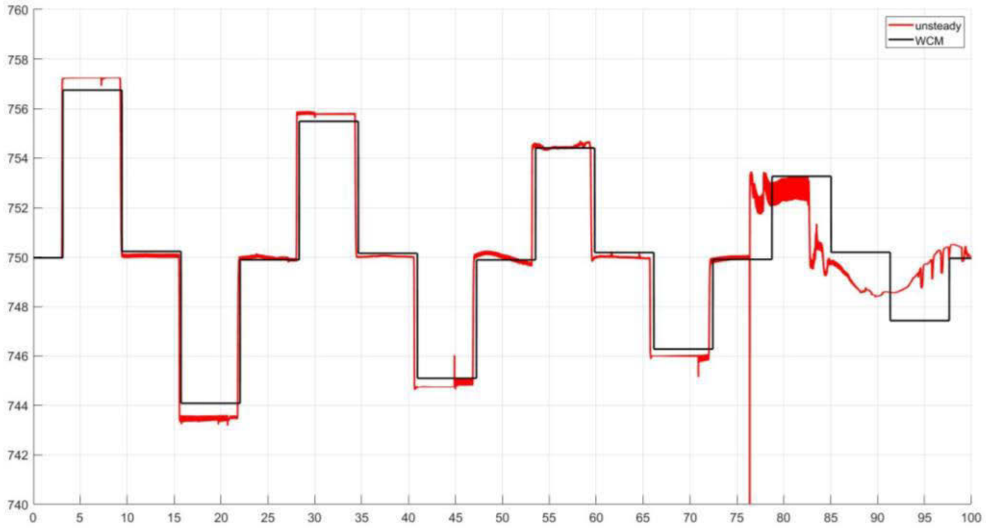


Figure 4: the transient response for the refined WCM and the unsteady TSNET model for the modified second case study.

cc BY NC SA

As seen in Figure 5, the refined WCM can depict a similar transient response to that of the unsteady model, it is important to note that the WCM take less computational resources and therefore less computation time when compared with its counterpart. In addition, the unsteady model faces some instability around 75 seconds from the beginning of the simulation, probably due to mesh sensitivity or other sensitive boundary conditions, when the WCM do not have this problem due to its inherently Lagrangian approach.

5 CONCLUSIONS

Previous works have implemented numerous approaches for accurately simulating transient response and capturing the damping effect that occurs in real life scenarios. Approaches mostly included Eulerian based method, more specifically different variations and modifications of the method of characteristics. However, these approaches require dense mesh to get accurate result, hence increasing the computational capacity needed. This study builds on the wave characteristics method, which is much less computationally intensive and requires no mesh for performing transient analysis. However, it generates less accurate result in general, unless numerous friction orifices are introduced. I this study, we introduced modification for the wave celerity and the friction factor. New suggested parameters (α, β) are introduced to the model and are calibrated in the first case study, and further used in the second case study, generating promising results.

In line with the hypothesis, the introduced Celerity and friction linear attenuation resulted in a better transient response that is, to an extent, similar to the response received by the more complex unsteady model. It is worthy noting that the running time of the refined WCM is significantly lower than the unsteady MOC based model.

6 ACKNOWLEDGMENTS

This research was supported by the Israel Science Foundation (grant No. 555/18).

7 REFERENCES

[1] Daily, J.W., Hankey Jr, W.L., Olive, R.W. and Jordaan Jr, J.M., 1955. Resistance coefficients for accelerated and decelerated flows through smooth tubes and orifices. MASSACHUSETTS INST OF TECH CAMBRIDGE.

[2] Brunone, B., 1999. Transient test-based technique for leak detection in outfall pipes. Journal of water resources planning and management, 125(5), pp.302-306.

[3] Wylie, E.B., Streeter, V.L. and Suo, L., 1993. Fluid transients in systems (Vol. 1, p. 464). Englewood Cliffs, NJ: Prentice Hall.

[4] Lee, P.J., Duan, H.F., Ghidaoui, M. and Karney, B., 2013. Frequency domain analysis of pipe fluid transient behaviour. Journal of hydraulic research, 51(6), pp.609-622.

[5] Meniconi, S., Duan, H.F., Lee, P.J., Brunone, B., Ghidaoui, M.S. and Ferrante, M., 2013. Experimental investigation of coupled frequency and time-domain transient test–based techniques for partial blockage detection in pipelines. Journal of hydraulic engineering, 139(10), pp.1033-1040.

[6] Gong, J., Lambert, M.F., Simpson, A.R. and Zecchin, A.C., 2014. Detection of localized deterioration distributed along single pipelines by reconstructive MOC analysis. Journal of hydraulic engineering, 140(2), pp.190-198.

[7] Gong, J., Lambert, M.F., Zecchin, A.C. and Simpson, A.R., 2016. Experimental verification of pipeline frequency response extraction and leak detection using the inverse repeat signal. Journal of Hydraulic Research, 54(2), pp.210-219.

[8] Gong, J., Lambert, M.F., Nguyen, S.T., Zecchin, A.C. and Simpson, A.R., 2018. Detecting thinner-walled pipe sections using a spark transient pressure wave generator. Journal of Hydraulic Engineering, 144(2), p.06017027.

[9] Duan, H.F. and Lee, P.J., 2016. Transient-based frequency domain method for dead-end side branch detection in reservoir pipeline-valve systems. Journal of Hydraulic Engineering, 142(2), p.04015042.

[10] Kim, S., 2016. Impedance method for abnormality detection of a branched pipeline system. Water resources management, 30(3), pp.1101-1115.

[11] Kim, S.H., 2020. Multiple leak detection algorithm for pipe network. Mechanical Systems and Signal Processing, 139, p.106645.

[12] Wang, X. and Ghidaoui, M.S., 2018. Pipeline leak detection using the matched-field processing method. Journal of Hydraulic Engineering, 144(6), p.04018030.

[13] Che, T.C., Duan, H.F., Lee, P.J., Pan, B. and Ghidaoui, M.S., 2018. Transient frequency responses for pressurized water pipelines containing blockages with linearly varying diameters. Journal of Hydraulic Engineering, 144(8), p.04018054.

[14] Che, T.C., Duan, H.F., Pan, B., Lee, P.J. and Ghidaoui, M.S., 2019. Energy analysis of the resonant frequency shift pattern induced by nonuniform blockages in pressurized water pipes. Journal of Hydraulic Engineering, 145(7), p.04019027.

[15] Wang, X., Palomar, D.P., Zhao, L., Ghidaoui, M.S. and Murch, R.D., 2019. Spectral-based methods for pipeline leakage localization. Journal of Hydraulic Engineering, 145(3), p.04018089.

[16] Keramat, A., Ghidaoui, M.S., Wang, X. and Louati, M., 2019. Cramer-Rao lower bound for performance analysis of leak detection. Journal of Hydraulic Engineering, 145(6), p.04019018.

[17] Zhao, M. and Ghidaoui, M.S., 2003. Efficient quasi-two-dimensional model for water hammer problems. Journal of Hydraulic Engineering, 129(12), pp.1007-1013.

[18] Zouari, F., Louati, M., Meniconi, S., Blåsten, E., Ghidaoui, M.S. and Brunone, B., 2020. Experimental verification of the accuracy and robustness of Area Reconstruction Method for Pressurized Water Pipe System. Journal of Hydraulic Engineering, 146(3), p.04020004.

[19] Xing, Lu, and Lina Sela. “Transient simulations in water distribution networks: TSNet python package.” Advances in Engineering Software 149 (2020): 102884.

[20] Wood, D.J., Dorsch, R.G. and Lightner, C., 1965. Digital Distributed Parameter Model for Analysis of Unsteady Flow in Liquid-Filled Lines. National Aeronautics and Space

[21] Duan, H.F., 2020. Development of a TFR-based method for the simultaneous detection of leakage and partial blockage in water supply pipelines. Journal of Hydraulic Engineering, 146(7), p.04020051.

[22] Wan, W. and Mao, X., 2016. Shock wave speed and transient response of PE pipe with steel-mesh reinforcement. Shock and Vibration, 2016.

[23] Jung, B.S., Boulos, P.F., Wood, D.J. and Bros, C.M., 2009. A Lagrangian wave characteristic method for simulating transient water column separation. Journal-American Water Works Association, 101(6), pp.64-73.

[24] Brunone, B., Karney, B.W., Mecarelli, M. and Ferrante, M., 2000. Velocity profiles and unsteady pipe friction in transient flow. Journal of water resources planning and management, 126(4), pp.236-244.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14784

THE SHAPE OF WATER DISTRIBUTION SYSTEMS - DESCRIBING LOCAL STRUCTURES OF WATER NETWORKS VIA GRAPHLET ANALYSIS

Bulat Kerimov[1], Franz Tscheikner-Gratl[2], Riccardo Taormina[3] and David B. Steffelbauer[4]

[1, 2]Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, (Norway)

[3] Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, (The Netherlands)

[4] KWB - Kompetenzzentrum Wasser Berlin, Berlin (Germany)

[1] *bulat.kerimov@ntnu.no*, [2] *franz.tscheikner-gratl@ntnu.no*, [3] *r.taormina@tudelft.nl*,
[4] *david.steffelbauer@kompetenz-wasser.de*

Abstract

The performance, vulnerability, and resilience of water distribution systems (WDS) depend, to varying degrees, on its underlying topological structure, herein referred to as its shape. Literature mostly differentiates between two main shapes of networks - branched or looped. However, the shape of real networks lies in between the two extremes of purely branched and looped systems. Although these networks are globally topologically different, they may show high similarity at the local scale of a borough or a neighbourhood. Recent studies focused on describing WDS via links and nodes by using graph theory. These first attempts at graph-theoretical applications showed promising results in describing the global structure and estimating the global resilience of WDS, but there are a limited number of measures that take the importance of local topology into consideration.

This research enters the new terrain of local WDS investigations using graphlet analysis to describe this local topology in more detail. Graphlets are small connected subgraphs of a large network which have recently gathered much attention as a useful concept to characterise local topology and uncover structural design principles of complex networks. Consequently, these novel analysis techniques can provide new insights into how local WDS structures influence their overall behaviour.

In this work, we first provide a framework to describe local and global topology with graphlets. We then employ the framework to assess the local criticality of two benchmark WDS, linking the results to topological metrics already adopted in the literature. Additionally, we analyse the potential gain of graphlet analysis in the prediction of local vulnerabilities by including graphlet features into a random forest regression setting. As a result, we observe a positive trend in performance in comparison to a similar model without graphlet features.

Keywords
Graphlet, graph theory, resilience, water distribution, topological structure.

1. INTRODUCTION

WDSs naturally relate to graphs, associating the pipes and junctions in the water system with links and nodes on a mathematical object. That is why graph theory can be utilised to describe and compare WDSs with different topologies in a unified mathematical language and serve as sound support in decision making in planning and optimization of networks. Water system research successfully adopted graph-theoretical measures in support of various tasks. The current set of instruments includes a large variety of graph-theoretical measures that are widely used in resilience and redundancy analysis [1], [2], [3], identifying critical components [4], [5], [6], and sectorisation [7], [8]. The resilience of a network, defined as the ability of the system to recover after adversarial effects [3], partially depends on the topology of the WDS. Earlier works showed that more resilient networks can be designed by graph theoretical approaches, for example, by taking spectral and statistical graph measures into consideration. [2] systematically reviewed and evaluated the correlation between common topological measures and network resilience. Additionally, topological and spectral properties were used in [9] and [10] to assess global and local vulnerabilities. These works highlight the importance of such metrics as algebraic and spectral connectivities. The utility of centrality measures for the assessment of local vulnerability has been indicated in [11] and [12].

Nevertheless, traditional graph theoretical measures (e.g., average node degree, graph diameter, link density) do not provide detailed information on local neighbourhood structures of networks but rather a global description. However, this local information can be of high importance in automatic generation of WDS [13], and applications that require a more precise description of those networks (e.g., node and link criticality analysis) [4], [5]. There are two sides to the criticality of elements of WDSs: fault probability and the impact severity. In this work, we introspect the latter part of the equation by looking only at the local topological structure. With the terms of local *importance,* and *criticality* we refer to the same phenomena, particularly - the total impact of the fault in the element on the whole system. Utilities and engineers estimate the impact by simulating pipe breakages, abrupt increases in demands, and other faults with hydraulic engines such as EPANET. The main issue of a hydraulic-based critical element analysis is computational costs for larger WDSs. On the flip side of the coin, pure topological measures will fail to take into account energy dissipation and the physical nature of the water distribution process. Some works attempt to bridge energy-based measures of local criticality with topological indicators. For example, water-flow centrality [14] and node demand centrality [15] assign hydraulically informed weights to centrality metrics. Furthermore, energy-informed shortest routes gently introduce energy loss properties into the graph structure [16],

With this work, we open the way for a more detailed description of WDSs by introducing graphlet representation to address the aforementioned challenges. Graphlet analysis aims to extract small induced subgraphs that appear in the network. Graphlets have been successfully applied in network alignment, description of brain networks [17] and used as an early precursors for in social networks [18]. The ultimate goal is to provide a relevant and comparable numerical representation of local connectivities of water networks, i.e. a topological fingerprint of the WDS. This research additionally covers the assessment of the descriptive power of graphlets in the estimation of local (i.e. node) criticality. Analysis of global network vulnerability and various resilience metrics is outside the scope of this project.

2. METHODOLOGY

In the first step of our methodology we introduce the reader to graphlet decomposition, the products of graphlet analysis, and the relationship between graphlets and traditional topological metrics. We show how the coarseness of the network influences the results of the analysis by comparing graphlet representation of the network versus the representation of the skeletonized version of the network. Next, we evaluate the intrinsic relationship between the graphlet features with existing topological measures. In this step, we carry out a graphlet analysis on a set of generated networks and estimate the correlations with the topological metrics on a global (network-level) and local (node-level) scale. Specifically, we carry out a Principal Component Analysis (PCA) on extracted graphlet features and assess the correlations of principal components with those features.

Second, we assess the descriptive power of graphlet analysis in the prediction of the importance of a node. We perform the analysis by incorporating the results of the graphlet into the random forest regression model and estimate the performance gain by comparing it with the performance of the same model without graphlet features.

2.1 Graphlet analysis

Graphlets are small subgraphs that collectively comprise the structure of the network. One can think of graphlets as (atomic) building blocks of the graph. A plethora of research in graphlet analysis focuses on extracting the graphlet level description of the network with a relatively low computational effort. This is usually performed by some form of a counting algorithm. Since direct enumeration of the graphlets is rather computationally expensive, researchers attempt to speed up the calculations by exploiting the symmetry properties of graphs [19] or employing sampling strategies.

In this work, we employed the ORCA algorithm due to the availability of the source code and particular suitability for sparse graphs. Due to the computational constraints, the size of the graphlets included in ORCA is limited to 5. Theoretically, the number of potential distinct graphlets is infinite and grows exponentially with its size. However, this may reduce the applicability of the graphlet representation due to a sparse representation. Current rules to find certain graphlets are mostly designed by hand, although there are advancements towards the automatic generation of those rules [20].

For a finer representation of graphlets, counting algorithms operate on the level of *orbits*. An orbit is an automorphism of a graphlet, e.g. isomorphism of the graphlet to itself. In other words, the parts of the graphlet that are symmetrical to others will belong to the same orbit. Orbit counting algorithms thus identify how many times a node touches the corresponding orbit. In total, 29 graphlets of sizes up to 5 contain 73 distinct orbits.

As a result, ORCA produces vectors on 2 scales - a local of size 73 and assigned to each node level, and a global, i.e. a vector of 29 which is assigned to a whole graph. The values on the former vector correspond to the number of times an orbit occurs on each node. We further denote this vector as an orbit count vector (OCV). OCVs are comparable between the nodes of different water networks. The global vector is calculated by summing up OCVs over nodes of a graph and results in the total graphlet count in that network. We denote this vector as a graphlet count vector (GCV).

One important aspect of the graphlet representation is the influence of the coarseness of the graph layout. Commonly, the number of nodes and connections in the topology varies depending on the ultimate task. Skeletonization is a technique to simplify networks that simultaneously alters the granularity of the graph [21]. As a consequence, the graphlet analysis is sensitive to this granularity. In our work, we assess the influence after 2 steps of skeletonization. The first step is a unification of adjacent consecutive pipes. The second is trimming dead ends and parallel connections.

After retrieving graphlet and orbit counts from the network we perform a PCA. PCA is a widely used method of dimensionality reduction that decomposes the variables into a new set of uncorrelated features through a singular value decomposition. As the result, we obtain a low dimensional representation of the original data, while preserving most of the variance. These new features are called Principal Components (PC_i). Each of them is composed as a result of a linear combination of original features and explains a certain share of variance. In the case of GCV and OCV, we use PCA to find the inner correlations between the features and to compare the key components altogether rather than each graphlet or orbit individually. To extend the analysis, for each principal component we leverage the information about contributions of original features and explained variances.

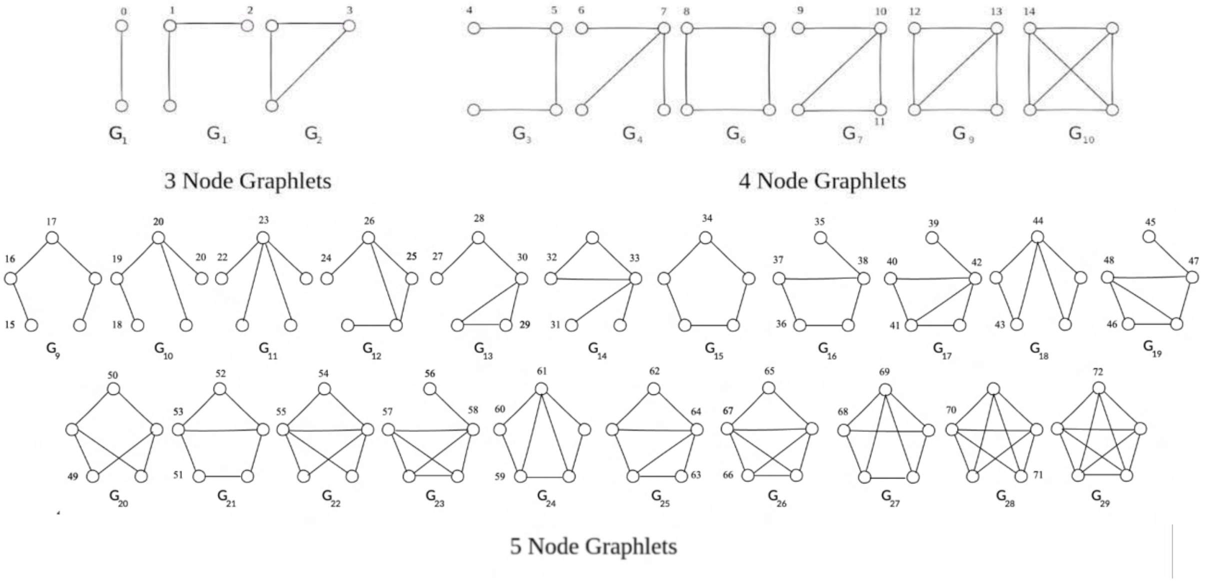


Figure 1. Existing graphlets (with sizes up to 5 nodes) and their orbits.

2.2 Graph-theory based measures

In this section we summarize topological measures that describe graph networks. As was mentioned above, graphlet analysis produces vectors on 2 scales - a local (node) level, and a global (network) level. The former representation can be viewed as a heatmap on the nodes network, while the latter assigns a vector to a whole graph. We thus compare both representations with the corresponding topological measures separately. These measures are widely used in WDS analysis, mostly in the context of vulnerability and resilience analysis.

2.2.1 Local measures

Node degree

Node degree is a basic metric that measures the number of incidental links (or pipes to the network.

$$d(v) = \sum_{\substack{j \in V \\ j \neq i}} A_{ij} \tag{1}$$

Here A_{ij} is the element of an adjacency matrix.

Betweenness centrality

Betweenness centrality is a statistical value that measures the relative position of a node on the network. A higher value of centrality is usually assigned to the “hubs” of the network. It indicates how the rest of the network “depends” on the node.

$$c_b(v) = \sum_{\substack{j \in V \\ j \neq i}} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \tag{2}$$

where σ_{ij} is the total number of shortest paths from node *i* to node *j*, $\sigma_{ij}(v)$ is a number of those paths that pass through the node *v.*

BY NC SA

Closeness centrality
Closeness centrality is another centrality metric that assigns a value to a node depending on the distance from the node to the rest of the network. In contrast to betweenness, closeness centrality indicates the independence of the node from the rest of the network.

$$c_c(v) = \sum_{j \neq i}^{j \in V} \frac{n-1}{d_{ij}} \tag{3}$$

Here d_{ij} denotes the distance from node i to node j, while n is the total number of nodes.

PageRank
PageRank is a ranking algorithm originally proposed by Google research and designed to assign relative importance to nodes on the graph based on number of incidental links and the "quality" of incidental links. This quality is likewise defined by the importance of the source nodes.

2.2.2 Global measures

For the comparison we chose the following set of descriptors, that are used in the estimation of redundancy and resilience of WDSs. [2] provides a systematic review over the vast majority of common topological metrics and their correlation with network resilience.

Distribution of node degrees
The distribution of node degrees is simply the normalized distribution of all node degrees over the entire graph *G.* Similarly to the average node degree, it indicates spatial organization of a network. In the context of WDS, where the node degree rarely exceeds 4, a right skew in the distribution corresponds to a highly connected structure. On the contrary, left-skewed distribution is related mostly to a tree-like structure.

Number of cycles
Number of cycles in the network is an indicator of redundancy. For example, clustering coefficient is defined by the total number of triangles in the loop. Intuitively, a higher number of 3-cycles suggests availability of alternative paths from supply to demand nodes.

Meshedness
Meshedness measures a fraction of the actual number of cycles of any order (or loops of any size) to the maximum possible number of cycles in the network. Meshedness is directly related to the number of alternative paths in a supply system and hence to the redundancy of the network [1].

$$c_m(G) = \frac{f}{2n-5} \tag{4}$$

Here, f denotes the number of total independent loops and n denotes the number of nodes.

Connectivity
Although connectivity metrics (e.g. spectral gap, algebraic connectivity) and modularity are important factors to the resilience of the network [2], we deliberately omit these measures from the comparison with GCV. We motivate it by the fact that connectivity and modularity measure properties of a larger scale rather than the one of graphlet analysis.

2.3 Local vulnerability analysis

A system is as resilient as its most vulnerable element, therefore there is a strong demand from utilities and planners to identify which elements have the highest chance of failing. One of the paradigms to approach this problem is based on simulating various types of faults in a simulator such as EPANET or others. A drastic change in the operations of the WDS, such as a burst or a pipe break, will impact the overall distribution of water pressure. Consequently, low pressure on the consumer end results in the system's incapability to supply the required volumes of water. We simulate a pipe break by closing the pipe during a 72 hour simulation and measure the total amount of water that was not supplied to end consumers. The final indicator of the importance of the pipe is measured by the amount of unsupplied demand conditioned by the fault in this element in a pressure-driven simulation. Here $\underline{q}_j$ and q_j denote demands during pressure driven simulations and base demand during normal conditions correspondingly. To facilitate comparability across WDSs, we evaluate this metric on a logarithmic scale.

$$c_{crit}(e) = \sum_{j \neq i}^{j \in V} \frac{(q_j - \underline{q}_j)}{\underline{q}_j} \tag{5}$$

We additionally constrain the lowest values at -10. As a result, the values of criticality are distributed in the negative range up to 0. Higher values correspond to higher criticality.

It is important to mention that criticality analysis yields pipe-level information, i.e. a value assigned to a link. In contrast, OCV is a junction-level vector. We thus transform the pipe criticality into a junction criticality by averaging the values over incidental links of the node.

$$C^{e}_{crit} = C^{v}_{crit} B D^{-1}, \tag{6}$$

where C_{crit} is vectors of criticality assigned to vertices or edges, B denotes an incidence matrix, and D is a degree matrix.

3. CASE STUDY

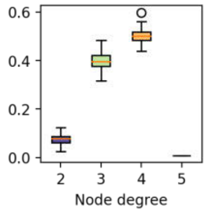
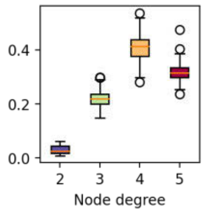
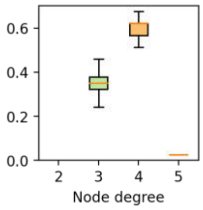
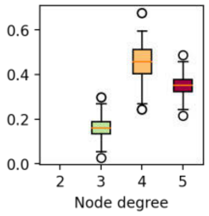
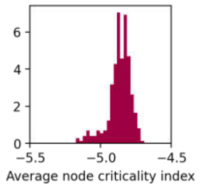
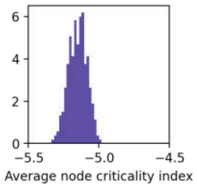
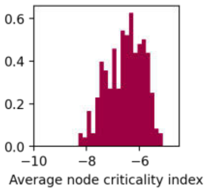
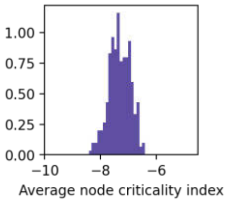
We selected ZJ and FOS networks due to their homogeneity in network geometry (e.g. elevations, diameters, and pipe length) and locations of the reservoirs, which limits the influence of network hydraulics on our graph-based analysis. In order to operate on a comparable dataset, we generated 1200 artificial variations of those networks. We introduced several surgical morphological transformations on the local connectivity and geometry during data generation.

3.1 Network generation

We first define the proximity radius *r* by calculating the median geographical distance between connected junctions in the network. Next, for every junction, we define unconnected candidates in the local neighbourhood located within the radius. We also define the upper limit to the degree of each junction as 4 and 5 and evaluate both cases separately. With a probability *p = 0.5* a pipe is generated for each possible candidate node.

Table 1. Summary of network generation.

	ZJ		FOS	
Number of networks	300	300	300	300
Maximum node connectivity	4	5	4	5
Number of nodes	114	114	38	38

Relative frequencies of node degrees				
Distribution of node criticalities (higher values relate to higher criticality)				

In Table 1, we observe a shift in the distribution of average node criticality towards less critical conditions (i.e., lower values) with the increase of the maximum connectivity (from 4 to 5) during data generation.

It is important to mention that some of the generated connections are likely to be unfeasible in real-life scenarios. Some of the connections intersect or double existing pipes. We leave the investigation of the feasibility and associated costs as a potential for future work.

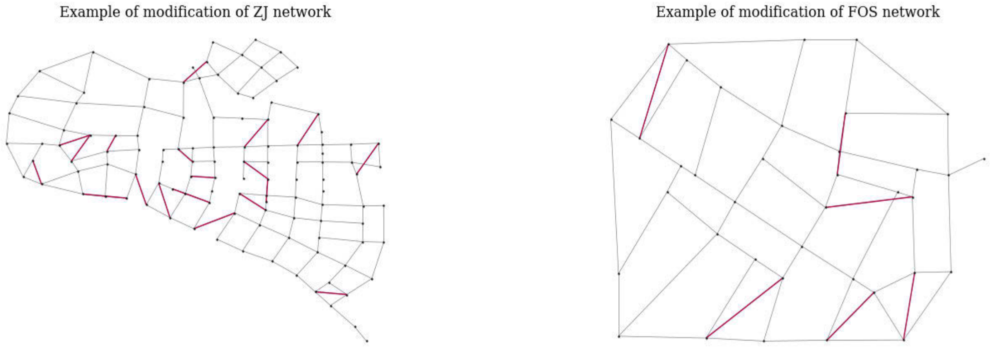

Example of modification of ZJ network
Example of modification of FOS network

Figure 2. Examples of topology modification of generated network variants (red links denote generated pipes). Both have a limit of a maximum of 4 connections.

4. RESULTS AND DISCUSSION

In this section we provide the result of the analysis of the correlation of graphlet counts with the topological measures described above. Additionally, to inspect the relationship of the graphlet count to the local criticality we show the effect of including the graphlet representation in the random forest regression model.

4.1 The influence of skeletonization on graphlet representation

On the example of L-Town network, we observe that reducing the number of adjacent essentially reduces the order of cycles on the graph and increases the average degree. Consequently, this increases the count of higher-level orbits related to the graphlets with cycles (e.g. G_2, G_5, G_{12}, etc.). We show in Figure 2 the example that skeletonization allows the graphlet counting algorithm to capture higher-order graphlets that wouldn't be captured otherwise. The second step continues the trend increasing the count of higher orbits, such as G_{37} and $G_{38.}$ and results in more dense GCV. However, we argue that for criticality analysis, this level of coarseness might be detrimental.

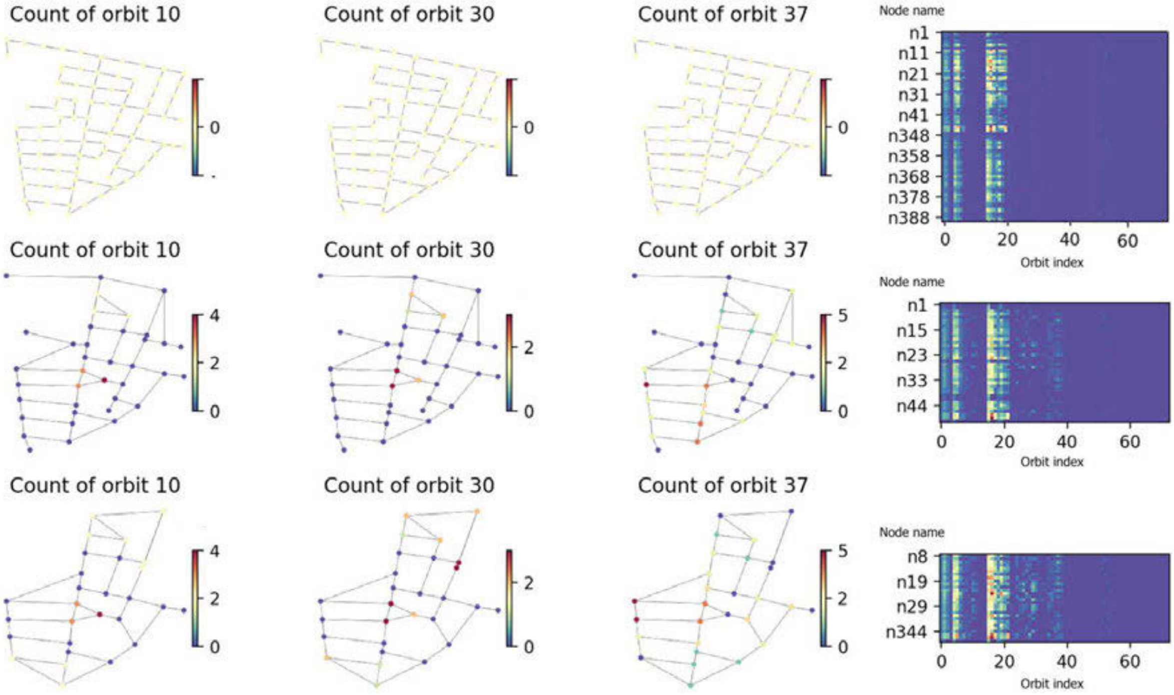
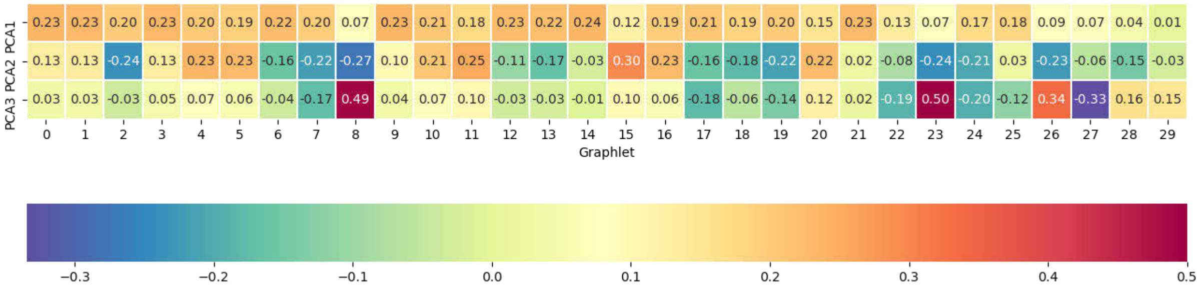


Figure 3. Example of influence of skeletonization on the graphlet count on the example of area C in L-Town water network (top - no skeletonization, center - unification of adjacent pipes, bottom - removal of parallel connections and dead-ends)

4.2 Principal component analysis

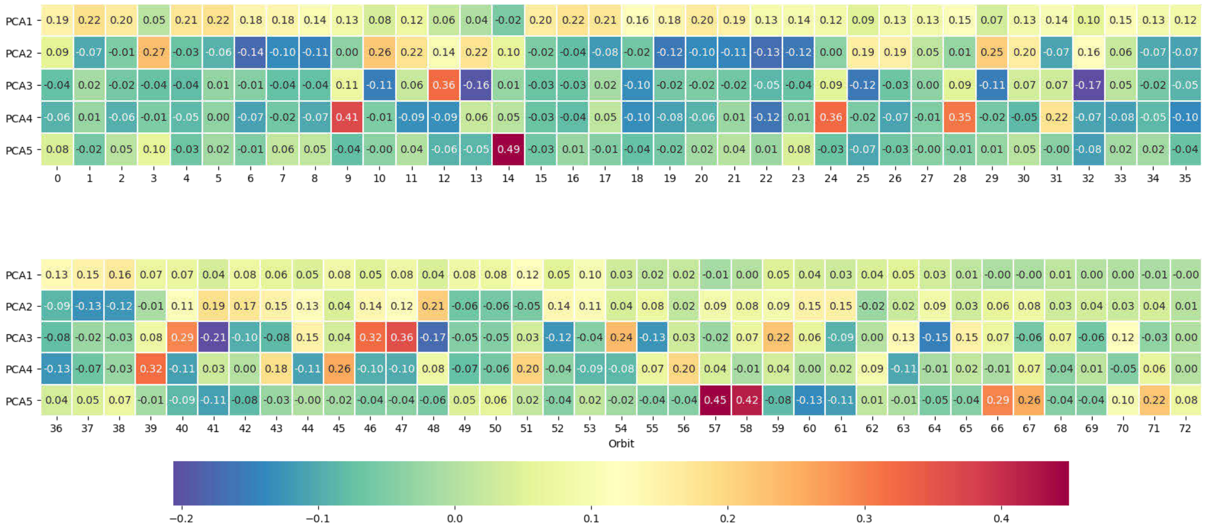
High granularity of the orbit representation comes at the cost of interpretability. Specifically, some of the graphlets can be present as components of the others, which can lead to certain autocorrelation in the orbit representation vector. We hypothesize that it is possible to translate OCV into a lower dimensionality representation via principal component decomposition.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PCA1	0.23	0.23	0.20	0.23	0.20	0.19	0.22	0.20	0.07	0.23	0.21	0.18	0.23	0.22	0.24
PCA2	0.13	0.13	-0.24	0.13	0.23	0.23	-0.16	-0.22	-0.27	0.10	0.21	0.25	-0.11	-0.17	-0.03
PCA3	0.03	0.03	-0.03	0.05	0.07	0.06	-0.04	-0.17	0.49	0.04	0.07	0.10	-0.03	-0.03	-0.01

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
PCA1	0.12	0.19	0.21	0.19	0.20	0.15	0.23	0.13	0.07	0.17	0.18	0.09	0.07	0.04	0.01
PCA2	0.30	0.23	-0.16	-0.18	-0.22	0.22	0.02	-0.08	-0.24	-0.21	0.03	-0.23	-0.06	-0.15	-0.03
PCA3	0.10	0.06	-0.18	-0.06	-0.14	0.12	0.02	-0.19	0.50	-0.20	-0.12	0.34	-0.33	0.16	0.15

Figure 4 (A). Contribution of each graphlet to the principal components of GCV of the overall dataset. A higher magnitude indicates a higher contribution of that graphlet with the principal components.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
PCA1	0.19	0.22	0.20	0.05	0.21	0.22	0.18	0.18	0.14	0.13	0.08	0.12	0.06	0.04	-0.02	0.20	0.22	0.21	0.16	0.18	0.20	0.19	0.13	0.14	0.12	0.09	0.13	0.13	0.15	0.07	0.13	0.14	0.10	0.15	0.13	0.12
PCA2	0.09	-0.07	-0.01	0.27	-0.03	-0.06	-0.14	-0.10	-0.11	0.00	0.26	0.22	0.14	0.22	0.10	-0.02	-0.04	-0.08	-0.02	-0.12	-0.10	-0.11	-0.13	-0.12	0.00	0.19	0.19	0.05	0.01	0.25	0.20	-0.07	0.16	0.06	-0.07	-0.07
PCA3	-0.04	0.02	-0.02	-0.04	-0.04	0.01	-0.01	-0.04	-0.04	0.11	-0.11	0.06	0.36	-0.16	0.01	-0.03	-0.03	0.02	-0.10	-0.02	-0.02	-0.02	-0.05	-0.04	0.09	-0.12	-0.03	0.00	0.09	-0.11	0.07	0.07	-0.17	0.05	-0.02	-0.05
PCA4	-0.06	0.01	-0.06	-0.01	-0.05	0.00	-0.07	-0.02	-0.07	0.41	-0.01	-0.09	-0.09	0.06	0.05	-0.03	-0.04	0.05	-0.10	-0.08	-0.06	0.01	-0.12	0.01	0.36	-0.02	-0.07	-0.01	0.35	-0.02	-0.05	0.22	-0.07	-0.08	-0.05	-0.10
PCA5	0.08	-0.02	0.05	0.10	-0.03	0.02	-0.01	0.06	0.05	-0.04	-0.00	0.04	-0.06	-0.05	0.49	-0.03	0.01	-0.01	-0.04	-0.02	0.02	0.04	0.01	0.08	-0.03	-0.07	-0.03	-0.00	-0.01	0.01	0.05	-0.00	-0.08	0.02	0.02	-0.04

	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
PCA1	0.13	0.15	0.16	0.07	0.07	0.04	0.08	0.06	0.05	0.08	0.05	0.08	0.04	0.08	0.08	0.12	0.05	0.10	0.03	0.02	0.02	-0.01	0.00	0.05	0.04	0.03	0.04	0.05	0.03	0.01	-0.00	-0.00	0.01	0.00	0.00	-0.01	-0.00
PCA2	-0.09	-0.13	-0.12	-0.01	0.11	0.19	0.17	0.15	0.13	0.04	0.14	0.12	0.21	-0.06	-0.06	-0.05	0.14	0.11	0.04	0.08	0.02	0.09	0.08	0.09	0.15	0.15	-0.02	0.02	0.09	0.03	0.06	0.08	0.03	0.04	0.03	0.04	0.01
PCA3	-0.08	-0.02	-0.03	0.08	0.29	-0.21	-0.10	-0.08	0.15	0.04	0.32	0.36	-0.17	-0.05	-0.05	0.03	-0.12	-0.04	0.24	-0.13	0.03	-0.02	0.07	0.22	0.06	-0.09	0.00	0.13	-0.15	0.15	0.07	-0.06	0.07	-0.06	0.12	-0.03	0.00
PCA4	-0.13	-0.07	-0.03	0.32	-0.11	0.03	0.00	0.18	-0.11	0.26	-0.10	-0.10	0.08	-0.07	-0.06	0.20	-0.04	-0.09	-0.08	0.07	0.20	0.04	-0.01	0.04	0.00	0.02	0.09	-0.11	-0.01	0.02	-0.01	0.07	-0.04	0.01	-0.05	0.06	0.00
PCA5	0.04	0.05	0.07	-0.01	-0.09	-0.11	-0.08	-0.03	-0.00	-0.02	-0.04	-0.04	-0.06	0.05	0.06	0.02	-0.04	0.02	-0.02	-0.04	-0.04	0.45	0.42	-0.08	-0.13	-0.11	0.01	-0.01	-0.05	-0.04	0.29	0.26	-0.04	-0.04	0.10	0.22	0.08

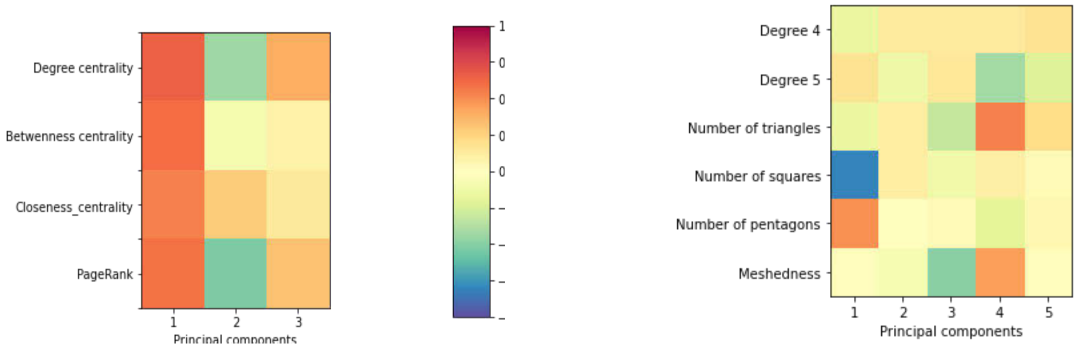
Figure 4 (B) Contribution of each orbit to the principal components of OCVs of the overall dataset. A higher magnitude indicates a higher contribution of that orbit with the principal components. Explained variances are 24%, 16%, 5.8%, 5.1, and 4.5% correspondingly.

According to weighting in Figure 5 (A), most of the graphlets collectively contribute to the first principal component (PC_1). This might be a result of intercorrelation between the counts of most graphlets. PC_2 is correlated with the graphlets that contain cycles in the structure and can be translated to the "loopedness" of networks. Lastly, PC_3 is predominantly described as graphlets that contain 4 fully connected nodes in the structure. Another observation is that orbit 14 (which belongs to G_8) has comparably low contribution to PC_1 in OCV as well. Since 4 fully connected nodes is a rare structure, these features explain the smaller share of variance (PC_3 in GCV and PC_5 in OCV).

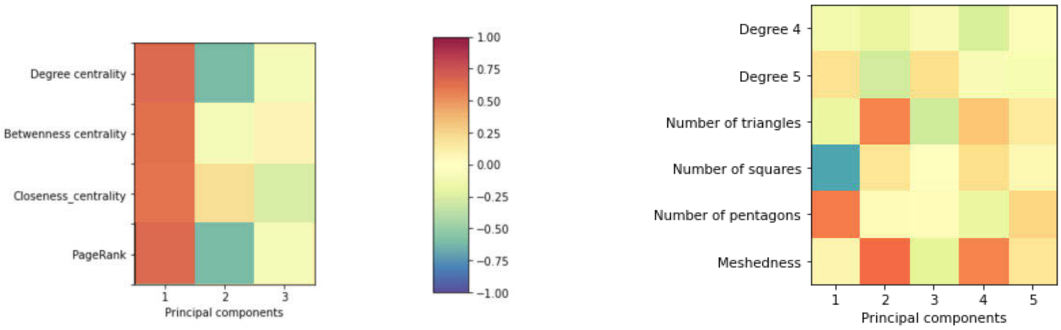
Likewise we observe that a large proportion of graphlets contribute to the first principal component of OCV. We can also note that orbits 1, 4, 5, 15, 16, 17, 20, and 21 have slightly higher weighting than the others. All these orbits belong to 4 graphlets of the same shape - a sequence of nodes (2, 3, 4, and 5 nodes) and additionally to G_{19}. We extend the discussion in the section . Likewise, orbits that correspond to "corners" of the triangles are included in PC_2. At the same time, orbits 19-23 have a negative weight, which emphasises the focus on the triangular structures. PC_4 is highly correlated with orbits 9, 24, and 28. All these orbits are located in the neighbourhood of triangular structures.

Below we summarize the correlation of the principal components of the entire OCV and GCV datasets for two networks separately. In each case, we carry out PCA separately in order to assess the difference in the main PCs.

4.2.1 Relationship of PCs to graph-theoretical metrics

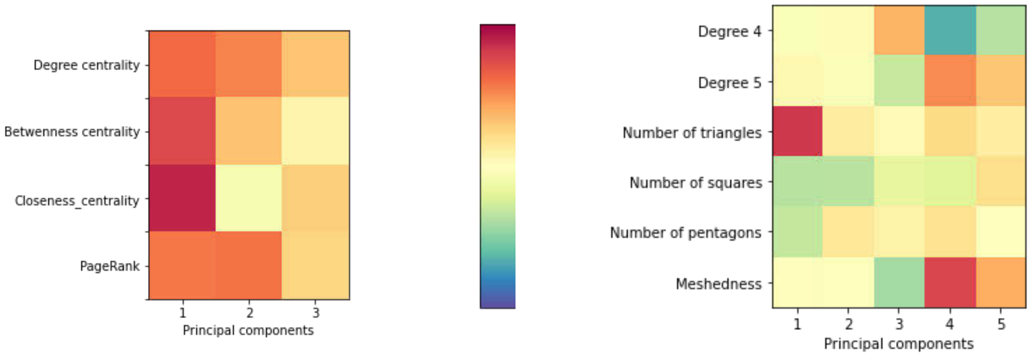


(a) Generated networks with maximum connectivity 4



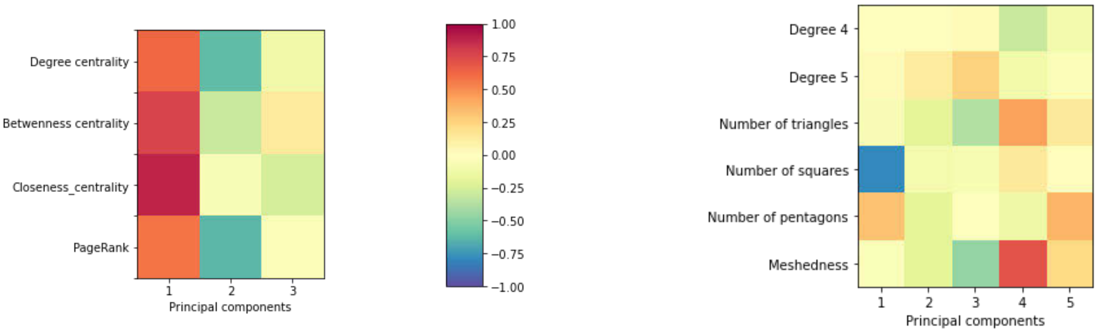
(b) Generated networks with maximum connectivity 5

Figure 5. Correlations of local graph-theoretical metrics with the principal components of OCV (left) and global graph-theoretical metric with GCV (right) on the example of modifications of the ZJ network.



(a) Generated networks with maximum connectivity 4

CC BY NC SA



(b) Generated networks with maximum connectivity 5

Figure 6. Correlations of local graph-theoretical metrics with the principal components of OCV (left) and global graph-theoretical metric with GCV (right) on the example of modifications of FOS network

Particular orbital indices are inherently related to graph-theoretical measures that characterize the structure and topology of a graph network. For example, an orbital count of the orbit 0 is equivalent to node degree. In other words, the number of times a node “touches” orbit 0 equals to a number of incidental links. Graph orbits of graphlets sequence of nodes (e.g. orbits 4, 5, 15, 16, and 17) are likely to appear in the parts of the network that are located in between larger neighbourhoods as a hub. One can imagine this as a hyperbolic increase in the number of nodes in the k-hop neighbourhood of a node as k increases. In other words, the number of combinations a graphlet G_9 can be placed in that neighbourhood will be high, which increases the count in its turn. In all of the cases, these orbits showed the highest contribution to the first principal component of OCVs (left pictures in Figures 5 and 6) and a strong correlation with centrality metrics.

The number of loops (or n-cycles) such as triangles, squares, or pentagons, in the network, is natively related to the number of identified graphlets that have loops in them. Since the graphlets in the ORCA counting algorithm are limited to the size of 5 nodes, higher-order cycles are omitted. Some of the graphlets, e.g. G_9, G_7, G_{14}, and G_{22} contain triangular structures in themselves, which means that their orbit counts might be correlated. As we show in the principal component analysis, some of the principal components are highly correlated with the number of identified squares and triangles. We further observe in the case of ZJ a stronger correlation of its PC_1 with the number of squares and pentagons than in the case of FOS. We argue that the original structure of ZJ contains higher-order cycles which contribute to the variability in GCVs, hence higher correlation with the first principal component.

4.3 Random Forest Regression model

To assess the informational gain from incorporating the graphlet analysis we introduce the feature vector from orbit count in a random forest model. Here we hypothesize that incorporating graphlet features into the regression model that estimates the criticality of the node will increase the performance of a regression model.

We then set up the task as a regression model with the importance of a node as the target variable of the i. We evaluate 3 cases of used features in the model: solely traditional graph measures *(a)*, solely OCV *(b)*, and traditional features and OCV altogether *(c)*. Further on, we compare the performance measured by coefficient determination. Traditional metrics include degree, betweenness and closeness centrality, and distance to reservoirs.

Table 2. Summary of network generation

	ZJ		**FOS**	
Maximum node connectivity	4	5	4	5
Topological features, R^2	0.36	0.25	0.30	0.18
OCV, R^2	0.36	0.21	0.32	0.15
Topological features + OCV, R^2	**0.47**	**0.34**	**0.39**	**0.22**

The random forest model has been set up with the same set of hyperparameters and evaluated with a coefficient of determination on 10% of the dataset.

We can see in the example above the improvement in the predictive power of the regression model as we include GCV in the analysis. This hints at the conclusion that graphlet representation contains local information that assists in the prediction.

5. CONCLUSIONS & FUTURE WORK

Consequently, this work shows that graphlet representation is beneficial as a descriptor of WDS. GCV captures local patterns and implicitly contains information about the global structure. PCA showed that GCV captures well the number of loops of different sizes and meshedness of a network. Simultaneously, OCV aids in the identification of small hubs and contains information about the degrees of the network. It is important to note that although the graphlet analysis enriches the set of descriptors of the graph structure and provides higher resolution to a local neighbourhood, it is not an exhaustive descriptor of the network structure. A key factor that limits the expressive power of graphlet analysis appears from the influence of global connectivity. Although graphlet decomposition is capable of identifying hubs in the influence of connectivity and network modularity on a global scale might not be captured.

Furthermore, this research introduces graphlets as an assisting factor in assessing local criticality. As it was mentioned, solely topological representation is not a piece of sufficient but necessary information. This problem implies the need for a holistic approach that includes both hydraulics and geometry. For example, the location of the reservoir and the main artery of the WDS is definitive for the importance of the elements of the network. Nevertheless, incorporating graphlet features into the analysis shows promising first results.

For future work, we suggest extending the dataset with more hydraulically “independent” and less modular networks in order to further mitigate the influence of these factors. For example, an artificial WDS with many homogeneously placed reservoirs can aid in the analysis.

Adding hydraulically informed edge weights to the graphlet count could be a potential direction for the analysis. Orbit count might be weighted by hydraulical costs, e.g. by multiplication of node counts by a weight factor. This can serve as a bridge between hydraulics and topological description of the local neighbourhood and increase the expressiveness of graphlets.

6. REFERENCES

[1] Yazdani, A., & Jeffrey, P. (2011). Complex network analysis of water distribution systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 21(1), 016111. https://doi.org/10.1063/1.3540339

[2] Meng, F., Fu, G., Farmani, R., Sweetapple, C., & Butler, D. (2018). Topological attributes of network resilience: A study in water distribution systems. Water Research, 143, 376–386. https://doi.org/10.1016/J.WATRES.2018.06.048

[3] Herrera, M., Abraham, · Edo, Stoianov, I., & Abraham, E. (2016). A Graph-Theoretic Framework for Assessing the Resilience of Sectorised Water Distribution Networks. Water Resour Manage, 30, 1685–1699. https://doi.org/10.1007/s11269-016-1245-6

[4] Tornyeviadzi, H. M., Neba, F. A., Mohammed, H., & Seidu, R. (2021). Nodal vulnerability assessment of water distribution networks: An integrated Fuzzy AHP-TOPSIS approach. International Journal of Critical Infrastructure Protection, 34, 100434. https://doi.org/10.1016/J.IJCIP.2021.100434

[5] Giudicianni, C., di Nardo, A., di Natale, M., Greco, R., Francesco Santonastaso, G., & Scala, A. (n.d.). Topological Taxonomy of Water Distribution Networks. https://doi.org/10.3390/w10040444

[6] Pinto, J., Varum, H., Bentes, I., Agarwal, J., Pinto, J., Bentes, · I, Bentes, I., Varum, H., & Agarwal, J. (2010). A Theory of Vulnerability of Water Pipe Network (TVWPN). Water Resources Management 2010 24:15, 24(15), 4237–4254. https://doi.org/10.1007/S11269-010-9655-3

[7] Perelman, L., & Ostfeld, A. (2011). Topological clustering for water distribution systems analysis. Environmental Modelling & Software, 26(7), 969–972. https://doi.org/10.1016/J.ENVSOFT.2011.01.006

[8] Hajebi, S., Roshani, E., Cardozo, N., Barrett, S., Clarke, A., & Clarke, S. (2016). Water distribution network sectorisation using graph theory and many-objective optimisation. Journal of Hydroinformatics, 18(1), 77–95. https://doi.org/10.2166/HYDRO.2015.144

[9] Gutiérrez-Pérez, J. A., Herrera, M., Pérez-García, R., & Ramos-Martínez, E. (2013). Application of graph-spectral methods in the vulnerability assessment of water supply networks. Mathematical and Computer Modelling, 57(7–8), 1853–1859. https://doi.org/10.1016/J.MCM.2011.12.008

[10] Phan, H. C., Dhar, A. S., & Bui, N. D. (2021). Accounting for Source Location on the Vulnerability Assessment of Water Distribution Network. Journal of Infrastructure Systems, 27(3), 04021024. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000620

[11] Giustolisi, O., Ridolfi, L., & Simone, A. (2019). Tailoring Centrality Metrics for Water Distribution Networks. Water Resources Research, 55(3), 2348–2369. https://doi.org/10.1029/2018WR023966

[12] Agathokleous, A., Christodoulou, C., Christodoulou, S. E., & Christodoulou schristo, S. E. (2017). Topological Robustness and Vulnerability Assessment of Water Distribution Networks. Water Resour Manage, 31, 4007–4021. https://doi.org/10.1007/s11269-017-1721-7

[13] Sitzenfrei, R., Möderl, M., & Rauch, W. (2013). Automatic generation of water distribution systems based on GIS data. Environmental Modelling and Software, 47, 138–147. https://doi.org/10.1016/J.ENVSOFT.2013.05.006

[14] Herrera, M., Abraham, E., & Stoianov, I. (2015). Graph-theoretic surrogate measures for analysing the resilience of water distribution networks. Procedia Engineering, 119(1), 1241–1248. https://doi.org/10.1016/J.PROENG.2015.08.985

[15] Zarghami, S. A., & Gunawan, I. (2020). A domain-specific measure of centrality for water distribution networks. Engineering, Construction and Architectural Management, 27(2), 341–355. https://doi.org/10.1108/ECAM-03-2019-0176/FULL/PDF

[16] Ulusoy, A.-J., Stoianov, I., & Chazerain, A. (n.d.). Hydraulically informed graph theoretic measure of link criticality for the resilience analysis of water distribution networks. https://doi.org/10.1007/s41109-018-0079-y

[17] Finotelli, P., Piccardi, C., Miglio, E., & Dulio, P. (2021). A Graphlet-Based Topological Characterization of the Resting-State Network in Healthy People. Frontiers in Neuroscience, 15, 376. https://doi.org/10.3389/FNINS.2021.665544/BIBTEX

[18] Jamra, H., Savonnet, M., & Leclercq, E. (2021). Detection of Event Precursors in Social Networks: A Graphlet-Based Method. Research Challenges in Information Science, 415, 10. https://doi.org/10.1007/978-3-030

BY NC SA

[19] Hočevar, T., & Demšar, J. (2014). A combinatorial approach to graphlet counting. Bioinformatics, 30(4), 559–565. https://doi.org/10.1093/BIOINFORMATICS/BTT717

[20] Melckenbeeck, I., Audenaert, P., Michoel, T., Colle, D., & Pickavet, M. (2016). An Algorithm to Automatically Generate the Combinatorial Orbit Counting Equations. PLoS ONE, 11(1). https://doi.org/10.1371/JOURNAL.PONE.0147078

[21] Saldarriaga, J. G., Ochoa, S., Rodríguez, D., & Arbeláez, J. (2009). Water distribution network skeletonization using the resilience concept. Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008, 852–864. https://doi.org/10.1061/41024(340)74

[22] The PageRank Citation Ranking: Bringing Order to the Web. (1998). www.yahoo.com

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14787

A STOCHASTIC SEWER MODEL TO PREDICT PIPE FLOWS AND POLLUTANT LOADS IN AN URBAN DRAINAGE SYSTEM

William Addison-Atkinson [1], Albert S. Chen[2], Fayyaz Memon[3], Jan Hofman[4] and Mirjam Blokker[5]

[1,3] University of Exeter, Exeter (The U.K)
[2,4] University of Bath, Bath (The U.K)
[4, 5] KWR Water Research Institute (The Netherlands)

[1] *wa259@exeter.ac.uk*, [2] *A.S.Chen@exeter.ac.uk*,[3] *f.a.memon@exeter.ac.uk*,
[4]*jamhh20@bath.ac.uk*, [5] *mirjam.Blokker@kwrwater.nl*

Abstract

This work implemented a stochastic sewer model (SIMDEUM-WW) to forecast dry weather sewer flows and pollutant loading, from probabilistic household demand patterns based on information about inhabitants and appliance usage. The probabilistic outputs were fed into MIKE URBAN (DHI) for hydrodynamic and water quality simulations. The MIKE URBAN model consists of a 1D sewer network model. The model was validated against field measurement data and the results show that the SIMDEUM-WW can adequately calculate wastewater and pollutant loading. However, the SIMDEUM-WW was originally calibrated on households in the Netherlands such that errors were observed in the UK application. The uncertainties in actual flow and pollutant loading also contributed to the inaccuracy of modelling results.

Keywords
Flood model, stochastic modelling, urban drainage.

1 INTRODUCTION

Climate change and population growth has been responsible for placing a burden on water resource management, particularly in recent decades (1–3). Changes in urbanisation and land use has complicated this problem, as the natural hydrological cycle is disrupted by non-permeable surfaces. This leads to increased run off, and an upsurge in flood risk. Urban flood models play a crucial role in flood risk management, as they aid in the geospatial analysis of flood hazards. Contemporary flood management requires reliable numerical models to better understand the performance of drainage systems. Stochastic methods are important in urban drainage modelling as they introduce probability distributions, rather than unique values to modelling outcomes (4). Therefore, the stochastic modelling results could better reflect the possible scenarios than single deterministic modelling result. This could a more realistic modelling approach, as some have argued that drainage models should always have some element of probability due to their complexity (5).

This paper intends to validate a stochastic sewer model (SIMDEUM-WW) against a new data set (*E. coli,* suspended solids, total nitrogen, and total phosphorus) collected in a U.K catchment. Further validation of the model will include the comparison of spill volumes at a combined sewer overflow (CSO) in the downstream part of the catchment.

2 METHODS

2.1 Study location and data sets

The case study is situated in a small town called Sandford, the Poole Harbour region of the U.K. A small section of Sandford Town was chosen for the analysis, which included 500 households. Data were provided by the water utility company Wessex Water. Data included a sewer map and CSO monitoring data. The data contained pollutant loads of *E. coli,* suspended solids, total nitrogen, and total phosphorus. Hydraulic data included spill volumes and durations at the outflow. Figures 1 and Figure *2* shows the concentration measurements.

2.2 Hydraulic discharge

SIMDEUM was originally developed in the Netherlands as a water demand tool (6), though has since been developed and calibrated to simulate wastewater discharge based on household and appliance usage data, using Monte Carlo Simulations (7,8) Calibration took place in a Dutch case study (Prinseneiland). Input variables (household occupancy, home–presence, individual details of household water consumption and average household occupancy) were adjusted in the calibration procedure. Validation included reviewing model performance over an average week using the Nash–Sutcliffe efficiency (NSE), the root mean squared error (RMSE) and correlation coefficient (R2). Dry weather flow data was chosen at several points of the year (2 weeks from each season) to produce mean water use patterns of the catchment.

2.3 Wastewater

SIMDEUM-WW links water quality from the stochastic flow patterns. A variety of pollutants and water quality indicators can be modelled, including *E. coli*, biological oxygen demand (BOD), chemical oxygen demand (COD), nitrogen (N), phosphorous (P) and suspended solids (SS). Appropriate inputs for nutrient values in SIMDEUM-WW were conducted in previous research (8).

2.4 Sewer model

MIKE URBAN (DHI) will simulate 1D sewer flow. Here, outputs from SIMDEUM-WW will be integrated into the 1D MIKE model. The sewer model is used to analyse the water movement within sewer systems that flow is confined by the drainage network and can only move along the pipes such that it is regarded as a 1D modelling practice. The storm event period was set for simulation period.

2.5 Water quality model

MIKE ECO Lab is an integrated module in the DHI software. The module is coupled to the Advection-Dispersion Modules of the hydrodynamic flow model, so that transport mechanisms based on advection-dispersion can be integrated in the MIKE ECO Lab simulation. *E. coli*, N, P and SS were simulated over the storm duration period.

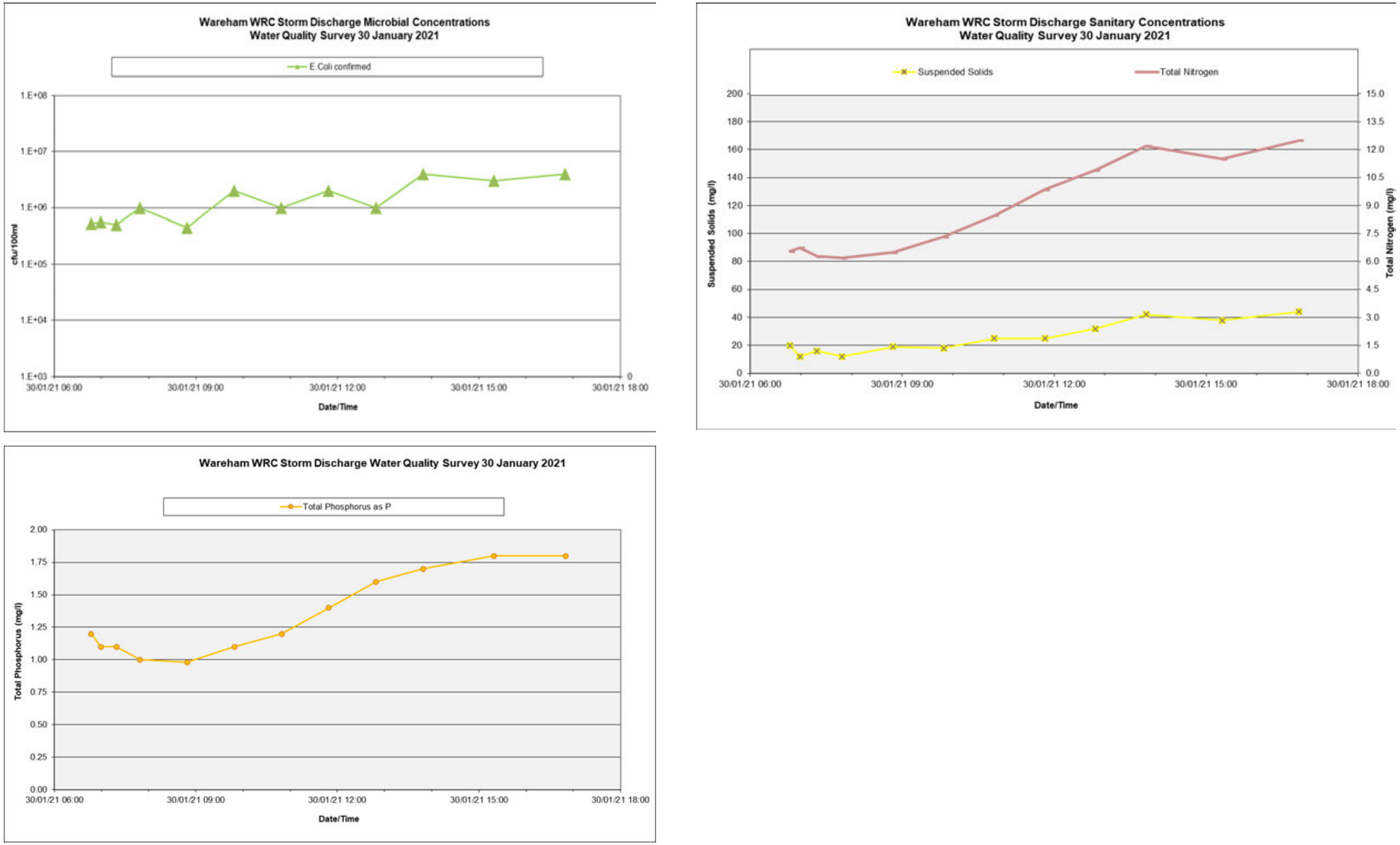


Figure 1: E. coli, suspended solids, total nitrogen, and total phosphorus data

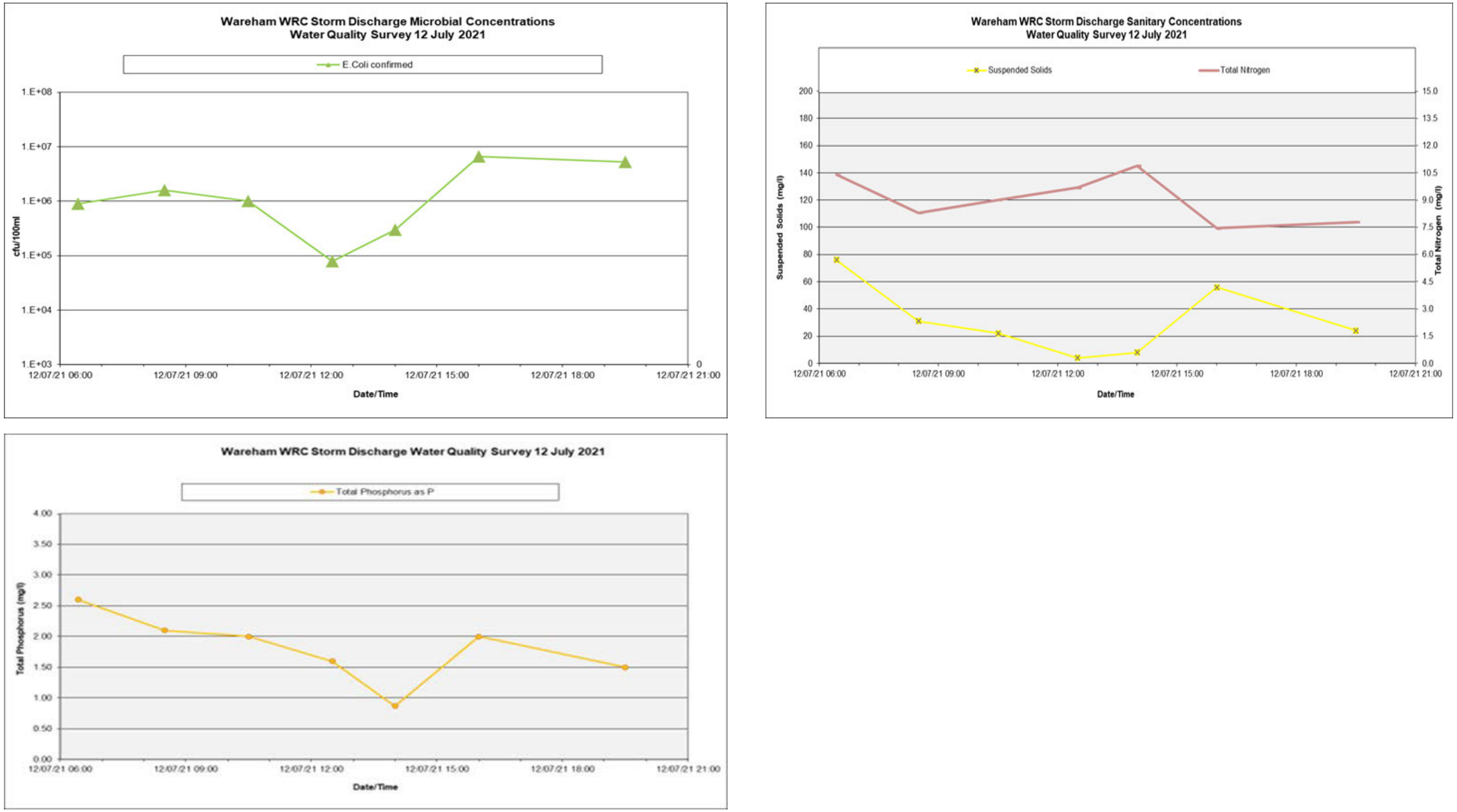


Figure 2: E. coli, suspended solids, total nitrogen, and total phosphorus data

cc BY NC SA

3 RESULTS AND DISCUSSION

As this is ongoing work preliminary results will be discussed. It is predicted that the stochastic sewer model can properly calculate wastewater flows and pollutant loading. However, small errors exist due to SIMDEUM-WW being calibrated on households in the Netherlands. The water consumption patterns in the UK household are not the same as the Netherlands households, though water demands in both countries are similar. As drainage systems are complicated other errors may include flow and pollutant loading at the pipe outflow. Combined sewer systems are very complicated, so flows and pollutants from other parts of the system is likely.

4 CONCLUSIONS

It was predicted that the numerical model was able to replicate accurate hydraulic and wastewater discharges, validated against CSO monitoring data. Though small error was present due to the complexity of dual drainage systems and the fact SIMDEUM was calibrated in The Netherlands.

5 REFERENCES

[1] Jongman B, Hochrainer-Stigler S, Feyen L, Aerts JCJH, Mechler R, Botzen WJW, et al. Increasing stress on disaster-risk finance due to large floods. Nature Clim Change. 2014 Apr;4(4):264–8.

[2] Hall J, Arheimer B, Borga M, Brázdil R, Claps P, Kiss A, et al. Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrol Earth Syst Sci. 2014 Jul 30;18(7):2735–72.

[3] Rubinato M, Nichols A, Peng Y, Zhang J min, Lashford C, Cai Y peng, et al. Urban and river flooding: Comparison of flood risk management approaches in the UK and China and an assessment of future knowledge needs. Water Science and Engineering. 2019 Dec;12(4):274–83.

[4] Leitao JP, Simoes NE, Pina RD, Ochoa-Rodriguez S, Onof C, Marques AS. Stochastic evaluation of the impact of sewer inlets' hydraulic capacity on urban pluvial flooding. Stochastic environmental research and risk assessment. 2017 Oct;31(8):1907–22.

[5] Butler D, Digman C, Makropoulos C, Davies JW. Urban drainage. Boca Raton: CRC Press, Taylor & Francis; 2018. 1 p.

[6] Blokker EJM, Vreeburg JHG, van Dijk JC. Simulating Residential Water Demand with a Stochastic End-Use Model. J Water Resour Plann Manage. 2010 Jan;136(1):19–26.

[7] Bailey O, Arnot TC, Blokker EJM, Kapelan Z, Vreeburg J, Hofman JAMH. Developing a stochastic sewer model to support sewer design under water conservation measures. Journal of Hydrology. 2019 Jun;573:908–17.

[8] Bailey O, Zlatanovic L, van der Hoek JP, Kapelan Z, Blokker M, Arnot T, et al. A Stochastic Model to Predict Flow, Nutrient and Temperature Changes in a Sewer under Water Conservation Scenarios. Water. 2020 Apr 21;12(4):1187.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14788

IOT.H2O - SUPERVISION AND CONTROL FOR WATER SYSTEMS BASED ON LOW-COST HARDWARE AND OPEN-SOURCE SOFTWARE

Harald Roclawski[1], Laura Sterle[2], Martin Böhle[3] and Thomas Krätzig[4]

[1,2,3] TU Kaiserslautern, Institute of Fluid Mechanics and Turbomachinery, Gottlieb-Daimler-Str. 44,
67663 Kaiserslautern, Germany
[4]Dr.Krätzig Ingenieurgesellschaft mbH - Süsterfeldstr. 25 - D-52072 Aachen, Germany

[1]roclawsk@mv.uni-kl.de, [2]laura.sterle@mv.uni-kl.de, [3]roclawsk@mv.uni-kl.de, [4]infoline@dr-kraetzig.com

Abstract

An adequate monitoring and control of water production and distribution systems is essential for operating the system efficiently, avoiding water losses and preventing damage to the system. Usually this is achieved by SCADA systems. But SCADA systems are technically complex and expensive. They are often not affordable for small water utilities. Their systems are operated with a very low level of automatization and monitoring causing inefficient operation, water loss or even interruption of service. The Internet of Things concept is very promising to realize a cost-effective tool for monitoring and control of water systems. The potential of this approach is investigated in the research project IoT.H2O funded in the WaterJPI call IC4WATER.

In the paper, the concept of the IoT system for monitoring and control of water supply systems is explained. Special gateways and IoT devices have been developed which communicate with the IoT platform by the LoRaWAN network protocol or WiFi. It is shown how a pump test rig can be controlled with these devices. Also, results of field installations in a water distribution system in Germany are presented. Moreover, it is shown how data from existing SCADA systems and results from network simulations with EPANET can be included in the IoT system and how the data can be processed and evaluated by the IoT platform. By combining the data of different systems, the developed IoT system provides the user a cost-effective tool for monitoring and control of a water system.

Keywords
Internet of Things, LoRaWAN, smart water, digital twin, Thingsboard, EPANET.

1 INTRODUCTION

Usually SCADA (Supervisory Control and Data Acquisition) systems are used for monitoring and operation of major water transport and distribution systems. However, because of their complexity and expensiveness, they are often not achievable for small water utilities and therefore their distribution network has to be operated with a very low level of automatization or even manually. The results are high water losses by frequent reservoir overflows or interruption of water supply because of empty storage tanks and pipe bursts caused by overpressure [1], [2], [3].

Also, the number of sensors for monitoring the water system which are connected to the SCADA system is usually limited. Based on our experience from previous research projects with water supply companies, we found that very often some data is monitored but not evaluated or the dataset is incomplete. Therefore, it is not possible to ensure a reliable operation of the system and to judge for example, the efficiency of pumps, which are the biggest energy consumers in water supply systems. Studies of pump manufacturers show that many pumps are very often not operated close to their point of best efficiency. In [4] 2500 well pumps were investigated, showing that about 37% of these pumps were operated under off-design conditions. In a press information [5] Grundfos states that for more than two thirds of the pumps installed in industry energy savings of up to 60% are possible. Based on our experience from energy audits with water supply companies we found that these numbers are realistic and not marketing numbers.

For judging pump operation, a complete set of data is essential. Then, based on engineering knowledge, pumps can be selected guaranteeing high energy efficiency. On top, by applying advanced optimization tools, additional savings are possible. But there is more than pumps in water supply systems. As an example, tank levels, water demand and quality or pressure in the network have to be monitored and controlled.

One of the United Nations sustainable development goals is to "Ensure availability and sustainable management of water and sanitation for all". Thus, in order to realize monitoring and control systems for small water utilities or systems in developing countries, the technology has to be very cost-effective. With the technological progress in ICT (Information and Communications Technology), new possibilities for technologically advanced but low-cost solutions for monitoring and control of water systems are available. Especially the Internet of Things (IoT) concepts is very promising.

Considering open-source software and cost-effective hardware there are already reports available in the literature. For monitoring and control of a water system, in [6] a system based on a Raspberry PI, a GSM module and Arduino UNO microcontroller is presented. With IoT devices, it is possible to measure pressure and flow rate, control a motorized electric valve and start a pump using a solid state relay switch. For interacting with the IoT system, a web application was created. By means of a fuzzy logic algorithm the operators are informed in case of events where specific actions should be performed as starting or stopping the pump.

An IoT system for pressure monitoring in water supply networks is proposed in [7]. IoT devices based on Arduino microcontrollers send pressure data using the Sigfox communication protocol to the IoT platform ThingSpeak where the data is analyzed and visualized. Alarms can be generated for the case that anomalous pressure data is detected. The system is in operation in a real water distribution system in Spain. In contrast to the LoRaWAN communication which is used in this work, for using Sigfox, the customers pay per year and per connected device. Also, customers relie on network coverage which is in hand of the operator. For ThingSpeak, according to the authors it is only possible to connect four devices free of charge. For more devices estimated costs of 70€ to 580€ apply depending on the subscribed functionalities.

The monitoring of water consumption of households and commercial buildings by an IoT system based on low-cost hardware is described in [8]. On each faucet, Hall-Effect flow sensors which are connected to a NodeMCU are installed. The data is sent using WiFi to a Heroku cloud database and can be accessed by a web browser.

In [9], the authors present the IoT framework REFlexWater based on the FIWARE platform. The focus of the paper is more on the processing of data than on the technology of IoT devices. REFlexWater uses declarative processes to model water systems which allows operators to handle routine situations and unforeseen occurrences. The data from low-cost IoT devices installed in a water system is sent to servers. The servers use Complex Event Processing methods to process the data in real-time for identifying water loss. In this case alarms are generated so that the operators can take suitable actions for solving the issue.

In the project IoT.H2O, which is funded by the European initiative „Water Challenges for a Changing World“ (Joint Programming Initiative, JPI Water), the potential of the Internet of Things concept for supervision and control of water systems is investigated. In the project, an IoT-system based on low-cost hardware and open-source software has been developed. The central element of the system is an open-source IoT platform. With the platform it is possible to receive, visualize and store data from IoT devices and to control components like pumps and valves. In addition, software models like network models, optimization tools or digital twins of pumping stations as well as additional data like a weather forecast can be virtually connected to the platform. In the paper, the design of the IoT system and the IoT devices are presented in detail and several use cases are discussed.

2 IOT SYSTEM

Figure 1 shows the concept of the proposed IoT architecture based on a three-layer model. Physical objects like pumps, valves or storage tanks equipped with appropriate sensors or actuators belong to the field layer. Microprocessors are connected to the sensors and actuators and transmit data from measurements or commands for actuators. In the transport level, several different communication technologies are possible. In IoT.H2O, the microcontrollers communicate via WiFi in case of short distances. For longer distances and when there is no WiFi available the LoRaWAN network protocol is used to transmit data from remote locations. In this case, the data is sent via LoRaWAN to a gateway which is connected to the internet. In the third level, the application level, all data is organized by an IoT platform. In IoT.H2O, the open-source platform Thingsboard is used. The users can interact with Thingsboard using mobile devices or computers through a web browser and visualize measurement data or control devices in dashboards.

In contrast to centralized SCADA systems, which receive information from remote sensors and control components, the IoT concept relies on a decentralized approach. Additional IoT devices can be easily added to the IoT platform by adding their specification in the software. Hardware changes on a centralized computer, which is often the reason for a limited amount of sensor data for SCADA systems, are not necessary. The IoT platform offers the possibility to combine data from different systems. Since the communication between the IoT platform and the devices is based on software protocols, a software for network calculation or optimization of pump schedules can be connected to the IoT platform. Also, data from already existing SCADA systems can be included. So, it is not absolutely necessary to completely replace a SCADA system by an IoT system. Instead, the functionality of the SCADA system can be extended by the IoT system. In the following sections the functionality of single components of the IoT system will be presented in more detail.

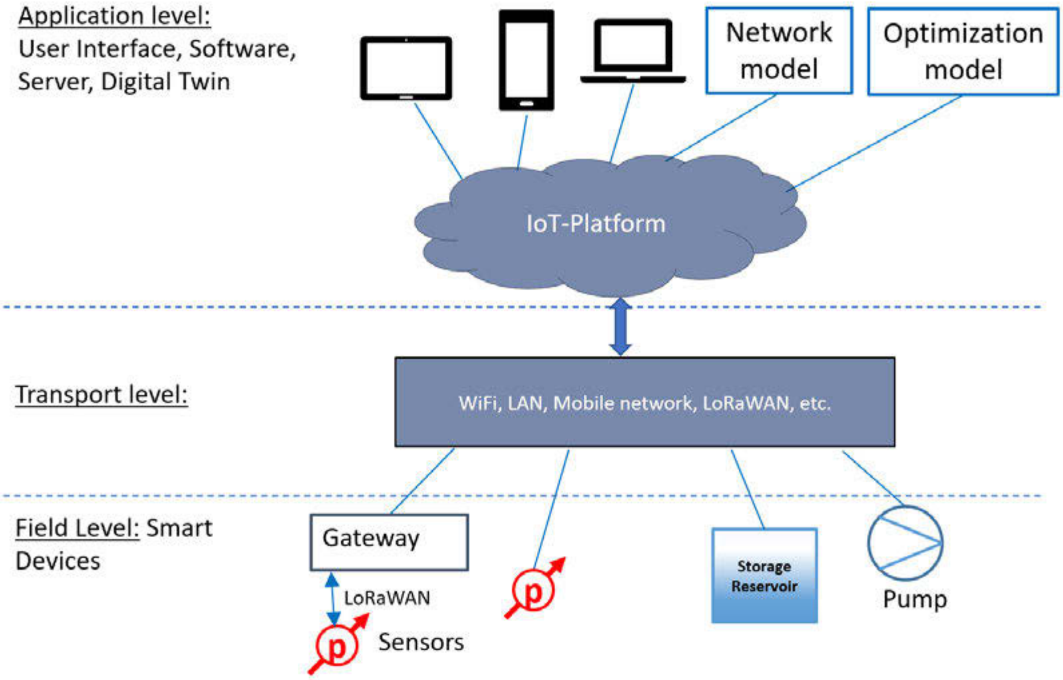


Figure 1: IoT architecture (IoT.H2O)

BY NC SA

2.1 IoT platform

In IoT.H2O, the open source IoT platform Thingsboard is used for data storage, data visualization and control of devices with dashboards. In Thingsboard, the components of the water distribution system, e.g. pumps, valves, tanks, pipes or sensors are mapped as devices. Thingsboard offers several possibilities for connecting the devices. In IoT.H2O, the open-source Message Queuing Telemetry Transport protocol (MQTT) is used. In this case, the Thingsboard server acts as MQTT broker which transports messages between sender and receiver. MQTT is a so called publish-subscribe network model. This means, that the sender publishes a message with a specific topic and the receiver has to subscribe to this topic to ensure that only the desired data is received.

For MQTT, there are many software libraries available. For the IoT field devices, the Arduino PubSubClient is used. The communication with actuators, like a control valve or a frequency converter, is realized with Remote Procedure Calls (RPC) based on MQTT. With RPC it is possible for the IoT platform to run programs on a different computer and return the results of the code execution to the IoT platform. This can be, for example, a microcontroller controlling the control valve position or any other code. For RPC, specific control widgets like an on-off switch or a knob control arranged in a dashboard communicate with the actuator. To ensure that each RPC command is received by its corresponding actuator, the control widget is mapped to a device. In case of RPC, the message is sent in JavaScript Object Notation (JSON) format and consists of a method name and parameters.

The method name indicates whether the “RPC get value” or “RPC set value” method is used. With the set value method, changes in the Thingsboard control widget are passed to the device. This is the case when the user changes for example the position of the control valve or the frequency of the motor. With the get value method, parameters can be passed from the device to Thingsboard for transmitting the current status of a device to the IoT platform. It is also useful for checking if a device is online. For each control widget, the function names for the get value and set value method can be specified in Thingsboard so that the correct device can be identified by the method in the callback function.

All telemetry data which is received by Thingsboard is stored in a database and can be visualized in a dashboard. Dashboards can be created by the user by drag & drop of predefined widgets. So, no programming of a web interface is necessary. However, it is possible to modify these widgets or to create new widgets. The dashboard will be described in a later section in more detail.

Another important feature of Thingsboard is the rule engine. With the rule engine, it is possible to automatically process or filter incoming data or generate alarms according to defined thresholds. In case of the water distribution system, it is possible to automatically generate an alarm for the case that a tank level is too high or too low. In our case we are also using the rule chain for calculating the hydraulic head and efficiency of a pump based on the incoming data of pressure, flow rate, torque and speed.

2.2 IoT-devices

In the IoT.H2O project, devices based on microcontrollers have been developed which can be connected to sensors for measuring e.g. pressure, flow or temperature. For cases where the IoT devices belong to a local wireless network and are powered by mains, the sensor data is transmitted by WiFi. For cases where there are no local WiFi network and wired power supply available, devices are battery powered and belong to a wireless Low-Power Wide-Area Network (LPWAN). In IoT.H2O the LPWAN is set up with Semtech’s LoRa radio technology in combination with the LoRaWAN network protocol where gateways relay data packages between the IoT devices in the field and the central network server and the IoT platform.

A requirement for the development of the IoT devices was to use as much as possible already available, cost-effective and functional modules which are usually in stock of suppliers. In case of LoRaWAN, the gateways were built based on a Raspberry PI and several different LoRaWAN Concentrators. Figure 2 shows a prototype of a LoRaWAN gateway with an iC880A - LoRaWAN® Concentrator 868MHz of IMST GmbH. The design is based on the GitHub project of Gonzalo Casas [10].

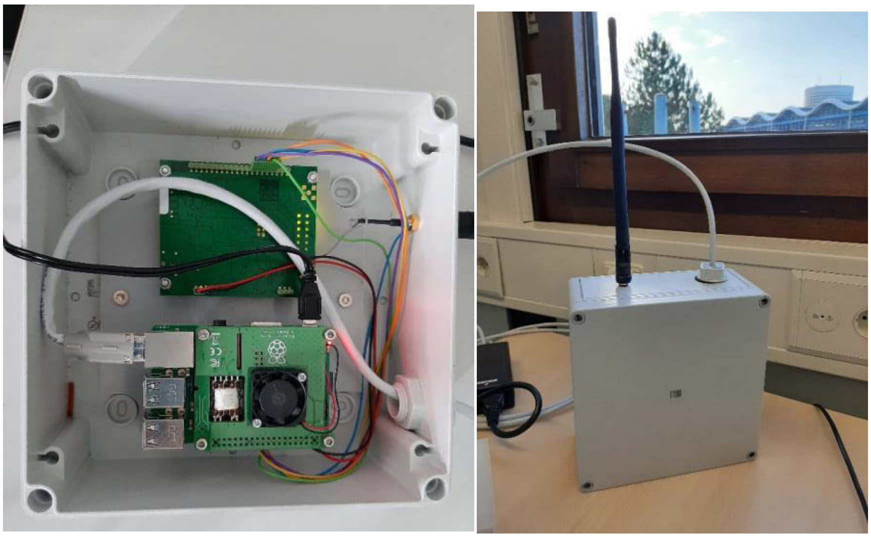


Figure 2: Prototype of LoRaWAN Gateway

While it was possible to use standard components for the design of the gateway, for the design of the field units, special printed circuit boards (PCB) had to be developed. The field units are directly connected to the devices or sensors and are responsible to transmit measurement data or receive commands for controlling actuators.

Figure 3 shows, as an example, a field unit of the first generation which can be used for pressure measurements. On the left side, the microcontroller is located. For the first prototype a Heltec ESP32WiFi LoRa32 (V2) module was installed. This module offers the possibility to use either LoRaWAN or WiFi for the communication with the IoT platform. The microcontroller is programmed in C using the Arduino IDE. On right side, exchangeable PCB modules can be installed according to the application of the field unit. Several options are available. There are modules with an analog to digital converter (ADC) for directly connecting sensors with a voltage output of 0-5 V or 0-10 V. In the later case, a voltage divider is integrated into the PCB of the ADC. For sensors with current output, an additional PCB with a current-voltage converter can be installed.

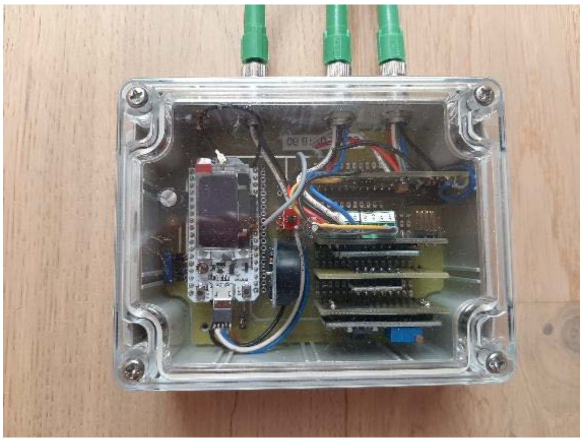


Figure 3: IoT.H2O Field Unit

Currently, we are working on the development of a second prototype where we want to replace the Heltec ESP32 Development Kit by a bare Espressif ESP32/WROOM microcontroller module and a separate LoRa modem in order to improve the energy consumption of the device and extend the timespan of operation when the device is battery powered.

Central piece of the IoT field units is the microcontroller (MCU). The design provides for different low-power MCUs, however main focus is on Espressif's ESP32 MCU family. Due to the ESP32 relatively high power consumption for battery operated devices, the MCU is put in "deep sleep mode" or disconnected from power supply during idle phases in order to extend battery lifetime. For connection to the LoRaWAN network a LoRa radio module is implemented. The IoT.H2O project team decided to implement time synchronization for all actions (measurements, actuator operation) of the IoT devices within the network. For WiFi connected networks, the synchronization can be achieved by a "system clock signal" broadcasted to all network participants, time granularity can be set to low values in the range of seconds. LoRa systems transmit in a "free" industrial, scientific and medical (ISM) radio band and must follow technical and legal requirements demanding special provisions to guarantee synchronous operation. Most important is the inclusion of a high precision real-time clock (RTC) in the hardware design. Even with best available RTCs additional clock adjustment over the network must be implemented. A 60 seconds time granularity for the IoT.H2O LoRaWAN devices was established.

Most commercially available sensors for the water industry focus on the classic automation sector. In practical, these sensors are designed for 4-20mA current loop, 0-5V or 0-10V interfaces. Battery powered systems preferentially operate in lower voltage ranges, e.g. 3.6 V. For this reason, the IoT.H2O Field units are provided with multi voltage capacities including voltage booster system (from voltages down to 2 V up to 10 V) and additional "level shifter" for in-system data communication.

In order to reduce power consumption in addition to MCU deep sleep or power off, "load switches" are used to cut power supply to all peripheral electronic components when not in use. A functional block diagram of the second-generation field unit which will be manufactured shortly is depicted in Figure 4.

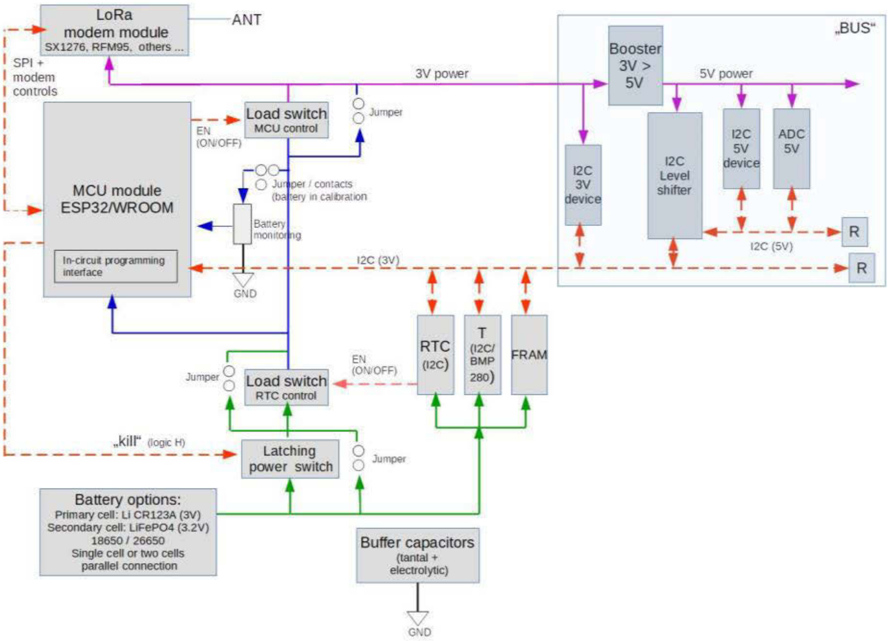

LoRa modem module
SX1276, RFM95, others ...
ANT
SPI + modem controls
3V power
MCU module ESP32/WROOM
In-circuit programming interface
EN (ON/OFF)
Load switch
MCU control
Jumper
Jumper / contacts (battery in calibration
Battery monitoring
GND
I2C (3V)
„BUS"
Booster 3V > 5V
5V power
I2C 3V device
I2C Level shifter
I2C 5V device
ADC 5V
R
I2C (5V)
R
RTC (I2C)
T (I2C/ BMP 280)
FRAM
EN (ON/OFF)
Jumper
Load switch
RTC control
„kill" (logic H)
Latching power switch
Jumper
Battery options:
Primary cell: Li CR123A (3V)
Secondary cell: LiFePO4 (3.2V)
18650 / 26650
Single cell or two cells
parallel connection
Buffer capacitors
(tantal + electrolytic)
GND

Figure 4: Functional block diagram IoT.H2O Field Unit V2.0

CC BY NC SA

With the field units, every minute the data is sent by LoRaWAN to the Gateway which is connected through the internet with a LoRaWAN network server. In IoT.H2O, the “The Things Stack” LoRaWAN server (V3) is used. It is operated by The Things Network (TTN) providing an open network. Everyone is able to register gateways and devices under the premise that all gateways can be used by the community. In the IoT.H2O project, a private server based on the community edition of The Things Stack has also been installed on a local machine and operates a completely independent network.

The LoRaWAN server does not store the data permanently. For this reason, the data has to be collected from the server. In our case, we established a MQTT connection to the LoRaWAN server using the graphical programming tool Node-RED. Then, in Node-RED the LoRaWAN data is converted to a format which can be sent as JSON message by MQTT to the IoT platform.

An important advantage of LoRaWAN is that it is possible to send data over large distances. So far, with the IoT.H2O field units, we were able to send data over a distance of 23km. The range strongly depends on the topography of the landscape. If there is a direct line of sight the LoRaWAN range should be even larger. Hills and buildings block the LoRaWAN signal. Data can be transmitted over hills, but in valleys there is no coverage. In this case additional gateways have to be installed. In urban areas we achieved a coverage of 1.4 km and the LoRaWAN signal can pass several buildings. The range is also dependent on the combination of antennas which are used for gateway and IoT device [11].

3 USE CASES

3.1 Monitoring and control of pumps

The development and testing of the IoT devices was first performed on a laboratory scale. An already existing pump test rig was equipped with the IoT devices. With these devices, it is possible to change the control valve position and set the speed of the pump through the IoT platform. Also, the data of the sensors for measuring pump performance are collected by IoT devices and can be monitored in the IoT platform.

Figure 5 (left) shows a 3D CAD model of the test rig. All input values are indicated with blue arrows, all measurement values with red arrows. The measurement data are acquired with the IoT devices. Then the data is sent either by LoRaWAN or by a WiFi connection to the IoT platform. In the laboratory application we rely on WiFi, since then we can achieve a higher temporal resolution of the measurements.

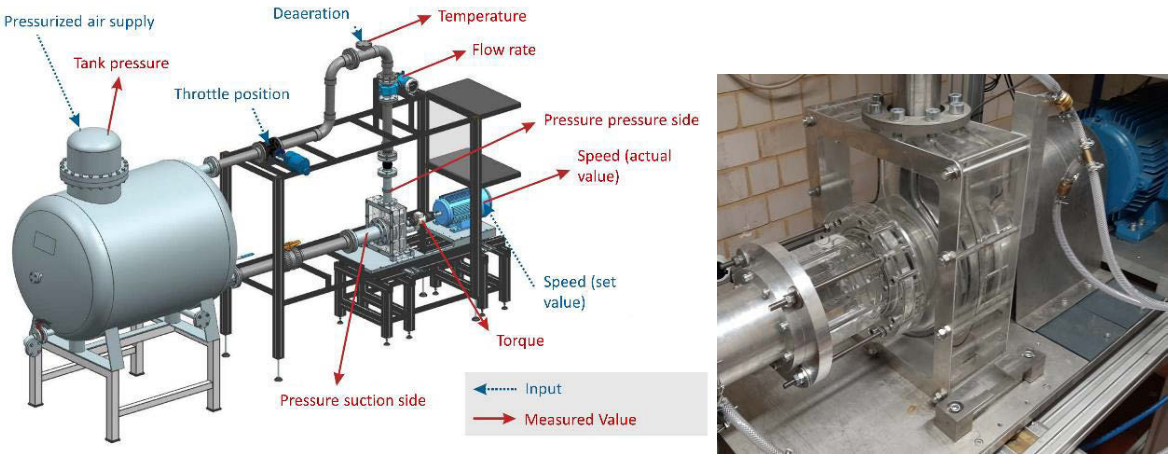
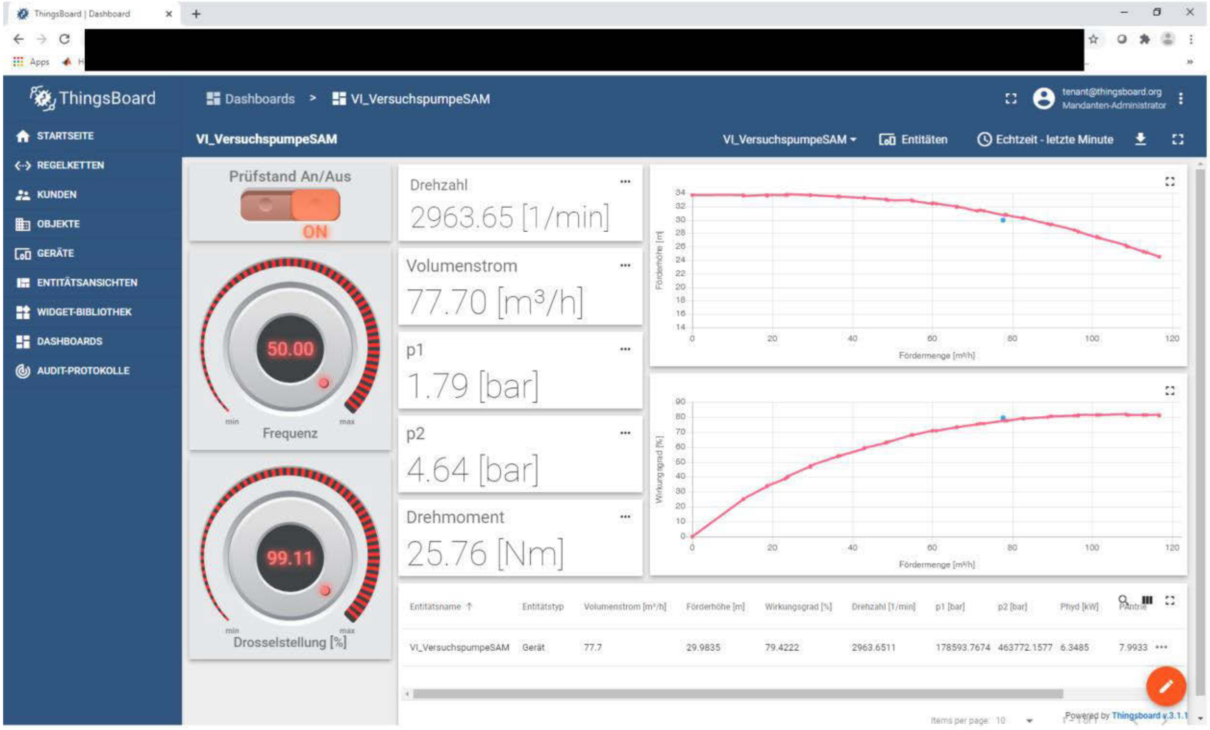


Figure 5: CAD model of pump test rig [12] and centrifugal pump

The centrifugal pump was developed at the Institute for Fluid Mechanics and Turbomachinery of the Technical University of Kaiserslautern and manufactured by rapid prototyping. The casing is made of acrylic for flow visualization experiments (Figure 5 (right)).

For controlling the test rig, a 16-bit digital to analog converter is connected using the I2C interface to an ESP32 microcontroller. With this setup it is possible to send a signal between 4 to 20mA to the control valve or the frequency converter, respectively for defining the valve position and the pump speed.

The test rig can be operated by a dashboard which is defined in Thingsboard (Figure 6). The test rig can be started with the on/off switch in the upper left part of the dashboard. The settings for control valve position and the frequency converter are applied by control knobs. The communication is based on RPC calls. The measured values are shown on cards and in a table. The chart widgets show the current operating point of the pump on the head and on the efficiency curve. Usually, in Thingsboard all data is plotted versus time in the predefined dashboard widgets. We modified the chart widget by JavaScript programming so that it is possible to plot any other quantity on the x-axis.


[redacted]
ThingsBoard | Dashboard
Apps
ThingsBoard
Dashboards > VI_VersuchspumpeSAM
tenant@thingsboard.org
Mandanten-Administrator
STARTSEITE
REGELKETTEN
KUNDEN
OBJEKTE
GERÄTE
ENTITÄTSANSICHTEN
WIDGET-BIBLIOTHEK
DASHBOARDS
AUDIT-PROTOKOLLE
VI_VersuchspumpeSAM
VI_VersuchspumpeSAM
Entitäten
Echtzeit - letzte Minute
Prüfstand An/Aus
ON
50.00
min
max
Frequenz
99.11
min
max
Drosselstellung [%]
Förderhöhe [m]
Fördermenge [m³/h]
Wirkungsgrad [%]
Fördermenge [m³/h]
Items per page: 10
Powered by Thingsboard v.3.1.1

Drehzahl
2963.65 [1/min]

Volumenstrom
77.70 [m³/h]

p1
1.79 [bar]

p2
4.64 [bar]

Drehmoment
25.76 [Nm]

Entitätsname ↑	Entitätstyp	Volumenstrom [m³/h]	Förderhöhe [m]	Wirkungsgrad [%]	Drehzahl [1/min]	p1 [bar]	p2 [bar]	Phyd [kW]	PAntrie
VI_VersuchspumpeSAM	Gerät	77.7	29.9835	79.4222	2963.6511	178593.7674	463772.1577	6.3485	7.9933

Figure 6: Dashboard for pump operation

The IoT controlled test rig is also used in laboratory classes for mechanical engineering students. Because of the COVID19 pandemic, the access of the students to the test rig was restricted. But with the IoT system, they were able to run their test online in a web browser. Because of safety issues they are not operating the real test rig. For the laboratory classes a virtual model of the test rig is operated based on a script running on the programming and numeric computing platform MATLAB. With these virtual tests, the students can measure head and efficiency curves at different speeds and investigate the behavior of pumps in different systems and with different control strategies. Another advantage of the virtual model is that different pumps can be investigated by just exchanging the pump characteristic curves in the software model. The biggest benefit however is, that now each student is able to run the test by himself. Before, the tests were run in groups because of time constraints. In the future the students will run the virtual test and work

BY NC SA

on the real test rig to ensure that they also get an understanding of the test rig setup and instrumentation. The design of the virtual test rig setup is described in detail in [13].

3.2 Field deployment in a water distribution system

After the first successful tests in the laboratory, prototypes of the IoT devices were installed in the water system of the “Zweckverband für Wasserversorgung Germersheimer Südgruppe” (ZWGS) in Germany. The ZWGS is supplying 13 small townships in the southwest of Germany with a total of 60000 inhabitants and, in addition, an industrial complex. In the network with a length of 330 km, two water utilities, a booster station and three water towers are operated.

One of the test devices is measuring the pressure of a well pump, a second device is used for measuring the pressure at the suction and pressure side of a water distribution pump (Figure 7). The devices are in continuous operation sending data at a time interval of one minute since 2020. Over time, more devices were installed for measuring pressure at two other well pumps and recently, a device was installed at the water tower Rülzheim for acquiring the water level. For the later one, the system is used as backup for an existing system where the data is transmitted by a mobile network which has been lately subject to several interruptions.

The data of the devices are currently sent to the LoRaWAN network server operated by The Things Network. From there the data is collected using Node-RED as described in section 2.2. As an example, for the visualization in Thingsboard, in Figure 8 the dashboard for a well pump is shown. On the upper left the time series of the pressure is plotted. In the table below, the pressure readings are listed for the past 15 minutes and in the OpenStreetMap widget provided by Thingsboard on the right, the locations of water utilities and water towers are shown.

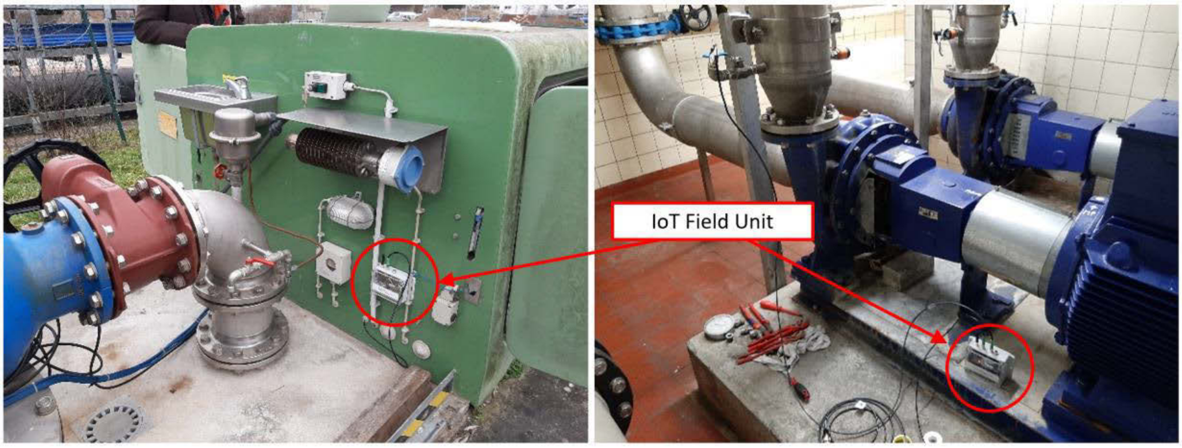
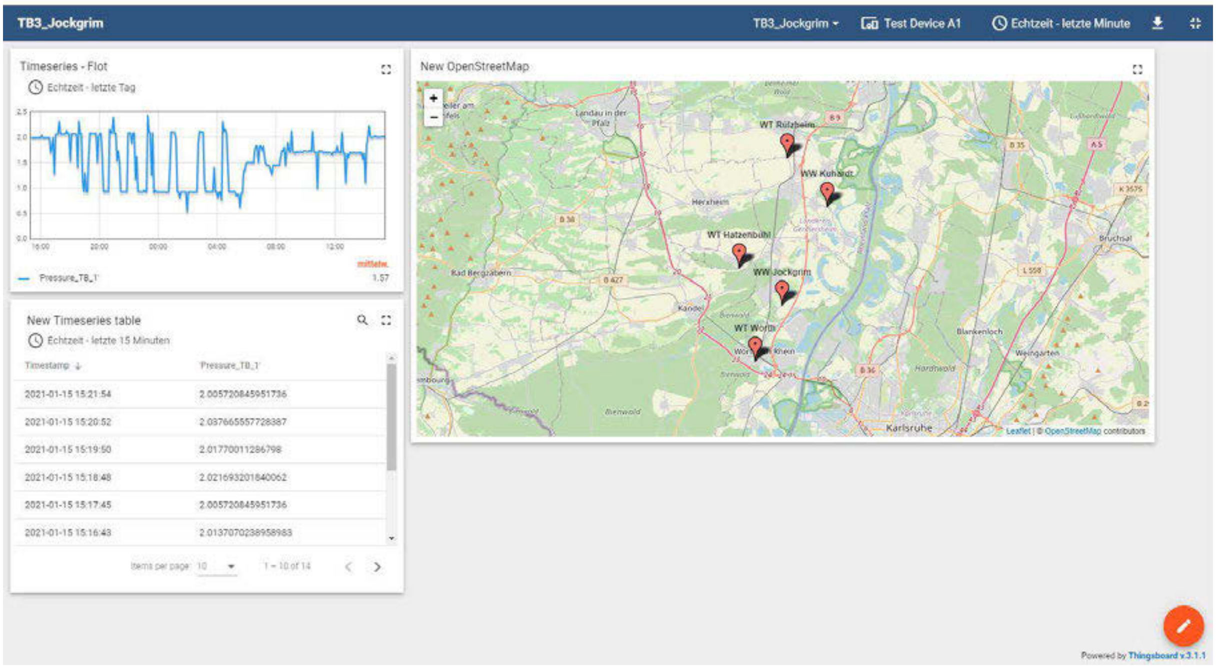


Figure 7: Test installations of IoT field units at a well pump (left) and a water distribution pump (right)

Usually, in existing water distribution systems there are already many sensors installed. These sensors are not IoT-ready. In order to receive a full set of data from the ZWGS system, data from the SCADA system is also sent to the IoT system by connecting the SCADA to the IoT platform. In this case a C-code running on a Raspberry PI was developed with support of the system operator. The program uses the library snap7 which is open source and supports a direct communication with the Siemens S7 Programmable Logic Controller (PLC) which is installed in the system. With the C-code, the data is read from the S7 PLC every minute and sent to Thingsboard as a JSON message by MQTT. There, the data is stored in a database and visualized in a dashboard. Figure 9 shows the SCADA data for the measurements of flow rate at two water utilities (upper left), the water level in the three water towers (middle), the flow rate and pressure of a booster station (upper right) and the pressure at the discharge pipe of a water utility (lower left).

CC BY NC SA


TB3_Jockgrim
TB3_Jockgrim
Test Device A1
Echtzeit - letzte Minute
Timeseries - Flot
Echtzeit - letzte Tag
16:00
20:00
00:00
04:00
08:00
12:00
mittelw.
Pressure_TB_1'
1.57
New OpenStreetMap
WT Rülzheim
WW Kuhardt
WT Hatzenbühl
WW Jockgrim
WT Wörth
Karlsruhe
Leaflet | © OpenStreetMap contributors
Powered by Thingsboard v.3.1.1

New Timeseries table

Echtzeit - letzte 15 Minuten

Timestamp ↓	'Pressure_TB_1'
2021-01-15 15:21:54	2.005720845951736
2021-01-15 15:20:52	2.037665557728387
2021-01-15 15:19:50	2.01770011286798
2021-01-15 15:18:48	2.021693201840062
2021-01-15 15:17:45	2.005720845951736
2021-01-15 15:16:43	2.0137070238958983

Items per page: 10 1 – 10 of 14

Figure 8: Visualization of LoRaWAN data of a well pump

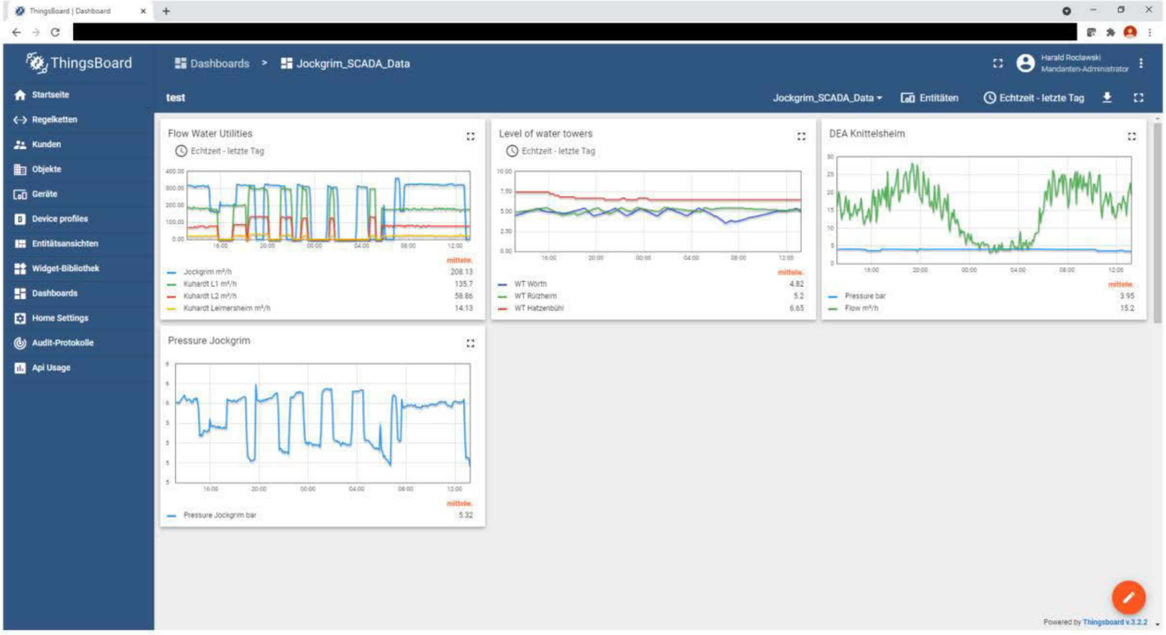

ThingsBoard | Dashboard
ThingsBoard
Dashboards
Jockgrim_SCADA_Data
Harald Roclawski
Mandanten-Administrator
Startseite
Regelketten
Kunden
Objekte
Geräte
Device profiles
Entitätsansichten
Widget-Bibliothek
Dashboards
Home Settings
Audit-Protokolle
Api Usage
test
Jockgrim_SCADA_Data
Entitäten
Echtzeit - letzte Tag
Flow Water Utilities
Echtzeit - letzte Tag
mittelw.
Jockgrim m³/h 208.13
Kuhardt L1 m³/h 135.7
Kuhardt L2 m³/h 58.86
Kuhardt Leimersheim m³/h 14.13
Level of water towers
Echtzeit - letzte Tag
mittelw.
WT Wörth 4.82
WT Rülzheim 5.2
WT Hatzenbühl 6.65
DEA Knittelsheim
mittelw.
Pressure bar 3.95
Flow m³/h 15.2
Pressure Jockgrim
mittelw.
Pressure Jockgrim bar 5.32
16:00
20:00
00:00
04:00
08:00
12:00
Powered by Thingsboard v.3.2.2

Figure 9: Visualization of SCADA of ZWGS in a dashboard

By this method already existing data can be included in the IoT platform. The existing dataset can then be easily extended using new IoT devices. This is especially useful in case of the ZWGS system since the current SCADA system does not allow to include more sensors without major changes in the hardware. It is also possible to combine this complete dataset with data from simulation software e.g. with results from network simulations. This is described in the next section.

BY NC SA

3.3 Integration of simulation software to the IoT platform

The RPC functionality of Thingsboard allows the execution of an arbitrary code on any other machine which is connected by MQTT to Thingsboard. So, as a consequence, it is possible to start network simulations of water distribution systems with Thingsboard. Live data from the IoT system can be used as boundary condition and the simulation results can then be sent back to the IoT platform for comparison with the measured data and detection of abnormal system behavior.

For this objective, a method was developed on a physical model of a water distribution system of our project partner from The Centro de Pesquisas Hidráulicas e Recursos Hídricos of the Federal University of Minas Gerais, Brazil (CPH). The test rig is depicted in Figure 10 in the picture on the lower left. On each node of the network pressure sensors and flow meters are installed. In addition, it is possible to simulate leakages by opening valves. All measurement data is acquired using IoT technology developed in the IoT. H2O project and sent by MQTT to a local Thingsboard server for data storage and visualization. Because of the decentralized Internet of Things approach it is also possible to operate the Thingsboard server and run the network simulations in completely different locations as long as they are accessible through the Internet. The test rig of the water system is described in more detail in [14]. In IoT.H2O, the water network modelling software EPANET is used. Figure 10 shows a dashboard for combining data from EPANET simulations and measurements. The test rig layout is represented by a JPEG image in the background of the dashboard. Then the location of each pressure sensor was added as a red marker on the dashboard. Each sensor is defined as device in Thingsboard.

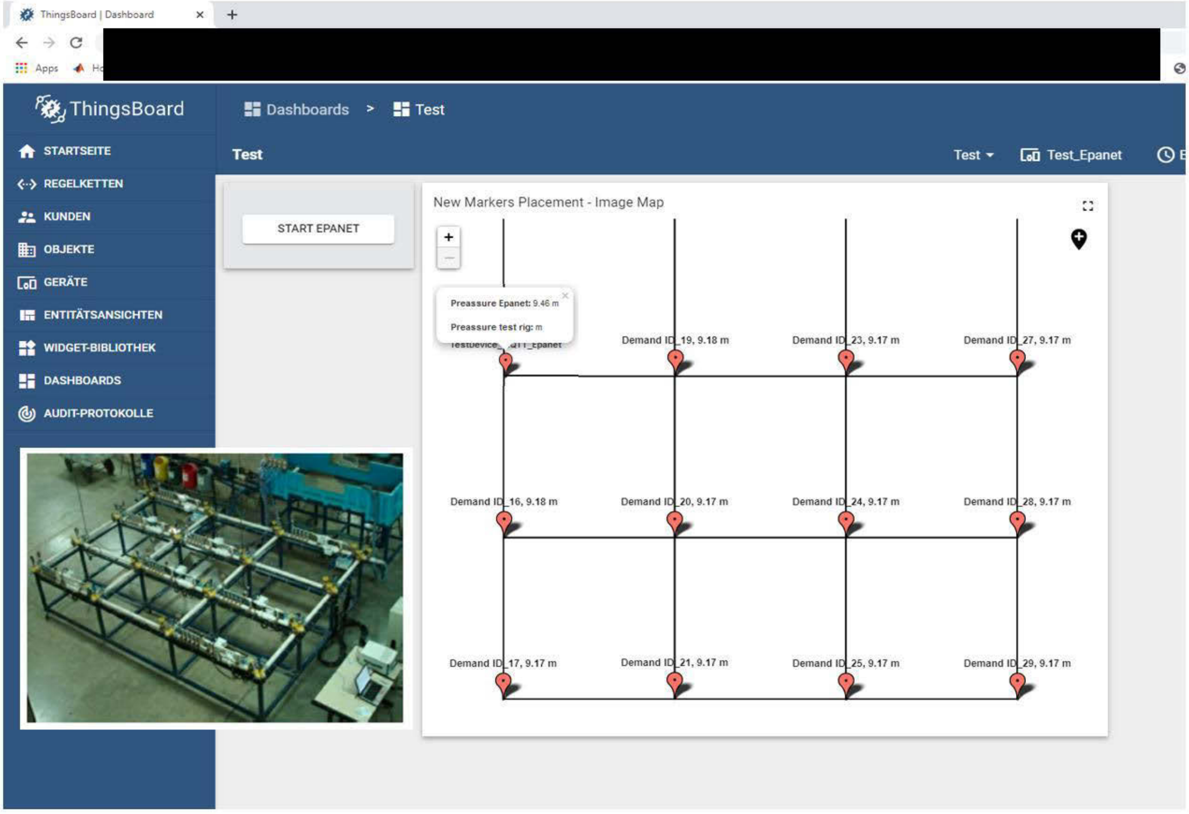

ThingsBoard | Dashboard
[redacted]
Apps
ThingsBoard
Dashboards > Test
Test
Test
Test_Epanet
STARTSEITE
REGELKETTEN
KUNDEN
OBJEKTE
GERÄTE
ENTITÄTSANSICHTEN
WIDGET-BIBLIOTHEK
DASHBOARDS
AUDIT-PROTOKOLLE
START EPANET
New Markers Placement - Image Map
Preassure Epanet: 9.46 m
Preassure test rig: m
Demand ID_19, 9.18 m
Demand ID_23, 9.17 m
Demand ID_27, 9.17 m
Demand ID_16, 9.18 m
Demand ID_20, 9.17 m
Demand ID_24, 9.17 m
Demand ID_28, 9.17 m
Demand ID_17, 9.17 m
Demand ID_21, 9.17 m
Demand ID_25, 9.17 m
Demand ID_29, 9.17 m

Figure 10: Test stand for water networks (Federal University of Minas Gerais, Brazil)

Currently, the EPANET simulation is started manually by clicking on the RPC button “START EPANET” on the upper left part of the dashboard. By clicking the button, an MQTT message is sent to a python script. The Python script is continuously checking if the button was pressed. Then the EPANET simulation is started from the Python script by executing the Windows executable of

CC BY NC SA

EPANET using the Windows command line. At the end, the EPANET output is evaluated with Python and the pressure at each node is returned to Thingsboard. In this case both values of measurement and simulation are shown in a card in the Dashboard. Other forms of visualization in tables or charts are possible.

Currently the EPANET Programmer's Toolkit and a JavaScript version of EPANET called epanet-js [15] is tested for replacing the Windows executable so that the boundary conditions can be edited dynamically according to the live data from the water system of ZWGS. Epanet-js has the advantage that a graphical representation of the network can be included in the Thingsboard dashboard (Figure 11). It is our vision to be able to represent the current status of the system by the EPANET simulations for improving the monitoring of the system. This requires in the future the installation of additional sensors in the network to improve the accuracy of the model. For testing this method, we plan to install temporarily pressure sensors in a small town of the network and use data of flow meters which are already permanently installed at all pipes supplying the town.

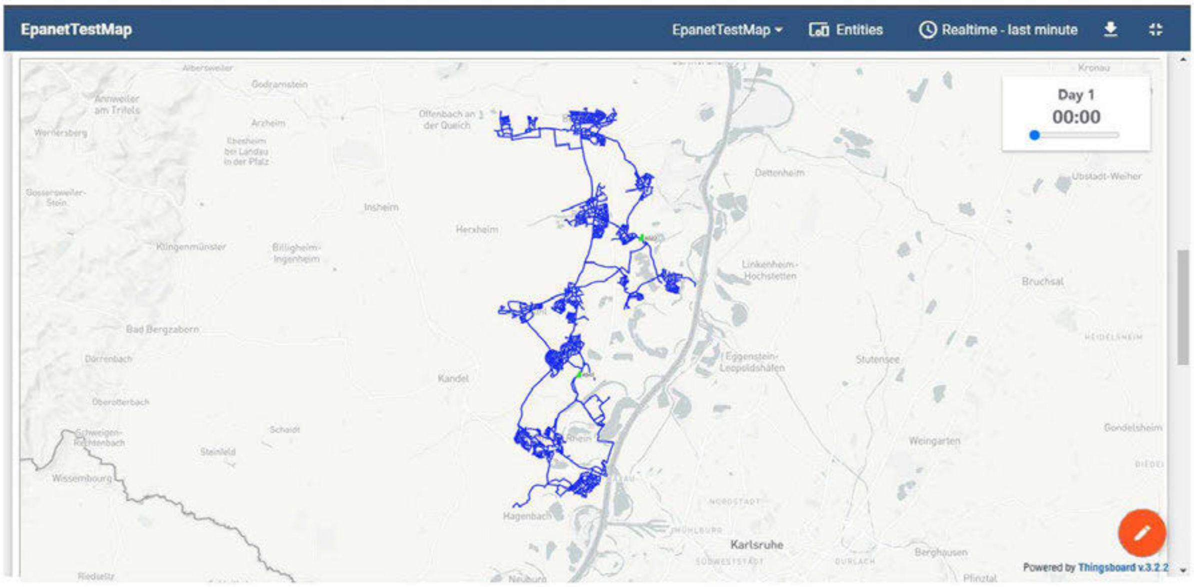


Figure 11: Representation of the network of ZWGS in Thingsboard using EPANETJS

For combining the data from SCADA system, IoT field units and network simulation a main dashboard was created (Figure 12). The main dashboard shows the current levels of the three water towers, the current flow rate of the water utilities and the booster station, the current water consumption, the status of the well and distribution pumps and a map of the network.

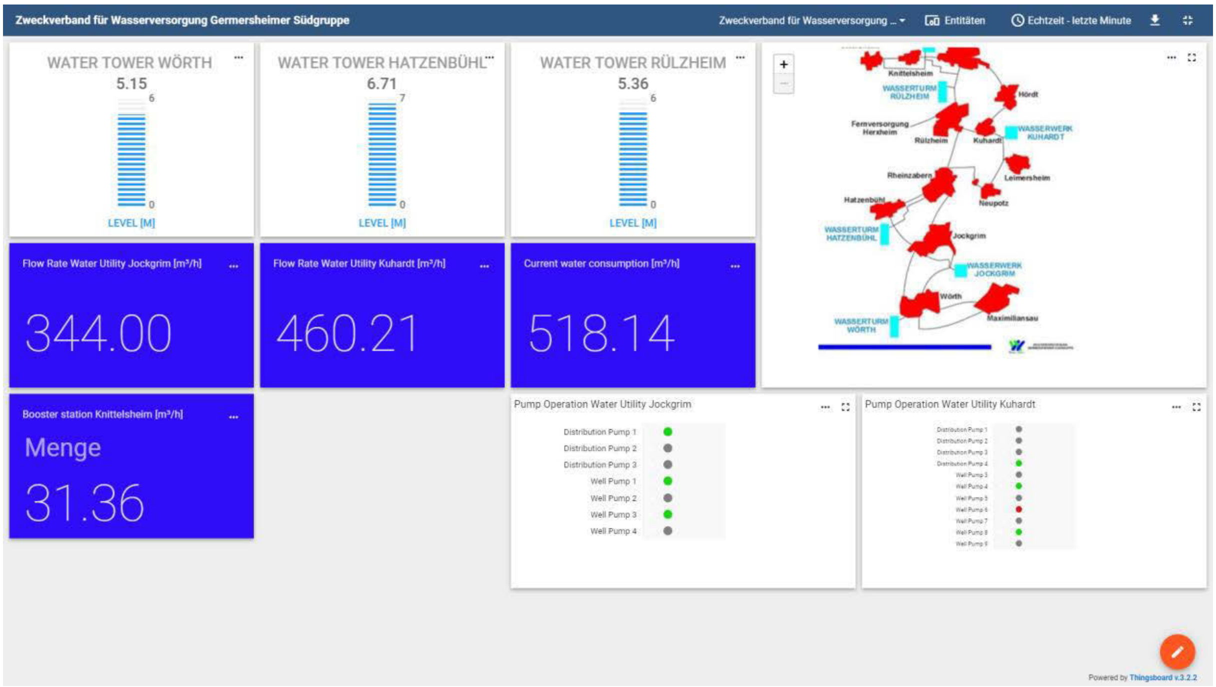

Zweckverband für Wasserversorgung Germersheimer Südgruppe
Zweckverband für Wasserversorgung ...
Entitäten
Echtzeit - letzte Minute
WATER TOWER WÖRTH
5.15
LEVEL [M]
WATER TOWER HATZENBÜHL
6.71
WATER TOWER RÜLZHEIM
5.36
Flow Rate Water Utility Jockgrim [m³/h]
344.00
Flow Rate Water Utility Kuhardt [m³/h]
460.21
Current water consumption [m³/h]
518.14
Booster station Knittelsheim [m³/h]
Menge
31.36
Pump Operation Water Utility Jockgrim
Distribution Pump 1
Distribution Pump 2
Distribution Pump 3
Well Pump 1
Well Pump 2
Well Pump 3
Well Pump 4
Pump Operation Water Utility Kuhardt
Knittelsheim
WASSERTURM RÜLZHEIM
Hördt
Fernversorgung Herxheim
Rülzheim
Kuhardt
WASSERWERK KUHARDT
Rheinzabern
Leimersheim
Hatzenbühl
Neupotz
WASSERTURM HATZENBÜHL
Jockgrim
WASSERWERK JOCKGRIM
Wörth
Maximiliansau
WASSERTURM WÖRTH
Powered by Thingsboard v.3.2.2

Figure 12: Main dashboard for ZWGS water distribution system

The current water consumption is actually not measured. Instead it can be calculated from the flow rates of the water utilities, the change in water level of the water towers and the flow rate of the booster station. These calculations are performed in Thingsboard using the rule engine. By clicking on the three dots, which are shown in the upper right corner of each dashboard widget, it is possible to get more detailed information by navigating to additional dashboards.

In case of the water towers, an additional dashboard is shown, showing charts with time series of water levels and in- and outflow as well as cards with the current values of these quantities. For the flow rate cards, also time series of these values are plotted and additional values as e.g. pressure of the water utilities are shown as in Figure 9. By clicking on the pump status, it is planned to show dashboards with the current operating data and plot the operating point on the head and efficiency curve similar to Figure 6. This work is currently in progress. By clicking on the map widget, the dashboard related to the network simulation (Figure 11) is shown.

4 CONCLUSION

An IoT system based on open-source software and cost-effective hardware has been presented. It was shown that with the IoT platform Thingsboard data from different systems can be easily combined. For the communication with the IoT platform, IoT devices were developed which can be connected to sensors with voltage or current output and send the measurements by LoRaWAN or WiFi to the IoT platform using the MQTT protocol. In case of LoRaWAN, Gateways based on a Raspberry PI are used.

It was demonstrated that it is possible to control and monitor a pump test rig with the IoT devices. The system can also be used for educating students in pump operation.

In field tests, the IoT devices were installed at pumps and a water tower in a water distribution system in Germany. By including the data of the SCADA system into the IoT platform, already existing sensors can be used to combine the data from IoT devices with the SCADA data. With the IoT devices it is possible to extend the number of sensors in a cost-effective way. Also, it was shown that simulation tools like EPANET can be included in the IoT platform. The method was developed on model of a water distribution network on a laboratory scale. For the ZWGS network, it is planned to install additional sensors within the network for improving the model quality in the future. We want to be able to analyze the current status of the network by EPANET simulations and to detect abnormal system behavior at an early stage by comparing simulation data and measurements.

All data, from SCADA system, IoT devices and network simulation, can be organized in different dashboards. Starting from a main dashboard, users can navigate to sub-dashboards for getting more detailed information on individual devices.

The IoT system offers the possibility to combine data from different systems and provide the user a complete and detailed set of data for analyzing system operation. By using the rule engine functionality, the incoming data can be automatically evaluated and processed for calculating other quantities. It is not necessary to export data to a third-party software which ensures that important quantities can be monitored and evaluated continuously.

5 ACKNOWLEDGMENTS

We would like to thank our project partners Aloysio P. M. Saliba (The Centro de Pesquisas Hidráulicas e Recursos Hídricos of the Federal University of Minas Gerais, Brazil), Benjamin Dewals (Research group Hydraulics in Environmental and Civil Engineering, University of Liege, Belgium), Thomas Pirard (Research group Hydraulics in Environmental and Civil Engineering, University of Liege, Belgium), Laurent Vercouter (LITIS LAB, MIND Group, INSA Rouen France) and Henrique Donancio Nunes Rodrigues (LITIS LAB, MIND Group, INSA Rouen France) for their comments and suggestions regarding this work. The authors acknowledge Minas Gerais State Agency for Research and Development – FAPEMIG, Brasil, Federal Ministry of Education and Research of Germany, Agence Nationale de la Recherce de France and Fonds de la Recherce Scientifique Belge, for funding this research by the project IoT.H2O on the IC4Water JPI call. We would like to express our gratitude to Ralf Friedmann (ZWGS) and Markus Justen (ZWGS) for the possibility to realize the field tests in the water distribution system of ZWGS.

6 REFERENCES

[1] F. Ramos da Fonseca, Modelo de Sistema de automação aplicado a setorização de redes de abastecimento hídrico. Tese de doutorado., Escola Politécnica da Universidade de São Paulo, 2011.

[2] A. S. Spolaor, „Automação nos sistemas de abastecimento de água. Caso do controle da reservação de distribuição. Tese de mestrado.," 2001.

[3] M. P. dos Santos und J. K. C. de Oliveira, „Automação de baixo custo para reservatórios de água.," in Revista Princípia N°. 25, João Pessoa, 2014.

[4] S. Brodersen, „Lebenszykluskosten für Förderanlagen – Auswahlkriterien für wirtschaftliche Entscheidungen," in wat2006, Berlin, 2006.

[5] Grundfos GmbH, „Presseinformation Energieeinsparung," 2017.

[6] J. G. Natividad und T. D. Palaoag, „IoT based model for monitoring and controlling," IOP Conf. Ser.: Mater. Sci. Eng. 482 012045, 2019.

[7] J. Pérez-Padillo, J. García Morillo, J. Ramirez-Faz, M. Torres Roldán und P. Montesinos, „Design and Implementation of a Pressure Monitoring System Based on IoT for Water Supply Networks.," Sensors, 20, 4247. https://doi.org/10.3390/s20154247, 2020.

[8] A. C. Tasong und R. P. Abao, „Design and Development of an IoT Application with Visual Analytics for Water Consumption Monitoring".Procedia Computer Science, Volume 157, 2019, Pages 205-213,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2019.08.159..

[9] R. Gonçalves, J. J. M. Soares und R. M. F. Lima, „An IoT-Based Framework for SmartWater Supply,“ Future Internet , Bd. 114, Nr. https://doi.org/10.3390/fi12070114, p. 12, 2020.

[10] G. Casas, „From zero to LoRaWAN in a weekend,“ 06 11 2018. [Online]. Available: https://github.com/ttn-zh/ic880a-gateway/wiki. [Zugriff am 10 05 2022].

[11] H. Roclawski, T. Krätzig, L. Sterle und M. Böhle, „Influence on antenna design on range of LoraWAN devices – a practical test,“ in EGU General Assembly, online, 19–30 Apr 2021, EGU21-458, 2021.

[12] A. Knapp, Numerische und experimentelle Untersuchung der Auswirkungen von Auslegungsmethoden für Spiralgehäuse auf die Performance einer Kreiselpumpe, Düren: Shaker Verlag, 2019.

[13] H. Roclawski, L. Sterle und M. Böhle, „VIRTUAL CENTRIFUGAL PUMP TEST RIG FOR LABORATORY CLASSES BASED ON IOT TECHNOLOGY,“ in Proceedings of ASME Turbo Expo 2022 Turbomachinery Technical Conference and Exposition GT2022, Rotterdam, Netherlands, 2022.

[14] D. Moraes, L. Sterle, A. Saliba, G. M. LIMA, H. Roclawski, T. Kratzig, B. Dewals und L. Vercouter, „LOW-COST MONITORING FOR WATER DISTRIBUTION SYSTEMS,“ in XXIV SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS, Belo Horizonte, Brazil, 2021.

[15] „Water distribution modelling in the browser using JavaScript,“ [Online]. Available: https://epanetjs.com/. [Zugriff am 12 05 2022].

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14789

GREY-BOX DYNAMIC MODEL OF A DRINKING WATER TREATMENT PLANT

Blaž Korotaj[1], Hrvoje Novak[2] and Mario Vašak[3]

[1,2,3]University of Zagreb, Faculty of Electrical Engineering and Computing, Laboratory for Renewable Energy Systems (url: www.lares.fer.hr), Unska 3, HR-10000, Zagreb, Croatia

[1] *blaz.korotaj@fer.hr*, [2] *hrvoje.novak@fer.hr*, [3] *mario.vasak@fer.hr*

Abstract

In the paper a concept of a general reaction volume in a drinking water treatment plant (DWTP) is introduced, where several such volumes need to be stacked in series with corresponding input chemicals used in them to finally represent the entire plant. A general case in which a reaction volume is not fixed is elaborated. The overall derived plant model is of a grey-box form with a small number of tunable parameters, which is beneficial for adaptation to a particular DWTP based on its laboratory and on-line measured data.

The developed mathematical models of reaction volumes are stacked and inputs in different volumes selected such that they represent a concrete DWTP for which a decision-support system is being developed. A preliminary validation is performed against laboratory-determined plant input-output data.

Keywords
DWTP, reaction volume, fast reactions, dynamical model.

1 INTRODUCTION

The extremes of raw water shortage and abundance tend to raise frequency due to climate change effects, with the corresponding chemical content considerably varying over shortening time periods. On the other hand, the drinking water produced from raw water must meet the dynamic quantity requirements of water distribution systems and at the same time the quality requirements that are getting stricter and stricter. This makes the drinking water production in drinking water treatment plants (DWTPs) a dynamically demanding process in which the operators' experience for running the plant, related often to steady-state recipes, should be additionally supported for a more resilient water service. The DWTP management task is further aggravated with non-availability of on-line sensory data for different water quality parameters in the plant. A first-hand solution to the raised situation comes in a form of a decision-support system for the DWTP operation relied on model predictive control (MPC) principles, while in a long-run it should evolve to a closed-loop optimal control system where the operators are left out of the control loop and just have a supervisory and monitoring role. In order to create a decision-support system for highly dynamical plant conditions described above and also enable soft-sensing for plant monitoring, a dynamical model of the DWTP is needed.

Different software tools exist to relate raw and treated water quality parameters for a DWTP. A very popular one is EPA WTP [1] which in the back-end uses static or quasi-dynamic empirical models, recently enriched also with stochastic modelling to represent relations between spreads of inputs and outputs. There are also attempts to build-up detailed dynamical models for specific parts of a DWTP with a high number of parameters that require complex tuning [2]. The framework of Mixed Logical Dynamical (MLD) systems [3] is also tried for modeling and control of a DWTP [4]. The main drawback here is the computation time.

Significant research focus is put on making DWTP models using artificial neural networks [5], [6]. The main disadvantage in that case is ensuring the appropriate structure of the network and of course the need for a large amount of process data for parameters tuning. Dynamical model and system identification of coagulant dosage system for DWTPs have been presented as well [7]. In [7] the model is based only on mixing different concentrations of elements entering the tank, while neglecting a very important issue which are time-transients in concentration due to chemical reactions that take place. There exists a lack of systematic approaches to modelling the entire DWTP with a simply parametrizable dynamic model and this article stems from the intention to fill this gap and enable dynamic decision-support for DWTPs based on MPC.

In this paper, a major simplifying assumption is that all chemical reactions that take place in DWTP reaction volumes are performed instantaneously when all reactants are present, in the amount dictated by the scarcest reactant that is annihilated. The dynamic DWTP model consists of a series of generalized reaction volumes where the volume can be fixed or varying. While deriving a dynamic model for a single generalized reaction volume, the number of tuneable parameters is kept low by exploiting mathematical descriptions of the underlying physical and chemical processes (grey-box approach). The DWTP considered in this paper consists of 4 reaction volumes, but the modelling framework is readily adaptable to different numbers of reaction volumes. In the considered DWTP model, the pre-oxidation processes take place in the first and second reaction volumes, the third reaction volume is used for coagulation, flocculation, sedimentation and filtering processes, while the fourth reaction volume is used for final chlorination and pH correction. The model is created and tested in the Matlab/Simulink software environment. For the considered DWTP in Spain, the values of the water quality variables simulated by the model at the outlet of the final reaction volume are compared with the actual laboratory measurements for a preliminary validation.

The paper is structured as follows: Section 2 describes the modelling of a DWTP reaction volume; the dynamical model for the full plant is set in Section 3, while Section 4 reports the simulation results in Matlab/Simulink and the laboratory data comparison.

2 MODELLING OF A DRINKING WATER TREATMENT PLANT

Drinking water treatment plants (DWTPs) receive raw water and perform necessary chemical and mechanical treatment upon it to achieve drinking water standards. The focus of the paper is put on the chemical treatments in different reaction volumes of DWTPs. Thus, the whole model includes the models of individual reaction volumes as building elements in which different chemical reactions of the water treatment plant occur.

2.1 Model for fast reactions with two reactants

Before modelling the concrete chemistry of a DWTP, the extensively used modelling framework of fast chemical reactions is introduced.

Assume that the following chemical reaction is given:

$$a\mathrm{A} + b\mathrm{B} \rightarrow \cdots \tag{1}$$

where a and b are stoichiometric coefficients. Using the ratio of stoichiometric coefficients, one can write differential equations of changes in the concentrations of reactants A and B:

$$V\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = \left(q_{A_{\mathrm{in}}}[\mathrm{A}]_{\mathrm{in}} - \frac{a}{b}q_{B_{\mathrm{in}}}[\mathrm{B}]_{\mathrm{in}}\right) - \left(q_{A_{\mathrm{in}}} + q_{B_{\mathrm{in}}}\right)[\mathrm{A}] - V\cdot\frac{a}{b}\cdot[\mathrm{B}]\cdot\delta(f([\mathrm{B}]), \tag{2}$$

$$V\frac{\mathrm{d}[\mathrm{B}]}{\mathrm{d}t} = \left(q_{B_{\mathrm{in}}}[\mathrm{B}]_{\mathrm{in}} - \frac{b}{a}q_{A_{\mathrm{in}}}[\mathrm{A}]_{\mathrm{in}}\right) - \left(q_{A_{\mathrm{in}}} + q_{B_{\mathrm{in}}}\right)[\mathrm{B}] - V\cdot\frac{b}{a}\cdot[\mathrm{A}]\cdot\delta(f([\mathrm{A}]), \tag{3}$$

where $[\mathrm{A}]_{\mathrm{in}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ and $[\mathrm{B}]_{\mathrm{in}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ are the concentrations in the input carrier fluids of A and B, V (L) is the volume of the tank where the reaction takes place, $q_{\mathrm{A}_{\mathrm{in}}}$ $(\frac{\mathrm{L}}{\mathrm{s}})$ and $q_{\mathrm{B}_{\mathrm{in}}}$ $(\frac{\mathrm{L}}{\mathrm{s}})$ are the volume flows of carrier fluids for solutes A and B, δ is the Dirac delta function, while the function $f(x)$ is defined as:

$$f(x) = \begin{cases} \neq 0 & \text{for } \; x = 0, \\ 0 & \text{for } \; x > 0. \end{cases} \tag{4}$$

Since the concentrations cannot be negative, the following inequalities hold:

$$[\mathrm{A}] \geq 0, [\mathrm{B}] \geq 0. \tag{5}$$

The corresponding time-discretized system is:

$$[\mathrm{A}]_{k+1} = \max\left(0, \alpha[A]_k + (1-\alpha)([\mathrm{A}]_{\mathrm{in},k} - \frac{a}{b}[\mathrm{B}]_{\mathrm{in},k}) - \frac{a}{b}[\mathrm{B}]_k\right), \tag{6}$$

$$[\mathrm{B}]_{k+1} = \max\left(0, \alpha[\mathrm{B}]_k + (1-\alpha)([\mathrm{B}]_{\mathrm{in},k} - \frac{b}{a}[\mathrm{A}]_{\mathrm{in},k}) - \frac{b}{a}[\mathrm{A}]_k\right), \tag{7}$$

where $\alpha = e^{-\frac{T_s}{T}}$, T_s is the model sampling time, $[\mathrm{X}]_k = [\mathrm{X}](kT_\mathrm{S})$ and $T = \frac{V}{q_{A_\mathrm{in}} + q_{B_\mathrm{in}}}$.

2.2 Chemical reactions in a DWTP

The considered chemical reactants in the reaction volumes of a DWTP are presented in Table 1. First two columns represent raw water chemical impurities, and the third and fourth columns represent reactants appearing in the reaction volumes due to the chemical treatments, with the exception of $[\mathrm{Ca}^{2+}]$ that can come both from the raw water and from the applied chemicals.

Table 1. The list of considered concentrations of chemical elements with their corresponding units. The variables are expressed in (mol/L).

Description	Variable	Description	Variable
Manganese	$[\mathrm{Mn}^{2+}]$	Sulfate	$[\mathrm{SO}_4^{2-}]$
Iron	$[\mathrm{Fe}^{2+}]$	Total free chlorine	$[\mathrm{HOCl}]_{\mathrm{eq}}$
Bicarbonate	$[\mathrm{HCO}_3^-]$	Permanganate	$[\mathrm{MnO}_4^-]$
Magnesium	$[\mathrm{Mg}^{2+}]$	Potassium	$[\mathrm{K}^+]$
Calcium	$[\mathrm{Ca}^{2+}]$	Chloride ion	$[\mathrm{Cl}^-]$
Total organic carbon	[TOC]	Aluminium	$[\mathrm{Al}^{3+}]$

The chemical reactions considered for the DWTP are as follows [7], [8], [9], [10]:

$$3\mathrm{Mn}^{2+} + 2\mathrm{MnO}_4^- + 4\mathrm{OH}^- \rightarrow 5\mathrm{MnO}_2 \downarrow + 2\mathrm{H}_2\mathrm{O}, \tag{8}$$

$$3\mathrm{Fe}^{2+} + \mathrm{MnO}_4^- + 5\mathrm{OH}^- + 2\mathrm{H}_2\mathrm{O} \rightarrow \mathrm{MnO}_2 \downarrow + 3\mathrm{Fe(OH)}_3 \downarrow \tag{9}$$

$$\mathrm{Mn}^{2+} + \mathrm{HOCl}_{\mathrm{eq}} + \mathrm{H}_2\mathrm{O} \rightarrow \mathrm{MnO}_2 \downarrow + 3\mathrm{H}^+ + \mathrm{Cl}^-, \tag{10}$$

$$2\mathrm{Fe}^{2+} + \mathrm{HOCl}_{\mathrm{eq}} + 5\mathrm{H}_2\mathrm{O} \rightarrow 2\mathrm{Fe(OH)}_3 \downarrow + \mathrm{Cl}^- + 5\mathrm{H}^+, \tag{11}$$

$$\mathrm{Al}_2(\mathrm{SO}_4)_3 \cdot 18\mathrm{H}_2\mathrm{O} + 3\mathrm{Ca}^{2+} + 6\mathrm{HCO}_3^- \rightarrow 2\mathrm{Al(OH)}_3 \downarrow + 3\mathrm{Ca}^{2+} + 3\mathrm{SO}_4^{2-} + 6\mathrm{CO}_{2(\mathrm{g})} + 18\mathrm{H}_2\mathrm{O}. \tag{12}$$

Notation $[\mathrm{HOCl}]_{\mathrm{eq}}$ represents the equivalent amount of total free chlorine and it is given as [11]:

$$[\mathrm{HOCl}]_{\mathrm{eq}} = [\mathrm{HOCl}] + [\mathrm{ClO}^-]. \tag{13}$$

Reactions (10) and (11) occur as long as one of the compounds in equation (13) exists, that is, as long as $[\mathrm{HOCl}]_{\mathrm{eq}}$ is larger than zero. Further, all of the concentrations in the index marked by the letters 'eq' represent the total concentration of all species of that element, which are in a certain balance, and cause a reaction as long as the total concentration is larger than zero since due to the balance established this implies that also the reacting species is present. In case free ions appear on the right-hand side of the reaction, then this should be considered as in the case of (10) because they affect the pH value. In general, also other reactions can be modelled as far as the stoichiometry and reaction rates are known (the latter is needed only for slow reactions). The following bidirectional chemical reactions are also considered [8]:

$$\mathrm{Al(OH)_3} \downarrow + 3\mathrm{H^+} \rightleftharpoons \mathrm{Al^{3+}} + 3\mathrm{H_2O}\,;\, K_{\mathrm{S0}} = 10^{10.8} \tag{14}$$

$$\mathrm{Al(OH)_3} \downarrow + 2\mathrm{H^+} \rightleftharpoons \mathrm{Al(OH)^{2+}} + 2\mathrm{H_2O}\,;\, K_{\mathrm{S1}} = 10^{5.8} \tag{15}$$

$$\mathrm{Al(OH)_3} \downarrow + \mathrm{H^+} \rightleftharpoons \mathrm{Al(OH)_2}^{+} + \mathrm{H_2O}\,;\, K_{\mathrm{S2}} = 10^{1.5} \tag{16}$$

$$\mathrm{Al(OH)_3} \downarrow \rightleftharpoons \mathrm{Al(OH)_3}^{0};\, K_{\mathrm{S3}} = 10^{-4.2} \tag{17}$$

$$\mathrm{Al(OH)_3} \downarrow + \mathrm{H_2O} \rightleftharpoons \mathrm{Al(OH)_4}^{-} + \mathrm{H^+}\,;\, K_{\mathrm{S4}} = 10^{-12.2} \tag{18}$$

where the constants $K_{\mathrm{S0}} - K_{\mathrm{S4}}$ represent the solubility constants for each of the hydroxide species listed in reactions (14) – (18).

Chemical reaction (8) is an oxidation-reduction (redox) reaction [12] where ($\mathrm{Mn^{2+}}$) is a reducing agent and ($\mathrm{MnO_4^-}$) is an oxidizing agent. The product is manganese dioxide ($\mathrm{MnO_2}$) which precipitates. The next chemical reaction (9) is analogous to the previous one except that in it now ($\mathrm{Fe^{2+}}$) is the reducing agent. Here the products are manganese dioxide ($\mathrm{MnO_2}$) and iron hydroxide ($\mathrm{Fe(OH)_3}$) which precipitates. Chemical reactions (10) and (11) show how manganese ($\mathrm{Mn^{2+}}$) and iron ($\mathrm{Fe^{2+}}$) react with the hypochlorous acid (HOCl). The products are manganese dioxide ($\mathrm{MnO_2}$) and iron hydroxide ($\mathrm{Fe(OH)_3}$) along with the chloride ion ($\mathrm{Cl^-}$). Chemical reaction (12) is a gas evolution reaction because ($\mathrm{CO_2}$) gas is formed. It is a precipitation reaction as well because aluminium hydroxide ($\mathrm{Al(OH)_3}$) precipitates in the amount that cannot be dissolved. Aluminium hydroxide ($\mathrm{Al(OH)_3}$) is formed when aluminium sulfate is injected to water for coagulation and particles destabilization. Calcium and sulfate in (12) react and form calcium sulfate only if the following condition according to [13] is met:

$$[\mathrm{Ca^{2+}}][\mathrm{SO_4^{2-}}] > K_{\mathrm{so,CaSO_4}}, \tag{19}$$

where $K_{\mathrm{so,CaSO_4}}$ is the solubility constant for gypsum. If the above condition is not met, calcium and sulfate remain dissolved in water. A similar condition applies to the precipitation of calcite ($\mathrm{CaCO_3}$). That is, if ($\mathrm{Ca^{2+}}$) and ($\mathrm{HCO_3^-}$) are present, ($\mathrm{CaCO_3}$) will precipitate if the following condition is met:

$$[\mathrm{Ca^{2+}}][\mathrm{HCO_3^-}]^2 > K_{\mathrm{so,CaCO_3}}, \tag{20}$$

where $K_{\mathrm{so,CaCO_3}}$ is the solubility constant for calcium carbonate. If condition (20) is not met, ($\mathrm{Ca^{2+}}$) and ($\mathrm{HCO_3^-}$) remain dissolved in water.

The total amount of dissolved aluminium that enters the reaction volume is calculated as follows:

$$[\mathrm{Al^{3+}}]_{\mathrm{eq,in}} = \min\left([\mathrm{Al^{3+}}]_{\mathrm{in}}, K_{S0}[\mathrm{H^+}]_{\mathrm{prv}}^3 + K_{\mathrm{S1}}[\mathrm{H^+}]_{\mathrm{prv}}^2 + K_{\mathrm{S2}}[\mathrm{H^+}]_{\mathrm{prv}} + K_{S3} + \frac{K_{S4}}{[\mathrm{H^+}]}\right), \tag{21}$$

where $[\mathrm{Al^{3+}}]_{\mathrm{in}}$ is the input concentration of aluminium and $[\mathrm{H^+}]_{\mathrm{prv}}$ is the hydrogen-ion concentration from the previous reaction volume. In (12) only a part of aluminium that enters the reaction volume reacts with bicarbonate, i.e. the aluminium that enters the reaction volume reduced by dissolved aluminium which is represented as aluminium ion and various aluminium hydroxides. The non-dissolved aluminium hydroxide ($\mathrm{Al(OH)_3}$) forms aluminium hydroxide flocs used to adsorb a part of total organic carbon (TOC) present in the treated water. Its concentration is included in the DOSE variable explained in the next subsection that deals with TOC removal by coagulation (and consequent flocculation).

2.3 Edwards model

Besides the chemical equations provided, another mechanism considered is the removal of total organic carbon by coagulation, for which the Edward's model [14] is used. It is a Langmuir-based semiempirical model for quantification of TOC removal by coagulation, flocculation and subsequent sedimentation and filtering. The model is a nonlinear function derived from physical relationships, primarily from the process of Langmuir sorptive removal. It conceptually divides the initial TOC ($\mathrm{TOC_{initial}}$) into sorbable and non-sorbable fractions, whose proportion is modelled as a function of the specific UV absorbance (SUVA). The final sorbable TOC concentration $\mathrm{TOC_{final,sorb}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ satisfies the quadratic equation [8]:

$$[\mathrm{TOC}]^2_{\mathrm{final,sorb}} - [\mathrm{TOC_{final,sorb}}]\left([\mathrm{TOC_{initial,sorb}}] - \frac{1}{bM_{\mathrm{TOC}}} - \frac{a \cdot \mathrm{DOSE}}{M_{\mathrm{TOC}}}\right) - \frac{[\mathrm{TOC_{initial,sorb}}]}{bM^2_{\mathrm{TOC}}} = 0, \quad (22)$$

where M_{TOC} $(\frac{\mathrm{mg}}{\mathrm{mol}})$ is the molar mass of TOC, DOSE $(\frac{\mathrm{mmol}}{\mathrm{L}})$ is the coagulant dose, i.e. the concentration of aluminium hydroxide ($\mathrm{Al(OH)_3}$) formed by the addition of aluminium sulfate, b $(\frac{\mathrm{L}}{\mathrm{mg}})$ is the Langmuir equilibrium constant, a $(\frac{\mathrm{mg\ TOC}}{\mathrm{mmol\ coagulant}})$ is the total adsorbent capacity at monolayer coverage, approximated as a cubic polynomial function of pH:

$$a = x_3 \cdot \mathrm{pH}^3 + x_2 \cdot \mathrm{pH}^2 + x_1 \cdot \mathrm{pH}, \quad (23)$$

and x_1, x_2 and x_3 are empirical coefficients. Additionally, $\mathrm{TOC_{initial,sorb}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ is the input water concentration of sorbable total organic carbon approximated as a function of the input water SUVA:

$$\mathrm{TOC_{inital,sorb}} = \mathrm{TOC_{initial}}(1 - \mathrm{SUVA} \cdot K_1 - K_2), \quad (24)$$

where K_1 and K_2 are empirical coefficients and $\mathrm{TOC_{initial}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ is the initial concentration of TOC. SUVA can be expressed as:

$$\mathrm{SUVA} = \frac{\mathrm{UV_{254}}}{\mathrm{TOC_{initial}}}, \quad (25)$$

where $\mathrm{UV_{254}}$ $(\frac{1}{\mathrm{cm}})$ is the UV absorbance at a wavelength of 254 nm. According to [15], $\mathrm{UV_{254}}$ is linearly dependent on TOC and the expression for $\mathrm{TOC_{initial,sorb}}$becomes:

$$\mathrm{TOC_{initial,sorb}} = \mathrm{TOC_{initial}}(1 - K_2) - K_1 \cdot \frac{M_{\mathrm{TOC}}\mathrm{TOC_{initial}} - l}{kM_{\mathrm{TOC}}}, \quad (26)$$

where k and l are coefficients of equation of a line which connects TOC and $\mathrm{UV_{254}}$.

Finally, the concentration of TOC in the corresponding reaction volume $\mathrm{TOC_{final}}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$ can be calculated as:

$$\mathrm{TOC_{final}} = \mathrm{TOC_{final,sorb}} + K_1 \cdot \frac{M_{\mathrm{TOC}}\mathrm{TOC_{initial}} - l}{kM_{\mathrm{TOC}}} + K_2 \cdot \mathrm{TOC_{initial}}, \quad (27)$$

where $\mathrm{TOC}_{\mathrm{final,sorb}}$ can be determined by combining the quadratic equation (22) and the expression for $\mathrm{TOC}_{\mathrm{initial,sorb}}$ (24). If DOSE is equal to zero, the combination of all the previous expressions easily gives that:

$$\mathrm{TOC}_{\mathrm{final}} = \mathrm{TOC}_{\mathrm{initial}}. \tag{28}$$

To write a differential equation for TOC, the following function is used:

$$h(K_1, K_2, x_1, x_2, x_3, \mathrm{DOSE}, \mathrm{pH}, \mathrm{TOC}_{\mathrm{initial}}) = \begin{cases} \mathrm{TOC}_{\mathrm{final}} & \text{if } \ \mathrm{DOSE} > 0 \\ \mathrm{TOC}_{\mathrm{initial}} & \text{if } \ \mathrm{DOSE} = 0 \end{cases} \tag{29}$$

which summarizes the Edwards model explained above.

2.4 Bicarbonate buffer system

Dissolved carbon dioxide (CO_2) reacts with water (H_2O) to form carbonic acid (H_2CO_3), which in turn quickly dissociates to form a bicarbonate ion (HCO_3^-) and a hydrogen ion (H^+) as shown in the following reactions:

$$\mathrm{CO_2{+}H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^-{+}H^+}, \tag{30}$$

$$\mathrm{HCO_3^- \rightleftharpoons CO_3^{2-}{+}H^+}. \tag{31}$$

The presence of both a weak acid (for example, H_2CO_3) and its conjugate base (for example, HCO_3^-) balances the pH as a buffer system, neutralizing any excess acid or base given to the system. To relate the pH of water to bicarbonate buffer system elements, a modified version of the Henderson–Hasselbalch equation can be used [16]:

$$[\mathrm{H_2CO_3}] = \frac{[\mathrm{H^+}][\mathrm{HCO_3^-}]}{K_{\mathrm{H_2CO_3}}}, \tag{32}$$

$$[\mathrm{HCO_3^-}] = \frac{[\mathrm{H^+}][\mathrm{CO_3^{2-}}]}{K_{\mathrm{HCO_3^-}}}, \tag{33}$$

where $K_{\mathrm{H_2CO_3}}$ ($10^{-6.35}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$) and $K_{\mathrm{HCO_3^-}}$ ($10^{-10.22}$ $(\frac{\mathrm{mol}}{\mathrm{L}})$) are the acidity constants for carbonic acid and bicarbonate ion, respectively.

The sum of the concentration of the total inorganic carbon species in the water, including carbonic acid (H_2CO_3), bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}) is called dissolved inorganic carbon. Concentration of carbonic acid $[H_2CO_3]$ is the sum of $[CO_2]$ and $[H_2CO_3]$ because the dissolved fraction of total (CO_2) in water is small and hydrolyses to (H_2CO_3). So, the total concentration of dissolved inorganic carbon $[H_2CO_3]_{\mathrm{eq}}$ is defined as:

$$[\mathrm{H_2CO_3}]_{\mathrm{eq}} = [\mathrm{H_2CO_3}] + [\mathrm{HCO_3^-}] + [\mathrm{CO_3}^{2-}]. \tag{34}$$

Combining expressions (32), (33) and (34) yields expressions for bicarbonate ion and carbonate ion depending only on dissolved inorganic carbon:

$$[\mathrm{HCO_3}^-] = \frac{[\mathrm{H_2CO_3}]_{\mathrm{eq}}[\mathrm{H^+}]K_{\mathrm{H_2CO_3}}}{[\mathrm{H^+}]^2 + K_{\mathrm{H_2CO_3}}[\mathrm{H^+}] + K_{\mathrm{H_2CO_3}}K_{\mathrm{HCO_3^-}}}, \tag{35}$$

$$[\mathrm{CO_3}^{2-}] = \frac{[\mathrm{H_2CO_3}]_{\mathrm{eq}}K_{\mathrm{H_2CO_3}}K_{\mathrm{HCO_3^-}}}{[\mathrm{H^+}]^2 + K_{\mathrm{H_2CO_3}}[\mathrm{H^+}] + K_{\mathrm{H_2CO_3}}K_{\mathrm{HCO_3^-}}}. \tag{36}$$

2.5 Dissolved free chlorine

The reaction of gaseous chlorine with water is instantaneous and is defined with the following reaction [11]:

$$Cl_{2(g)} + H_2O \rightarrow HOCl + H^+ + Cl^-. \tag{37}$$

The concentrations in equilibrium of the hypochlorous acid (HOCl) and the hypochlorite ion (ClO^-) with dissolved chlorine are connected in the following way:

$$[ClO^-] = \frac{[HOCl]_{eq}}{1+\frac{[H^+]}{A}}, \tag{38}$$

where A ($10^{-7.53}$ ($\frac{mol}{L}$)) is the acidity constant for the hypochlorite ion. From (38) it can be seen that for the current pH of water, i.e. for the current $[H^+]$, [HOCl] and $[ClO^-]$ are analytically tied together. It is important to model disinfectant decay ($[HOCl]_{eq}$) after a fast chemical reaction in this scenario, according to [8]. Disinfectant decay can be modeled as first order, that is:

$$r_d = -k_d\ [HOCl]_{eq}, \tag{39}$$

where r_d ($\frac{mol}{L\cdot s}$) is the reaction rate for the decline in disinfectant concentration with time, k_d ($\frac{1}{s}$) is the first-order decay rate.

2.6 Total dissolved calcium and magnesium

In the treatment of drinking water, lime is also used to enhance alkalinity and pH. Lime dissociates when it is added to water as follows:

$$Ca(OH)_2\downarrow \rightarrow Ca^{2+}+2OH^-. \tag{40}$$

The total dissolved calcium concentration is considered to be:

$$[Ca^{2+}]_{eq}=[Ca^{2+}]+[CaOH^+]+[Ca(OH)_2], \tag{41}$$

where $[CaOH^+]$ and $[Ca(OH)_2]$ are:

$$[CaOH^+]=\frac{[Ca^{2+}]K_{CaOH}}{[H^+]}, \tag{42}$$

$$[Ca(OH)_2]=\frac{[Ca^{2+}]K_{Ca(OH)_2}}{[H^+]^2}. \tag{43}$$

The coefficients K_{CaOH} and $K_{Ca(OH)_2}$ are the base dissociation constants for the calcium hydroxides.

The total dissolved magnesium concentration is expressed in a similar way:

$$[Mg^{2+}]_{eq}=[Mg^{2+}]+[MgOH^+]+[Mg(OH)_2], \tag{44}$$

where $[CaOH^+]$ and $[Ca(OH)_2]$ are:

$$[MgOH^+]=\frac{[Mg^{2+}]K_{MgOH}}{[H^+]}, \tag{45}$$

$$[Mg(OH)_2]=\frac{[Mg^{2+}]K_{Mg(OH)_2}}{[H^+]^2}, \tag{46}$$

and K_{MgOH} and $K_{Mg(OH)_2}$ are the base dissociation constants for the magnesium hydroxides as well.

3 MATHEMATICAL MODEL OF A DWTP

The drinking water treatment plant considered in this paper consists of four reaction volumes and is shown in Fig. 1.

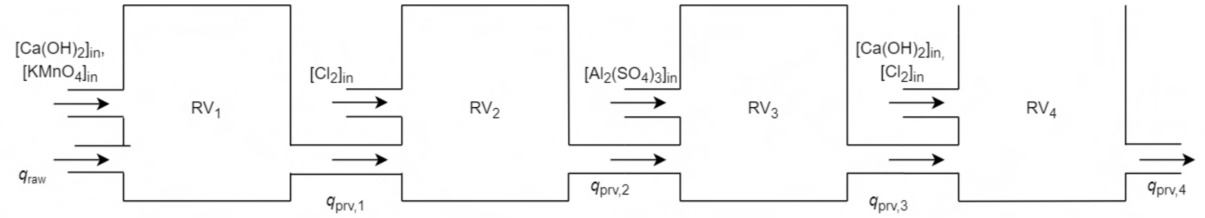


Figure 1 Scheme of the considered drinking water treatment plant

In the first reaction volume (RV_1), pre-oxidation processes take place by adding the reagent potassium permanganate ($KMnO_4$). Additionally, calcium hydroxide (lime – $Ca(OH)_2$) is added to keep the pH stable. In the second reaction volume (RV_2), chlorine gas (Cl_2) is added to reduce the concentration of manganese (Mn^{2+}) and iron (Fe^{2+}). In the third reaction volume (RV_3), coagulation, flocculation, sedimentation and filtering processes take place. In it aluminium sulfate ($Al_2(SO_4)_3$) is added to reduce the total organic carbon concentration. The fourth reaction volume (RV_4) is used for the free chlorine establishment and pH correction. The optional addition of chlorine gas (Cl_2) keeps the free chlorine on the desired level before entering the water distribution system while the optional addition of lime raises the pH to the desired value.

3.1 Chemical reactions modelling

A general case in which a reaction volume V is not fixed is elaborated. One water stream from the previous reaction volume $q_{\mathrm{in},1}$ is assumed, with the concentration $[\mathrm{X}]_{\mathrm{in},1}$ and also one input stream of the chemical dissolved in water carrier of flow $q_{\mathrm{in},2}$ with the chemical concentration equal $[\mathrm{X}]_{\mathrm{in},2}$. Now one can write the following differential equations for the substance X and the volume:

$$\frac{\mathrm{d}}{\mathrm{d}t}(V\cdot[\mathrm{X}]) = [\mathrm{X}]\frac{\mathrm{d}V}{\mathrm{d}t} + V\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = q_{\mathrm{in},1}[\mathrm{X}]_{\mathrm{in},1} + q_{\mathrm{in},2}[\mathrm{X}]_{\mathrm{in},2} - q_{\mathrm{out}}[\mathrm{X}], \tag{47}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = q_{\mathrm{in},1} + q_{\mathrm{in},2} - q_{\mathrm{out}}. \tag{48}$$

Combining these equations, one gets:

$$[\mathrm{X}](q_{\mathrm{in},1} + q_{\mathrm{in},2} - q_{\mathrm{out}}) + V\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = q_{\mathrm{in},1}[\mathrm{X}]_{\mathrm{in},1} + q_{\mathrm{in},2}[\mathrm{X}]_{\mathrm{in},2} - q_{\mathrm{out}}[\mathrm{X}], \tag{49}$$

$$V\frac{\mathrm{d}[X]}{\mathrm{d}t} = q_{\mathrm{in},1}([\mathrm{X}]_{\mathrm{in},1} - [\mathrm{X}]) + q_{\mathrm{in},2}([\mathrm{X}]_{\mathrm{in},2} - [\mathrm{X}]). \tag{50}$$

Equation (50) should be extended to account for the chemical reactions of appearance and disappearance of chemical element X. It has to be taken into account that the volume may change dynamically in case of imbalance between the input and output flows. The flow q_{out} is given externally or is defined with V in the case of hydraulically uncoupled tanks. Also, q_{out} is defined with current reaction volume V_{crv} and the next reaction volume in case of hydraulically coupled tanks.

For the case when the volume is fixed, one can write the following equation:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = 0 \Rightarrow V = \mathrm{const.} \tag{51}$$

In a similar way as described in section 2.1 one can obtain differential equations of concentrations of chemicals with respect to the previously introduced chemical reactions and coagulation/flocculation. The reaction rates are considered infinite to simplify the model parametrization. In that case the reactions happen instantaneously and stop with the scarcest reactant vanished.

The subscript ‘prv’ is introduced, with the meaning ‘previous reaction volume’. For presentation clarity, the following substitutions are introduced:

$$\dot{[\mathrm{MnO_4^-}]}_{\mathrm{in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{MnO_4^-}]_{\mathrm{prv}} + q_{\mathrm{KMnO4_{in}}} \cdot [\mathrm{KMnO4}]_{\mathrm{in}}, \tag{52}$$

$$\dot{[\mathrm{K^+}]}_{\mathrm{in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{K^+}]_{\mathrm{prv}} + q_{\mathrm{KMnO4_{in}}} \cdot [\mathrm{KMnO4}]_{\mathrm{in}}, \tag{53}$$

$$\dot{[\mathrm{Cl^-}]}_{\mathrm{in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{Cl^-}]_{\mathrm{prv}} + q_{\mathrm{Cl_{2_{in}}}} \cdot [\mathrm{Cl_2}]_{\mathrm{in}}, \tag{54}$$

$$\dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{HOCl}]_{\mathrm{eq,prv}} + q_{\mathrm{Cl_{2_{in}}}} \cdot [\mathrm{Cl_2}]_{\mathrm{in}}, \tag{55}$$

$$\dot{[\mathrm{Ca^{2+}}]}_{\mathrm{eq,in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{Ca^{2+}}]_{\mathrm{eq,prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}} \cdot [\mathrm{Ca(OH)_2}]_{\mathrm{in}} - (q_{\mathrm{prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}}) \cdot C, \tag{56}$$

$$\begin{aligned} C = {} & \max\left(0, \frac{q_{\mathrm{prv}} \cdot [\mathrm{Ca^{2+}}]_{\mathrm{eq,prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}} \cdot [\mathrm{Ca(OH)_2}]_{\mathrm{in}}}{q_{\mathrm{prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}}} - [\mathrm{CaCO_3}]_{\mathrm{limit}}\right) + \\ & \max\left(0, \max\left(0, \frac{q_{\mathrm{prv}} \cdot [\mathrm{Ca^{2+}}]_{\mathrm{eq,prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}} \cdot [\mathrm{Ca(OH)_2}]_{\mathrm{in}}}{q_{\mathrm{prv}} + q_{\mathrm{Ca(OH)_{2_{in}}}}} - [\mathrm{CaCO_3}]_{\mathrm{limit}}\right) \right. \\ & \left. - [\mathrm{CaSO_4}]_{\mathrm{limit}}\right), \end{aligned} \tag{57}$$

$$\dot{[\mathrm{H_2CO_3}]}_{\mathrm{eq,in,total}} = q_{\mathrm{prv}} \cdot [\mathrm{H_2CO_3}]_{\mathrm{eq,prv}} - q_{\mathrm{prv}} \cdot D, \tag{58}$$

$$D = \max(0, [\mathrm{H_2CO_3}]_{\mathrm{eq,prv}} - [\mathrm{CaCO_3}]_{\mathrm{limit}}), \tag{59}$$

$$\begin{aligned} \dot{[\mathrm{Al^{3+}}]}_{\mathrm{eq,in,total}} = {} & q_{\mathrm{prv}} \cdot [\mathrm{Al^{3+}}]_{\mathrm{eq,prv}} + 2 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}} \cdot [\mathrm{Al_2(SO_4)_3}]_{\mathrm{in}} \\ & - (q_{\mathrm{prv}} + 2 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}}) \cdot \mathrm{DOSE}, \end{aligned} \tag{60}$$

$$\begin{aligned} \mathrm{DOSE} = {} & \max\left(0, \frac{q_{\mathrm{prv}} \cdot [\mathrm{Al^{3+}}]_{\mathrm{prv}} + 2 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}} \cdot [\mathrm{Al_2(SO_4)_3}]_{\mathrm{in}}}{q_{\mathrm{prv}} + 2 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}}} \right. \\ & \left. - [\mathrm{Al^{3+}}]_{\mathrm{eq,in}}([\mathrm{H^+}]_{\mathrm{prv}})\right), \end{aligned} \tag{61}$$

$$\begin{aligned} \dot{[\mathrm{SO_4^{2-}}]}_{\mathrm{in,total}} = {} & q_{\mathrm{prv}} \cdot [\mathrm{SO_4^{2-}}]_{\mathrm{prv}} + 3 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}} \cdot [\mathrm{Al_2(SO_4)_3}]_{\mathrm{in}} - (q_{\mathrm{prv}} + 3 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}}) \\ & \cdot E, \end{aligned} \tag{62}$$

$$E = \max\left(0, \frac{q_{\mathrm{prv}} \cdot [\mathrm{SO_4^{2-}}]_{\mathrm{prv}} + 3 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}} \cdot [\mathrm{Al_2(SO_4)_3}]_{\mathrm{in}}}{q_{\mathrm{prv}} + 3 \cdot q_{\mathrm{Al_2(SO_4)_{3_{in}}}}} - [\mathrm{CaSO_4}]_{\mathrm{limit}}\right), \tag{63}$$

$$q_{\mathrm{in,total}} = q_{\mathrm{prv}} + q_{\mathrm{KMnO4_{in}}} + q_{\mathrm{Cl_{2_{in}}}} + q_{\mathrm{Ca(OH)_{2_{in}}}} + q_{\mathrm{Al_2(SO_4)_{3_{in}}}}. \tag{64}$$

Variable $[\mathrm{Al^{3+}}]_{\mathrm{eq,in}}$ is the function of $[\mathrm{H^+}]_{\mathrm{prv}}$ and represents the dissolved aluminium at the inlet to a reaction volume. Quantities $[\mathrm{CaCO_3}]_{\mathrm{limit}}$ and $[\mathrm{CaSO_4}]_{\mathrm{limit}}$ represent the limiting concentrations for precipitation to start, in accordance with (19) and (20). The precipitation conditions were not met for the set of living laboratory input concentrations evaluated in this paper, hence the variables C, D, and E are zero. The differential equations of the concentrations are provided next.

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{MnO}_4^-]}{\mathrm{d}t} = \frac{1}{V}((\max(0; \dot{[\mathrm{MnO}_4^-]}_{\mathrm{in,total}} - \frac{1}{3}[\mathrm{Fe}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}} - \frac{2}{3}[\mathrm{Mn}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}})\\ - q_{\mathrm{in,total}} \cdot [\mathrm{MnO}_4^-]) - \frac{1}{3}[\mathrm{Fe}^{2+}] \cdot \delta(f_2([\mathrm{Fe}^{2+}], [\mathrm{MnO}_4^-]))\\ - \frac{2}{3}[\mathrm{Mn}^{2+}] \cdot \delta(f_2([\mathrm{Mn}^{2+}], [\mathrm{MnO}_4^-])).\end{aligned} \tag{65}$$

The function $f_n(x)$ is defined as:

$$f_n(x_1, x_2, \dots, x_n) = \begin{cases} \neq 0 & \text{if any } x_i \leq 0, \\ 0 & \text{if all } \ x_i > 0. \end{cases} \tag{66}$$

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{Cl}^-]}{\mathrm{d}t} = \frac{1}{V}((\dot{[\mathrm{Cl}^-]}_{\mathrm{in,total}} + \min(\frac{1}{1}[\mathrm{Mn}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}}; \frac{1}{1}\dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}})\\ + \min(\frac{1}{2}[\mathrm{Fe}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}}; \frac{1}{1}\dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}}) - q_{\mathrm{in,total}} \cdot [\mathrm{Cl}^-])\\ + \min(\frac{1}{1}[\mathrm{Mn}^{2+}]; \frac{1}{1}[\mathrm{HOCl}]_{\mathrm{eq}}) \cdot \delta(f_2([\mathrm{Mn}^{2+}], [\mathrm{HOCl}]_{\mathrm{eq}}))\\ + \min(\frac{1}{2}[\mathrm{Fe}^{2+}]; \frac{1}{1}[\mathrm{HOCl}]_{\mathrm{eq}}) \cdot \delta(f_2([\mathrm{Fe}^{2+}], [\mathrm{HOCl}]_{\mathrm{eq}}))\\ + k_d[\mathrm{HOCl}]_{\mathrm{eq}},\end{aligned} \tag{67}$$

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{Mn}^{2+}]}{\mathrm{d}t} = \frac{1}{V}(\max(0; [\mathrm{Mn}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}} - \frac{1}{1}\dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}} - \frac{3}{2}\dot{[\mathrm{MnO}_4^-]}_{\mathrm{in,total}})\\ - q_{\mathrm{in,total}} \cdot [\mathrm{Mn}^{2+}]) - \frac{1}{1}[\mathrm{HOCl}]_{\mathrm{eq}} \cdot \delta(f_2([\mathrm{HOCl}]_{\mathrm{eq}}, [\mathrm{Mn}^{2+}]))\\ - \frac{3}{2}[\mathrm{MnO}_4^-] \cdot \delta(f_2([\mathrm{MnO}_4^-], [\mathrm{Mn}^{2+}])),\end{aligned} \tag{68}$$

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{Fe}^{2+}]}{\mathrm{d}t} = \frac{1}{V}(\max(0; [\mathrm{Fe}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}} - \frac{3}{1}\dot{[\mathrm{MnO}_4^-]}_{\mathrm{in,total}} - \frac{2}{1}\dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}}) - q_{\mathrm{in,total}}\\ \cdot [\mathrm{Fe}^{2+}]) - \frac{3}{1}[\mathrm{MnO}_4^-] \cdot \delta(f_2(\mathrm{MnO}_4^-], [\mathrm{Fe}^{2+}]))\\ - 2[\mathrm{HOCl}]_{\mathrm{eq}}\delta(f_2([\mathrm{HOCl}]_{\mathrm{eq}}, [\mathrm{Fe}^{2+}])),\end{aligned} \tag{69}$$

$$\frac{\mathrm{d}[\mathrm{Ca}^{2+}]_{\mathrm{eq}}}{\mathrm{d}t} = \frac{1}{V}(\dot{[\mathrm{Ca}^{2+}]}_{\mathrm{eq,in,total}} - q_{\mathrm{in,total}} \cdot [\mathrm{Ca}^{2+}]_{\mathrm{eq}}), \tag{70}$$

$$\frac{\mathrm{d}[\mathrm{Mg}^{2+}]_{\mathrm{eq}}}{\mathrm{d}t} = \frac{1}{V}([\mathrm{Mg}^{2+}]_{\mathrm{eq,in}} \cdot q_{\mathrm{prv}} - q_{\mathrm{in,total}} \cdot [\mathrm{Mg}^{2+}]_{\mathrm{eq}}), \tag{71}$$

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{H}_2\mathrm{CO}_3]_{\mathrm{eq}}}{\mathrm{d}t} = \frac{1}{V}(\max(0; \dot{[\mathrm{H}_2\mathrm{CO}_3]}_{\mathrm{eq,in,total}} \cdot q_{\mathrm{prv}} - 3 \cdot (q_{\mathrm{prv}} + 2 \cdot q_{\mathrm{Al}_2(\mathrm{SO}_4)_{3_{\mathrm{in}}}}) \cdot \mathrm{DOSE})\\ - q_{\mathrm{in,total}} \cdot [\mathrm{H}_2\mathrm{CO}_3]_{\mathrm{eq}}),\end{aligned} \tag{72}$$

$$\frac{\mathrm{d}[\mathrm{K}^+]}{\mathrm{d}t} = \frac{1}{V}(\dot{[\mathrm{K}^+]}_{\mathrm{in,total}} - q_{\mathrm{in,total}} \cdot [\mathrm{K}^+]), \tag{73}$$

$$\frac{\mathrm{d}[\mathrm{TOC}]}{\mathrm{d}t} = \frac{1}{V}\Big[[\mathrm{TOC}]_{\mathrm{in}} \cdot q_{\mathrm{prv}} - \ h\,(K_1, K_2, x_1, x_2, x_3, \mathrm{DOSE}, \mathrm{pH}, [\mathrm{TOC}]_{\mathrm{in}} \cdot \frac{q_{\mathrm{prv}}}{q_{\mathrm{prv}} + q_{\mathrm{Al_2(SO_4)_3}_{\mathrm{in}}}}) \cdot q_{\mathrm{in,total}}\Big]\,, \tag{74}$$

where h denotes the function defined in subsection 2.3,

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{HOCl}]_{\mathrm{eq}}}{\mathrm{d}t} = \frac{1}{V}\big(\max\ (0; \dot{[\mathrm{HOCl}]}_{\mathrm{eq,in,total}} - \frac{1}{1}[\mathrm{Mn}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}} - \frac{1}{2}[\mathrm{Fe}^{2+}]_{\mathrm{in}} \cdot q_{\mathrm{prv}})\\ - q_{\mathrm{in,total}} \cdot [\mathrm{HOCl}]_{\mathrm{eq}}\big)\ - \frac{1}{1}[\mathrm{Mn}^{2+}]\delta\,(f_2([\mathrm{Mn}^{2+}],[\mathrm{HOCl}]_{\mathrm{eq}}))\\ - \frac{1}{2}[\mathrm{Fe}^{2+}]\delta\,(f_2([\mathrm{Fe}^{2+}],[\mathrm{HOCl}]_{\mathrm{eq}})) - k_d[\mathrm{HOCl}]_{\mathrm{eq}},\end{aligned} \tag{75}$$

$$\frac{\mathrm{d}[\mathrm{Al}^{3+}]_{\mathrm{eq}}}{\mathrm{d}t} = \frac{1}{V}(\dot{[\mathrm{Al}^{3+}]}_{\mathrm{eq,in,total}} - q_{\mathrm{in,total}} \cdot [\mathrm{Al}^{3+}]_{\mathrm{eq}}), \tag{76}$$

$$\frac{\mathrm{d}[\mathrm{SO}_4^{2-}]}{\mathrm{d}t} = \frac{1}{V}(\dot{[\mathrm{SO}_4^{2-}]}_{\mathrm{in,total}} - q_{\mathrm{in,total}} \cdot [\mathrm{SO}_4^{2-}]), \tag{77}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = q_{\mathrm{in,total}} - q_{\mathrm{out}}. \tag{78}$$

3.2 pH modelling

Besides the concentrations, an important output of the model will also be the pH value established in the reaction volume. It is modelled by starting with the dissociation equation for water [17]:

$$[\mathrm{H}^+] \cdot [\mathrm{OH}^-] = k_w = 10^{-14}. \tag{79}$$

From charge balance one can obtain:

$$[\mathrm{H}^+] - [\mathrm{OH}^-] = r_{\mathrm{net}}, \tag{80}$$

where:

$$\begin{aligned}r_{\mathrm{net}} = [\mathrm{HCO}_3^-]([\mathrm{H}^+], [\mathrm{H_2CO_3}]_{\mathrm{eq}}) + 2[\mathrm{CO_3}^{2-}]]([\mathrm{H}^+], [\mathrm{H_2CO_3}]_{\mathrm{eq}}) + [\mathrm{Cl}^-] + [\mathrm{MnO}_4^-]\\ + 2[\mathrm{SO}_4^{2-}] + [\mathrm{ClO}^-]([\mathrm{H}^+], [\mathrm{HOCl}_{\mathrm{eq}}])\ - 2[\mathrm{Ca}^{2+}]([\mathrm{H}^+],[\mathrm{Ca}^{2+}]_{\mathrm{eq}})\\ - 2[\mathrm{Mg}^{2+}]([\mathrm{H}^+],[\mathrm{Mg}^{2+}]_{\mathrm{eq}}) - 2[\mathrm{Fe}^{2+}] - 2[\mathrm{Mn}^{2+}] - [\mathrm{K}^+]\\ - [\mathrm{CaOH}^+]([\mathrm{H}^+],[\mathrm{Ca}^{2+}]_{\mathrm{eq}}) - [\mathrm{MgOH}^+]([\mathrm{H}^+],[\mathrm{Mg}^{2+}]_{\mathrm{eq}}) - c_{\mathrm{avg}}[\mathrm{Al}^{3+}]_{\mathrm{eq}},\end{aligned} \tag{81}$$

where $c_{\mathrm{avg}} = -0.1937$, denoting the average charge of the dissolved aluminium species at pH 7 to simplify the calculation. After substituting (35), (36), (38), (42), (43), (45), (46), (79) and (81) into (80) one gets the equation for $[\mathrm{H}^+]$ with a ninth order polynomial which is not written here because of the length of its notation. The concentration of (H^+) ions obtained from the solution to it is used to calculate the pH value.

Now one can determine pH using the following expression:

$$\mathrm{pH} = -\log[\,\mathrm{H}^+]. \tag{82}$$

4 VALIDATION ON CASE STUDY

4.1 Description

Dynamical model for fast reactions was implemented and simulated in the Matlab/Simulink development environment. The raw water quality, i.e. the concentration of each element in raw water obtained by laboratory analysis of samples from a real DWTP is set as in Table 2. All subsequent concentrations will be presented in (mg/L) for easier comparison of results with the measurements.

Table 2 Raw water quality. The concentrations are expressed in mg/L

$[Mn^{2+}]$	$[Fe^{2+}]$	$[Ca^{2+}]$	$[Mg^{2+}]$	$[H_2CO_3]_{eq}$	[TOC]
0.543	0.308	41	4.5	175	6.12

The doses of each of the reagents added to each reaction volume are shown in Table 3.

Table 3 Doses of each reagent. The reagent dosages are expressed in mg/L (the unit L refers to the litre of the treated water stream here)

	RV_1	RV_2	RV_3	RV_4
$[KMnO_4]_{in}$	0.5	-	-	-
$[Ca(OH)_2]_{in}$	5	-	-	3
$[Cl_2]_{in}$	-	6.8	-	1
$[Al_2(SO_4)_3]_{in}$	-	-	80	-

4.2 Mathematical model simulation

The obtained concentration of each element and pH value responses at the outlet of the fourth reaction volume are shown in Fig. 2. In it the first two responses show how much reagent is dosed and when in each reaction volume. Here it can be observed that the iron concentration was reduced to zero while the manganese concentration was reduced to approximately half of the initial value after the addition of potassium permanganate. By adding chlorine in the second reaction volume, the manganese concentration is reduced to zero. In the third reaction volume, coagulation/flocculation processes take place since aluminium sulfate is added as a coagulant with the concentration from Table 3. The main goal of adding aluminium sulfate is to reduce the TOC concentration, i.e. to reduce the presence of organic matter and fifth response in Fig. 2 clearly shows that this has been achieved. In the fourth reaction volume free chlorine is achieved in the required concentration, while the addition of lime regulates the pH also on the required level.

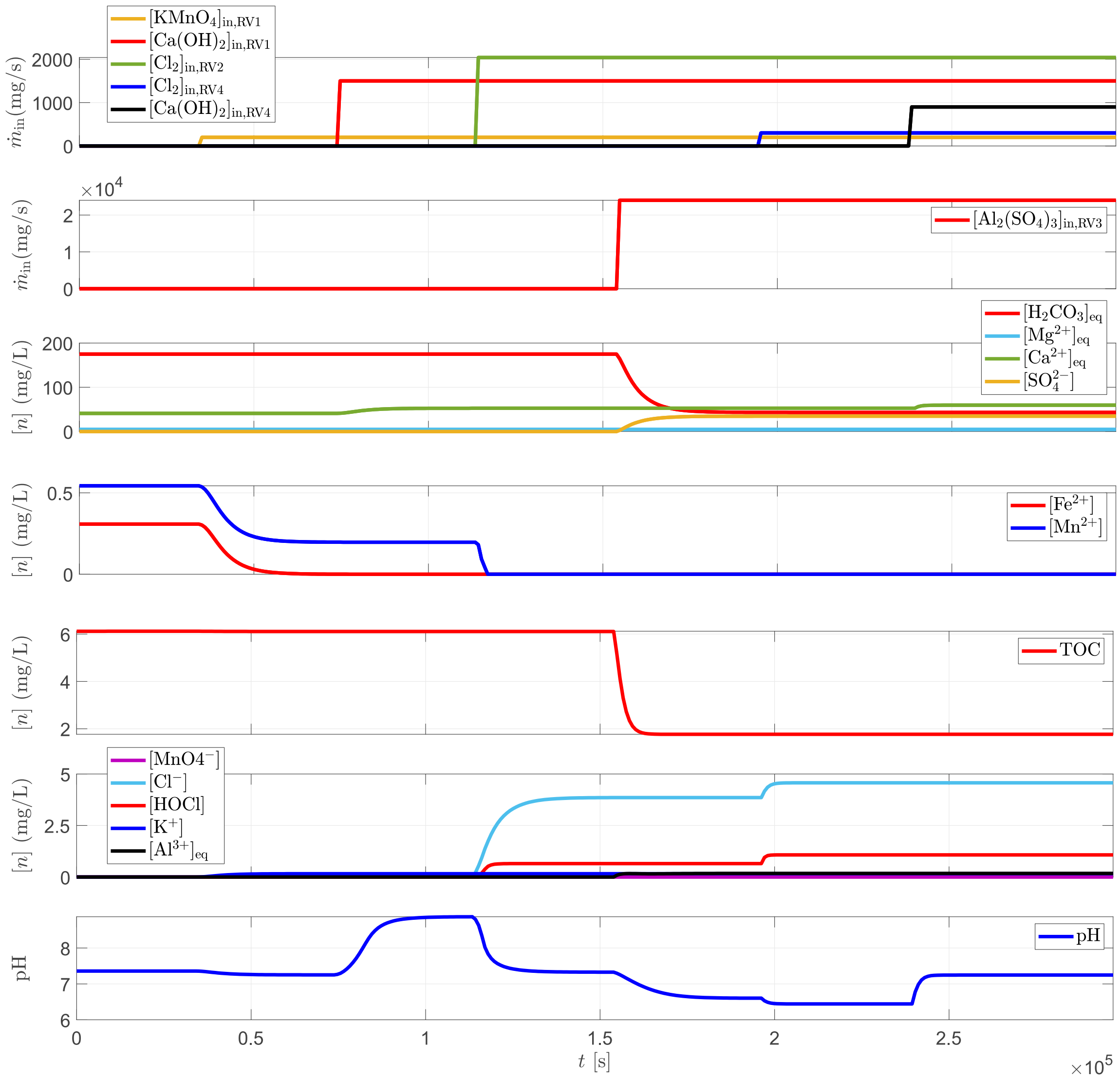


Figure 2 The concentration of each element and pH value at the outlet of the fourth reaction volume

4.3 Comparison of the model results with living lab measurements

Table 4 compares the measurements taken at the DWTP's output, i.e. at the fourth reaction volume's output, to the values calculated using the proposed model.

Table 4 Comparison of measurements to the model results. The concentrations are expressed in (mg/L)

	$[HOCl]_{eq}$	$[Fe^{2+}]$	$[Mn^{2+}]$	[TOC]	pH
Measurement	1.1	<0.032	0.009	1.63	7.3
Model	1.08	0	0	1.77	7.3

The model yielded a concentration of 0.177 (mg/L) of aluminium at the fourth reaction volume's outlet. Because the concentration of dissolved aluminium is highly dependent on pH, and its value is irrelevant in determining the model's accuracy, there was no comparison of aluminium with the real measurements. However, the concentration is within the DWTP's allowed limits, with a maximum dissolved aluminium of 0.2 (mg/L). A comparison of the results for this preliminary validation shows a satisfactory model accuracy with the maximum deviation from the actual measurements being 8.6%, for the case of TOC.

5 CONCLUSIONS

A model of a general reaction volume of a drinking water treatment plant (DWTP) is presented in the paper. The DWTP under consideration is divided into four reaction volumes, each with its own set of chemical processes, all considered as fast chemical reactions. The model implementation is simplified with only a couple of known parameters required. The model is simulated in the Matlab/Simulink development environment and the expected values of drinking water quality parameters are obtained at the output. It includes a bicarbonate buffer system that prevents pH levels from fluctuating drastically when chemicals are added. The Edwards model was also used, which resembles the TOC removal with good accuracy when compared to the real measurement. A comparison of measurements on an actual DWTP and the model results was made. For this preliminary validation, the comparison shows a good model accuracy with the maximum deviation from the actual measurement of 8.6%, presenting a good indicator of the model reliability.

6 ACKNOWLEDGMENT

This work was supported by the European Union via Horizon 2020 programme through the project Resilient Water Innovation for Smart Economy (REWAISE, Grant agreement ID: 869496). The work has also been supported by the Croatian Science Foundation through the Young Researchers' Career Development Project - Training New Doctoral Students under contract No. DOK-2020-01-4118.

7 REFERENCES

[1] J. Yang, “An Update on Modifications to Water Treatment Plant Model,” in WQTC 2017, Portland, Oregon, 2017.

[2] M. Yurong and H. Wenbao, “A new Fuzzy-Smith control algorithm used in oilfield produced water treatment control system,” in 2010 Sixth International Conference on Natural Computation, Yantai, China, 2010.

[3] A. Bemporad and M. Morari, “Control of systems integrating logic, dynamics, and constraints,” Automatica, vol. 35, no. 3, pp. 407-427, 1999.

[4] J. L. Villa, M. Duque, A. Gauthier and N. Rakoto-Ravalontsalama, “Modeling and control of a water treatment plant,” in SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics., Washington, DC, USA, 2003.

[5] W. Xiaojie, J. Yunzhe and L. Xiaojing, “Research on the prediction of water treatment plant coagulant dosage based on feed-forward artificial neutral network,” in International Conference on Consumer Electronics, Communications and Networks (CECNet), Xianning, China, 2011.

[6] Z. Song, Y. Zhao, X. Song and C. Liu, “Research on prediction model of optimal coagulant dosage in water purifying plant based on nerual network,” ISECS International Colloquium on Computing, Communication, Control, and Management, pp. 258-261, 2009.

[7] O. Bello, Y. Hamam and K. Djouani, “Dynamic Modelling and System Identification of Coagulant Dosage,” in Proceedings of the 3rd International Conference on Systems and Control, Algiers, Algeria, 2013.

[8] J. C. Crittenden, R. R. Trussell, D. W. Hand, K. J. Howe and G. Tchobanoglous, MWH's Water Treatment: Principles and Design, New Jersey: John Wiley & Sons, Inc., 2012.

[9] A. H. Lundquist, “Chemical and Biological Removal of Iron and Manganese from Drinking Water,” Montana State University, Bozeman, Montana, 1999.

[10] R. D. Letterman, M. Tabatabaie and R. S. Ames., “The Effect of the Bicarbonate Ion Concentration on Flocculation With Aluminum Sulfate,” Journal (American Water Works Association), vol. 71, no. 8, pp. 467-472, 1979.

[11] Center for Drinking Water Optimization, University of Colorado at Boulder, “Water Treatment Plant Model,” University of Colorado at Boulder, Boulder, 2001.

[12] D. W. Hendricks, Water Treatment Unit Processes, Physical and Chemical, Boca Raton, USA: CRC Press, Taylor & Francis Group, 2006.

[13] D. Rai and W. T. Franklin, “Program for Computing Equilibrium Solution Composition in CaCO3 and CaSO4 Systems from Irrigation Water Compositions,” Agronomy Department Colorado State University Fort Collins, Colorado, 1973.

[14] D. W. Dunnington, B. F. Trueman, W. J. Raseman, L. E. Anderson and G. A. Gagnon, “Comparing the Predictive Performance, Interpretability, and Accessibility of Machine Learning and Physically Based Models for Water Treatment,” ACS ES&T Engineering, pp. 348-356, 2021.

[15] S. Szerzyna, M. Molczan, M. Wolska, W. Adamski and J. Wisniewski, “Absorbance based water quality indicators as parameters for treatment process control with respect to organic substance removal,” in E3S Web of Conferences, Web, 2017.

[16] F. Liu, H. A. Merchant, R. P. Kulkarni, M. Alkademi and A. W. Basit, “Evolution of a physiological pH6.8 bicarbonate buffer system: Application to the dissolution testing of enteric coated products,” European Journal of Pharmaceutics and Biopharmaceutics, vol. 78, no. 1, pp. 151-157, 2011.

[17] A.-M. O. Mohamed and E. K. Paleologos, Fundamentals of Geoenvironmental Engineering., Oxford, United Kingdom: Elsevier Inc., 2018.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14793

A THOROUGH CYBERSECURITY DATASET FOR INTRUSION DETECTION IN SMART WATER NETWORKS

Andrés F. Murillo[1], Riccardo Taormina[2], Nils Ole Tippenhauer[3] and Stefano Galelli[4]

[1]Singapore University of Technology and Design, Singapore
[2]Delft University of Technology, Delft, Netherlands
[3]CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
[4]Singapore University of Technology and Design, Singapore

[1]andres_murillo@sutd.edu.sg, [2]r.taormina@tudelft.nl, [3]tippenhauer@cispa.de, [4] stefano_galelli@sutd.edu.sg

Abstract

The increase in the number and complexity of cyber-physical attacks on water distribution systems requires better intrusion detection systems. So far, the design and validation of such systems has relied on datasets, such as the EWRI 2017 BATtle of the Attack Detection ALgorithms (BATADAL), that provide a detailed representation of hydraulic processes in response to cyber-physical attacks. However, the BATADAL, generated with epanetCPA, does not include an equivalent and detailed representation of the processes occurring within the industrial communication system. Here, we fill in this gap by presenting the BATADAL 2.0 dataset, generated with the DHALSIM simulator, a novel co-simulation environment that can represent the hydraulic processes, digital control, and network communication of smart water networks. The dataset includes a broad variety of attacks and network anomalies. Most importantly, the availability of both process and network data is expected to pave the way to more advanced and accurate detection algorithms.

Keywords
Cyber-security, cyber-physical security, cyber-attacks, DHALSIM, water distribution systems, smart water networks, BATADAL, dataset, SCADA.

1 INTRODUCTION

With the widespread digitalization of the water sector, the world is witnessing an increasing trend in the number and complexity of cyber-physical attacks against critical water infrastructures [1]. This trend requires extra measures to harden smart water networks so that they can operate in a hostile cyber-security environment. Such hardening includes the integration of intrusion detection and protection mechanisms into the infrastructure management systems. Designing such mechanisms usually requires vast amounts of data that represent the behaviour of both physical and cyber layers during normal and under attack conditions.

In recent years, researchers and practitioners used the EWRI 2017 BATtle of the Attack Detection ALgorithms (BATADAL) dataset generated with epanetCPA to develop intrusion detection techniques for water distribution systems [2]. While tools like epanetCPA [3] or RISKNOUGHT [4] can simulate the response of physical processes to cyber-physical attacks, they are not able to replicate traffic data in the cyber layer of smart water networks. Yet, complementing process data with features extracted from network traffic would enable the development of advanced attack detection and localization techniques discovering anomalies across the entire cyber-physical system [5].

In this work, we fill this gap by presenting an open dataset generated by the Digital HydrAuLic SIMulator (DHALSIM) simulator [5]. DHALSIM is a novel co-simulation environment that can represent the hydraulic processes, digital control, and network communication of smart water networks by interfacing EPANET with miniCPS, an industrial control network emulator. DHALSIM thus outputs both process variables (e.g., pressures, water levels, flows) as well as records of packet data pulled from network scans (e.g., PCAP files). The proposed dataset "extends" the BATADAL dataset by implementing similar attacks on the C-Town benchmark system, with data on both normal operating conditions and anomalous conditions. The normal operating conditions were generated using multiple demand patterns, tank initial water levels, sensor noise, as well as small network events (e.g., loss of packets in the communication networks) that better resemble the real operation of water distribution networks. The anomalous conditions include not only malicious cyber-physical attacks, but also benign anomalies, such as disruptive network events.

In the remainder of the manuscript, we first introduce DHALSIM (Section 2) and describe the experimental setup for the generation of the dataset (Section 3). We then illustrate key aspects of the dataset (Section 4) and briefly outline the way forward (Section 5).

2 THE DIGITAL HYDRAULICSIMULATOR (DHALSIM)

DHALSIM is a simulation environment that combines the hydraulic simulation capabilities of EPANET with the network emulation capabilities provided by MiniCPS [6] and Mininet [7]. Mininet is a virtualization platform that allows users to easily create virtual networks to connect virtualized guests (or Mininet nodes). These virtual nodes have their own virtual network interfaces and can run any software installed in the host machine. MiniCPS is built on top of Mininet and provides an implementation of popular ICS communication protocols. DHALSIM uses MiniCPS and Mininet to create virtual industrial control networks. These virtual networks are composed of Mininet nodes and virtual network links. The Mininet nodes represent PLCs or SCADA and communicate using industrial communication protocols. As for the physical system, DHALSIM launches a process running an EPANET process simulation. Finally, DHALSIM experiments can simulate cyber-physical attacks or network events that impact the behaviour of the industrial communication layer, and, as a consequence, also impact the underlying physical processes.

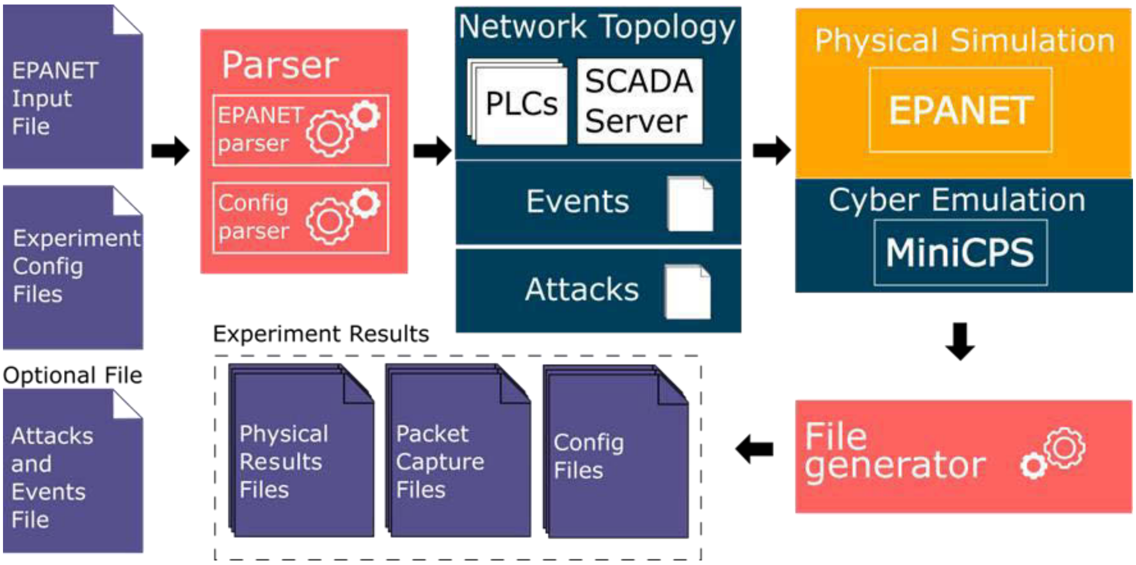


Figure 1. DHALSIM architecture. DHALSIM uses EPANET to simulate hydraulic processes and MiniCPS to emulate the cyber physical system.

The main components of DHALSIM include a parser, a physical simulator, a cyber emulator, and a file generator (Figure 1). Experiments in DHALSIM run in the following way: first, the configuration files are created for the experiment. Second, the Parser reads these configuration files and creates a Mininet network and concurrently launches a process to run the physical

CC BY NC SA

simulation. The Mininet network consists of Mininet nodes, running scripts with the behaviour of PLCs, or SCADA, and network links connecting these nodes. On the other hand, the physical simulation is an EPANET instance running in a step-by-step fashion. Third, during the simulation, different events or attacks can be launched. The experiment ends when the end of the physical simulation is reached, and the resulting output files are stored.

Three types of input files are used to launch an experiment:

- EPANET Input file: this is the EPANET input file of the water distribution used in the simulation.
- Experiment configuration files: two files constitute the configuration files. The first one is the general config file, which defines global parameters such as the path to the EPANET input file, the number of hydraulic time step iterations the experiment will run for, the type of simulation used (demand-driven or pressure-driven), and the paths to additional configuration files.
- Optional attacks and events files: DHALSIM can process these optional files in order to launch cyber physical attacks or network anomalies during an experiment. Network anomalies are network events that affect the way a network link behaves. For example, users could launch anomalies that cause a percentage of packets to be lost at a network link, or a network delay between each packet being sent. Cyber-physical attacks can be of two types: PLC attacks or network attacks. The former force a PLC to operate hydraulic actuators (such as pump) in a way that ignores the system control rules, while the latter launch an additional network node that executes a script running the network attack.

3 EXPERIMENTAL SETUP

The objective of DHALSIM is to provide a co-simulation environment for water distribution systems that combines physical simulation with network emulation capabilities. Providing network emulation is important, because the network behaviour might affect the physical response of a water distribution system. Using DHALSIM, we generated a dataset with 52 weeks of simulation of the C-Town water distribution system under normal operating conditions. In addition, the dataset contains 8 more weeks of network anomalies and cyber physical attacks. Here, we first describe the benchmark water distribution system and then illustrate the cyber-physical attacks included in the dataset.

3.1 CASE STUDY: C-TOWN AND BATADAL

To create the dataset, we began by extending the traditional C-Town topology to include 9 PLCs controlling the system behaviour, as illustrated in [2,5]. Figure 2 shows the resulting C-Town cyber physical system. The left panel illustrates the topology of the water distribution system along with the location of the PLCs controlling the system. The right panel illustrates the corresponding network topology deployed in MiniCPS. In the network topology, each PLC is located in a local area network, and all local networks are connected through a central router. In addition, a SCADA server is located at another substation. This SCADA server polls all the PLCs for information regarding the physical state of the system.

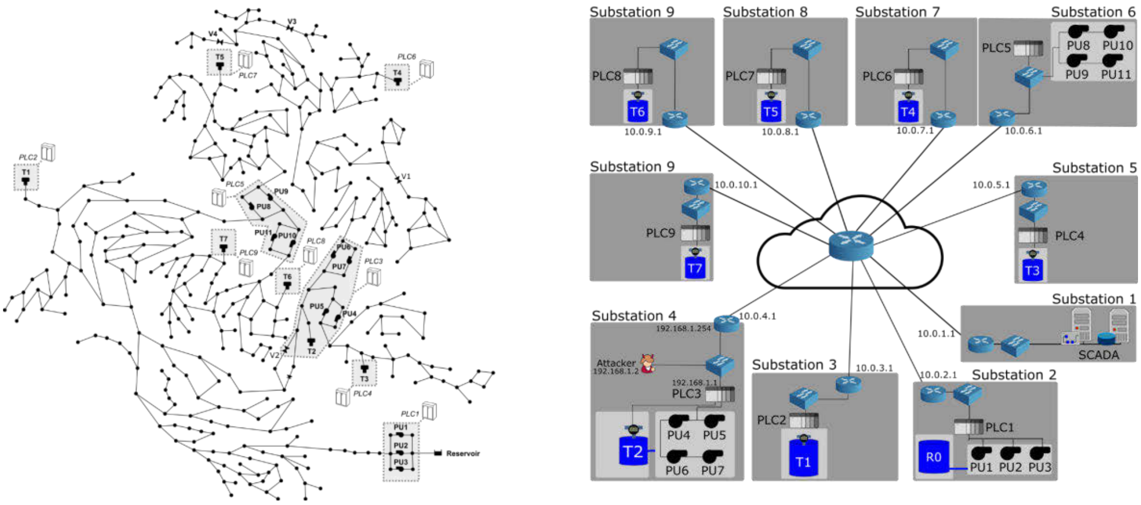

V3
V4
T5
PLC7
PLC6
T4
PLC2
T1
V1
PLC5
PU9
PU8
PU11
PU10
PLC8
T7
PLC9
T6
PU6
PU7
PLC3
PU5
PU4
T2
V2
T3
PLC4
PLC1
PU1
PU2
PU3
Reservoir
Substation 9
Substation 8
Substation 7
Substation 6
PLC8
PLC7
PLC6
PLC5
PU8
PU10
PU9
PU11
T6
T5
T4
10.0.9.1
10.0.8.1
10.0.7.1
10.0.6.1
Substation 9
10.0.10.1
10.0.5.1
Substation 5
PLC9
PLC4
T7
T3
Substation 1
Substation 4
10.0.4.1
192.168.1.254
10.0.1.1
SCADA
Attacker
192.168.1.2
192.168.1.1
PLC3
Substation 3
10.0.3.1
10.0.2.1
Substation 2
PLC2
PLC1
PU4
PU5
T2
PU6
PU7
T1
R0
PU1
PU2
PU3

Figure 2. C-Town Cyber-physical system. The left panel shows the topology of the water distribution system. The right panel shows the network topology and the local area networks with PLCs, sensors, and actuators

As mentioned above, the dataset presented in this paper is inspired by the BATADAL [2], which we aim to extend by providing a thorough representation of both physical and cyber processes. The no attack conditions were generated using 52 different weekly demand patterns, with a pattern timestep of 15 minutes. In addition, for each week, different initial tank level conditions were configured. Finally, for each week, we introduced small network benign anomalies in the links connecting all PLCs and SCADA. These included a percentage of packet loss lower than 1% and a network delay of less than 30ms in all links (thus representing typical operating conditions of the cyber layer). We run all simulations using a hydraulic timestep of five minutes, and by resorting to pressure-driven hydraulic simulations [8] with EPANET 2.2 and epynet, a Python wrapper developed by Vitens[1].

3.2 CYBER-PHYSICAL ATTACKS AND NETWORK ANOMALIES

In addition to the 52 normal operating conditions, we ran 8 weeks of simulation under cyber-physical attacks and network anomalies. All simulations were run using the hydraulic boundary conditions of Week 6, to compare the physical and network impact of the anomalies under the same physical and network conditions (i.e., tank levels, demand patterns, percentage of packet loss, and network delay). Table 1 shows the details of the simulated attacks and benign anomalies, collectively named anomalies for simplicity. We grouped the four types of anomalies reported in Table 1 with respect to two different effects (or objectives). All the anomalies with an odd ID number may cause Tank T3 to run empty, while all anomalies with an even ID number may cause Tank T3 to overflow.

Attack ID	Duration (hours)	Anomaly Description	Anomaly Type	Objective/Effect
1	31	MiTM on PLC4 to manipulate T3	MiTM Attack	Empty T3
2	31	MiTM on PLC4 to manipulate T3	MiTM Attack	Overflow T3

---

[1] https://github.com/Vitens/epynet

BY NC SA

3	31	Malicious activation Pump PU4	PLC Attack	Empty T3
4	31	Malicious activation Pump PU4	PLC Attack	Overflow T3
5	31	DoS on PLC3	Dos Attack	Empty T3
6	31	DoS on PLC3	Dos Attack	Overflow T3
7	31	Network anomaly on PLC3	Network Anomaly	Empty T3
8	31	Network anomaly on PLC3	Network Anomaly	Overflow T3

Table 1. Attack details. The attacks have two objectives, that is, to empty or to overflow Tank 3. The acronyms MiTM and DoS refer to Man-in-The-Middle and Denial-of-Service respectively.

Figure 3 shows an overview of the events run during these experiments. The first anomaly (top left panel) is a Man-in-the-Middle attack (MiTM). In this case, an attacker manipulates the sensor readings reaching a PLC. Specifically, the attacker manipulates Tank T3 readings sent by PLC4 and received by PLC3. Since PLC3 activates Pumps PU3 and PU4 (according to the T3 reading), this manipulation can cause PLC3 to make the wrong control decision, emptying or overflowing the tank. The second anomaly (top right panel) is a PLC attack. An attacker gains control of PLC3 and causes a malicious activation of Pump PU4. This manipulation can lead PU4 to being closed or open during the entire anomaly duration, causing Tank T3 to empty or overflow. The third anomaly (bottom left) is a Denial-of-Service Attack (DoS). In this attack, the attacker intercepts Tank T3 readings destined to PLC3, causing these messages not to be received by PLC3. This causes PLC3 to operate the pumps with outdated information. That is, Pumps PU3 and PU4 are activated with the last value before the attack was launched. Depending on the state of the system (T3 emptying or filling), this might cause T3 to empty or overflow. Finally, the fourth anomaly (bottom right) is a Network Anomaly. In this case, there is no attacker present, and the anomaly causes PLC3 to be disconnected from the network and to stop receiving updates on Tank T3 water level. This causes a similar physical response to the previously describe DoS attack.

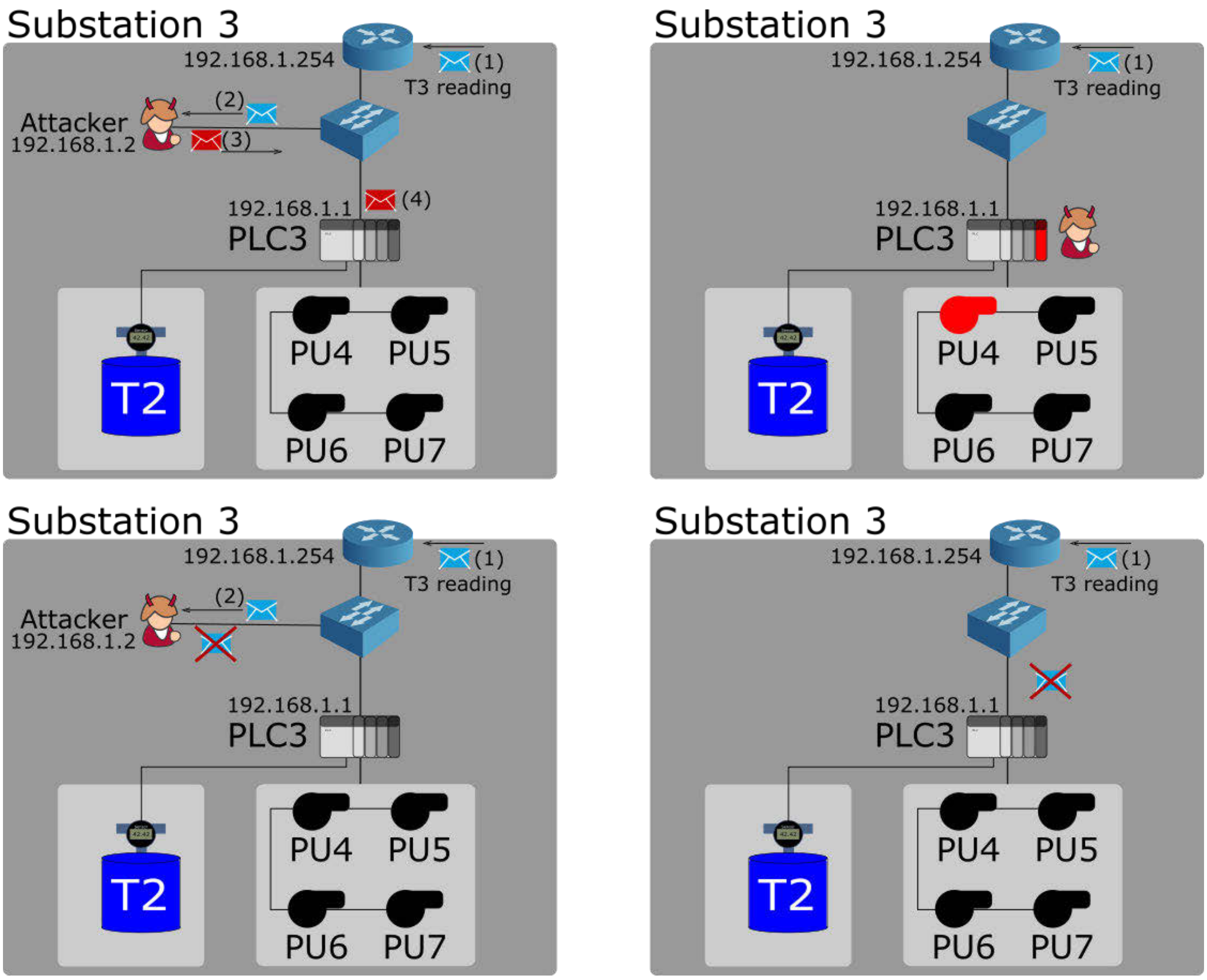

Substation 3
192.168.1.254
(1)
T3 reading
Attacker
192.168.1.2
(2)
(3)
192.168.1.1
(4)
PLC3
T2
PU4
PU5
PU6
PU7
Substation 3
192.168.1.254
(1)
T3 reading
192.168.1.1
PLC3
T2
PU4
PU5
PU6
PU7
Substation 3
192.168.1.254
(1)
T3 reading
Attacker
192.168.1.2
(2)
192.168.1.1
PLC3
T2
PU4
PU5
PU6
PU7
Substation 3
192.168.1.254
(1)
T3 reading
192.168.1.1
PLC3
T2
PU4
PU5
PU6
PU7

Figure 3. Cyber-physical attacks and network anomalies used in the experiments. All anomalies have PLC3 as the target. The top left panel shows a Man-in-the-Middle Attack. The top right panel shows a PLC attack. The bottom left panel shows a Denial of Service attack. Finally, the bottom right panel shows a network anomaly.

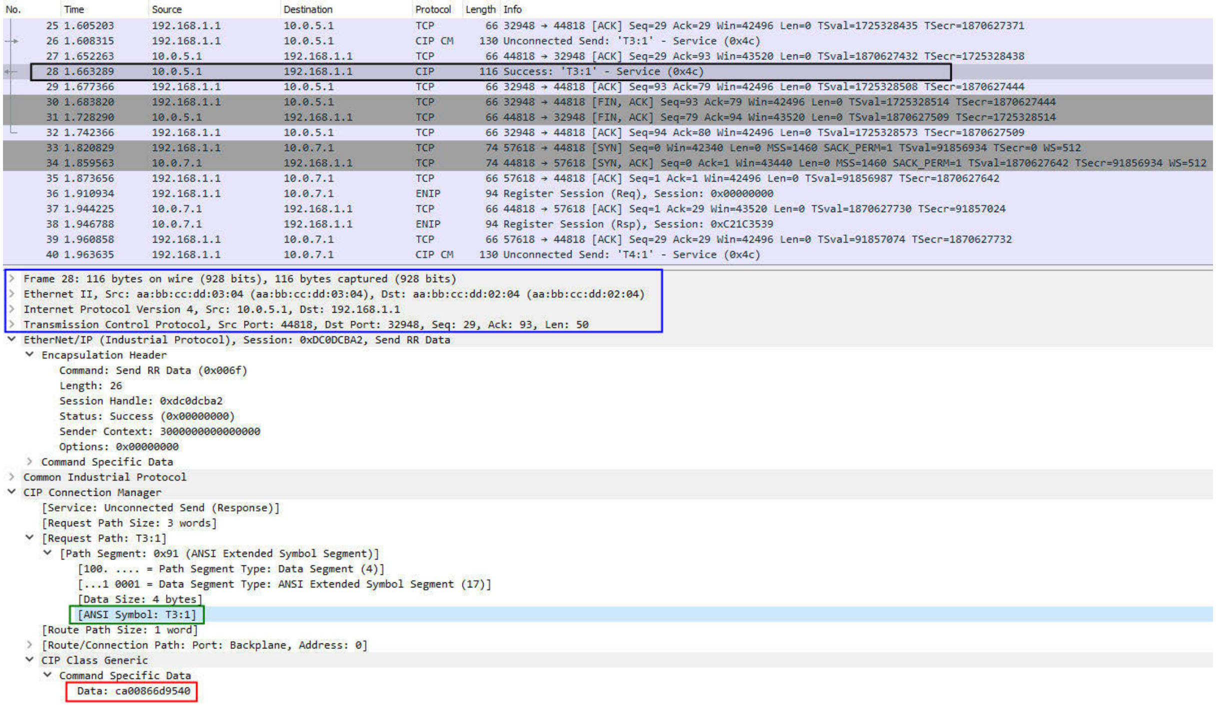
4 KEY FEATURES

4.1 ANALYSIS OF NETWORK DATA

In addition to the physical results, DHALSIM generates network captures of all the messages exchanged between PLCs and SCADA during the simulation. DHALSIM has the capability to generate these files, because MiniCPS and Mininet offer a full implementation of common industrial protocols used in cyber-physical systems, including water distribution systems. Figure 4 shows a screenshot of a CIP[2] packet dissection carried out with the network analyzer Wireshark. The dissected message was received by PLC3 and sent by PLC4 (PLC4 has an outbound IP of 10.0.5.1). This particular message carries the value of Tank T3 level. The general information of this message is highlighted in the black box in the figure. In addition, the blue box highlights the stack of protocols supporting the CIP protocol. These are: Ethernet, IIP, TCP, and ENIP. Finally, the dissection shows additional details of the CIP message, including the hexadecimal reading of the tank level (highlighted in red). This value is decoded by the MiniCPS library into the float value reported by sensor T3. Accurately generating this network information is important because the network behaviour can have impacts on the physical system. In a similar way, network information can help identify the root cause of physical anomalies or cyber-physical attacks.

---

[2] The Common Industrial Protocol (CIP) is a common industrial protocol for industrial automation applications.



```
No.    Time        Source        Destination   Protocol  Length Info
    25 1.605203    192.168.1.1   10.0.5.1      TCP         66 32948 → 44818 [ACK] Seq=29 Ack=29 Win=42496 Len=0 TSval=1725328435 TSecr=1870627371
    26 1.608315    192.168.1.1   10.0.5.1      CIP CM     130 Unconnected Send: 'T3:1' - Service (0x4c)
    27 1.652263    10.0.5.1      192.168.1.1   TCP         66 44818 → 32948 [ACK] Seq=29 Ack=93 Win=43520 Len=0 TSval=1870627432 TSecr=1725328438
    28 1.663289    10.0.5.1      192.168.1.1   CIP        116 Success: 'T3:1' - Service (0x4c)
    29 1.677366    192.168.1.1   10.0.5.1      TCP         66 32948 → 44818 [ACK] Seq=93 Ack=79 Win=42496 Len=0 TSval=1725328508 TSecr=1870627444
    30 1.683820    192.168.1.1   10.0.5.1      TCP         66 32948 → 44818 [FIN, ACK] Seq=93 Ack=79 Win=42496 Len=0 TSval=1725328514 TSecr=1870627444
    31 1.728290    10.0.5.1      192.168.1.1   TCP         66 44818 → 32948 [FIN, ACK] Seq=79 Ack=94 Win=43520 Len=0 TSval=1870627509 TSecr=1725328514
    32 1.742366    192.168.1.1   10.0.5.1      TCP         66 32948 → 44818 [ACK] Seq=94 Ack=80 Win=42496 Len=0 TSval=1725328573 TSecr=1870627509
    33 1.820829    192.168.1.1   10.0.7.1      TCP         74 57618 → 44818 [SYN] Seq=0 Win=42340 Len=0 MSS=1460 SACK_PERM=1 TSval=91856934 TSecr=0 WS=512
    34 1.859563    10.0.7.1      192.168.1.1   TCP         74 44818 → 57618 [SYN, ACK] Seq=0 Ack=1 Win=43440 Len=0 MSS=1460 SACK_PERM=1 TSval=1870627642 TSecr=91856934 WS=512
    35 1.873656    192.168.1.1   10.0.7.1      TCP         66 57618 → 44818 [ACK] Seq=1 Ack=1 Win=42496 Len=0 TSval=91856987 TSecr=1870627642
    36 1.910934    192.168.1.1   10.0.7.1      ENIP        94 Register Session (Req), Session: 0x00000000
    37 1.944225    10.0.7.1      192.168.1.1   TCP         66 44818 → 57618 [ACK] Seq=1 Ack=29 Win=43520 Len=0 TSval=1870627730 TSecr=91857024
    38 1.946788    10.0.7.1      192.168.1.1   ENIP        94 Register Session (Rsp), Session: 0xC21C3539
    39 1.960858    192.168.1.1   10.0.7.1      TCP         66 57618 → 44818 [ACK] Seq=29 Ack=29 Win=42496 Len=0 TSval=91857074 TSecr=1870627732
    40 1.963635    192.168.1.1   10.0.7.1      CIP CM     130 Unconnected Send: 'T4:1' - Service (0x4c)

> Frame 28: 116 bytes on wire (928 bits), 116 bytes captured (928 bits)
> Ethernet II, Src: aa:bb:cc:dd:03:04 (aa:bb:cc:dd:03:04), Dst: aa:bb:cc:dd:02:04 (aa:bb:cc:dd:02:04)
> Internet Protocol Version 4, Src: 10.0.5.1, Dst: 192.168.1.1
> Transmission Control Protocol, Src Port: 44818, Dst Port: 32948, Seq: 29, Ack: 93, Len: 50
v EtherNet/IP (Industrial Protocol), Session: 0xDC0DCBA2, Send RR Data
  v Encapsulation Header
      Command: Send RR Data (0x006f)
      Length: 26
      Session Handle: 0xdc0dcba2
      Status: Success (0x00000000)
      Sender Context: 3000000000000000
      Options: 0x00000000
  > Command Specific Data
> Common Industrial Protocol
v CIP Connection Manager
    [Service: Unconnected Send (Response)]
    [Request Path Size: 3 words]
  v [Request Path: T3:1]
    v [Path Segment: 0x91 (ANSI Extended Symbol Segment)]
        [100. .... = Path Segment Type: Data Segment (4)]
        [...1 0001 = Data Segment Type: ANSI Extended Symbol Segment (17)]
        [Data Size: 4 bytes]
        [ANSI Symbol: T3:1]
    [Route Path Size: 1 word]
  > [Route/Connection Path: Port: Backplane, Address: 0]
  v CIP Class Generic
    v Command Specific Data
        Data: ca00866d9540
```

Figure 4. Wireshark packet retrieved for one of the routers in C-Town. The dissected packet shows a CIP packet and its payload (i.e., the value of Tank 3 level, black/green boxes). The message also shows the stack of protocols supporting CIP (blue box) and the hexadecimal reading of the tank level (red box).

Recall that DHALSIM generates both physical and network data. For the physical data, the model creates one file with ground truth values and one file with SCADA values. The former contains the values of variables handled by the EPANET simulator, namely, tank level, pressure at junctions, status and flow of valves and pumps, time-stamp, and a variable indicating the status of the water system (i.e., normal operating conditions or under attack / anomaly). The latter stores the values of the variables received by the SCADA server (as specified in the PLCs configuration files). Both files have a *.csv* format. For a single one-week simulation they have a size of about 20 MB. As for the network data, DHALSIM creates one network capture file for each PLC and SCADA server. Such file stores all network messages sent and received by that node in *.pcap* format, which allows software libraries to retrieve and process the network packets. In our case, the *.pcap* files processing is carried out with a software library named *scapy*. In addition to *scapy*, we used two specific parsers for ENIP and CIP messages [9]. For a single one-week simulation, a PLC *.pcap* file has a size of 30MB.

4.2 NO ATTACK CONDITIONS

Figure 5 shows the envelope of variability of the 52 weeks of normal operating conditions for the water levels of Tanks T1. The envelope of variability was calculated by obtaining the minimum and maximum value at each iteration (or hydraulic timestep). The behaviour of the envelope is explained by the different initial conditions (tank levels) and by the different demand patterns, as those are the ones driving the system behaviour. The blue line highlights the trajectory of the water level in T1 for Week 6.

BY NC SA

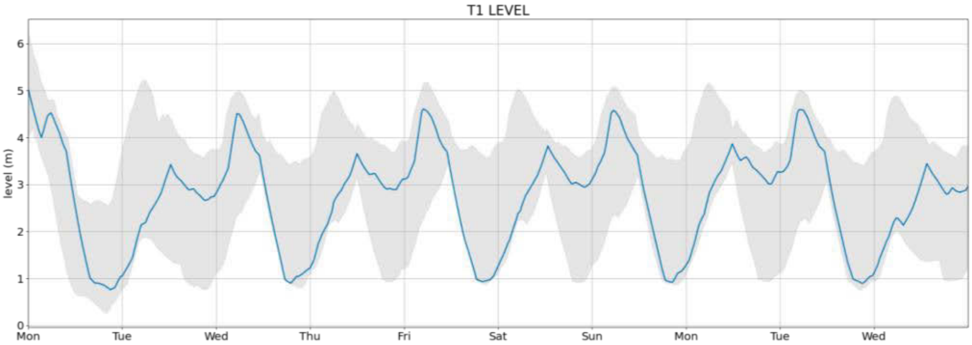


Figure 5. Envelop of variability for Tank T1 for all normal conditions. For the 52 weeks of normal operating conditions, different initial tank levels and demand patterns were used. The blue line shows the trajectory of the water level in T1 for Week 6.

4.3 PHYSICAL AND NETWORK RESULTS OF THE ANOMALIES

The main objective of all the anomalies presented in this dataset is to show how DHALSIM extends the type of data included in the BATADAL dataset. Since DHALSIM has network emulation capabilities, similar anomalies in the physical processes may derive not only from different cyber-physical attacks, but also from benign network anomalies.

The results of the anomalies are shown in Figures 6 and 7. In particular, Figure 6 shows the physical and network results of all the anomalies resulting in an empty T3, while Figure 7 shows the physical and network results of all the anomalies leading to an overflow in T3. Beginning with Figure 6, the left panels show the water level of Tanks T1 and T3 during the weeks featuring different anomalies. The anomalies occur during the grey highlighted area in the plots. The figure shows that the physical responses are similar for almost all the anomalies, except for the PLC3 Attack. The process followed by the victim PLC in all anomalies can be summarized as follows: for Anomaly 1 (MiTM Attack), PLC3 receives a modified value of T3. This modified value (7.0m), causes PLC3 to turn off both pumps (PU3 and PU4) and as a consequence, empties T3. For Anomaly 3 (PLC Attack), the PLC3 operates maliciously on PU4, causing its shut down. Note that PU3 keeps working normally and, consequently, Tank T3 does not run empty, although water levels are lower than expected. Anomalies 5 and 7 have a similar physical effect—PLC3 stops receiving updated information regarding T3 and does not activate the pumps properly. By timing the beginning of these anomalies, we can control whether the tank will be emptied or overflown.

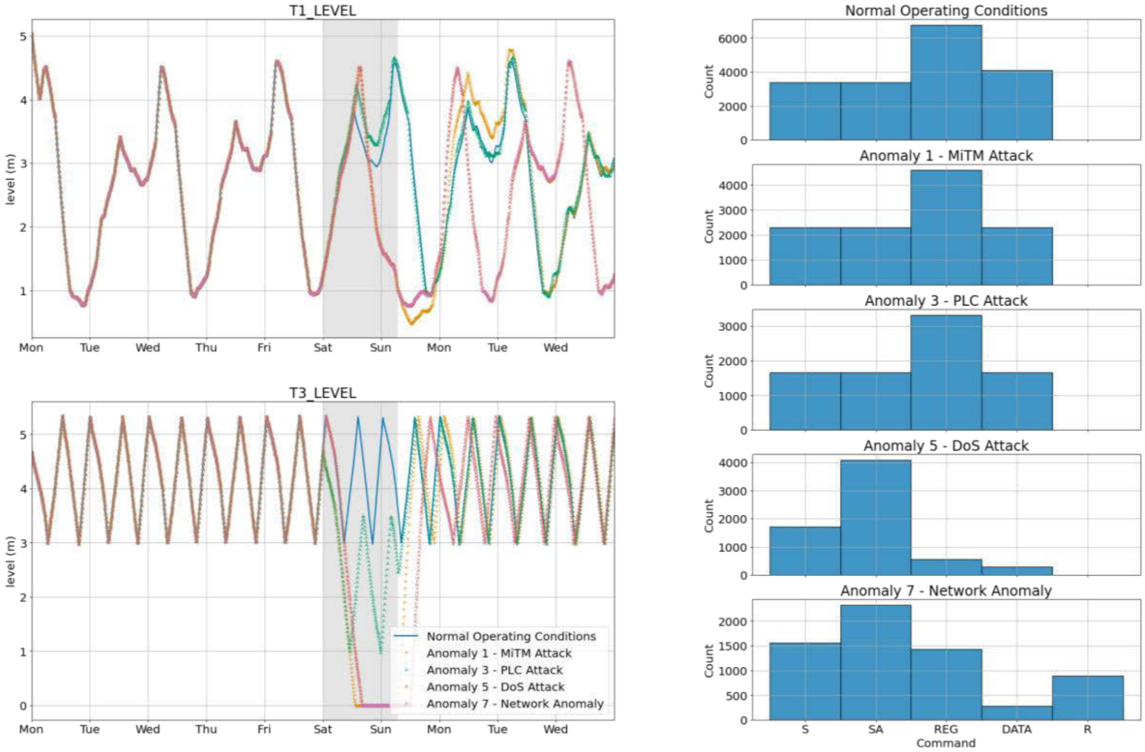


Figure 6. The panels on the left show the water tank levels of Tank T1 and T3. All physical responses are similar, but the anomalies used to obtain them have different vectors. The panels on the right show a histogram of the number of messages received by PLC3 during the anomalies.

The right panels of Figure 6 show a histogram of the different types of Command messages received by PLC3 during the anomalies. A "Command" is a network feature extracted using the libraries specified in Section 4.1. This feature indicates the objective or type of network message. For example, a TCP-SYN message (a special type of message used by the TCP protocol to establish a network connection), would receive the "Command" value of "S". An "SA" is the acknowledgement of an "S" message. "REG" messages are messages requesting for a sensor or actuator reading. "DATA" messages are the messages carrying the sensor or actuator readings, and "R" messages are messages sent when a connection is abruptly closed or cannot be established. The network behaviour depends on the type of anomaly generated, as the panels show. For example, for Normal operating conditions and anomalies 1 and 3 there are no changes in the patterns of the packets received by PLC3. This is because the PLC attack does not generate changes in the network patterns. Instead, PLC3 operates Pump PU4 ignoring the configured control rules. In a similar way, Anomaly 1 (MiTM Attack) simply manipulates the payload of the messages with the T3 readings—all other messages and communication behaviour are unaffected.

The more significant network differences happen with anomalies 5 and 7. The DoS attack causing anomaly 5 stops PLC3 from receiving all "REG" and "DATA" commands. This happens because the DoS stops all messages arriving at PLC3; the attacker drops even the messages that are sent back by the switch connected to PLC3. In anomaly 7, the network issue happens between the router and the PLC3, which allows for some messages to reach PLC3. Nevertheless, PLC3 has a significant drop in the "DATA" messages received and "R" messages are sent by the PLC3, because none of the TCP connections are established successfully. We can draw similar conclusions by analysing the remaining four anomalies in Figure 7.

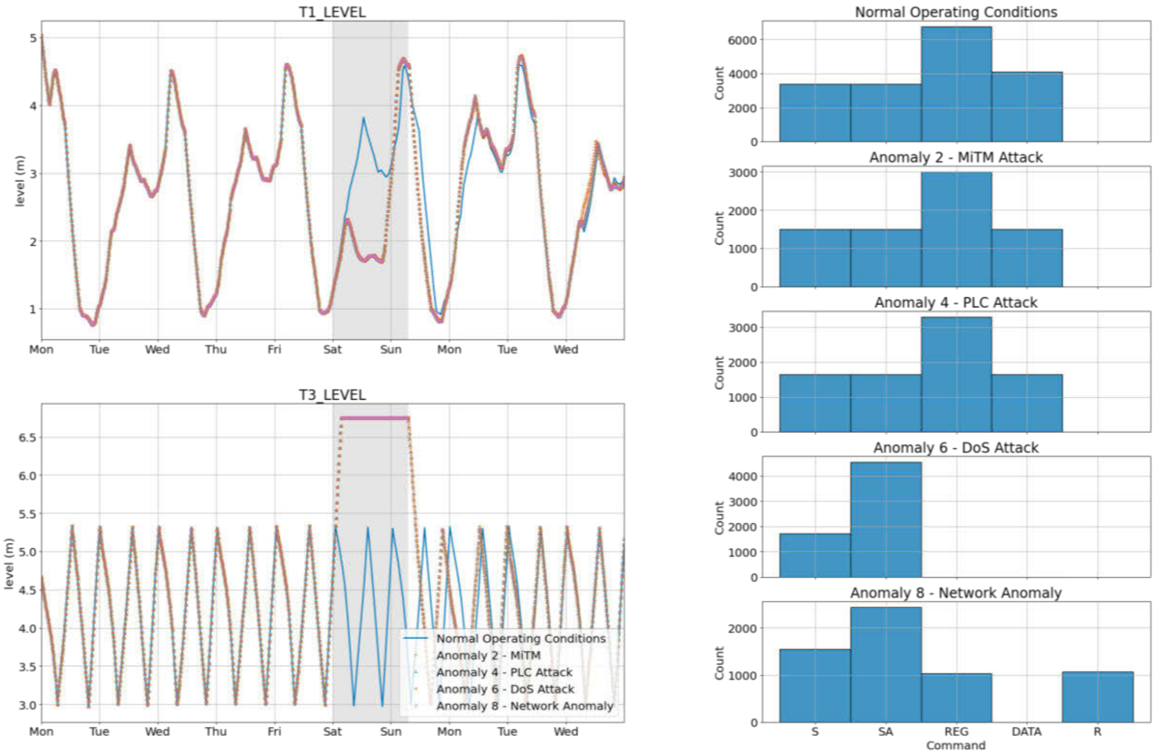


Figure 7. The panels on the left show the water levels of Tanks T1 and T3. All physical responses are similar, but the anomalies used to obtain them have different vectors. The panels on the right show a histogram of the number of messages received by PLC3 during the anomalies.

5 OUTLOOK

This paper presents the first version of the BATADAL 2.0 dataset generated with the DHALSIM simulator. As explained above, our ultimate goal is to provide a thorough dataset that can support the development of a new generation of intrusion detection systems. The next step of our research will therefore focus on such systems. Another important step would be to extend DHALSIM with attack concealment capabilities, to fully replicate the attacks generated in BATADAL and to extend those attacks with more attack vectors, as enabled by DHALSIM. We finally note that the dataset presented in this paper is available online at https://zenodo.org/record/6545035.

6 ACKNOWLEDGMENTS

This research is supported by Singapore's NATIONAL SATELLITE OF EXCELLENCE, DESIGN SCIENCE AND TECHNOLOGY FOR SECURE CRITICAL INFRASTRUCTURE (NSoE DeST-SCI) through the project "LEarning from Network and Process data to secure Water Distribution Systems (LENP-WDS)" (Award No. NSoE_DeST-SCI2019-0003). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

7 REFERENCES

[1] Hassanzadeh, A., Rasekh, A., Galelli, S., Aghashahi, M., Taormina, R., Ostfeld, A., and Banks, K. (2020). "A Review of Cybersecurity Incidents in the Water Sector". Journal of Environmental Engineering 146, 5, 03120003.

[2] Taormina, R., Galelli, S., Tippenhauer, N.O., Salomons, E., Ostfeld, A., Eliades, D., et al. (2018). "Battle of the Attack Detection Algorithms: Disclosing Cyber Attacks on Water Distribution Networks." Journal of Water Resources Planning and Management 144, 8, 04018048.

[3] Taormina, R., Galelli, S., Douglas, H.C., Tippenhauer, N.O, Salomons, E., and Ostfeld, A. (2019). "A Toolbox for Assessing the Impacts of Cyber-physical Attacks on Water Distribution Systems". Environmental Modelling & Software 112 (2019), 46-51.

[4] Nikolopoulos, D., Moraitis, G., Bouziotas, D., Lykou, A., Karavokiros, G., and Makropoulos, C. (2020). "Cyber-Physical Stress-Testing Platform for Water Distribution Networks." Journal of Environmental Engineering, 146(7), 04020061.

[5] Murillo, A., Taormina, R., Tippenhauer, N., and Galelli, S. (2020). "Co-Simulating Physical Processes and Network Data for High-Fidelity Cyber-Security Experiments." Sixth Annual Industrial Control System Security (ICSS) Workshop, 13–20.

[6] Antonioli, D. and Tippenhauer, N. O. (2015). "MiniCPS: A Toolkit for Security Research on CPS Networks." Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy, CPS-SPC '15, New York, NY, USA, Association for Computing Machinery, 91–100.

[7] Lantz, B., Heller, B., and McKeown, N. (2010). "A Network in a Laptop: Rapid Prototyping for Software-Defined Networks." Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics

[8] Douglas, H.C., Taormina, R., and Galelli, S. (2019). "Pressure-driven modeling of cyber-physical attacks on water distribution systems." Journal of Water Resources Planning and Management 145, 3, 06019001.

[9] Urbina, D. I., Giraldo, J. A., Tippenhauer, N. O., and Cárdenas, A. A. (2016). "Attacking Fieldbus Communications in ICS: Applications to the SWaT Testbed." SG-CRC.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14799

PATHOINVEST: PATHOGEN CONTAMINATION INVESTIGATIONS DURING EMERGENCIES

Sotirios Paraskevopoulos[1], Stelios Vrachimis[2], Marios Kyriakou[3], Pavlos Pavlou[4], Dimitris Kouzapas[5], George Milis[6], Patrick Smeets[7], Demetrios G. Eliades[8], Gertjan Medema[9], Marios Polycarpou[10] and Christos Panayiotou[11]

[1,2,3,4,5,8,10,11]KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
[10,11]Electrical and Computer Engineering Department, University of Cyprus, Nicosia, Cyprus
[7,9]KWR Watercycle Research Institute, Nieuwegein, The Netherlands
[6] Department of Water Management, Delft University of Technology, Delft, The Netherlands
PHOEBE Research and Innovations Ltd, Cyprus

[1] *Sotirios.Paraskevopoulos@kwrwater.nl*, [2] *vrachimis.stelios@ucy.ac.cy*, [3] *kiriakou.marios@ucy.ac.cy*, [4] *pavlou.v.pavlos@ucy.ac.cy*, [5] *kouzapas.dimitrios@ucy.ac.cy*, [6] *george.milis@phoebeinnovations.com*, [7] *patrick.smeets@kwrwater.nl*, [8] *eldemet@ucy.ac.cy*, [9] *gertjan.medema@kwrwater.nl*, [10] *mpolycar@ucy.ac.cy*, [11] *christosp@ucy.ac.cy.*

Abstract

Emergencies and disasters (such as earthquakes and floods), may contaminate drinking water systems with pathogens, that can affect the health of both First Responders and Citizens. As part of the Horizon 2020 "Pathogen Contamination Emergency Response Technologies" (PathoCERT) project, we are developing a Digital Twin tool (PathoINVEST) to assist First Responders and Water Authorities in investigating and responding efficiently to drinking water contamination events. In this paper, we present preliminary work on PathoINVEST, its architecture, and how it operates with the PathoCERT ecosystem of technologies. Moreover, using an illustrative case study, we demonstrate how PathoINVEST will process data and produce useful insights for the First Responders during a realistic contamination event. This work demonstrates how different research results can be integrated into a holistic water contamination emergency management system, in accordance with the needs of First Responders who need to make decisions within a limited time frame and to reduce the impact of a contamination event.

Keywords
Water contamination, first responders, emergency response, digital twins.

1 INTRODUCTION

During disasters and emergencies, water systems can be unexpectedly exposed to pathogens [1, 2]. For example, during an earthquake, the drinking and sewerage system of a city may be affected by pipe breaks, which may cause infiltration of sewage water into the drinking water network. Floods after intense rainfall may carry away toxic substances, whereas, in the case of technological accidents, or even bioterrorism attacks, unknown pathogens may be injected into the drinking water supply.

During these emergencies, citizens, as well as First Responders operating in the area, may become exposed to contaminated water, through skin contact, ingestion, or inhalation. This can pose a significant risk of illness, disease, or even death. In these situations, water may be contaminated with pathogens such as Cryptosporidium, E. coli O157:H7, Norovirus, and Vibrio cholerae [3]. Some recent examples of these emergency events include an earthquake causing infiltration of dirt in the Drinking Water Network (DWN) (Larisa, Greece, March 2021), flash-flood water overflowing a wastewater treatment plant which went through the city of Merrit (British

Columbia, Canada, November 2021), and hospital sewage infiltrating a water distribution network (Prague, May 2015).

To effectively manage these situations, and to reduce the loss of human lives, First Responders and Water Authorities need to be equipped with the appropriate emergency response technologies. This is also the goal of the project "Pathogen Contamination Emergency Response Technologies" (PathoCERT), which is funded by the European Union under the Horizon 2020 programme. PathoCERT aims to develop new technologies for fast alerting and detection of pathogens in surface and drinking water, using smart portable sensors, as well as develop new technologies for improving situational awareness using drones, satellites, social media, and smart cameras. A new intelligent IoT gateway and an open-standards (FIWARE-based) platform will be designed to support the collection and analysis of these heterogeneous data. The platform is connected with tools for threat and risk assessment, as well as for investigating contamination events and proposing the most appropriate mitigation actions. Eventually, these tools will increase the First Responder and Water Operator capabilities and reduce the exposure to pathogens during an emergency, which can have a significant impact on human health.

This paper presents an overview of related research on the topic of water contamination emergency response management and introduces PathoINVEST, a Digital Twin Platform for investigating urban drinking water contamination events. PathoINVEST is a collection of tools that incorporates sensor telemetry and hydraulic models to create an up-to-date state-estimation of the network model. Additionally, it uses a set of modeling tools for forecasting contaminant evolution during emergencies, health risk assessment, mobile/portable sensor deployment, and evaluation of mitigation measures to reduce exposure and health risk.

The main contributions of this work are a) the presentation of the PathoINVEST reference architecture, and b) the release of a complete contamination emergency management case study using the L-Town benchmark network.

This paper is structured as follows: In Section 2, a literature review on pathogen contamination emergency response technologies is provided highlighting the key elements that constitute an emergency response (from preparedness and contamination modeling to assessing the risk and applying mitigation measures). In Section 3, the architecture of the PathoINVEST digital twin system and the PathoCERT ecosystem is presented, whereas Section 4 presents an illustrative case study using the L-Town benchmark and example implementations, of how a possible contamination event could be managed using the proposed tool.

2 BACKGROUND ON PATHOGEN CONTAMINATION EMERGENCY RESPONSE TECHNOLOGIES

To deal with emergencies, such as floods, earthquakes, accidents, or even attacks, First Responders and Water Authorities need to coordinate to effectively manage the situation. To achieve this, procedures and technologies need to be in place to capture the different phases of the event. These phases include: 1) monitoring and detecting contamination events, 2) assessing the threat and risk, 3) identifying the source, and 4) mitigating the contamination event. For each phase, significant research has been conducted during the previous years, and this section summarizes some important works.

2.1 Monitoring and detecting contamination events

Due to their high costs, a key challenge is to decide what type of sensors to install, how many, and in which locations in the network, considering all the uncertainties. Another related challenge is where and when to conduct manual sampling for lab analysis. These problems have been studied extensively, and are typically expressed as optimization problems, sometimes with multiple objectives which may include impact risk, detection time, and coverage [4, 5, 6, 7].

After the placement of the sensors, online monitoring tools can be employed to continuously analyze the readings and determine changes in water quality, which could be attributed to contamination events. In the literature, both model-based and data-driven approaches have been proposed, considering one or more water quality parameters [8, 9]. In addition to sensors monitoring physical parameters within water systems, researchers in social networks study how citizens can act as sensors. They focus on what information can be extracted from their networks to improve the ability to determine the extent of a contamination event and to allocate the appropriate response mechanism [10, 11, 12].

2.2 Threat and Risk Assessment

Water utilities are increasingly required to establish Water Security Plans (WSP). That includes preparations to manage emergencies that threaten their system. According to the World Health Organization (WHO) [13], risk assessment is an integral part of developing and implementing a WSP. Many factors can cause adverse effects to this critical infrastructure through the introduction of microbiological, chemical, or radiological hazards. Therefore, risk assessment is considered imperative to facilitate the evaluation of health risks associated with contamination of water supply and to assist responsible authorities in controlling and mitigating an event. One form of risk assessment that has received attention over the last two decades and has been embedded in the WHO water-related guidelines [14] is the Quantitative Microbial Risk Assessment (QMRA). QMRA is a mathematical framework for evaluating infectious risks by combining scientific knowledge about pathogens (fate, transport, route of exposure, and health effects of human pathogens) with the effect of physical/mechanical barriers and mitigation actions [13]. There are 4 steps associated with QMRA, summarized in Table 1.

Table 1. Steps of QMRA (adapted from WHO [13]).

Step	***Description***
1. Problem formulation	Highlight the reference pathogens, exposure pathways, contamination events, and health outcomes of interest
2. Exposure assessment	Measure the dose and frequency of contaminant to which people are exposed (via the identified exposure pathway)
3. Health effects assessment	Identify a dose-response relationship (linking dose to the probability of infection/illness) for the reference pathogen.
4. Risk characterization	Combine the information from the exposure and health effects assessment to generate a quantitative measure of risk

Scientists are researching the combination of QMRA with modeling tools to assess the risk of a drinking water contamination event for the population. The research ranges from main breaks and wastewater intrusion in the network due to pressure transients (unintentional contamination) to hypothetical intentional wastewater injection at key locations. The work from [15] assessed the risk due to wastewater intrusion after negative pressure transients. [16, 17] combined modeling with QMRA using stochastic water demands to account for contamination events after main repairs in the DWN. The authors concluded that the initial contaminant concentration determines highly the exposure, while the infection risk is determined by the most infectious pathogen dose-response. Finally, other works assessed the effects and exposure scenarios after deliberate microbiological contamination in a DWN [18, 19]. Specifically, the work in [18] used a variation of QMRA by modeling the number of affected consumers and highlighting the most critical areas (in terms of exposure) within a network, while [19] investigated the effects of duration, concentration, exposure pathway, and pathogen infectivity on exposure and infection risk.

2.3 Contamination source identification

During chemical or microbial contamination in a DWN, the ability of a water utility to have an early indication of the potential source of contamination is of utmost importance since it can facilitate mitigation measures to stop the spreading and isolate the contaminant. The *Contamination Source Identification* (CSI) problem is mainly considered a deterministic inverse problem, where, using hydraulic calculations, water parcels are backtracked to reach the source of contamination [20]. It is considered a challenging problem due to the computational burden associated with hydraulic calculations, hydraulic uncertainties, and the non-uniqueness of the solutions in identifying the source [20].

A first attempt to solve the CSI was in [21], which introduced the particle backtracking approach. The authors used a particle backtracking algorithm based on a Lagrangian model where the contaminants were considered as particles that run in reverse time from the detection node to the source of contamination.

An alternative response action to tackle a contamination event in the DWN is the option of expanded sampling. Water utilities can focus on examining water quality at specific locations within a DWN (after initial detection of contamination), to help evaluate the contamination impact and identify the potential source area. [22] proposed a computational approach based on decision trees to select a sequence of, as few as possible, nodes for expanded sampling, during which the contamination impact is evaluated, and the source of contamination is isolated/identified.

2.4 Mitigation Measures

After the source identification, a water utility wants to minimize the impact of confirmed contamination by using mitigation measures such as network operational interventions or Optimal Booster Chlorination Placement (OBCP).

Regarding network operational interventions, the most common is valve manipulation. For example, [23] proposed an active contamination detection system by manipulating valves to drive flows to designated nodes within a DWN, thus enabling sensors to monitor water quality. The objective function of this study was to minimize the impact on the population by detecting the contaminant as fast as possible.

Regarding OBCP, [24] presented an optimization model to tackle a (un)intentional contamination event in the DWN. The authors included the reactions of chlorine with unknown contaminants, their fate, and transport, as well as the time delay between detection and application of booster chlorination. The objective was to minimize the impact on the population specified as the number of people who ingest contaminated water above a specified mass threshold.

2.5 Modeling water quality and contaminants in drinking water networks

In most DWN, disinfection is performed to provide (microbiologically) safe drinking water and prevent water quality deterioration. For some of the methodologies described previously, it is assumed that a mathematical/computational model describes the reaction dynamics of the disinfectant agent with contaminants in the water. Both chlorine reactions and pathogen inactivation kinetics can be modeled using *EPANET-MSX*, an open-source multi-species simulator used in conjunction with the *EPANET* hydraulic simulator. EPANET-MSX considers multiple interacting species and enables the modeling of fate, transport, and reaction dynamics of biological species [25].

Various researchers have investigated the interactions of chlorine with microbial contaminants in a DWN. In [18], the authors assessed the vulnerability of a network under deliberate

BY NC SA

contamination attacks using a parallel first-order model to describe chlorine decay (coupled with a bacterial regrowth model). Moreover, in [26], the authors also modeled parallel first-order (fast and slow) chlorine reactions with microbial contaminants in a DWN. In [27], the researchers modeled chlorine reactions with pathogens (wastewater intrusion) incorporating water quality parameters (pH, temperature), while using *Giardia* and *Escherichia coli* 0157: H7 as the intrusion pathogens. The Chick-Watson equation and first-order kinetics were used to describe the inactivation of the microbial contaminants and chlorine decay respectively.

3 SOFTWARE ARCHITECTURE

The PathoCERT integrated solution is composed of different technological modules (such as sensors) as well as information systems, to assist first responders and other relevant stakeholders in communicating information, sharing knowledge, and effectively managing contamination threats and events. In PathoCERT, all software modules are connected to a FIWARE-enabled backend, the PathoWARE. The vision is for this platform to be installed beforehand and activated when an emergency occurs, following Standard Operating Procedures. Even though the details may be different in the various countries, as a general framework, we will assume that during emergencies, a Command-and-Control Center is set up and there, and a monitoring area with multiple screens is set up and operated by experts. The Center is managed by the Incident Commander, who has the role of coordinating the activities of the different First Responders, as well as the experts and the other relevant stakeholders. Other First Responders and Utility personnel may operate on the field, whereas other experts may provide specialized knowledge and support. Figure 1 illustrates the proposed architecture for the integrated PathoCERT modules and technologies, and in more detail the PathoINVEST module. In the following paragraphs, the different modules are be explained in more depth:

PathoWARE is the core of the PathoCERT Platform, based on FIWARE. It is a cloud-based solution, responsible for collecting data from heterogeneous data sources, harmonizing and processing the data to generate useful information, as well as serving data and results to the other modules of the PathoCERT platform. Data interoperability is achieved using a common PathoCERT ontology. PathoWARE will operate as a service facilitating the integration of different modules from the PathoCERT ecosystem, including the IoT gateways, situation awareness technologies for processing data from social media, as well as interfacing with wearables, mobile apps, GIS, and Decision Support tools for threat and risk assessment. In addition, PathoWARE can provide access to DWN models (provided by a water utility, that can be extracted from the GIS using middleware software).

PathoINVEST is a digital twin of the DWN which implements functionalities that support decision-making during contamination emergencies (Figure 1). These functionalities will be supported by software tools that are integrated into PathoINVEST, including a) state estimation (for estimating the hydraulic states based on the available flow and pressure measurements), b) demand forecasting (to estimate the future hydraulic dynamics), c) simulation tools for multiple species and reactions, d) optimization tools for sampling and sensor placement, e) tools to estimate the health impacts using epidemiological data and population statistics, and f) models that simulate mitigation measures (valves closing, network flushing and booster chlorination). The User Interface of PathoINVEST allows the operator (typically a modeling expert), to receive requests from the Incident Commander, who is responsible for managing the emergency. The PathoINVEST operator will manually set up the software to produce the requested outputs, and decide which information (maps, animations, figures, etc.) to communicate back to the Command-and-Control Center. PathoINVEST is being developed as a QGIS plugin, and the analytics modules will be implemented on Python and MATLAB.

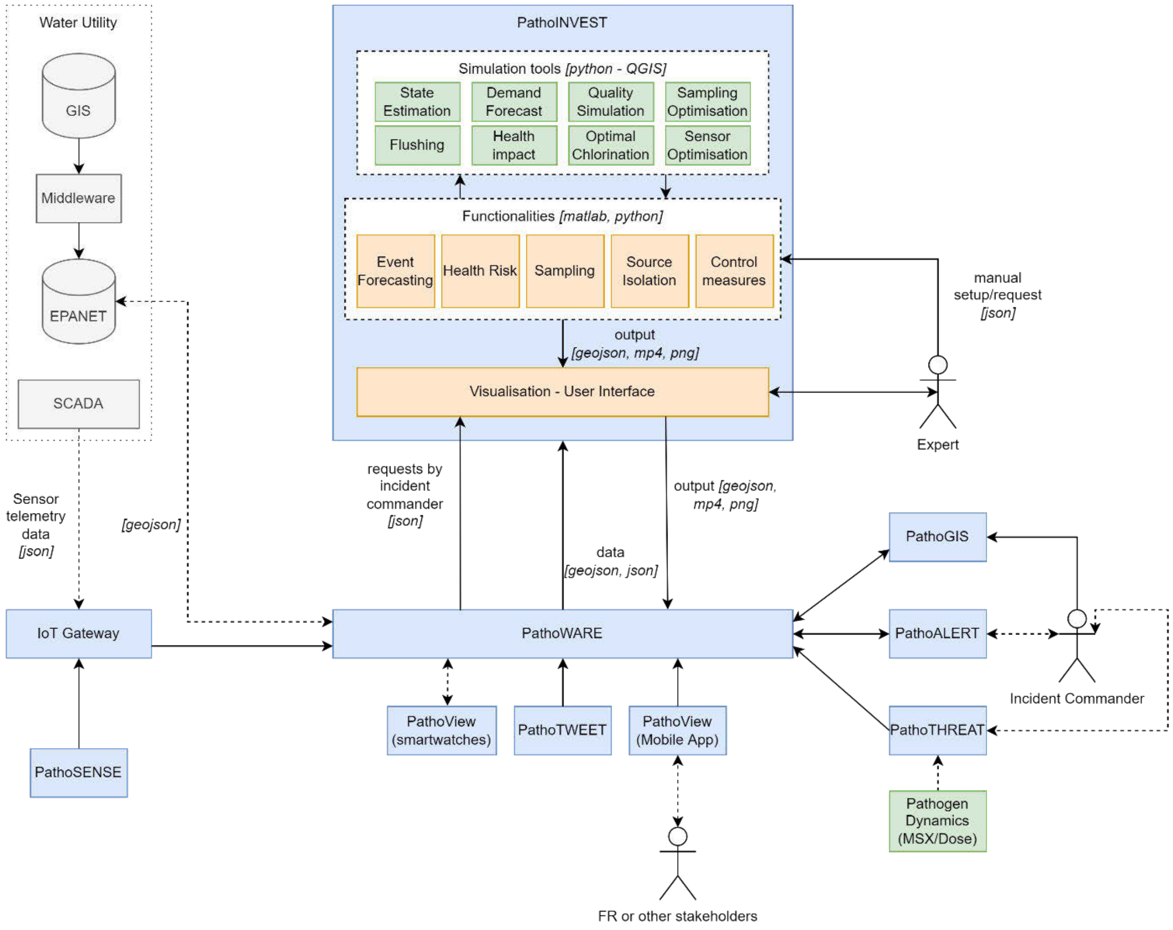

Water Utility
GIS
Middleware
EPANET
SCADA
Sensor telemetry data [json]
[geojson]
IoT Gateway
PathoSENSE
PathoINVEST
Simulation tools [python - QGIS]
State Estimation
Demand Forecast
Quality Simulation
Sampling Optimisation
Flushing
Health impact
Optimal Chlorination
Sensor Optimisation
Functionalities [matlab, python]
Event Forecasting
Health Risk
Sampling
Source Isolation
Control measures
output [geojson, mp4, png]
Visualisation - User Interface
manual setup/request [json]
Expert
requests by incident commander [json]
data [geojson, json]
output [geojson, mp4, png]
PathoWARE
PathoView (smartwatches)
PathoTWEET
PathoView (Mobile App)
FR or other stakeholders
PathoGIS
PathoALERT
PathoTHREAT
Incident Commander
Pathogen Dynamics (MSX/Dose)

Figure 1. The architecture of PathoINVEST within the PathoCERT ecosystem.

PathoSENSE is a collection of PathoCERT-compatible water quality sensors, which can be used for alerting and detecting pathogens within the DWN, in less than 1 hour. In addition to in-line sensors, portable sensors (**PathoTeSTICK**) which can be carried by First Responders and connected to their mobile phones, are also being developed, to detect specific pathogens within a few minutes [28]. These sensors can connect wirelessly to a specialized ***PathoSENSE IoT Gateway*** (via Wi-Fi, Bluetooth, or LoRaWAN), which manages the metadata of these sensors and communicates the relevant information to the PathoWARE platform. Moreover, the PathoSENSE IoT Gateway can also be connected to a SCADA service from the water utility authority, which will provide additional to the PathoSENSE sensors, real-time and historical telemetry data of both hydraulic and water quality dynamics, which are also essential for configuring the PathoINVEST simulations.

PathoTWEET is a cloud-based technology for monitoring anonymous data (text and photos) generated by citizens on social media platforms, that are relevant to the quality of water within the affected area. As a result, social media users can be considered “human sensors”, which can assist in identifying the scale of a contamination event, and the source area, faster.

PathoVIEW is a set of technologies that enables First Responders to utilize smart devices on the field. For example, a First Responder may be receiving an alert on their smartwatch if they are entering a neighborhood that is receiving contaminated water, or they may use Augmented Reality glasses to overlay a map of the area that is being affected, as computed by PathoINVEST.

PathoGIS is a web application for real-time representation of geospatial data relevant to the emergency. It also serves as a visualization tool for the outputs of PathoINVEST, as it provides maps illustrating contamination evolution, as well as locations for sampling and sensor deployment.

PathoALERT is a system integrated with PathoWARE, that implements algorithms for real-time data analytics received from the PathoSENSE IoT Gateway, the PathoTWEET service, and the PathoVIEW smartphone app.

PathoTHREAT is a knowledge database with historical and scientific information on water contamination events and pathogen characteristics. The Incident Commander, after gathering information from similar emergency events, creates requests for the PathoINVEST. Moreover, PathoTHREAT will provide an API that will allow interaction with the knowledge database, and PathoINVEST will be able to use pathogen-specific models (e.g., using the EPANET-MSX data structure).

4 ILLUSTRATIVE CASE STUDY

The case study presented in this section focuses on demonstrating how PathoINVEST could be used during a contamination emergency, within the framework of PathoCERT, for forecasting the evolution of the event and identifying its source, to respond appropriately to the emergency.

The scenario utilizes the benchmark network “L-Town”, created for the needs of the BattLeDIM (Battle of the Leakage Detection and Isolation Methods) competition [29] and it is based on a realistic network in Cyprus. The network has been suitably modified for security purposes. It is comprised of 782 junctions and 905 pipe segments, and it is assumed to provide water to around 10,000 citizens and industries. Each node has a randomly generated demand, synthesized from realistic data.

The L-Town network has two chlorination points at the inlets of the network, while the water supplied is assumed to have a constant concentration of Total Organic Carbon (TOC) of 1 mg/L. Water-quality dynamics, specifically the disinfection (chlorine) reactions in both the bulk and wall phase, as well as inactivation kinetics, are also incorporated in the benchmark. In the bulk phase, chlorine reacts with a series of reactants such as natural organic matter, and pathogens, whereas in the wall phase, chlorine reacts with biofilm. Finally, water quality parameters such as pH and temperature are also incorporated in the reaction model as they are factors that influence the chlorine demand and disinfection efficacy in the network.

We further assume that the PathoWARE service is already active in L-Town and that Standard Operating Procedures are already in place relevant to water contamination event management.

4.1 Emergency and establishment of Command & Control Center

An earthquake of 6.3 magnitude occurs near the city of L-Town. Damages on various buildings and infrastructures are being reported. The local authorities request the assistance of First Responders to set up a Command & Control (C&C) Center to manage the situation, coordinated by an Incident Commander. Following the Standard Operating Procedures, at the C&C Center, a dedicated area is assigned for collecting and managing information concerning water contaminations. The PathoGIS, PathoTHREAT, and PathoALERT are set up within the C&C Center, and PathoWARE connectivity is established.

4.2 Evaluating the risk

The Incident Commander requests an evaluation of the risk of possible waterborne contamination events due to the earthquake, using the PathoTHREAT tool. Similar past contamination events indicate that this is a likely scenario and that this can have a significant impact on the population within the next 12 hours.

The Incident Commander requests an evaluation of the situation from the utility operators, and both the drinking water and sewerage operators report abnormal pressure irregularities, which could be due to some leakage events. From past experiences, water and sewerage utilities have identified 5 "vulnerable" locations in their networks S = {S1, ..., S5}. At those points, there is no horizontal separation between water mains and sewer mains, resulting in an intersection of the two infrastructures with the former being below the latter. In the event of a severe earthquake, these 5 areas are potential sources of contamination due to wastewater leakage and subsequent infiltration into the DWN (Figure 2).

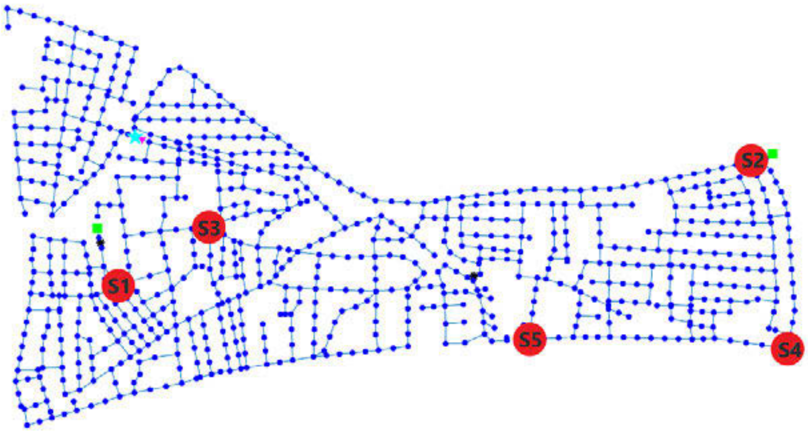


Figure 2. The 5 "vulnerable" areas within the network (S1-S5). The two reservoirs of the network are depicted as green squares and the tank as a cyan star.

As a result, the Incident Commander requests from the PathoINVEST operator, an impact assessment for all the different contamination scenarios (the 5 vulnerable locations) for the next 2-12 hours. Additionally, the Incident Commander requests maps that will show: a) contaminant evolution, b) optimal sensor placement, and c) sampling locations for source identification. For the contamination scenarios, single faults are considered as well as simultaneous (maximum two) faults with $\binom{5}{2} = 10$ maximum possible combinations. Although there are no results from sampling yet, it is suspected that wastewater has infiltrated the network, and therefore pathogen indicator *Escherichia coli* will be modeled in PathoINVEST. It is assumed that the contamination is continuous throughout the whole simulation and after exploring the literature, a conservative concentration of 3190 CFU/L *E.coli* and 140 mg/L (after 1% wastewater dilution) additional Total Organic Carbon (TOC) are modeled [30, 31]. Fast and slow reactions of chlorine with TOC and Chick-Watson inactivation kinetics are incorporated, with a pH of 7.5 and a steady temperature of 25°. Finally, no natural decay of *Escherichia coli* is included in the model, since this is insignificant in the time scale of the emergency event.

4.1.1 Forecast the evolution of possible contamination events

Figure 3 demonstrates the estimated *Escherichia coli* propagation through the network at 2, 4, 8, and 12 hours after the initial contamination using $S1$ and $S2$ as sources, whereas Figure 4 depicts the evolution of scenarios where there is a combination of simultaneous contamination sources $S3 + S4$ and $S4 + S5$ respectively. The PathoINVEST expert generates these maps (in an MP4 and GeoJSON format) and sends them through PathoWARE to the PathoGIS for visualization. Through this, the Incident Commander has a (near) real-time estimate of the possible propagation of contaminated water for the first 12 hours of each contamination scenario considered.

CC BY NC SA

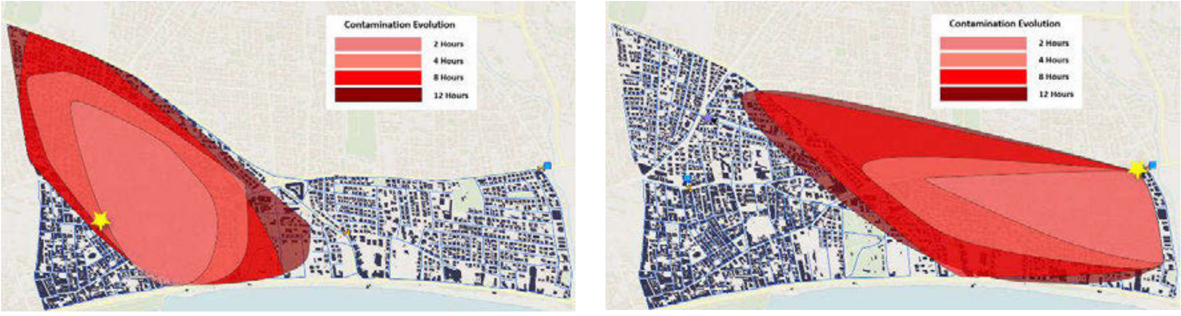


Figure 3. The contamination propagation for 2,4,8, and 12 hours when the source is S1 (left) and S2 (right).

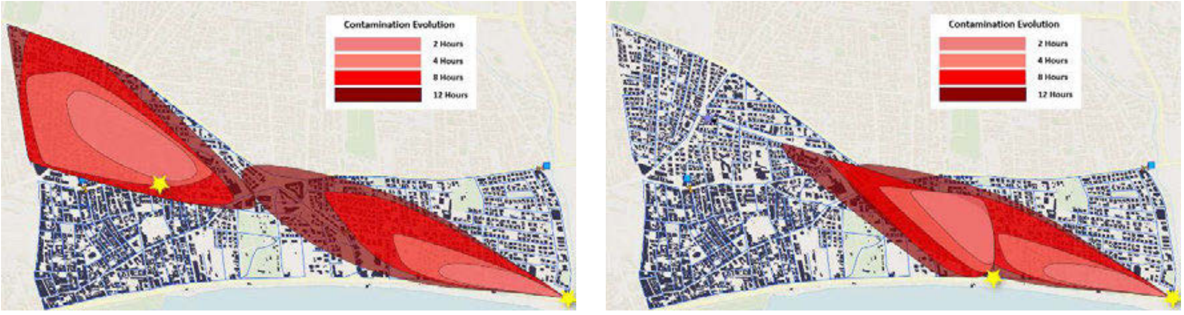


Figure 4. The contamination propagation for 2, 4, 8, and 12 hours for the combination of sources S3 + S4 (left) and S4 + S5 (right).

4.1.2 Impact assessment

The contamination Impact $I(k)$ in this case study is defined as the number of people affected by a contamination event, up until a specified discrete time k. Let the affected people be those who have ingested at least 1.0 Colony Forming Unit (CFU) of a microbiological contaminant mass until the considered time-step k. For this case study, we assume that the reference pathogen is *Escherichia coli* 0157. The impact is then calculated by simulating the reference pathogen concentration at each node of the network and calculating the total mass consumption at each node. The contaminant ingestion per person is calculated by considering the population estimate at each node, which is correlated to the base water demand at the node. The average water consumption in L-Town is assumed to be 150 L/person/day, where only 1% accounts for ingestion of tap water.

In PathoINVEST the operator is able to configure the contamination parameters and dynamics, using the pathogen reaction model extracted from the PathoTHREAT knowledge base (e.g., in an EPANET-MSX format), as well as other epidemiological parameters related to the pathogen, including the relevant exposure routes (i.e., ingestion, inhalation, dermal exposure). PathoINVEST can be used to compute an impact analysis for all 15 scenarios discussed previously (5 single plus 10 combinations), and the results are then communicated in a JSON format to PathoWARE, (an example can be found in Table 2). S1 and S2 have the highest impact and S5 the lowest.

Table 2. Impact of selected scenarios, defined as the percentage of affected people in the network.

	S1	S2	S3	S4	S5	S1&S2	S2&S3	S3&S4	S4&S5
2 hours	16.3%	9.7%	2.4%	1.8%	3.7%	26.1%	12.1%	4.2%	5.6%
4 hours	23.1%	19.4%	2.5%	3.2%	7.1%	42.5%	21.9%	5.7%	10.3%
8 hours	36.7%	32.9%	15.8%	9.2%	11.8%	69.6%	48.7%	24.9%	17.3%
12 hours	41.7%	37.4%	19.9%	18.7%	13.5%	76.8%	55.8%	38.1%	20.3%

4.3 Establish an early warning system for contamination

The Incident Commander reviews the impact assessment results, and following the recommendations from the expert operators of PathoTHREAT, decides to establish an early warning system by requesting the installation of PathoSENSE sensors within the DWN. The Incident Commander requests from the PathoINVEST operator to prepare a sensor placement map so that water utility staff together with First Responders can install and integrate them to PathoWARE through the PathoSENSE IoT Gateway.

For this illustrative case study, we assume that the water authorities are equipped with 3 PathoSENSE sensors for monitoring pathogens, which can be installed at any of the 20 sampling nodes (locations in the network that have been designed to allow water sampling and installation of mobile sensors during normal network operation). As a note, it is not possible to install mobile sensors everywhere in the network, due to physical and technological constraints.

As soon as the request reaches PathoINVEST operator, the Sensor Placement module is activated. To solve this problem, the tool needs to identify among the 20 possible sampling nodes, which 3 are the most suitable for monitoring a possible contamination event, considering that one (or more) of the 5 vulnerable locations initially identified may experience wastewater infiltration. For this problem, there are $\binom{20}{3} = 1140$ possible solutions. The goal is to identify the optimal combination of solutions (sensor locations), which minimizes the impact across all the possible contamination scenarios (i.e., all the combinations of the 5 potential sources). As these events have a low probability, the experts suggest considering up to two potential contamination sources. Moreover, it is assumed that the impact stops as soon as the contaminant is detected by any of the installed sensors, in the sense that mitigation actions could be taken after that point.

The sensor placement results are illustrated in Figure 5, with the 20 potential sensor nodes (left), as well as the selected sensor locations after performing the sensor placement analysis (right) — these {"413", "268", "206"}. The results are communicated in a GeoJSON format and are presented to the Incident Commander through PathoGIS. Instructions are then given to the appropriate teams for installing and setting up the PathoSENSE sensors as well as establishing communication links with the PathoSENSE IoT Gateway. Moreover, PathoALERT is configured to trigger an alarm, when the readings from these sensors deviate from normal.

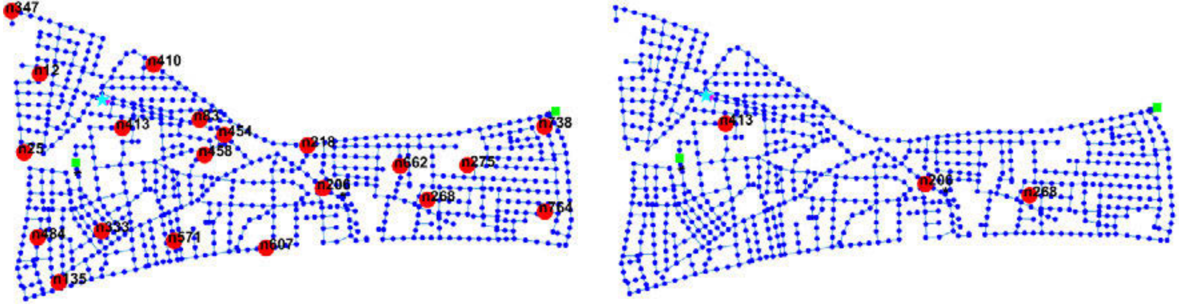


Figure 5. The 20 possible locations for sampling and sensor placement (left) and the 3 best suitable locations for sensor placement (right).

4.4 Identifying the source

A few hours after the early warning system has been set up, one of the PathoSENSE sensors installed in the network ("206") measures high pathogen concentrations which trigger a PathoALERT notification. The Incident Commander issues a request for PathoINVEST to analyze the data to isolate the source. The PathoINVEST operator configures the Source Identification tool, to indicate which of the sampling locations are suspected to be contaminated (as determined by sensor readings, manual samplings, or consumer complaints).

The analysis uses the installed sensor measurements to identify a set of possible source locations of the contamination event. As already mentioned, there are 15 possible contamination scenarios. By running simulations for each scenario, a binary indicator is computed for each installed sensor, depending on whether they have detected a contaminant (1) or not (0). The binary signatures for all possible scenarios are illustrated in Table 3.

Table 3. The binary signatures for all possible contamination scenarios were generated using the installed sensor measurements. In gray, are the scenarios that match the sensor observations.

Scenario	**S1**	**S2**	**S3**	***S4***	***S5***	**S1+S2**	**S1+S3**	**S1+S4**	**S1+S5**	**S2+S3**	**S2+S4**	**S2+S5**	***S3+S4***	***S3+S5***	***S4+S5***
Sensor 1 (node 413)	1	0	0	***0***	***0***	1	1	1	1	0	0	0	***0***	***0***	***0***
Sensor 2 (node 268)	0	1	0	***0***	***0***	1	0	0	0	1	1	1	***0***	***0***	***0***
Sensor 3 (node 206)	0	1	0	***1***	***1***	1	0	1	1	1	1	1	***1***	***1***	***1***

Let the observed binary signature in this case study, using the 3 installed sensors, be $[\ 0\ 0\ 1\ ]^T$. Given the observed signature and after examining Table 3, a reduced set of possible contamination scenarios with the same signature can be identified, denoted by $S_r = \{S4, S5, S3 + S4, S3 + S5, S4 + S5\}$.

At this point, the PathoINVEST operator cannot distinguish which of the scenarios found in S_r is the actual contamination's source. For identifying the actual source, PathoINVEST needs more information, which can be acquired by requesting a team to perform manual sampling in the field. The sampling can be performed using a portable rapid testing sensor, such as PathoTeSTICK. It is assumed that any of the 20 accessible locations in the network (see Figure 5) are also suitable for manual sampling. Through simulation, the binary signatures are computed considering each scenario in S_r, for each possible sampling location (Table 4). According to Table 4, including the results from manual water sampling at locations {458, 662} to the existing binary indicators of available sensors {413, 268, 206}, will provide a unique signature to each of the scenarios in S_r. The PathoINVEST source identification tool would then instruct the operator to take samples, first from 662 and then from 458, if the algorithm is set up to give priority to True Positives rather than True Negatives.

Table 4. The 20 sampling locations and their respective expected signatures for the reduced set of contamination scenarios. Note that sensors already exist at nodes {413, 268, 206}.

S_r	**Available sampling locations (Node)**																			
	413	***268***	***206***	**458**	**662**	**333**	**347**	**275**	**12**	**764**	**607**	**738**	**218**	**484**	**410**	**454**	**135**	**571**	**25**	**83**
S4	*0*	*0*	*1*	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S5	*0*	*0*	*1*	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S3+S4	*0*	*0*	*1*	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S3+S5	*0*	*0*	*1*	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S4+S5	*0*	*0*	*1*	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0

The first field analysis is performed at node 662 using PathoTeSTICK, indicating that there are no traces of pathogen. This result is communicated back to PathoWARE and PathoINVEST is updated. Therefore, the contamination scenarios $S4$, $S3 + S4$, and $S4 + S5$ are removed from the set of possible contamination scenarios S_r. At around the same time, citizen complaints increase on social media and in the Water Utility customer service, complaining about the taste, color, and smell, in the northwest part of L-Town. Social media are monitored via PathoTWEET, thus PathoALERT issues the alarm for the increasing consumer complaints. This information is communicated to PathoINVEST from the Command & Control Center, and PathoINVEST suggests requesting another sampling at node 458 to validate the “human sensors. The sampling analysis confirms contamination at that node, and therefore, PathoINVEST reasons that the contamination scenario that is most likely to have occurred is the infiltration of wastewater in both $S3$ and $S5$ locations (Figure 6).

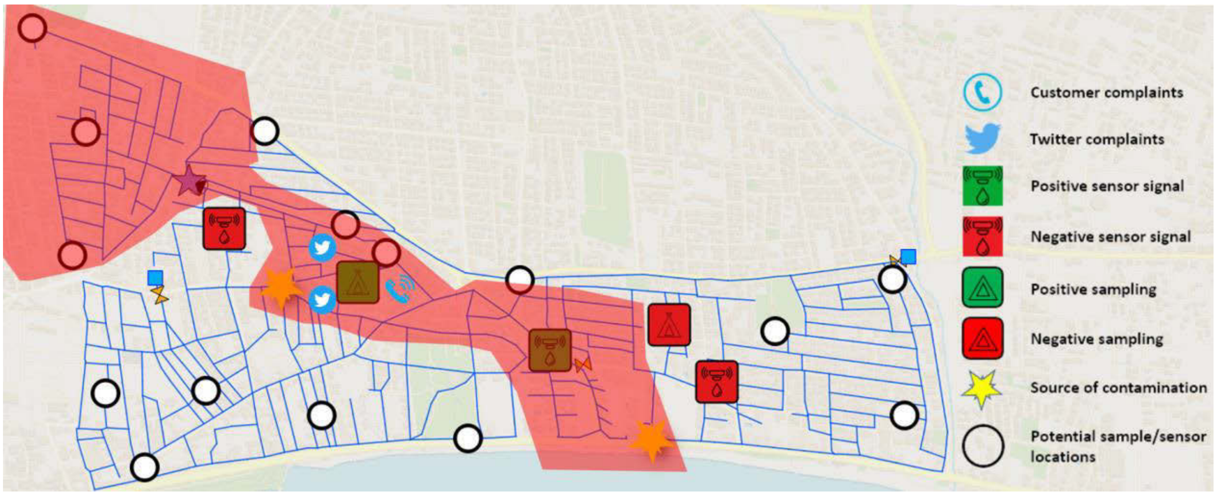


Figure 6. The final contamination map with all the necessary information.

Based on the verified source(s), the final contamination maps are computed and sent in GeoJSON format to PathWARE, to be then depicted through PathoGIS. Moreover, first responders on the field are receiving information whenever they are entering a contaminated area, or when the contamination has been extended to the area where they operate.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose an architecture for a decision support system (PathoINVEST), that aims to assist First Responders and Water Authorities in investigating and managing pathogen contamination events that could occur in a DWN after an emergency (such as an earthquake that causes damage to drinking water and sewerage infrastructure). An illustrative case study is described as a complete proof-of-concept, to demonstrate how the system could operate during a real emergency, considering the limitations in time and information. For this case study, simplified algorithms were implemented to demonstrate how each module can function. Our future work will investigate and propose new methodologies for modeling wastewater infiltration, as well as determining the most suitable mitigation and policies to minimize the impact of contamination. The case study highlighted that through the interoperability of the PathoCERT modules, a contamination event can be assessed and managed in a timely and effective manner.

PathoINVEST assumes that standard operating procedures are in place, to guide First Responders in establishing early warning systems and managing such an emergency event. Moreover, it assumes that the PathoWARE service is operational and that PathoSENSE sensors are available. Currently, all technologies described in this paper are under development and are expected to be evaluated and validated in field and tabletop pilot exercises by First Responders within 2023.

6 DATA AVAILABILITY

The data models and code generated and used for this case study are available in the following repository: https://github.com/KIOS-Research/PathoINVEST-WDSA-CCWI-2022

7 ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883484 (PathoCERT) and is supported by the European Union Horizon 2020 program Teaming under Grant Agreement No. 739551 (KIOS CoE) and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

8 REFERENCES

[1] T. Braeye, K. De Schrijver, E. Wollants, M. Van Ranst, and J. Verhaegen, "A large community outbreak of gastroenteritis associated with consumption of drinking water contaminated by river water, Belgium, 2010," Epidemiology & Infection, vol. 143, no. 4, pp. 711–719, 2015.

[2] J. Laine, E. Huovinen, M. Virtanen, M. Snellman, J. Lumio, P. Ruutu, E. Kujansuu, R. Vuento, T. Pitkanen, I. Miettinen, and J. Herrala, "An extensive gastroenteritis outbreak after drinking-water contamination by sewage effluent, Finland," Epidemiology & Infection, vol. 139, no. 7, pp. 1105–1113, 2011.

[3] F. Y. Ramirez-Castillo, A. Loera-Muro, M. Jacques, P. Garneau, F. J. Avelar-Gonzalez, J. Harel, and A. L. Guerrero-Barrera, "Waterborne pathogens: detection methods and challenges," Pathogens, vol. 4, no. 2, pp. 307–334, 2015.

[4] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks," Journal of Water Resources Planning and Management, vol. 134, no. 6, pp. 516–526, 11 2008.

[5] R. Murray, T. Haxton, R. Janke, W. E. Hart, J. Berry, and C. Phillips, Sensor Network Design for Drinking Water Contamination Warning Systems: A Compendium of Research Results and Case Studies Using the TEVA-SPOT Software. EPA, 2010.

[6] A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S.-T. Khu, D. Savic, D. Eliades, M. Polycar-pou, S. R. Ghimire, B. D. Barkdoll, R. Gueli, J. J. Huang, E. A. McBean, W. James, A. Krause, J. Leskovec, S. Isovitsch, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. B. Trachtman, Z. Y. Wu, and T. Walski, "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," Journal of Water Resources Planning and Management, vol. 134, no. 6, pp. 556–568, 2008.

[7] D. G. Eliades and M. M. Polycarpou, "A Fault Diagnosis and Security Framework for Water Systems," IEEE Transactions on Control Systems Technology, vol. 18, no. 6, pp. 1254–1265, 11 2010.

[8] L. Perelman, J. Arad, M. Housh, and A. Ostfeld, "Event detection in water distribution systems from multivariate water quality time series," Environmental Science & Technology, vol. 46, no. 15, pp. 8212–8219, 2012.

[9] D. G. Eliades, D. Stavrou, S. G. Vrachimis, C. G. Panayiotou, and M. M. Polycarpou, "Contamination Event Detection Using Multi-level Thresholds," in Computing and Control for the Water Industry (CCWI), vol. 119, no. 1, 2015, pp. 1429–1438.

[10] P. Fraternali, A. Castelletti, R. Soncini-Sessa, C. Vaca Ruiz, and A. E. Rizzoli, "Putting humans in the loop: Social computing for Water Resources Management," Environmental Modelling and Software, vol. 37, pp. 68–77, 11 2012.

[11] M. C. Getchell and T. L. Sellnow, "A network analysis of official Twitter accounts during the West Virginia water crisis," Computers in Human Behavior, vol. 54, pp. 597–606, 1 2016.

[12] S. Andreadis, G. Antzoulatos, T. Mavropoulos, P. Giannakeris, G. Tzionis, N. Pantelidis, K. Ioannidis, A. Karakostas, I. Gialampoukidis, S. Vrochidis, and I. Kompatsiaris, "A social media analytics platform

visualising the spread of covid-19 in Italy via exploitation of automatically geotagged tweets," Online Social Networks and Media, vol. 23, p. 100134, 2021.

[13] W.H.O., "Quantitative microbial risk assessment: application for water safety management," 2016.

[14] W.H.O, "Guidelines for drinking-water quality: first addendum to the fourth edition," 2017.

[15] P. Teunis, M. Xu, K. Fleming, J. Yang, C. Moe, and M. LeChevallier, "Enteric virus infection risk from intrusion of sewage into a drinking water distribution network," Environmental Science & Technology, vol. 44, no. 22, pp. 8561–8566, 2010.

[16] M. Blokker, J. Vreeburg, and V. Speight, "Residual chlorine in the extremities of the drinking water distribution system: the influence of stochastic water demands," International Conference on Computing and Control for the Water Industry (CCWI2013), vol. 70, pp. 172–180, 2014.

[17] M. Blokker, P. Smeets, and G. Medema, "Quantitative microbial risk assessment of repairs of the drinking water distribution system," Microbial Risk Analysis, vol. 8, pp. 22–31, 2018.

[18] N. Pelekanos, D. Nikolopoulos, and C. Makropoulos, "Simulation and vulnerability assessment of water distribution networks under deliberate contamination attacks," Urban Water Journal, vol. 18, no. 4, pp. 209–222, 2021.

[19] J. Schijven, J. M. Foret, J. Chardon, P. Teunis, M. Bouwknegt, and B. Tangena, "Evaluation of exposure scenarios on intentional microbiological contamination in a drinking water distribution network," Water Research, vol. 96, pp. 148–154, 2016.

[20] D. Jerez, H. Jensen, M. Beer, and M. Broggi, "Contaminant source identification in water distribution networks: A bayesian framework," Mechanical Systems and Signal Processing, vol. 159, p. 107834, 2021.

[21] F. Shang, J. G. Uber, and M. M. Polycarpou, "Particle backtracking algorithm for water distribution system analysis," Journal of Environmental Engineering, vol. 128, no. 5, pp. 441–450, 2002.

[22] D. G. Eliades and M. M. Polycarpou, "Water contamination impact evaluation and source-area isolation using decision trees," Journal of Water Resources Planning and Management, vol. 138, no. 5, pp. 562–570, 2012.

[23] S. G. Vrachimis, R. Lifshitz, D. G. Eliades, M. M. Polycarpou, and A. Ostfeld, "Active contamination detection in water-distribution systems," Journal of Water Resources Planning and Management, vol. 146, no. 4, p. 04020014, 2020.

[24] A. Seth, G. A. Hackebeil, T. Haxton, R. Murray, C. D. Laird, and K. A. Klise, "Evaluation of chlorine booster station placement for water security," in Computer Aided Chemical Engineering. Elsevier, 2019, vol. 47, pp. 463–468.

[25] F. Shang, J. Uber, and L. Rossman, "Epanet multi-species extension software and user's manual," U.S. Environmental Protection Agency, Washington, DC, EPA/600/C-10/002, 2011.

[26] D. E. Helbling and J. M. VanBriesen, "Modeling residual chlorine response to a microbial contamination event in drinking water distribution systems," Journal of Environmental Engineering, vol. 135, no. 10, pp. 918–927, 2009.

[27] E. W. Betanzo, R. Hofmann, Z. Hu, H. Baribeau, and Z. Alam, "Modeling the impact of microbial intrusion on secondary disinfection in a drinking water distribution system," Journal of Environmental Engineering, vol. 134, no. 4, pp. 231–237, 2008.

[28] A. Canciu, M. Tertis, O. Hosu, A. Cernat, C. Cristea, and F. Graur, "Modern analytical techniques for detection of bacteria in surface and wastewaters," Sustainability, vol. 13, no. 13, 2021.

[29] Vrachimis, Eliades, Taormina, Ostfeld, Kapelan, Liu, Kyriakou, Pavlou, Qiu, and Polycarpou, "BattLeDIM: Battle of the Leakage Detection and Isolation Methods," 2020.

[30] J. Yang, O. D. Schneider, P. K. Jjemba, and M. W. Lechevallier, "Microbial risk modeling for main breaks," Journal American Water Works Association, vol. 107, no. 2, pp. E97–E108, 2015.

[31] L. Metcalf, H. P. Eddy, and G. Tchobanoglous, Wastewater engineering: treatment, disposal, and reuse. McGraw-Hill New York, 1991, vol. 4.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14802

OPTIMIZATION OF RESERVOIR TREATMENT LEVELS CONSIDERING UNCERTAINTY IN MIXING AT CROSS JUNCTIONS IN WATER DISTRIBUTION SYSTEMS USING INFO-GAP DECISION THEORY

Sriman Pankaj Boindala[1], G Jaykrishnan[2] and Avi Ostfeld, F. ASCE[3]

[1] PhD Student, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[2] PhD Student, Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 32000, Israel
[3] Professor (Corresponding Author), Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

[1] srimanpankaj@gmail.com, [2] jaykrishnang@hotmail.com , [3] ostfeld@technion.ac.il

Abstract

Water distribution systems are affected by several uncertainties in multiple stages. This uncertainty makes solving the optimal design and management of WDS a multifaceted problem. Past research has focused only on solving design and management problems of system hydraulics. There have been very few studies that involve considering uncertainties that affect the water quality aspect of WDS. One of the major assumptions in solving the design and management problems of WDS is considering uniform and instantaneous mixing at the cross junctions. However, in reality, this is not true. This assumption is made due to the lack of computational power to accurately estimate the level of mixing at every junction in a water distribution network. This study focuses on considering this level of mixing as uncertain/unknown and provides the optimal treatment levels required at the reservoirs to ensure the system is immune to the level of mixing occurring at the junctions to satisfy the water quality requirements at the customer level. Info-gap decision theory-based optimization approach combined with the cuckoo search metaheuristic is proposed in this study to handle the uncertainty. The proposed methodology is applied to a 4x4 grid hypothetical network example. The study's objective is to provide the best designs that can handle the maximum variation of the level of mixing at junctions within the given budget by the designer. The maximum variation of the level of mixing is reported for different budget levels. The designs are compared with the deterministic case using Monte-Carlo simulations.

Keywords
Water Quality Uncertainty, Incomplete Mixing, Epanet-BAM, Info-Gap Decision Theory, Treatment Level Design Optimization, Water Distribution System (WDS).

1 INTRODUCTION

A system consisting of reservoirs (sources), tanks, pipes, pumps, and valves that work together to provide high quality (clean, odourless, and clear) water at sufficient pressure head for the customers downstream is called a water distribution system (WDS). With growing urban migration, the demanding stress on these systems increases, and hence more centralized systems are being developed. With limited potable water sources available, these systems' optimal resilient design and management are inevitable. Optimal design and management of WDS have always been focused on satisfying hydraulic requirements like demand and pressure constraints in the past. The passing of the safe drinking water act (SDWA) in 1990 motivated research to include water quality constraints in the optimal design and management of WDS. Today the importance of maintaining the water quality is equivalent to satisfying the hydraulic constraints (Pasha and Lansey 2010). Even with including water quality constraints, the performance of the WDS is not as expected. Not accounting for the uncertainty involved in the optimal design and management problem of WDS and the assumptions made for simplifying the problem are the major reasons for the low performance of the systems when applied practically.

(Lansey et al. 1989)were pioneers that proposed an optimal WDS design methodology that considered uncertainty in hydraulic parameters. The uncertain parameters like future demand are assumed to follow a probability distribution function (PDF) with a mean and standard deviation and formulated the hydraulic constraints as chance constraints. They used a non-linear programming method with a generalized reduced gradient-II (GRG-II) technique for optimization. They applied this method to a small network and proposed multiple designs with corresponding reliability values. (Babayan et al. 2004; Kapelan et al. 2005), proposed methodologies to solve optimal single and multi-objective WDS design problems under uncertainty. They used Latin hypercube sampling (LHS) for sampling and PDF generation instead of assuming. They used genetic algorithms as optimization algorithms. (Babayan et al. 2005) proposed a methodology that eliminates sampling techniques by adding a margin of safety factor in the required minimum pressure heads. Then the stochastic problem became a simple deterministic problem with additional constraints and was solved using a standard genetic algorithm (GA). (Kang et al. 2009)proposed a methodology to find the optimal design of WDS under uncertainty by using approximation techniques like first-order second moment (FOSM) and combined that with LHS to simplify the computational time for the stochastic problem.

Due to the high computation time of these probabilistic approaches, the research shifted towards using a non-probabilistic approach to handle uncertainty in WDS parameters. Info gap decision theory (IGDT) and robust optimization (RO) are a few of the popular non-probabilistic uncertainty handling techniques. (Chung et al. 2009)implemented robust optimization (RO) approach to solve the design of municipal WDS, considering demand is uncertain. (Perelman et al. 2013) used robust counterpart (RC) to solve the least-cost design problem of WDS with demand uncertainty. (Ghelichi et al. 2018; Naderi and Pishvaee 2017) also used investigated RO formulations to handle uncertainties in the water resources and stated their advantages over the probabilistic approach. (Korteling et al. 2013) stated the benefits of using info-gap decision theory in water resources planning under severe uncertainty. They showed the effectiveness of the info-gap decision theory in supporting adaptive management of water systems under severe uncertainty in supply or demand. (Roach et al. 2015) compared the use of RO and IGDT for water resource management under deep uncertainty. They applied both methods to a case study approximating the Susexx north water resource zone in England. They concluded that the IGDT produced more expensive designs than RO as it has a more rigorous robustness analysis.

The research works in the past related to WDS design are highly concentrated on hydraulic uncertainty. Very few studies have considered uncertainty related to water quality parameters. (Pasha and Lansey 2005, 2010) examined the effect of a few water quality uncertain parameters like bulk and wall reaction coefficients and pipe diameters in the distribution system's water quality analysis. In addition to these water quality parameters, the uncertainty in solute mixing at cross junctions can also affect the consumer nodes' water quality requirements.

Fowler and Jones (1991)were the first to investigate the practicality of complete and instantaneous mixing assumptions at junctions in WDSs. They stated that among many other concerns regarding the accuracy of the water quality (WQ) models, the assumption of instantaneous, complete mixing at junctions was regarded as a significant cause of erroneous outcomes in water quality modelling of WDSs. (Romero-Gomez et al. 2009) investigated the impact of incomplete mixing on sensor network designs. They found the perfect mixing assumption inefficient, which led to wrong sensor placements that led to some locations in the system not being covered. (Ho and Khalsa 2008)developed a Bulk Advective Mixing model (BAM) for addressing the non-uniform mixing behaviour at cross junctions. They incorporated a mixing parameter "$s \in [0,1]$" which governs the extent of mixing from incomplete mixing being the lower bound and complete mixing being the upper bound. (Song et al. 2009)also developed a non-uniform mixing model named (AZRED) to model incomplete mixing phenomena for specific junction types like T and Y junctions. (Paez et al. 2017)compared the water quality at the consumer nodes assuming complete mixing (EPANET) and non-uniform mixing BAM model for WDS of two cities and two large grid networks. They concluded that the mixing uncertainty is more predominant in a grid-type network containing many cross junctions than in the conventional city WDSs in the study.

The present design of WDS does not consider various uncertainties in water quality affecting parameters, especially the mixing uncertainty at junctions. This assumption may lead to unsatisfactory designs that may not satisfy the practical water quality constraints. Realizing the exact mixing phenomenon is quite complex and requires complex CFD simulation. The complexity only increases with the size of the WDS and the number of cross junctions. The current study incorporates the mixing uncertainty in designing the water distribution system. Present work emphasizes quantifying the significance of non-uniform mixing uncertainty in WDS design and using non-probabilistic uncertainty handling techniques like info-gap decision theory to solve the complex, uncertain design problem.

2 METHODOLOGY

2.1 Info Gap Decision Theory

Info Gap Decision Theory is a non-probabilistic decision-making technique that helps in prioritizing alternatives and making decisions under deep uncertainty. The IGDT analysis is governed by three components, system model, desired model performance, and uncertainty model. The system model defines the problem, and the understanding of the system. The performance requirements are the answer to the question: What do we need to achieve in order for the outcome of the decision to be acceptable? i.e. our goal and the conditions it should satisfy to say we reached our goal. Uncertain model is to realize the uncertainty involved in either our understanding (system model) or our uncertainty in our performance criteria.

For illustration:

Consider the problem $\mathcal{P}$. Assume that some information about ξ is known. For example, assume that ξ lies in the polyhedral set U, where $\mathrm{U} = \{\xi \in \mathbb{R}^n \ : \ \mathrm{A}\xi \leq \mathrm{b}\}$ where A is a matrix and b is a vector. The problem is now defined as:

$$\min_{x} f(x, \xi) \tag{1}$$

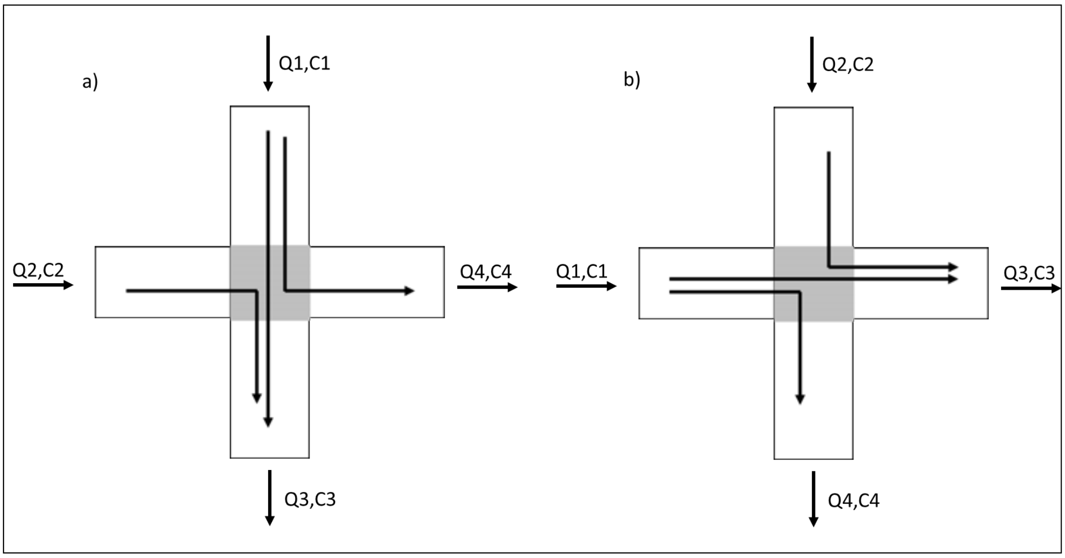
$$\mathrm{s.t.}\, g(x, \xi) \leq 0, \forall \xi \in \mathrm{U}. \tag{2}$$

The ideal case of an uncertain problem is that the constraint should be satisfied for all possible realizations of ξ. This uncertain version of the problem is not tractable since there are infinite constraints (Eq. 2).

Let the system model be problem P. This problem is solved by assuming both x, ξ are variables that can take any value, and the problem P is solved to obtain the minimum objective value $\#obj$, optimal decision variable $\#x$ and optimal parameter $\#\xi$. Using these as a base, we build our decision model. The uncertain model is created by taking the $\#\xi$ as a base. The performance criteria satisfy the constraints, and the maximum value the objective function can be allowed or acceptable. The result from the info-gap model is the uncertainty set, i.e. the maximum variation the system model can handle such that the objective function is within the performance criteria. As the maximum value of the objective function is increased, the uncertain sets will also increase, and the sets will be nested in each other (Ben-Haim 2006).

2.2 BAM Model

BAM model that was developed by (Ho et al. 2008) is used as the WQ model in the present study. Ho et al. (2008) assumed that when two fluid streams with different flows and concentrations enter a cross junction, the distribution of solute concentrations in the fluid is between complete and incomplete mixing/bulk mixing. The extent of deviation from the complete mixing is controlled by a mixing parameter "s" generally obtained through experimentation. In this present study, this mixing parameter is treated as uncertain as we cannot always find it through experimentation on real networks. We then aim to obtain optimal designs within the designer's budget that are immune to set of maximum variations of mixing. Equations associated with BAM model are described below



Legend: Q_1, Q_2, Q_3, Q_4 – Flows; C_1, C_2, C_3, C_4 – Concentrations where $Q_1 > Q_2$

Figure 1 Pictorial representation of bulk mixing (Ho et al, 2008), the shaded region is the non-uniform mixing region

The bulk mixing equations for the case are shown in Fig 1 where $Q_1 > Q_2$ can be written easily with mass balance,

$$C_4 = C_1 \tag{3}$$

$$C_3 = \frac{Q_2 C_2 + (Q_1 - Q_4) C_1}{Q_3} \tag{4}$$

Whereas complete mixing equations are,

$$C_3 = C_4 = \frac{Q_1C_1 + Q_2C_2}{Q_3 + Q_4} = \frac{Q_1C_1 + Q_2C_2}{Q_1 + Q_2} \tag{5}$$

The final BAM model equation is between

$$C_{BAM} = C_{incomplete} + s * (C_{complete} - C_{incomplete}) \tag{6}$$

3 TREATMENT COST OPTIMIZATION PROBLEM

Assume that every source for the WDS is associated with a treatment plant, the extent of treatment required at the source level is governed by the water quality requirements at the customer nodes and level of mixing at the junctions. The objective of the problem is to obtain optimal treatment levels that satisfy the water quality constraints at the customer nodes and minimize the overall treatment cost at the sources. We assume steady state conditions and the contaminant is non-reactive.

The problem structure can be written as follows equation (7-8):

$$\min \mathrm{TC}(\mathrm{IRC}, \mathrm{RCAT}) \tag{7}$$

$$\mathrm{s.t.}\ \mathrm{C^j}_{\mathrm{LL}} \leq \mathrm{C^j_{out}}\big(\mathrm{RCAT}, \mathrm{Q}, \mathrm{s}(\xi)\big)_{\mathrm{BAM}} \leq \mathrm{C^j_{UL}}, \forall \mathrm{j} \in \mathrm{J}, \xi \in \mathrm{U} \tag{8}$$

Here, TC -unit treatment cost, IRC – initial reservoir concentration before treatment, RCAT- reservoir concentration after the treatment, $\mathrm{C^j_{LL}}, \mathrm{C^j_{UL}}$ – lower limit and upper limit of desired outlet concentration at junction j, $\mathrm{C^j_{out}}$- outlet concertation at node j, Q- flows in pipes. Its states that the outlet concentration is dependent on the pipe flow, mixing parameter (which is uncertain) and reservoir concentration after treatment.

The explicit formulation of C^J_{out} for every node considering mixing parameter "s" as uncertain gets complex with increase in number of cross junctions, in order to eliminate the complexity, a data driven linear surrogate model is developed. The surrogate model is of the form shown in equation 7 where $\mathrm{C^j_{out}}$ is the outlet concentration and $\mathrm{a_j}$ is a vector and $\mathrm{E_j}$ is a matrix are obtained from regression modelling.

$$C^j_{out}\big(RCAT, Q, s(\xi)\big)_{BAM} = \boldsymbol{a_j^T} RCAT + RCAT^T \boldsymbol{E}_j s \tag{9}$$

Any convex model can be used instead of linear model if the accuracy is not significant

Using this surrogate model, the constraint (9) in treatment cost optimization problem can be re-written as

$$\mathrm{s.t.}\ \mathrm{C^j}_{\mathrm{LL}} \leq \boldsymbol{a_j^T} * RCAT + RCAT^T \boldsymbol{E}_j s \leq \mathrm{C^j_{UL}}, \forall \mathrm{j} \in \mathrm{J}, \xi \in \mathrm{U} \tag{10}$$

3.1 IGDT problem formulation

As mentioned earlier, for IGDT formulation, we need first to define the system model. The system model can be represented as shown in Equations 11-13. We also know that the mixin parameter can take any value in [0,1] where 1 -represents complete mixing and 0- represents bulk mixing (incomplete mixing).

$$\min_{\mathrm{RCAT}} \mathrm{co}^{\mathrm{T}}(\mathrm{IRC} - \mathrm{RCAT}) \tag{11}$$

$$\mathrm{s.t.}\, \mathrm{C}_{\mathrm{min}}^{\mathrm{j}} \leq \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT} + \mathrm{RCAT}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s} \leq \mathrm{C}_{\mathrm{max}}^{\mathrm{j}}\,, \forall \mathrm{j} \in \mathrm{J} \tag{12}$$

$$0 \leq \mathrm{RCAT} \leq \mathrm{IRC} \tag{13}$$

IRC-Initial reservoir concentration before treatment, RCAT – reservoir concentration after treatment. co – the cost of unit treatment of water per unit volume.

Suppose s is considered as an uncertain parameter. In that case, the constraint in (Eq. 12) becomes (Eq. 14), which should satisfy every realization of the parameter s∈U., which leads to an infinite number of constraints. The problem then becomes a semi-infinite optimization problem that is intractable.

$$\mathrm{C}_{\mathrm{min}}^{\mathrm{j}} \leq \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT} + \mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s} \leq \mathrm{C}_{\mathrm{max}}^{\mathrm{j}}\,, \forall \mathrm{j} \in \mathrm{J}, \forall \mathrm{s} \in \mathrm{U} \tag{14}$$

As the most common phenomenon and the previous assumption is complete mixing, and we need the system to satisfy the complete mixing case definitely, we solve the system model (equations 11-13) by assuming (s=e). Once we assume this, the problem is linear and can be easily solved to obtain a base objective value ($\#Obj$).

The second component is to define the uncertainty model. The uncertainty model is assumed to be an envelope-based model, $\mathrm{U}(\delta) = \{\mathrm{s} \in \mathbb{R}^{\mathrm{I}} : (1 - \delta\sigma)\mathrm{e} \leq \mathrm{s} \leq \mathrm{e}\}, \delta \geq 0$ (Ben-Haim, 2006). The size of which is controlled by the scaling parameter δ. Notice that the uncertainty set is a function of δ, and as δ is increased, the set increases in size, but the sets are all nested. Since there are bounds on s, i.e., $0 \leq \mathrm{s} \leq 1$, correspondingly, the bounds obtained for δ are:

$$1 - \delta\sigma \geq 0 \; \Rightarrow \delta \leq \frac{1}{\sigma} \tag{15}$$

The uncertain model can be equivalently written as $\mathrm{U}(\delta) = \{\mathrm{s} \in \mathbb{R}^{\mathrm{I}} : \mathrm{As} \leq \mathrm{b}\}, \delta \geq 0$ where $\mathrm{A} = \begin{bmatrix} \mathrm{I}^{\mathrm{I}\times\mathrm{I}} \\ -\mathrm{I}^{\mathrm{I}\times\mathrm{I}} \end{bmatrix}$ and $\mathrm{b} = \begin{pmatrix} \mathrm{e} \\ (1 - \alpha\sigma)\mathrm{e} \end{pmatrix}$, I is the identity matrix, e is the vector of all 1.

Now, based on the budget the decision maker can provide, the maximum objective value which is allowed is incorporated as a constraint $\mathrm{obj}' > \#\mathrm{obj}$, IGDT can be used to develop an uncertainty set and a solution such that the objective value corresponding to the robust solution is at most obj'. Thus, the info-gap formulation is below.

$$\max \delta \tag{16}$$

BY NC SA

$$\mathrm{s.t.}\ \mathrm{co}^{\mathrm{T}}(\mathrm{IRC}-\mathrm{RCAT}) \leq \mathrm{obj}' \tag{17}$$

$$\mathrm{C}_{\mathrm{min}}^{\mathrm{j}} \leq \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT}+\mathrm{RCAT}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s} \leq \mathrm{C}_{\mathrm{max}}^{\mathrm{j}}\ \forall \mathrm{s} \in \mathrm{U}(\delta), \forall \mathrm{j} \in \mathrm{J} \tag{18}$$

$$0 \leq \mathrm{RCAT} \leq \mathrm{IRC} \tag{19}$$

$$0 \leq \delta \leq \frac{1}{\sigma} \tag{20}$$

This formulation (Eq. 16-20) can be reformulated using the dualization technique,

Consider LHS of equation 18,

$$a_j^T CR + CR^T E_j s \geq C_{min}^j\ \forall s \in U(\delta), \forall j \in J \tag{21}$$

This constraint is equivalent to

$$\begin{gathered} \min a_j^T CR + CR^T E_j s \geq C_{min}^j \\ s.t.\, s \in \mathbb{R}^I : As \leq b, A = \begin{bmatrix} I^{I\times I} \\ -I^{I\times I} \end{bmatrix} and\ b = \begin{pmatrix} e \\ (1-\alpha\sigma)e \end{pmatrix} \end{gathered} \tag{22}$$

Writing lagrangian for the above problem,

Writing the Lagrangian and minimizing, the dual objective is:

$$\min_{\mathrm{s}\in\mathbb{R}^{\mathrm{I}}} \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}+\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s}+\mu_{\mathrm{j}}^{\mathrm{T}}(\mathrm{As}-\mathrm{b}) \tag{23}$$

$$= \begin{cases} \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}-\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b}, \mathrm{if}\ \mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}+\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A}=0 \\ -\infty, \mathrm{else} \end{cases} \tag{24}$$

Using strong duality:

$$\mathrm{C}_{\mathrm{min}}^{\mathrm{j}} \leq \min_{\mathrm{s}\in\mathrm{U}} \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}+\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s} = \max_{\mu\in\mathbb{R}_{+}^{2\mathrm{I}}} \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}-\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b} :\ \mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}+\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A}=0 \tag{25}$$

Then, by asserting that since the maximum of the argument is at least $\mathrm{C}_{\mathrm{min}}^{\mathrm{j}}$ there should exist at least one value of the dual variable (μ_{j}) that makes the argument at least $\mathrm{C}_{\mathrm{min}}^{\mathrm{j}}$. We drop the maximization. Thus, the infinite number of constraints can be replaced with constraints (26), (27), and (28) for each $\mathrm{j} \in \mathrm{J}$.

$$\mathrm{C}_{\mathrm{min}}^{\mathrm{j}} \leq \mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}-\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b} \tag{26}$$

$$\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}+\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A}=0 \tag{27}$$

$$\mu_{\mathrm{j}} \geq 0 \tag{28}$$

Similarly, the upper bound constraint (for some $\mathrm{j} \in \mathrm{J}$) can be reformulated to:

$$\mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{CR}+\nu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b} \leq \mathrm{C}_{\mathrm{max}}^{\mathrm{j}} \tag{29}$$

$$\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}} + \nu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A} = 0 \tag{30}$$

$$\nu_{\mathrm{j}} \geq 0 \tag{31}$$

So the final IGDT problem of equations 16-20 can be reformulated into equations (32-40)

$$\max_{\delta,\mathrm{RCAT},\mu_{\mathrm{j}},\nu_{\mathrm{j}},\mathrm{j}\in\mathrm{J}} \delta \tag{32}$$

$$\mathrm{s.t.\,co}^{\mathrm{T}}(\mathrm{IRC} - \mathrm{RCAT}) \leq \mathrm{obj}' \tag{33}$$

$$\mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT} - \mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b} \geq \mathrm{C}_{\mathrm{min}}^{\mathrm{j}}, \forall \mathrm{j} \in \mathrm{J} \tag{34}$$

$$\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}} + \mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A} = 0, \forall \mathrm{j} \in \mathrm{J} \tag{35}$$

$$\mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT} + \nu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b} \leq \mathrm{C}_{\mathrm{max}}^{\mathrm{j}}, \forall \mathrm{j} \in \mathrm{J} \tag{36}$$

$$\mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}} - \nu_{\mathrm{j}}^{\mathrm{T}}\mathrm{A} = 0, \forall \mathrm{j} \in \mathrm{J} \tag{37}$$

$$\mu_{\mathrm{j}}, \nu_{\mathrm{j}} \geq 0, \forall \mathrm{j} \in \mathrm{J} \tag{38}$$

$$0 \leq \mathrm{RCAT} \leq \mathrm{IRC} \tag{39}$$

$$0 \leq \delta \leq \frac{1}{\sigma} \tag{40}$$

where: $\mu_{\mathrm{j}}, \nu_{\mathrm{j}}\mathrm{J} \in \mathrm{J}$ is a dual variable that is required for writing the dual problem. The above formulation gives a larger and larger set as obj' is increased since a more robust solution will increase the objective value but will be more immunized against larger realizations of the mixing parameter s. The obtained formulation is non-convex because of the terms $\mu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b}$ and $\nu_{\mathrm{j}}^{\mathrm{T}}\mathrm{b}$. The cuckoo search algorithm is used to solve this problem.

3.2 Cuckoo Search Algorithm

Cuckoo search is a metaheuristic based on swarm intelligence inspired by the breeding behavior of a few species of cuckoos. Yang and Deb introduced this algorithm in the year 2009. The optimization algorithm has been applied in multiple areas in engineering as well as in the sciences. This algorithm and its variations have effectively solved the WDS design problem (Naveen Naidu et al. 2020; Pankaj et al. 2020; Wang et al. 2012).The working mechanism of the cuckoo search algorithm in (Figure 2) is extracted from (Pankaj et al. 2020). And the detailed explanation of the mechanism can be referred from (Yang and Deb 2009). This cuckoo search algorithm is used to obtain the robustness α for the info-gap problem formulation.

BY NC SA

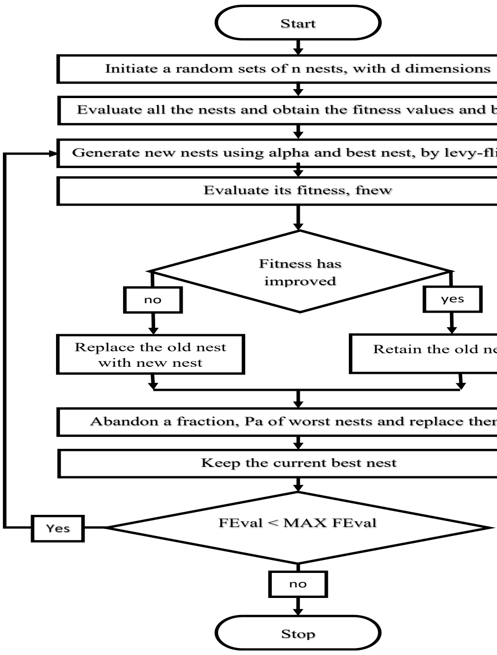

Start
Initiate a random sets of n nests, with d dimensions
Evaluate all the nests and obtain the fitness values and best
Generate new nests using alpha and best nest, by levy-flight
Evaluate its fitness, fnew
Fitness has improved
no
yes
Replace the old nest with new nest
Retain the old nest
Abandon a fraction, Pa of worst nests and replace them
Keep the current best nest
FEval < MAX FEval
Yes
no
Stop

Figure 2 Cuckoo search algorithm flow chart

3.3 Illustrative example

A 4x4 grid network is considered for this study, as shown in (Figure 3), where all the possible junctions where incomplete mixing is feasible are highlighted. The possibility of incomplete mixing is dependent on flow direction. The objective is to obtain optimal treatment costs. Sources 1 and 2 are at 0m elevation, and source-3 is 70m. The pump linking source-1 is of power 2, and the pump linking source 2 is of power 1.5. The base demand, elevation, and concentration limits of all the nodes are mentioned in Table 1.

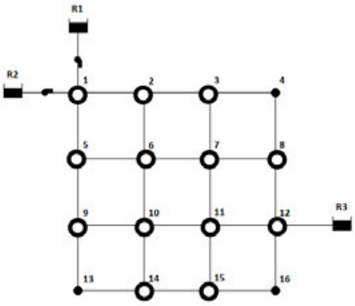

R1
R2
R3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Figure 3 A 4x4 grid network that is considered for this study with three reservoirs. The nodes where the non-uniform mixing can occur are highlighted

Initial concentrations at the sources before treatment IRC = (100,200,150). Treatment costs for each source co = (5000, 200,10000).

Objective function:

$$\min_{i=1,2,3} \sum \mathrm{co}^{\mathrm{T}} \, (\mathrm{IRC} - \mathrm{RCAT}) \qquad (56)$$

Here RCAT is the reservoir concentration after treatment entering the network.

Constraints:

$$\mathrm{C}_{\mathrm{j}}^{\mathrm{min}} \leq \mathrm{C}_{\mathrm{out}}^{\mathrm{j}}\big(\mathrm{RCAT}, \mathrm{Q}, \mathrm{s}(\xi)\big)_{\mathrm{BAM}} \leq \mathrm{C}_{\mathrm{j}}^{\mathrm{max}} \qquad (57)$$

C_{j} = concentration of the flow from node j, $\mathrm{C}_{\mathrm{j}}^{\mathrm{min}}, \mathrm{C}_{\mathrm{j}}^{\mathrm{max}}$ are the lower and upper bounds for the concentration at node j.

Surrogate model:

As mentioned earlier, to apply RO and IGDT, a surrogate model is needed to replace the BAM model-based equations for outlet concentrations. A unique surrogate equation is developed for each outlet concentration (10 nodes- Illustrative example). After evaluating various forms, $\mathrm{a}_{\mathrm{j}}^{\mathrm{T}}\mathrm{RCAT} + \mathrm{CR}^{\mathrm{T}}\mathrm{E}_{\mathrm{j}}\mathrm{s}$ form has been selected based on the model accuracy. Here, a_{j} is a vector and E_{j} is a matrix corresponding to node j. These equations are developed for a particular design of WDN (i.e., constant flow direction and flow values). The procedure to obtain the surrogate model is explained as a flowchart in (Figure 4).

Table 1 Illustrative example node data

Node-ID	Base Demand (GPM)	Min Co-Limit	Max Co-Limit
1	0	0	0
2	10	40	70
3	5	40	70
4	10	40	70
5	50	20	50
6	0	0	0
7	0	0	0
8	10	20	50
9	30	30	60
10	0	0	0
11	0	0	0
12	0	0	0
13	40	30	60
14	80	30	60
15	60	20	50
16	60	40	70

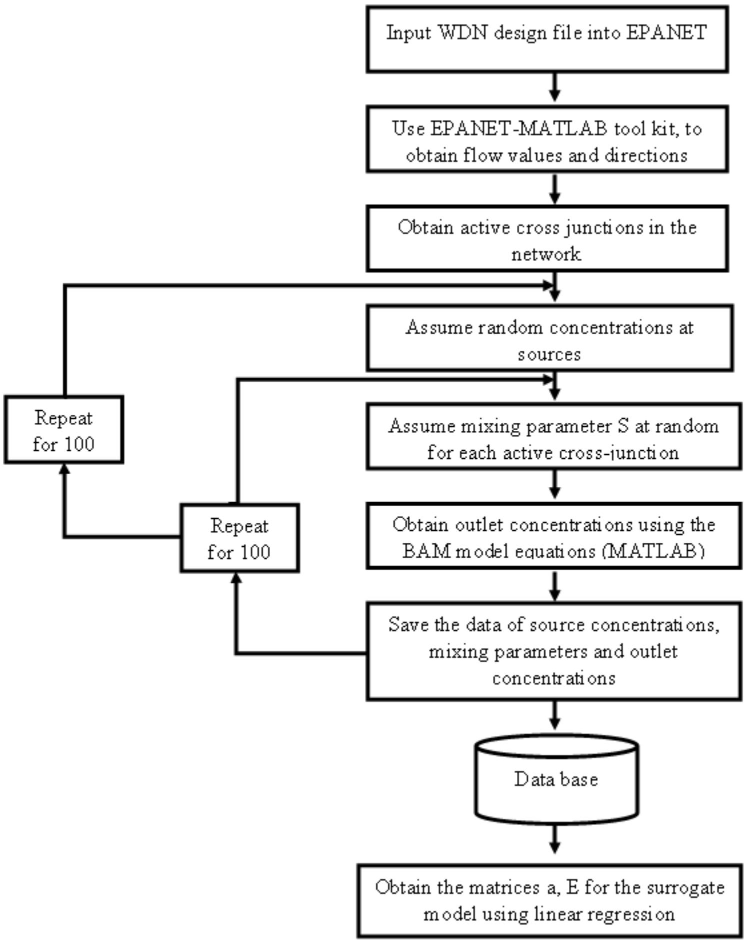

Input WDN design file into EPANET
Use EPANET-MATLAB tool kit, to obtain flow values and directions
Obtain active cross junctions in the network
Assume random concentrations at sources
Assume mixing parameter S at random for each active cross-junction
Obtain outlet concentrations using the BAM model equations (MATLAB)
Save the data of source concentrations, mixing parameters and outlet concentrations
Data base
Obtain the matrices a, E for the surrogate model using linear regression
Repeat for 100
Repeat for 100

Figure 4 Procedure for obtaining surrogate model

The surrogate model's accuracy depends on the flow directions, and the node location, the maximum RMSE obtained for the cases studied is 2.876, with a minimum being $R^2 = 0.989$. This shows that the linear approximation is decently accurate. Monte Carlo simulations are used to evaluate the reliability of the obtained design.

After obtaining the surrogate models for each scenario, the nominal or deterministic problem is evaluated assuming complete mixing (i.e., s=e) for all cross junctions. The obtained treatment costs are assumed to be base treatment costs, and the mixing parameter set (s=e) as the nominal mixing parameter value. Using this nominal treatment cost, the maximum budget which can be provided is approximated by linearly increasing the nominal budget (10 times). For each of these approximated budgets, the info-gap problem is solved to obtain the least cost and the corresponding maximum robustness (δ). For solving this problem, a cuckoo search optimization algorithm is used. The initial nests (search agents) are taken to be 25, and the Lévy Flight parameters $\alpha_{\text{levy}} = 0.01$, $P_a = 0.25$ and $\beta_{\text{levy}} = 2/3$ as suggested in (Yang and Deb 2009) are used. The algorithm is run for 200 iterations or 25*2*200 = 10000 function evaluations. For all the cases, the problem converged at (100 ± 10) iterations. The uncertainty set obtained from this (δ) is reported. For each scenario, ten different linearly increasing budgets (Green) are solved, and their corresponding maximum treatment cost (Blue) and robust uncertainty set (X-axis) are reported in the graphs (Figure 5).

4 CONCLUSIONS

The present work is to understand the effect of non-uniform mixing as a water quality uncertain parameter. A simple treatment cost optimization problem is devised to understand the optimal design of the non-uniform mixing in a water distribution system. Non-uniform mixing uncertainty cannot assume any probability distribution. Handling this uncertainty requires a non-probabilistic approach. IGDT methodology is explored in this study. The complete methodology is explained with a simple grid network with three water sources. The complexity of the non-uniform mixing is clearly explained.

With varying flow patterns, the outlet concentration patterns have changed. The worst-case cannot be assumed as it changes with a change in flow directions and values. Considering incomplete mixing at all junctions did not lead to worst-case (i.e., maximum constraint violations). The change in flow patterns changes the locations of cross junctions, changing the non-uniform mixing junctions. A new set of surrogate models are required for a change in each flow direction and flow value. A convex surrogate model which includes flow values and direction with reasonable accuracy is difficult to achieve.

The treatment cost increases with an increase in the uncertainty set is observed from the results. Changes in flow direction changed the optimal treatment levels. The Monte Carlo simulation results show a small percentage of infeasibility is due to the surrogate model's approximation. This IGDT approach can be easily applied to any network and can solve this optimal treatment problem. Further study is to combine the water distribution network design problem with this treatment cost design problem. The emphasis was more on the theory, and less on using real data. This is why unitless cost data were used, and conservative water quality constituents. Applications to non-conservative water quality parameters, real cost data, and more complex systems are suggested as future work, as well as efficient extension to extended period simulations (EPS) loading conditions.

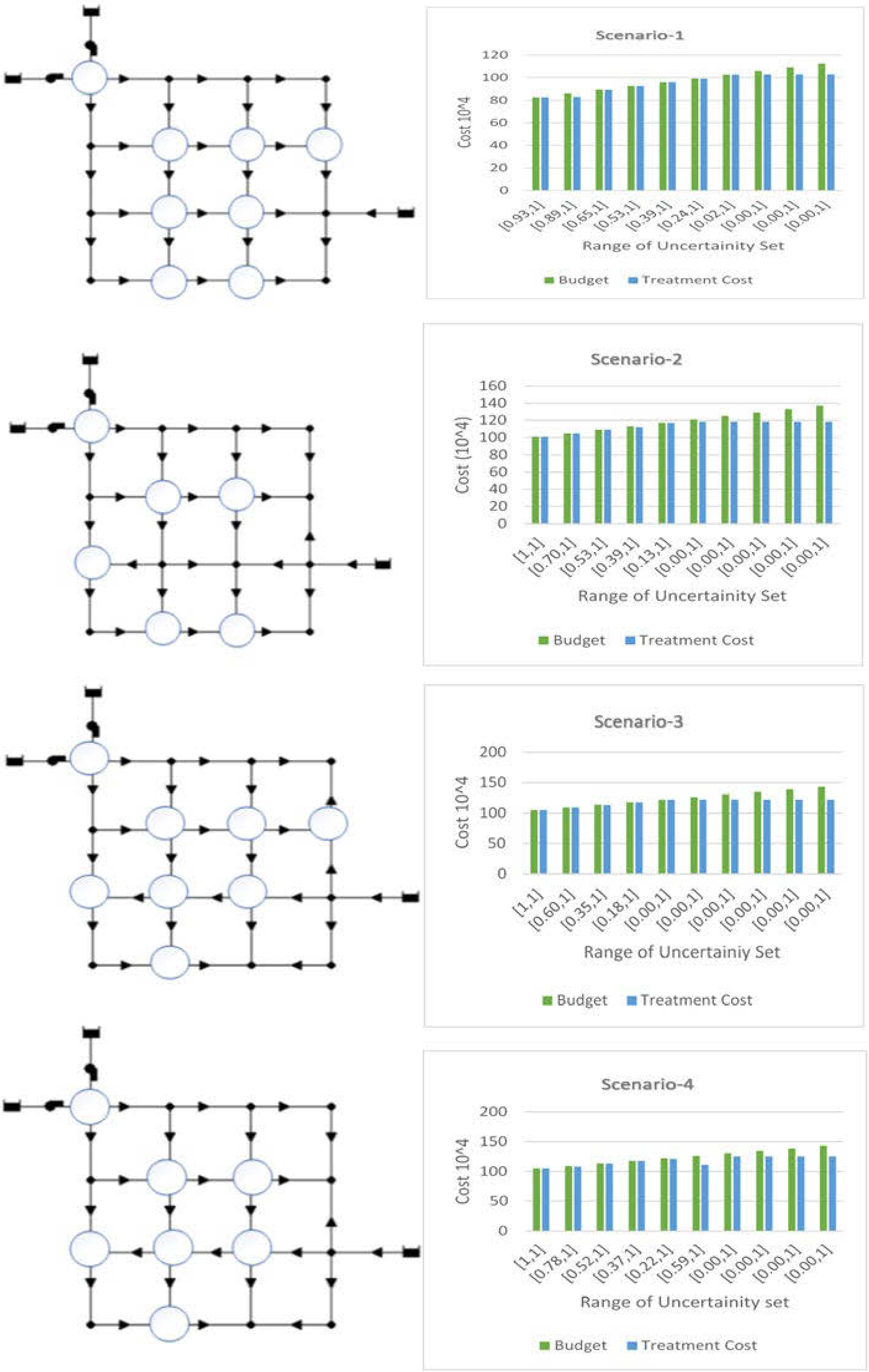


Figure 5 Results show the variation of uncertainty sets with an increase in budget values for four different flow directions, as shown in the network diagrams.

cc BY NC SA

5 ACKNOWLEDGEMENT

This research was supported by the Israel Science Foundation (Grant No. 555/18).

6 REFERENCES

[1] Babayan, A., Kapelan, Z., Savic, D., and Walters, G. (2005). "Least-Cost Design of Water Distribution Networks." (October), 375–382.

[2] Babayan, A. V, Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2004). "Comparison of two approaches for the least cost design of water distribution systems under uncertain demands." Critical Transitions in Water and Environmental Resources Management, 1–10.

[3] Ben-Haim, Y. (2006). Info-gap decision theory: decisions under severe uncertainty. Elsevier.

[4] Chung, G., Lansey, K., and Bayraksan, G. (2009). "Reliable water supply system design under uncertainty." Environmental Modelling and Software, Elsevier Ltd, 24(4), 449–462.

[5] Fowler, A. G., and Jones, P. (1991). "Simulation of water quality in water distribution systems." Proc., Water Quality Modeling in Distribution Systems, AWWA/EPA, Cincinnati

[6] Ghelichi, Z., Tajik, J., and Pishvaee, M. S. (2018). "A novel robust optimization approach for an integrated municipal water distribution system design under uncertainty: A case study of Mashhad." Computers and Chemical Engineering, Elsevier B.V., 110, 13–34.

[7] Ho, C. K., and Khalsa, S. S. (2008). "EPANET-BAM: water quality modeling with incomplete mixing in pipe junctions." Water Distribution Systems Analysis 2008, 1–11.

[8] Ho, C. K., Orear Jr, L., Wright, J. L., and McKenna, S. A. (2008). "Contaminant mixing at pipe joints: Comparison between laboratory flow experiments and computational fluid dynamics models." Water Distribution Systems Analysis Symposium 2006, 1–18.

[9] Kang, D. S., Pasha, M. F. K., and Lansey, K. (2009). "Approximate methods for uncertainty analysis of water distribution systems." Urban Water Journal, 6(3), 233–249.

[10] Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2005). "Multiobjective design of water distribution systems under uncertainty." Water Resources Research, 41(11), 1–15.

[11] Korteling, B., Dessai, S., and Kapelan, Z. (2013). "Using Information-Gap Decision Theory for Water Resources Planning Under Severe Uncertainty." Water Resources Management, Kluwer Academic Publishers, 27(4), 1149–1172.

[12] Lansey, K. E., Duan, N., Mays, L. W., and Tung, Y. (1989). "Water Distribution System Design Under Uncertainties." Journal of Water Resources Planning and Management, 115(5), 630–645.

[13] Naderi, M. J., and Pishvaee, M. S. (2017). "Robust bi-objective macroscopic municipal water supply network redesign and rehabilitation." Water Resources Management, 31(9), 2689–2711.

[14] Naveen Naidu, M., Boindala, P. S., Vasan, A., and Varma, M. R. R. (2020). Optimization of Water Distribution Networks Using Cuckoo Search Algorithm. Advances in Intelligent Systems and Computing, Springer Singapore.

[15] Paez, N., Saldarriaga, J., and Bohorquez, J. (2017). "Water Quality Modeling Considering Incomplete Mixing in Extended Periods." Procedia Engineering, The Author(s), 186(571), 54–60.

[16] Pankaj, B. S., Naidu, M. N., Vasan, A., and Varma, M. R. (2020). "Self-Adaptive Cuckoo Search Algorithm for Optimal Design of Water Distribution Systems." Water Resources Management, Water Resources Management, 34(10), 3129–3146.

[17] Pasha, M. F. K., and Lansey, K. (2005). "Analysis of uncertainty on water distribution hydraulics and water quality." World Water Congress 2005: Impacts of Global Climate Change - Proceedings of the 2005 World Water and Environmental Resources Congress, 10.

[18] Pasha, M. F. K., and Lansey, K. (2010). "Effect of parameter uncertainty on water quality predictions in distribution systems-case study." Journal of Hydroinformatics, 12(1), 1–21.

[19] Perelman, L., Housh, M., and Ostfeld, A. (2013). "Robust optimization for water distribution systems least cost design." Water Resources Research, 49(10), 6795–6809.

[20] Roach, T., Kapelan, Z., and Ledbetter, R. (2015). "Comparison of info-gap and robust optimization methods for integrated water resource management under severe uncertainty." Procedia Engineering, Elsevier B.V., 119(1), 874–883.

[21] Romero-Gomez, P., Choi, C. Y., Lansey, K. E., Preis, A., and Ostfeld, A. (2009). "Sensor network design with improved water quality models at cross junctions." Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008, (2004), 1056–1066.

[22] Song, I., Romero-Gomez, P., and Choi, C. Y. (2009). "Experimental Verification of Incomplete Solute Mixing in a Pressurized Pipe Network with Multiple Cross Junctions." Journal of Hydraulic Engineering, 135(11), 1005–1011.

[23] Wang, Q., Liu, S., Wang, H., and Savić, D. A. (2012). "Multi-objective cuckoo search for the optimal design of Water Distribution Systems." Civil Engineering and Urban Planning 2012 - Proceedings of the 2012 International Conference on Civil Engineering and Urban Planning, 2012(Ceup), 402–405.

[24] Yang, X. S., and Deb, S. (2009). "Cuckoo search via Lévy flights." 2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009 - Proceedings, IEEE, 210–214.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14812

NUMERICAL AND EXPERIMENTAL INVESTIGATION ON LEAK DISCHARGE-PRESSURE RELATIONSHIP FOR LINEAR ELASTIC DEFORMING PIPES

**Yuanzhe Li[1], Jinliang Gao[2], Jianxun Chen[3], Wenyan Wu[4],
Shihua Qi[5] and Jingyang Yu[6]**

[1,2,3]School of Environment, Harbin Institute of Technology, Harbin, Heilongjiang Province, China
[4]School of Engineering and the Built Environment, Birmingham City University, Birmingham, UK
[5,6]Heilongjiang College of Construction, Harbin 150500, China

[1] *yuanzhelihit@163.com* , [2] *gjl@hit.edu.cn,* [3] *cjianxun877@gmail.com,* [4] *wenyan.wu@bcu.ac.uk*
[5] *16678993@qq.com,* [6]*qishihua@126.com*

Abstract

Large amount of research investigated the flow rate from an individual leak as a function of the internal pressure, but few of them managed to separate the effect of pipe deformation and that of leak hydraulics. In this study, experiments were carried out simultaneously measuring leak area, flow rate and pressure head so that the discharge coefficient can be experimentally determined. The experiments under both free and submerged condition were conducted. A few anomalies were found. For example, the leakage flow rate of submerged discharge was significantly greater than that of free discharge under full turbulence regime. Meanwhile, for submerged discharge, the discharge coefficient was not constant, but increased first and then decreased. Through CFD simulation, this study provided preliminary expansions of the two anomalies. The less leak flow rate of free discharge may be caused by the friction of water flow and air, and the abnormal decrease of the discharge coefficient of submerged discharge may due to cavitation phenomenon. The phenomena observed in this study could offer more information for researchers about leak hydraulics.

Keywords
Leakage Management, Free Discharge, Submerged Discharge, Pressure Management, Water Distribution System.

1 INTRODUCTION [FIRST ORDER HEADING, CAMBRIA 12-POINT, BOLD, LEFT JUSTIFIED, CAPITALIZED, 12PT LINE SPACING BEFORE AND AFTER]

Leakage is the main component of non-revenue water (NRW) and an important index of water loss in water distribution system (WDS). The relationship between the leakage flow rate of a certain leak and the internal pressure of the pipe plays an important role in leakage management. The orifice equation, which was derived based on the Torricelli's law, was often used to calculate leakage flow rate in WDS:

$$Q = C_d A\sqrt{2gh} \tag{1}$$

where Q is the leakage flow rate; C_d is the discharge coefficient; A is the leak area, and h is the pressure drop. However, since the 1990s, numerous field tests and experiments have concluded that the leak flow rate was much more sensitive than the relationship predicted by the orifice equation[1-2], spanning from 0.5 to 2.79, with an average value of 1.15. a power equation was introduced and recommended by the Water Losses Task Force of the International Water Association(IWA):

$$Q = Ch^N \tag{2}$$

where C denotes the leakage coefficient and N denotes the leakage exponent. IWA's power equation was simple and easy to use, however, it can't reflect the real physical phenomenon. May [3] proposed a model assuming that the leak area increases linearly with increasing pressure:

$$Q = C_d(A_0 + mh)\sqrt{2gh} \tag{3}$$

where A_0 is the initial leak area under zero pressure conditions and m is the head-area slope.

Numerous researches were conducted to investigate the leak flow-pressure relationship under the effect of pipes' defamation [4-7]. However, due to the effect of leak discharge, few researches managed to measure leak area directly. Thus, the relationship between C_d and pressure can't be experimentally investigated.

In this study, through image analysis methods [7], leak area was measured experimentally together with leak flow rate and pressure. To provide more information for researchers about leak hydraulics, computational fluid dynamics (CFD) simulations were carried out.

2 METHOD AND MATERIALS

2.1 Laboratory Setup

A schematic of the experimental apparatus used in this study is illustrated in Fig. 1. The experiments were conducted in the Pipe Leakage Test Laboratory in Harbin Institute of Technology, China, composed by adjustable pipes and hydraulic pumping systems controlled by variable frequency drivers.

The pump could supply water at a speed of 7 L/s and a head of 75 m. An air vessel was installed downstream of the pipe to stabilize the upstream pressure. Two air valves were installed both upstream and downstream of the test section to remove the air from the pipe. Two electromagnetic meters with ±0.5% accuracy was used to measure and record the flow rate. The leakage flow rate was determined by calculating the difference between the upstream and downstream flowmeter readings. The leak flow discharge from the test tank was collected in a bucket with a capacity of 17 L. The leakage flow rate data were calibrated by simultaneously measuring the leaked water weight and the time through a weight sensor with 20 Kg full scale (FS) and 0.05% FS accuracy. Two series of manometers with 1.0 MPa and 0.10 MPa FS were utilized to measure the pressure the test section. The accuracy of the manometers was 0.2% FS. The manometers with 0.10 MPa FS was used to measure the pressure at low-pressure conditions with higher accuracy. The data collected by the weight sensor, flow meter, and manometers were recorded simultaneously by a computer at a frequency of 1 Hz.

A replaceable test pipe section with an artificial crack was installed in a tank with 0.5 m^3 capacity. As depicted in Fig. 1b, one side of this tank was transparent in order to facilitate the leak area measurements. The length of the test pipe section between two flange was 1m. The parameters of the test sections were summarized in Table.1. On the other side of the tank, a 0.6-m-high overflow weir was installed. The test pipe section was installed horizontally, facing the transparent side of the tank. The average height of the leak on the test section with respect to the bottom of the test tank was 0.1 m, so that the external water pressure head could be stabilized at 0.5 m.

Table 1. Parameters of test sections

Number of test sections	Geometry of test sections	Diameter/Length of the leak (mm)	Initial Area (mm²)	Pipe material	Nominal Diameter (mm)	Pipe Thickness (mm)
1	Longitudinal crack	21.05	26.59	HDPE	50	4.3
2	Longitudinal crack	20	20	Steel	50	3.5
3	Round hole	5	19.64	Steel	50	3.5
4	Circumferential crack	20	20	Steel	50	3.5

Similar to the procedure reported by van Zyl and Malde [8], the pressure of the test section was increased and decreased in a stepwise manner. Each test step lasted about 30 s, and was proved to be long enough for the flow and pressure data to be stabilized. To eliminate the effect of viscoelastic behaviour of pipes, after each test step the pump would turn off for 1 minute to provide enough time for pipe material recovery. Once the total time of an experiment run exceeded 2700s, the experiment would be terminated and restarted after at least 4 hours.

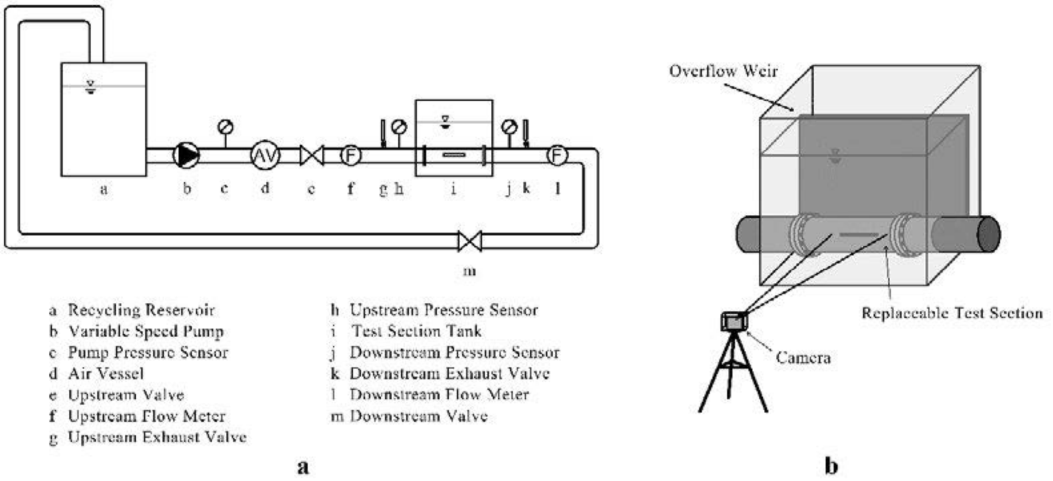

a
b
c
d
e
f
g h
i
j k
l
m
a Recycling Reservoir
b Variable Speed Pump
c Pump Pressure Sensor
d Air Vessel
e Upstream Valve
f Upstream Flow Meter
g Upstream Exhaust Valve
h Upstream Pressure Sensor
i Test Section Tank
j Downstream Pressure Sensor
k Downstream Exhaust Valve
l Downstream Flow Meter
m Downstream Valve
a
Overflow Weir
Replaceable Test Section
Camera
b

Figure 1. Layout of experimental setup.

The objective of the experiment was to simultaneously acquire data concerning the leak area-pressure head (A-h) and leak flow rate-pressure (Q-h). Image analysis methods, similar to those used in (Fox et al., 2016b), were adopted to measure the area under the effect of leakage flow. Images of the cracks were recorded using a camera with a resolution of 48 MP. The external surface around the crack was painted with waterproof red paint to provide a clear distinction line between the red pipe surface and the black leak area under the effect of leak discharge. Subsequently, an image processing procedure was followed to separate the leak area from the pipe edges based on the RGB value of the pixels. The initial length and average width of the crack on each test section were measured using a vernier caliper with an accuracy of 0.1 mm. Before

BY NC SA

each experiment, a ruler with known length and width was installed and photographed with the camera, in order to calibrate the ratio between pixel size and actual size. Under each pressure level, three images were recorded and the leak area value was the average of the three measurements. Due to the complex procedures involved in the image analysis methods adopted in this study, the leak area was recorded every 5 mH2O.

Furthermore, the wetted perimeter of the crack outlet section (χ) was measured through the image analysis method. By combining the measured leak area and flow rate, the Reynolds number of the crack outlet section under different pressures can be determined.

The experiments of this study were carried out under both submerged and free condition. When taking photos to record the leakage area, the pipe was submerged. To collect the data for free discharge, a submersible pump was installed at the bottom of the tank to discharge the leaked water from the tank. In this study, it was assumed that only the internal pressure had an impact on the leakage area, and the effect of surrounding media can be ignored. The discharge coefficient under free condition was calculated using the area measured under submerged condition.

2.2 Settings of CFD simulations

In order to obtain more information about the relationship between pressure and discharge coefficient, interaction of leak discharge and surrounding media near the leakage was investigated by finite element simulations. Model development and analyses were conducted in ANSYS Workbench (ANSYS, 2021).

Fig. 2 illustrates the 3D geometrical models of the fluid domain, which was composed of fluid field in the pipe, in the leak and in the water tank around the pipe. The geometry of computational domain of the finite element simulation were the same as those of the experimental device, and boundary conditions were the same as the experimental conditions. The height of the fluid domain in tank was set as 0.5m and the pressure of its outlet was set as 99500 Pa, which was the local atmosphere pressure in Harbin.

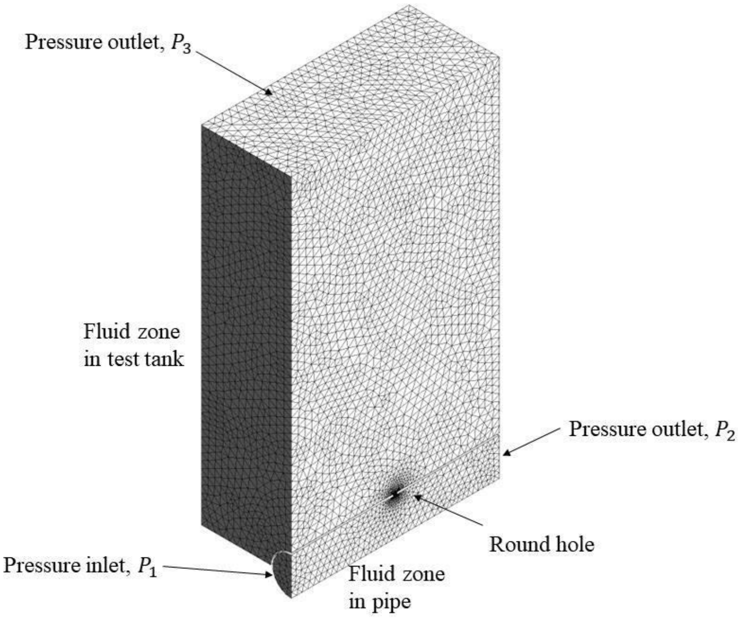


Figure 2. Standardized mesh distribution for finite-element analysis for a 5 mm round hole (test 3).

In order to avoid the influence of the side wall of the water tank on the calculation results, the direction of the leakage is set to be vertically upward. Experiments was conducted and proved that the orientation of the leakage has no effect on the leakage flow rate.

A plane of symmetry was used to reduce the model size and consequently increase calculation efficiency. Full and half fluid domain models were run and compared in order to confirm that this had a negligible impact on the leakage flow rate. The mesh near the leak is encrypted. A mesh invariance analysis was conducted and showed that increasing the resolution of the developed mesh did not significantly alter the simulation solution.

To simulate the submerged discharge, the material of fluid domain in tank was set as water. It was set as air when simulating free discharge. The turbulence model was set as realizable k-ϵ model, and the multiple flow model is set as mixture model. The measured leak flow rate and pressure data was used to calibrate the simulation results.

3 RESULTS AND DISSCUSSION

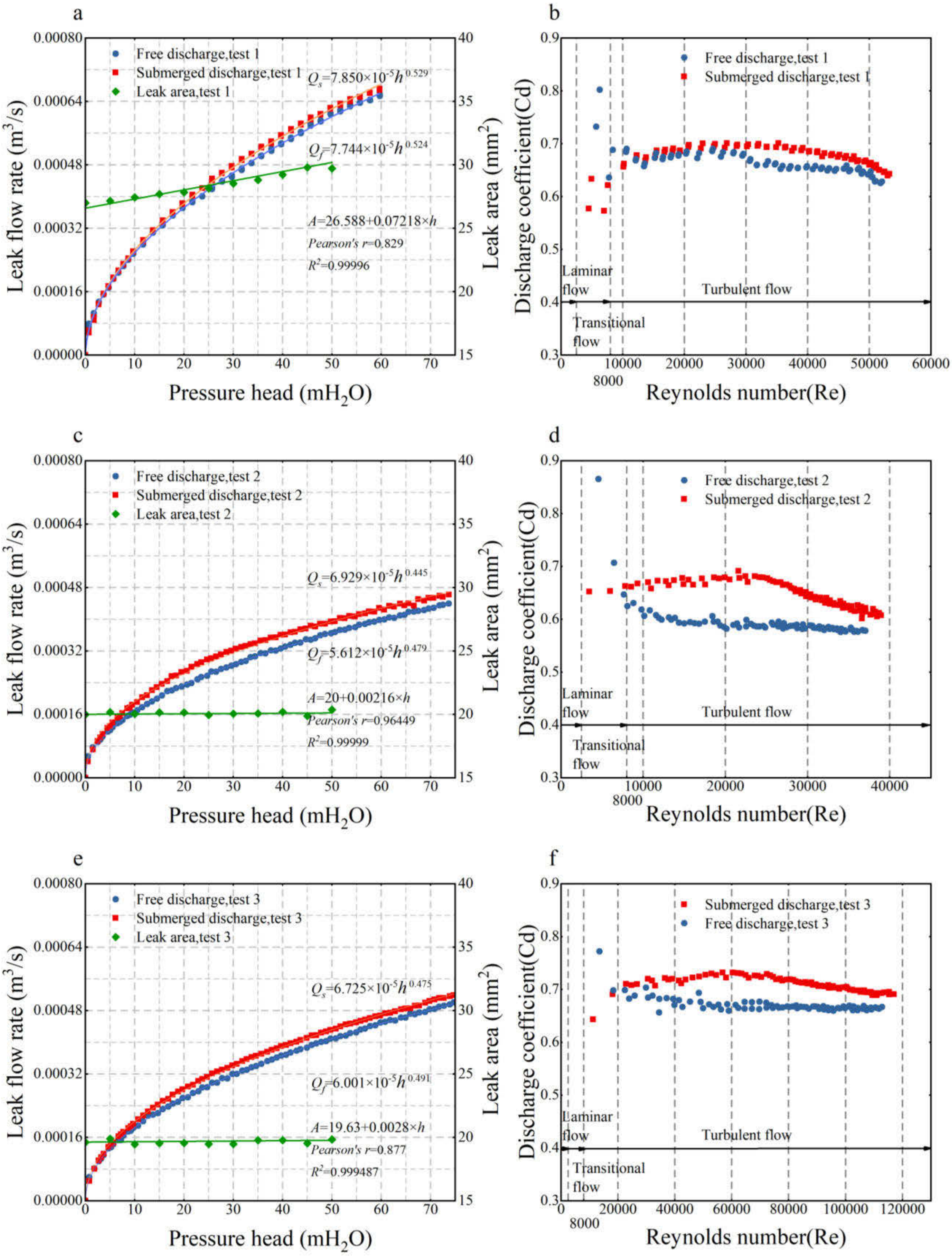
Variation of the measured leak area and flow rate with internal pressure and corresponding variation of discharge coefficient with Reynolds number for the experimental test 1~4 under the conditions of free discharge and submerged discharge are shown in Fig. 3.

It can be seen that there was an obvious linear correlation between leakage area and pressure when viscoelastic deformation effect was eliminated. In addition to the obvious linear deformation of HDPE pipe, the linear deformation of steel pipe was almost negligible. This is consistent with the study of van Zyl and Malde [8]. Combined with the data of initial area, the relationship between leakage area and pressure can be obtained by using the least square method. The discharge coefficient can be obtained by using the orifice equation:

$$C_d = \frac{Q}{A\sqrt{2gh}} = \frac{Q}{(A_0 + mh)\sqrt{2gh}} \tag{4}$$

The IWA's power equation was used to fit the pressure-leak flow data. It can be seen that only the leakage exponent of HDPE pipe is greater than 0.5. This was consistent with the conclusions of previous researches that the expansion of the leakage area could lead to the leakage exponent greater than 0.5.

In addition, interesting phenomena can be seen by comparing the leak flow rate-pressure curves of free discharge and submerged discharge under the same conditions. Since there is an external pressure of 0.5 mH_2O under submerged condition, theoretically, the leakage flow rate of the submerged discharge should be less than that of the free discharge when the internal pressure and discharge coefficient are the same. However, the data was not consistent with this prediction: in all 4 tests, when the internal pipe pressure was less than 2 mH_2O, the leakage flow rate of free discharge was greater than that of submerged discharge; When the pressure was between 2 mH_2O and 5 mH_2O, the leakage flow of the two were roughly the same; When the pressure was greater than 5 mH_2O, the leakage flow of submerged discharge was significantly greater than that of free discharge.


a
Free discharge,test 1
Submerged discharge,test 1
Leak area,test 1
Q_s=7.850×10$^{-5}h^{0.529}$
Q_f=7.744×10$^{-5}h^{0.524}$
A=26.588+0.07218×h
Pearson's r=0.829
R²=0.99996
Leak flow rate (m³/s)
Leak area (mm²)
Pressure head (mH₂O)
0.00000 0.00016 0.00032 0.00048 0.00064 0.00080
15 20 25 30 35 40
0 10 20 30 40 50 60 70
b
Free discharge,test 1
Submerged discharge,test 1
Laminar flow
Transitional flow
Turbulent flow
Discharge coefficient(Cd)
Reynolds number(Re)
0.3 0.4 0.5 0.6 0.7 0.8 0.9
0 8000 10000 20000 30000 40000 50000 60000
c
Free discharge,test 2
Submerged discharge,test 2
Leak area,test 2
Q_s=6.929×10$^{-5}h^{0.445}$
Q_f=5.612×10$^{-5}h^{0.479}$
A=20+0.00216×h
Pearson's r=0.96449
R²=0.99999
Leak flow rate (m³/s)
Leak area (mm²)
Pressure head (mH₂O)
d
Free discharge,test 2
Submerged discharge,test 2
Laminar flow
Transitional flow
Turbulent flow
Discharge coefficient(Cd)
Reynolds number(Re)
0 8000 10000 20000 30000 40000
e
Free discharge,test 3
Submerged discharge,test 3
Leak area,test 3
Q_s=6.725×10$^{-5}h^{0.475}$
Q_f=6.001×10$^{-5}h^{0.491}$
A=19.63+0.0028×h
Pearson's r=0.877
R²=0.999487
Leak flow rate (m³/s)
Leak area (mm²)
Pressure head (mH₂O)
f
Submerged discharge,test 3
Free discharge,test 3
Laminar flow
Transitional flow
Turbulent flow
Discharge coefficient(Cd)
Reynolds number(Re)
0 8000 20000 40000 60000 80000 100000 120000

cc BY NC SA

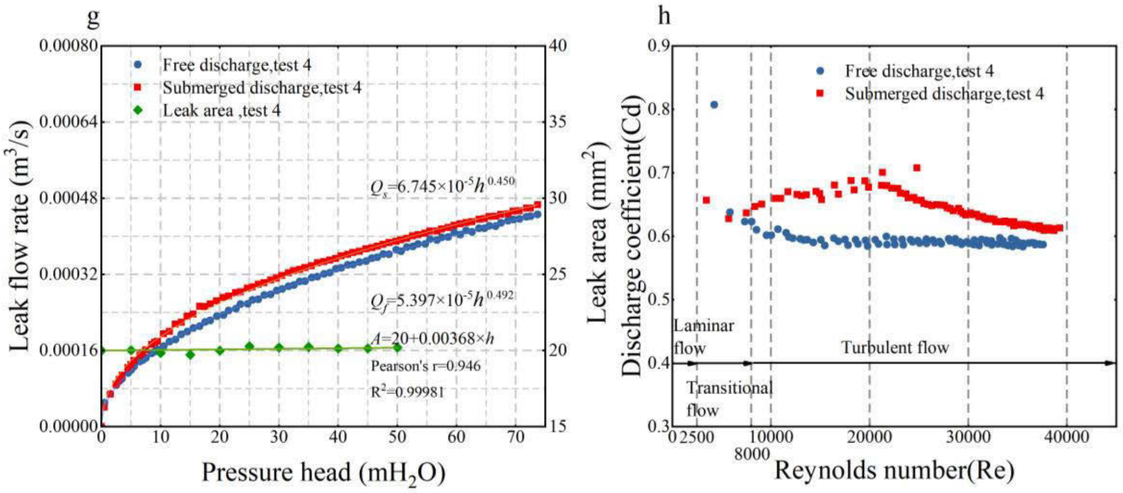


Figure 3. Variation of the measured leak area and flow rate with internal pressure and corresponding variation of discharge coefficient with Reynolds number for test 1~4

By analysing the relationship between discharge coefficient and Reynolds number, it is not difficult to see that all the flows in experiment tests were under the transition regime or fully turbulent regime. Since the normal operating pressure of the real WDS is usually greater than 10 mH_2O, it can be inferred that most of the leakage flow in real WDS is under fully turbulent regime. The leakage coefficients of free discharge and submerged discharge showed different characteristics. For free discharge, the discharge coefficient decreased monotonically with Reynolds number, while at high Reynolds number, it decreased lowlier with the increase of pressure or even remained unchanged, which is consistent with to Iderik's inverse proportional model. However, for submerged discharge, the discharge coefficient was not constant, but increased first and then decreased.

To authors' knowledge, there lacks detailed research explaining the difference between free discharge and submerged discharge. This study tried to give some explanations through the information obtained by CFD simulations. This study only simulated leak hydraulics without considering the variation of leak area caused by pressure. Therefore, this study only simulates the working conditions of test2, 3 and 4.

Under turbulence regime, taking the leakage of 5mm Round hole (test 3) under 40 mH_2O as an example, the streamline near the hole is shown in Figure 4. The lower part of the fluid domain was the water flow inside the pipe, the upper part was the media around the pipe, the middle part is fluid in the hole. Obvious vena contracta can be observed in the hole, and there was a clear eddy current near the vena contracta. The velocity of eddy in free discharge was significantly higher than that of submerged discharge, which means that the disturbance of air was greater than that of water in the orifice. In addition, by observing the volume fraction of air for free discharge near the round hole shown in Figure. 5, due to the friction between water flow and air, the water column will gradually diffuse, resulting in a much more dispersed velocity distribution of free discharge than that of submerged discharge. More sufficient mixing and friction with external media may lead to the dissipation of flow kinetic energy, so that the velocity of free discharge is less than that of submerged discharge, resulting in less flow rate.

For the flow with low Reynolds number, the impact by this friction may be less than that of external pressure, so that the leakage flow of free discharge is greater than that of submerged discharge.

in addition, one possible explanation for the abnormal decrease of the discharge coefficient of submerged discharge is that cavitation occurred when the pressure in the pipe exceeded a certain threshold. Bubbles generated by cavitation hindered the leakage flow.

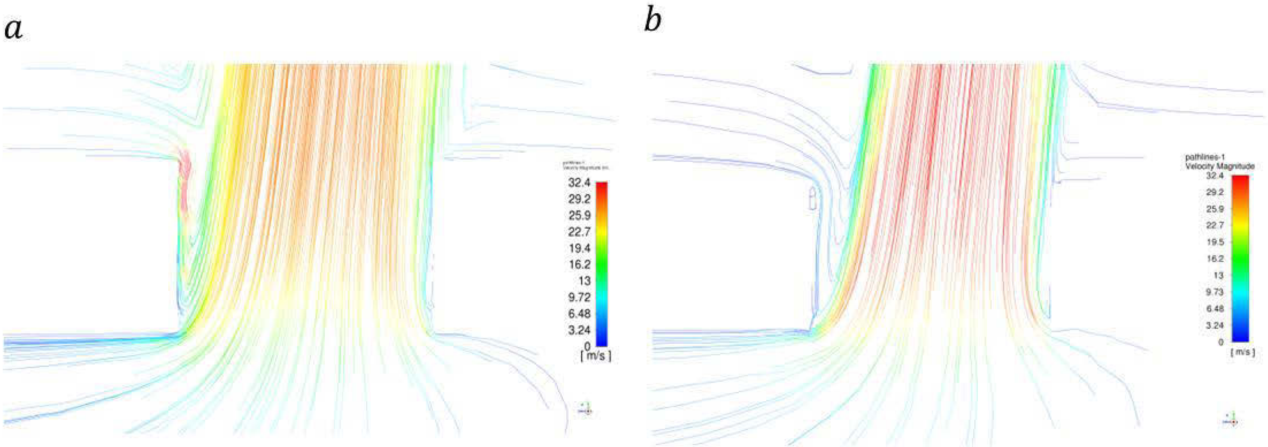


Figure 4. Streamline of leak flow near a 5mm round hole with 40mH₂O pipe internal pressure under (a)free and (b) submerged condition

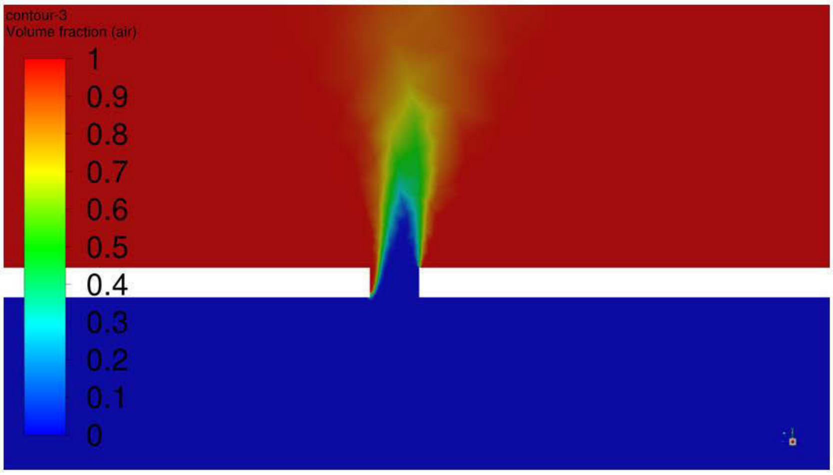


Figure 5. the volume fraction of air for free discharge near the round hole

4 CONCLUSION

For clear understanding of physical characteristics of the relationship between leakage flow rate and pipe internal pressure, experiments were carried out simultaneously measuring leak area, flow rate and pressure head so that the discharge coefficient can be experimentally determined. The experiments under both free and submerged condition were conducted. Two anomalies were observed:

1. The leakage flow rate of submerged discharge was significantly greater than that of free discharge under full turbulence regime, even the external water pressure could have negative effect on submerged discharge.

2. For submerged discharge, the discharge coefficient was not constant, but increased first and then decreased.

Through CFD simulation, this study provided premilitary expansions of the two anomalies. The less leak flow rate of free discharge may due to the friction of water flow and air, and the abnormal decrease of the discharge coefficient of submerged discharge may due to cavitation phenomenon.

The phenomena observed in this study could offer more information for researchers about leak hydraulics. In the future, more research will be carried out to investigate the effect of soil and more efforts will be made to find more reliable theoretical explanations for these anomalies.

5 ACKNOWLEDGEMENTS

This study is jointly supported by the National Natural Science Foundation of China (51778178, 51978203); Natural Science Foundation of Heilongjiang Province of China (No.LH2019E044) and EU Horizon 2020 Marie Skłodowska-Curie Actions-ITN-IOT4Win under Grant 765921.

6 REFERENCES

[1] Lambert, A., “What do we know about pressure: Leakage relationships in distribution systems?” In: System Approach to Leakage Control and Water Distribution Systems Management, Brno, Czech Republic, 2001.

[2] Thornton, J., & Lambert, A. “Progress in practical prediction of pressure: leakage, pressure: burst frequency and pressure: consumption relationships,” in Proc. IWA Special Conference' Leakage, 2005, pp. 12-14.

[3] May, J., “Pressure dependent leakage,” in World Water and Environmental Engineering, London,1994.

[4] Cassa, A. M., J. E. Van Zyl, and R. F. Laubscher. "A numerical investigation into the effect of pressure on holes and cracks in water supply pipes." Urban Water Journal 7.2 2010: 109-120.

[5] Ferrante, M. "Experimental investigation of the effects of pipe material on the leak head-discharge relationship." Journal of Hydraulic Engineering 138.8., 2012, pp. 736-743.

[6] De Marchis, M., Milici, B., "Leakage estimation in water distribution network: effect of the shape and size cracks." Water Resources Management 33.3 2019: 1167-1183.

[7] Fox, Sam, Richard Collins, and Joby Boxall. "Experimental study exploring the interaction of structural and leakage dynamics." Journal of Hydraulic Engineering 143.2,2017,04016080.

[8] Van Zyl, J. E., and R. Malde. "Evaluating the pressure-leakage behaviour of leaks in water pipes." Journal of Water Supply: Research and Technology—AQUA 66.5,2017, 287-299.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14813

HIGH RESOLUTION WATER PRESSURE MONITORING FOR THE ASSESSMENT OF FATIGUE DAMAGE IN WATER DISTRIBUTION PIPES

Carlos Jara-Arriagada[1], Ivan Stoianov[2]

[1,2]InfraSense Labs, Department of Civil and Environmental Engineering, Imperial College London
SW7 2AZ , London, (United Kingdom)

[1] *cij18@imperial.ac.uk*, [2] *ivan.stoianov@imperial.ac.uk*

Abstract

In the last decade, the increasing use of high frequency water pressure monitoring has shown that water distribution networks can be frequently affected by pressure transients. In some instances, thousands of high amplitude cyclic loadings per year have been exerted on pipes posing a risk of fatigue damage. However, despite this awareness of a possible fatigue risk on pipes, current research has not been able to show the extent of the relevance of this risk. To address this issue, this work analyses and evaluates several months to years of high frequency water pressure data at various locations in operational water distribution networks. We have acquired extensive time series datasets with detailed information of transient events, and utilised this information together with principles and knowledge from fracture mechanics to estimate the extent of fatigue damage on pipes. Results from this research suggest that at some logging locations, and depending on the pipe material and current deterioration state, fatigue damage due to constant cyclic loadings can be a major contributor to pipe breaks.

Keywords: Pressure transients, fatigue failure, high frequency pressure monitoring, cycle counting, pipe breaks, water pressure variations

1 INTRODUCTION

A better understanding of causes for pipe breaks in water distribution networks (WDNs) is of critical importance to optimise capital investment and burst reduction [1]. Over the last decade, several research have been performed in order to better understand and predict pipe deterioration and breaks of pipes [2]. As a result, a corpus of modelling approaches and covariates influencing pipe breaks have been presented in the literature [3]. Nevertheless, although much work have been performed, the relevance of some potential factors causing pipe breaks is not clearly understood. Among these potential factors, hydraulic pressure fluctuations are a topic of current attention [4].

Suggestions that water pressure fluctuations could have a an impact on fatigue failure in WDNs, were made in early investigations utilising high frequency water pressure monitoring. These investigations were initially motivated by the need to localize and understand the propagation of pressure transients for leak localisation and pipe break detection [5], [6]. In addition, high frequency pressure monitoring was also used to understand low and negative pressure transients [7]–[9]. Although the purpose of these investigations was not the structural analysis of pipes, they showed early empirical evidence of the frequency and magnitude of transients in WDNs. In the work of Gullick et al. [7], various examples of pressure monitoring for periods up to 5 days in pumping stations showed several pressure transients (in the order of 5 a day) of cycle range close to 50 mH$_2$O. Further evidence of cycles were also reported by Fleming et al. [8], were pressure cycles of range between 20 up to 50 mH$_2$O were measured at two hydrants in a period of 5 days.

The monitoring of water pressure at high frequencies in water distribution networks has presented a challenge that has been investigated and gradually solved over years. Early investigation on the use of high frequency monitoring were presented by Stoianov et al. [10]. These investigations were a decade later further expanded to technologies that allow for more extended monitoring of WDNs with various examples such as [11] and [12]. The use of such technology allowed the extension of monitoring periods from days to weeks up to months, from which further research and characterisation of the frequency of pressure transients have been performed. Examples of recent published investigation include the works of [4], [13], [14]. These works have provided characterisation and a better understanding of how pressure transients occur.

Nevertheless, although the literature has reported several cases of pressure fluctuations, these reports have been obtained only over brief periods of pressure monitoring, thus limiting a generalisation of results. In addition, previous analysis in the literature have focused on the description of the pressure fluctuations and transient events, but have lacked a mechanistic evaluation of the relevance of water pressure fluctuations on the propagation of cracks. To address these gaps in research, the present work analyses and mechanistically evaluates time series of water pressure data acquired over several months of high frequency pressure monitoring. In particular, the obtained time series have been processed utilising a cycle counting algorithm in order to extract the magnitude and frequency of cyclic loading at several locations in WDNs. We have assessed optimal sampling rates for the purpose of cycle counting, and, on the basis of the optimal sampling rate, categorised different water pressure profiles according to the frequency and magnitude of measured cycles. We also present some useful characterisation of the obtained water pressure profiles as input parameters for fatigue crack propagation models and provided how to utilise this data for such models. We further present an example for the analysis of a particular material. The main goal of this paper is to provide fundamental understanding of operating cyclic loading conditions in pipes and discuss the impact of their occurrence using fracture mechanics.

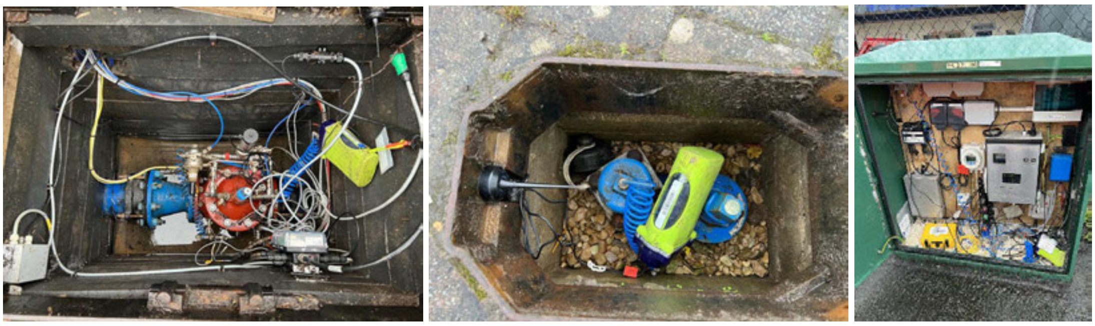
Our results show empirical evidence of the magnitude and frequency of cyclic loadings due to water pressure fluctuations. At some logging locations, thousands of cycles over 20 mH_2O were recorded over a one-year-period, and the results were relatively constant between two different years. When analysing the implications of the magnitude and frequency of the measured cyclic loadings from a mechanistic point of view, depending on the material analysed, a reduction of 10 mH_2O can substantially increase the life of an already deteriorated pipe. Therefore, suggesting the importance of reducing pressure transient activity.

2 METHODS

This investigation involved the analysis of high frequency water pressure monitoring data measured from operational WDNs. The analyses were performed to characterise cyclic loading conditions in pipes and assess their implication on the propagation of cracks and eventual fracture of pipes. This section describes the data and fundamental background and methodologies used to perform such analyses.

2.1 Pressure monitoring technologies

The data used in this study was acquired utilising high frequency pressure monitoring devices [12] installed in various locations. The sampling rate was 128 S/s (Samples per second), although some of the data was only available for our analysis at 1 S/s. This is because battery operated devices shown in Figure 1b were setup to send 1 S/s sub-sampled data from 128S/s locally stored data. The locations for the installation of the pressure monitoring devices were pressure reducing valves (PRVs), fire hydrants, and telemetry bollards/kiosks. Example illustrations of the installation locations are presented in Figure 1.



a) b) c)

Figure 1. Installation of a pressure transducer in a) pressure reducing valve (PRV) b) fire hydrant c) telemetry kiosk

2.2 Assessment of the required sampling rate

Some pressure monitoring devices were set to communicate their raw data of 128 S/s, while others only communicated a sub-set of 1 S/s derived from the continuously acquired 128 S/s. Pressure data acquired at 128 S/s have been utilised to localise sources of transients. Pressure transients in WDNs are the consequence of travelling waves [6]. These waves are generated due to sudden changes in flow conditions, and can reach velocities up to 1200 m/s in rigid pipes [13]. The optimal sampling rate for measuring such fast propagating waves depends on the purpose of the analysis. If the intention is to fully characterise the shape of a wave and their exact arrival time in space, then a sampling rate of 128 S/s or more is required to achieve errors of less than 10 m [15]. Nevertheless, if the sole purpose of the analysis is to capture the maximum rising edge of wave in order to count cyclic loadings, then the sampling rate can be reduced without substantial loss of information. Optimising the sampling rate according to the required use is of critical importance because of the large computational resources needed to process and store billions of data points obtained at high resolution. We have assessed the loss of information between counting cyclic loading at data sampled at 128 S/S, 10 S/s and 1 S/s. The three levels separated by one order of magnitude were selected on the basis of practical energy consumption and storage capacity savings. In addition we also evaluated the utilisation of counting cycles after the application of a noise removal algorithm on raw data. (It should be noted that a recent change to the firmware of the utilised InflowSense pressure monitoring devices has included the calculation of the pressure cycles on the devices from the 128 S/s data; e.g. edge processing for the pressure cycles. These pressure cycles are then communicated without the need to communicate raw data).

2.3 Cycle counting

Measuring water pressure at high sampling rates unveil loading conditions in pipes otherwise undetectable at lower frequencies. Figure 2 contrasts the difference between water pressure measured at 1 sample every 15 min for common SCADA systems versus a high speed logger at a rate of 1 S/s. The water pressure profile presented from the high speed logger shows clear pressure variations at different magnitudes. From a mechanistic point of view, a single load of such magnitude does not pose any threat to a pipe. However, it has been experimentally demonstrated that several repetitions of cyclic loadings can propagate cracks up to a point at which the pipe can become vulnerable to normal operating loadings [16], [17]. In order to evaluate such crack propagation due to repeated cyclic loadings, fatigue crack propagation theories from fracture mechanics can be utilised [18].

CC BY NC SA

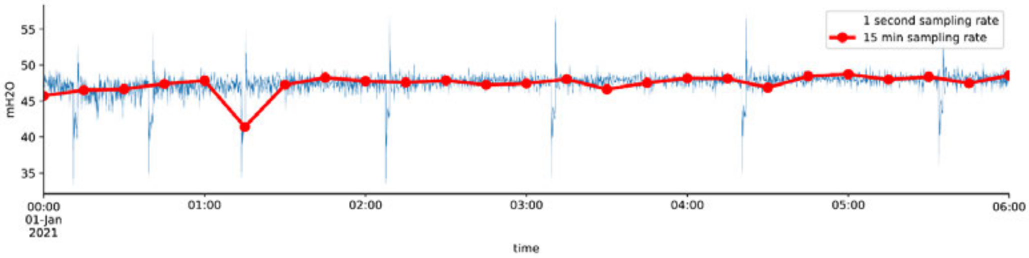


Figure 2. Comparison of pressure profiles obtained by a low and high frequency sampling rates

In order to apply methods from fracture mechanics to assess the propagation of fatigue cracks, it is first necessary to extract the information of the magnitude and number of cyclic loadings presented in the load history. This task is not trivial. In contrast to constant amplitude loadings, where the amplitudes of each cycle are well defined, there is no precise definition of what constitute a cycle under random load histories [19]. In such cases, cycle counting algorithms that perform heuristics to count cycles must be utilised [20]. Several techniques for such cycle counting exits, and the results can vary between one or other technique [20]. In the field of fatigue analysis there is a preference for the rainflow cycle counting technique [21]. The preference of this algorithm relies on its capability to count cycles within cycles, and the possibility to relate its results with stress and strain hysteresis curves [22]. It has been also empirically demonstrated that the use of this algorithm produces satisfactory loadings for prediction of actual fatigue life [23]. Given its preference and general performance acceptability, we have implemented this algorithm in the version of the ASTM E1049-85 [20] for cycle counting of our time series pressure data.

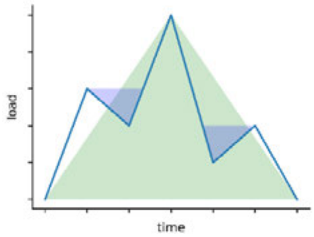


Figure 3. Visualisation of cycle detection by the rainflow cycle counting technique

The outcome of the rainflow cycle counting technique is a distribution or histogram that provides the amplitude of a cyclic load and the number of repetitions for each cycle. This information is, however, not provided in sequence, which means that information of history effects are missing, i.e., the occurrence of high load before a small one. Nevertheless, this information is usually required in detailed fatigue analysis performed "cycle by cycle" where crack closure effects can be considerable at low R ratios [24] (The R ratio is the ratio between minimum and maximum pressure). For the majority of cases where R ratios are high, a simple average of the cycle counts for a specific load history can provide good enough estimates for fatigue analysis [24]. In the results section we present some typical R ratios observed with the measured data.

2.4 Dynamic pressure profile assessment

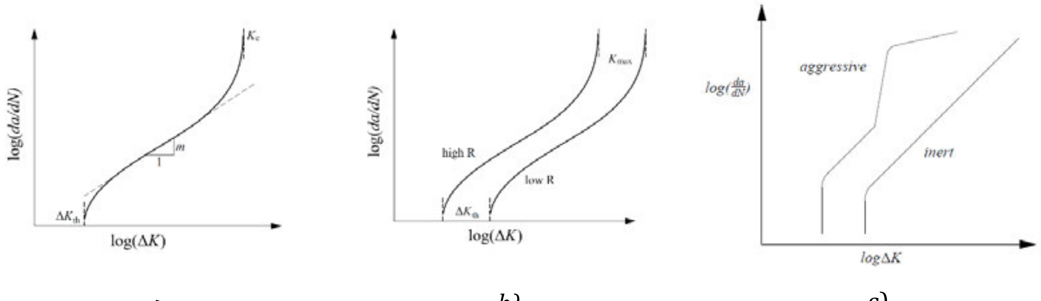
We have applied the rainflow cycle counting algorithm on several time series of water pressure measured over periods of one year. We then categorised different water pressure profiles within 3 categories of calm, regular and dynamic water pressure profiles according to the magnitude and number of cycle counts within a period of one year. This classification can serve in order to reach targets of pressure reductions according to the implications of the water pressure on specific materials.

2.5 Fatigue analysis

Once the assessment and characterisation of the water pressure fluctuations are performed, it is necessary to understand the implications of these cyclic loadings from a mechanistic point of view. As it is presented below, the effects that water pressure fluctuations can cause are highly material dependent, and therefore specific analysis must be performed for particular pipe materials. In this section, we present the basic theory of how to perform a general fatigue crack propagation analysis on pipes. To illustrate the proposed analysis, we present its application for a PVC pipe in order to assess the implications of the observed water pressure fluctuations for a particular material.

The assessment of fatigue damage can be performed by three main mechanistic procedures. These are the stress-life approach (S-N), the strain-life approach (ε-N) and the fatigue crack propagation approach [25]. The first two approaches rely on testing an intact component under different cyclic loadings and measuring its life until failure. The first procedure produce S-N curves. In applications where fairly constant amplitude loadings are expected and no cracks in the materials are present, or if the initial conditions of the element to be assessed coincide with the tested specimen, these procedures can be sufficient. However, if the material to be analysed is significantly different from the geometric configuration tested under laboratory conditions, then both methods can become unreliable. The S-N is empirical in nature and provides no insights into the mechanics of fatigue. Meanwhile, the ε-N approach provides insights into the mechanisms that lead to crack initiation [25]. In summary, both approaches provide little understanding into the mechanics of fatigue experienced on propagation of cracks on the tested element, especially in the S-N approach [25].

In contrast to the two previous methods, the fatigue crack propagation approach, which is also based on empirical results, provides direct insight into the mechanics of fatigue through estimating the rate of crack propagation using linear elastic fracture mechanics (LEFM) [18]. The method relies on the observance that the rate of propagation of cracks for a specific material is primarily a function of the amplitude of the stress intensity factor range produced by the applied cyclic loading, and the mean stress, expressed by the ratio between the minimum and maximum stress, as presented in Figure 4a, and in functional form in Equation 1. It is possible to observe from Figure 4 that for every material there is a particular minimum value of stress intensity range below which fatigue crack propagation does not occur.



a) b) c)

Figure 4. a) empirical crack growth rate curve. b) effect of mean pressure on the crack growth rate curve. C) effect of corrosive environment of fatigue crack propagation on metals (corrosion fatigue)

$$\frac{da}{dN} = f(\Delta K, R, \dots) \tag{1}$$

On the basis of Figure 4a and Figure 4b and the functional form of Equation 1, various "fatigue crack propagation laws" exist that attempt to model this curve. Common equations include the

Paris Law (Equation 2) [17], the Walker equation (Equation 3) [26] and the NASGRO equation (Equation 4) [27], presented below.

$$\frac{da}{dN} = C \cdot \Delta K^{m} \tag{2}$$

$$\frac{da}{dN} = C_0 \cdot \left[\frac{\Delta K}{(1-R)^{1-\gamma}}\right]^{m} \tag{3}$$

$$\frac{da}{dN} = \frac{C(1-f)^{n}\Delta K^{n}\left(1-\frac{\Delta K_{th}}{\Delta K}\right)^{p}}{(1-R)^{n}\left(1-\frac{\Delta K}{(1-R)K_c}\right)^{q}} \tag{4}$$

These equations differ in complexity and their use depends on the material to be analysed. Some materials are insensitive to the mean stress and therefore the R ratio, for which the use of the Paris Law is sufficient [25]. In case of materials that present more complex dependence to the R ratio and more advanced phenomena such as crack closure [28], then the Walker and NASGRO equation are applied. In order to estimate fatigue life, the number of cycles to reach a specific critical crack size can be estimated by integrating any of the above fatigue laws using Equation 5.

$$N = \int_{a_i}^{a_f} \frac{1}{f(\Delta K, R, \ldots)} da = \int_{a_i}^{a_f} \left(\frac{dN}{da}\right) da \tag{5}$$

In cases of pipes subjected to corrosive environment, corrosion can also produce an important cause for crack growth. Corrosion can increase a crack by its own and also can increase the rate of crack propagation generating corrosion fatigue, mathematically expressed by Equation 6. Corrosion fatigue has been shown to be of potential relevance in cast iron pipes [29].

$$\left(\frac{da}{dN}\right)_{aggressive} = \phi\left(\frac{da}{dN}\right)_{inert} + \frac{1}{f}\left(\frac{d\bar{a}}{dt}\right)_{EAC} \tag{6}$$

For illustration, we present an example of fatigue crack propagation analysis for a PVC-U pipe in order to show the effects of water pressure fluctuations on fatigue crack growth. The selected pipe configuration was a PVC-U pipe of nominal size 6”, (168mm mean outside diameter) according to the BS 3505:1986 [30]. In order to perform the analysis, we have fitted the Paris Law equation to data of fatigue crack propagation (Figure 5) on PVC-U, which was published in [31]. The obtained constants used for the fatigue analysis after fitting the data were C = 1.9318e-6 and m = 3.5578. We estimated the number of cyclic loading at constant amplitude required to increase various crack sizes up to a critical estimated crack of 92mm that would became unstable with the application of a load of 150 mH_2O in the pipe.

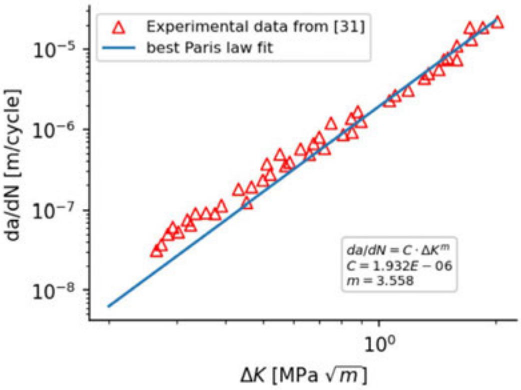


Figure 5) Fatigue crack propagation curve for PVC-U with data from [31]

3 RESULTS

This section presents key results based on the determination of cyclic loadings using the rainflow cycle counting algorithm on time series of high frequency water pressure data. In addition we include an example of the implications of reducing water pressure fluctuations on the fatigue life of a PVC pipe.

3.1 Sampling rate

This work required the utilisation of several months of water pressure monitoring in order to provide sensible results on the fatigue effect of cyclic loadings. Therefore, the trade-off of utilising a sampling rate detailed enough to accurately capture cycles generated by rapidly travelling waves, and a sampling rate that is low enough to allow for savings in computational resources and storage capacity was of critical importance. In order to assess an optimal sampling rate we have compared the difference in cycle counting between sampling data a 128 S/S, 10 S/s and 1 S/s. We have also incorporated the result of counting cycles from the raw data after application of a noise removal algorithm. We have prepared this analysis on data sampled in a single location during 2018 and 2021. The results of this analysis are presented in Table 1 and Table 2. To the best of the authors knowledge, this is the first published evaluation of the ideal sampling rates for purposes of cycle counting of water pressure fluctuations.

The pressure at the measured location had moderate pressure fluctuations. As Table 1 shows, most cycles are in a range below 2 m, with more than 86% of the presented data. This number can reach almost 99% in case of also counting cycles below 1 m. However, these cycles were not included as they were considered noise. It is observed that the application of a lower sampling rate resolution of 10 S/s can achieve better results closer to the counting of raw data. When applying the lowest sampling rate of 1 S/s, the errors increase considerably for lower cycle ranges. Nevertheless, the errors become acceptable when dealing with cycles over 5 m. Depending on the material, these could be a limitation of the 1 S/s sampling rate. Nevertheless for materials that possess higher stress intensity threshold, a sampling rate of 1 S/s seems to be adequate. The results over the two analysed years are consistent.

Table 1. Cycle counting comparison at different time resolutions, 1 year of data 2018

Stoke Lane	01/01/2018 count	19/12/2018					
range	128 S/s	1 m noise removal	relative error	10 S/s	relative error	1 S/s	relative error
(0.999, 2.0]	584769	440229	25%	485298	17%	226982	61%
(2.0, 3.0]	40527	31988	21%	32162	21%	19484	52%
(3.0, 4.0]	5179	4563	12%	4493	13%	3468	33%
(4.0, 5.0]	1748	1790	2%	1666	5%	1429	18%
(5.0, 10.0]	1537	1241	19%	1456	5%	1292	16%
(10.0, 20.0]	430	421	2%	413	4%	399	7%
(20.0, 40.0]	179	178	1%	150	16%	146	18%
(40.0, 60.0]	5	5	0%	2	60%	1	80%
cycles over 5 m	2151	1845	14%	2021	6%	1838	15%
cycles over 10 m	614	604	2%	565	8%	546	11%

Table 2. Cycle counting comparison at different time resolutions, 1 year of data 2021

Stoke Lane	01/01/2021 count	31/12/2021					
Range	128 S/s	1 m noise removal	relative error	10 S/s	relative error	1 S/s	relative error
(0.999, 2.0]	604109	324631	46%	473317	22%	281498	53%
(2.0, 3.0]	84528	48762	42%	65497	23%	38827	54%
(3.0, 4.0]	11406	7139	37%	8844	22%	5368	53%
(4.0, 5.0]	2270	1726	24%	1955	14%	1450	36%
(5.0, 10.0]	1580	1267	20%	1383	12%	1122	29%
(10.0, 20.0]	417	407	2%	392	6%	371	11%
(20.0, 40.0]	45	39	13%	37	18%	33	27%
(40.0, 60.0]	3	3	0%	3	0%	3	0%
Cycles > 5 m	2045	1716	16%	1815	11%	1529	25%
Cycles > 10 m	465	449	3%	432	7%	407	12%

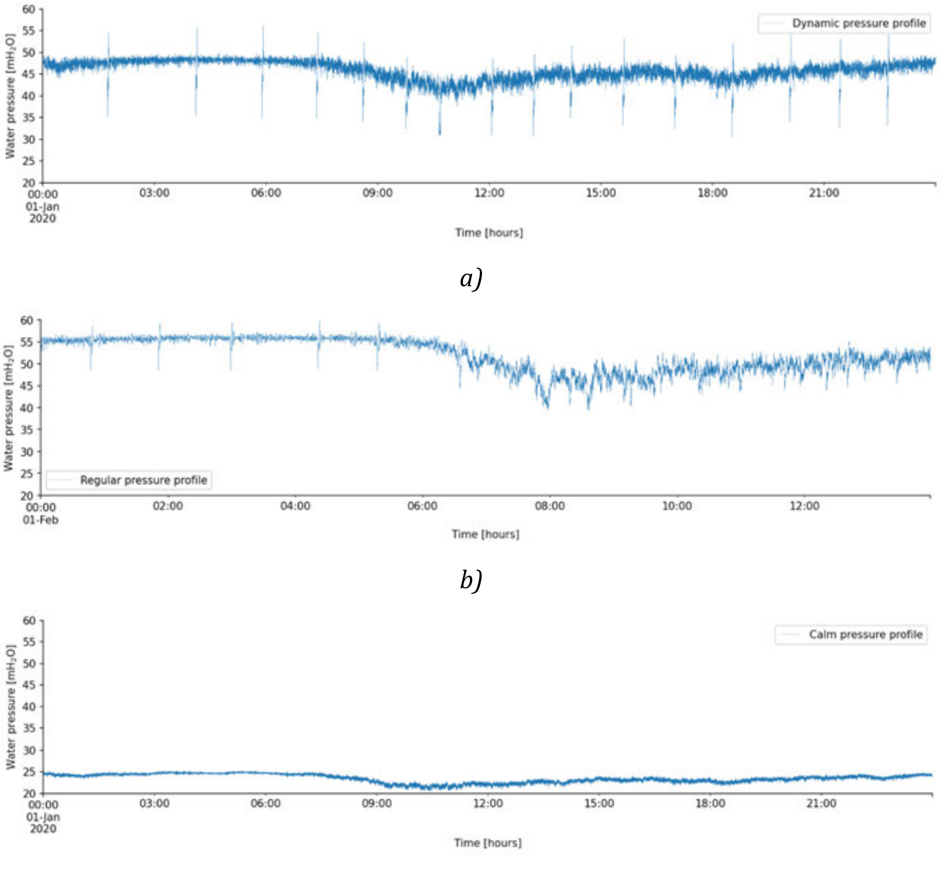
A point of consideration is that the cycle count presented in the table only shows the count of full cycles. However, there are also residual half cycles. Especially for very large, isolated cycles, it is possible that at lower resolutions, only half of a large cycle might be counted. Nevertheless, as it can be seen, these isolated events are unlikely to generate large changes in the overall estimations for fatigue in a fracture mechanics analysis. Unless these events are very large in order to reach the fracture toughness of the material, or if these events are capable of proving retardation or acceleration events, they should not pose a source of large errors.

On the basis of the presented calculations, it is clear that a minimum sampling rate of 10 S/s is closest to the original sampling rate. An important consideration is that in the analysed profiles, it can be seen that sampling at a rate of 1 S/s can be accurately enough for cycles above 5 m. For lower cycle magnitude, it seems that there could be a proportional loss of information that could be accounted when estimating cyclic loadings.

As it is presented below, depending on the material to be analysed, it might be required only to assess cycles above 5 m, for which a sampling rate of 1 S/s is sufficient in operational WDNs. Nevertheless, materials with very low stress intensity threshold such as asbestos cements might require a higher precision in counting cycles. Because of computational complexity and efficiency, the rest of the analyses are being performed on data sampled at 1 S/s. However, for further work, depending on the stress intensity threshold of the material analysed, it might be desirable to utilise the cycles calculated at the edge from the 128 S/s data, which has been recently implemented.

3.2 Dynamic vs calm profiles

Several time series of one year of water pressure monitoring were investigated. Given the size of the accumulated data over a period of 1 year, and computational processing time, the pressure profiles were analysed using a sampling rate of 1 S/s. From the analysed profiles, three categories based on the cycle counting results were generated. These categories are illustrated with the example pressure profiles of Figure 6 and the cycle counting summary of Figure 7.



a)

b)

c)

Figure 6) Measured water pressure profiles in operational water distribution networks, example of a regular day, a) Dynamic, b) regular, c) calm

BY NC SA

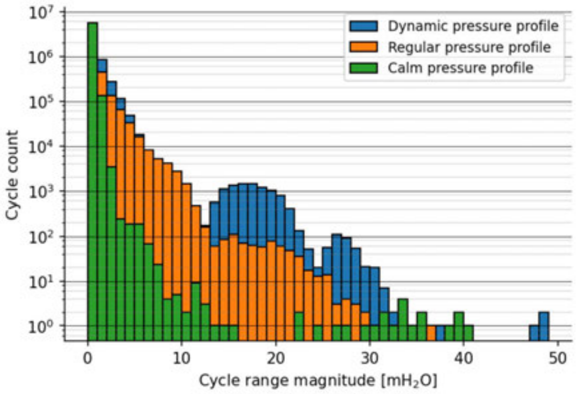


Figure 7) Cycle count histogram comparison of dynamic vs calm profiles over one year of pressure monitoring

Based on the application of the cycle counting algorithm on several datasets, we categorised the observed pressure profiles as shown in Figure 7. Our categorisation is based on the shape of the distribution of cycle count for the measure water pressure. This categorisation updates the one provided by Rezaei [32], which was based on the shape of the water pressure profiles and the visible apparent number of pressure fluctuations and magnitude. The utilisation of the shape of the cycle counting histogram provides better insight into the possible damage that a cyclic loading can generate, and it is easy to visually evaluate the number of cycles that occur over one-year-period.

The presented histogram is in a log scale, which shows an indication of the orders of magnitude of difference between the pressure profiles. We have observed these profiles as a summary of several location of the investigated WDN, nevertheless it is also possible to encounter larger profiles at transmission mains. For the largest measured pressure profile, more than 1700 cycles with a minimum magnitude of 20 m were counted. This number increase to more than 8000 cycles with a minimum magnitude of 10 m. Over a period of 50 years this can reach up to 400,000 cycles over 10 m. Depending on the material evaluated and the stress intensity factor generated by the combination of the stress and crack length, this level of cyclic loading can produce an accelerated deterioration of a pipe, as it is estimated in the next section.

Figure 8 shows a rainflow matrix, which presents the distribution of the cycles that occur from one initial magnitude until its peak. Cycles that spread over a diagonal from left to right show little damage, while cycles that spread over the perpendicular diagonal are of consideration. It is worth observing that although the loading conditions can be random in nature, most pressure profiles showed a symmetric shape as the one presented in Figure 8. The shape of the rainflow matrix can be an indication of the source of the pressure fluctuations. A symmetric shape for example, might be the result of a pump operation. An ideal shape of rainflow matrix showing little pressure fluctuations is indicated in Figure 8a.

The data presented in this section shows different levels of cyclic loading conditions that can be observed in water distribution networks. Our results show a more damaging water pressure cycles and in larger quantity that the one assumed in the study of Rajani and Kleiner [33] for example. These results reinforce the importance of continuously measuring pressure at high-sampling rates (above 1 S/s). However, the acquired high frequency data has little value if these data are not analysed by taking into consideration material properties and a fatigue crack propagation analysis. For example, what can be a highly damaging pressure fluctuation over time for a material can produce no damage for other material. The next section illustrates the effects of different water pressure fluctuations from a fracture mechanics perspective.

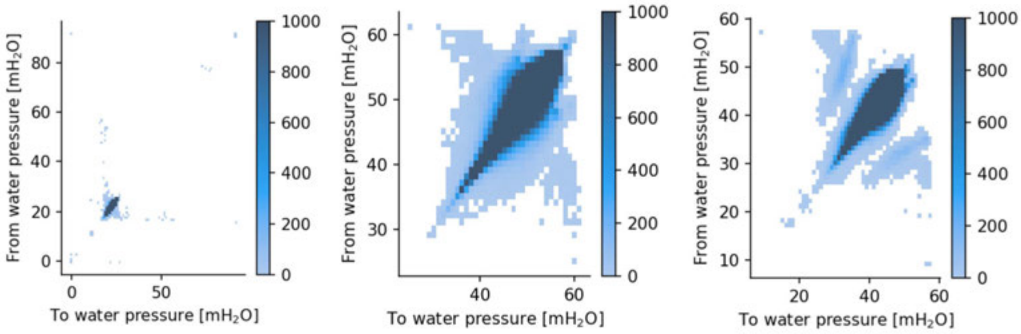


Figure 8) Rainflow matrix on water pressure profiles a) calm water pressure, b) regular water pressure c) dynamic water pressure

3.3 Fracture mechanics and fatigue implications from water pressure

As indicated, cycle counting data by itself provides little insight into the deterioration and fracture of a pipe if these data are not assessed in combination with knowledge from fracture mechanics and material properties of the pipe. As an starting point to perform a fatigue crack propagation analysis, Figure 9 shows the relation between the measured cycle ranges and their corresponding R values. As explained in the methods, larger R values can shift the crack growth curve of fatigue crack propagation of a material to the left, which indicates that lower values of stress intensity factor can cause fatigue. From the observed results it appears reasonable to utilise fracture mechanics curves obtained from testing at R ratios of 0.5 or more in case of lack of data for an specific material.

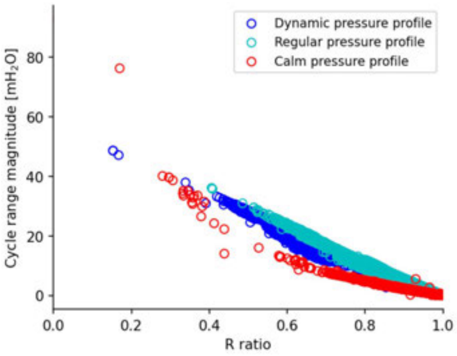


Figure 9) Cycle range and R ratio distribution

Material properties play an important role for fatigue crack propagation. The most important material properties include the stress intensity threshold and constants for the material fatigue crack propagation law. These values can largely differ for specific materials. For example, for cast iron it is common to observe fatigue thresholds in the order of 2-8 MPa√m [33], [34], while this number can be reduced almost an order of magnitude for PVC pipes up to 0.3 MPa√m and even further for asbestos cement pipes. The value of stress intensity threshold is the minimum value of stress intensity factor that a pipe must experience in order to be affected by fatigue crack propagation.

The stress intensity factor in a pipe is a function of the applied stress, pipe geometry and the size of the crack. For reference, Figure 10 illustrates how the stress intensity factor changes in different groups of cast iron pipes when these are being affected by different constant amplitude cyclic loadings at different crack sizes. As it can be observed, as a crack increases, the value of the stress intensity factor increases and more cycles start to affect the fatigue life of a pipe. In order to generate fatigue for a cast iron pipe with a stress intensity threshold of 2 MPa√m2, a minimum cycle range of 20 mH_2O for a crack of around 60 mm is required. This value substantially decreases for a PVC or asbestos cement pipes. Further work is currently conducted by the authors

CC BY NC SA

to gather evidence and analyse fatigue crack propagation for specific pipe materials, and the stochastic nature of pipe failures.

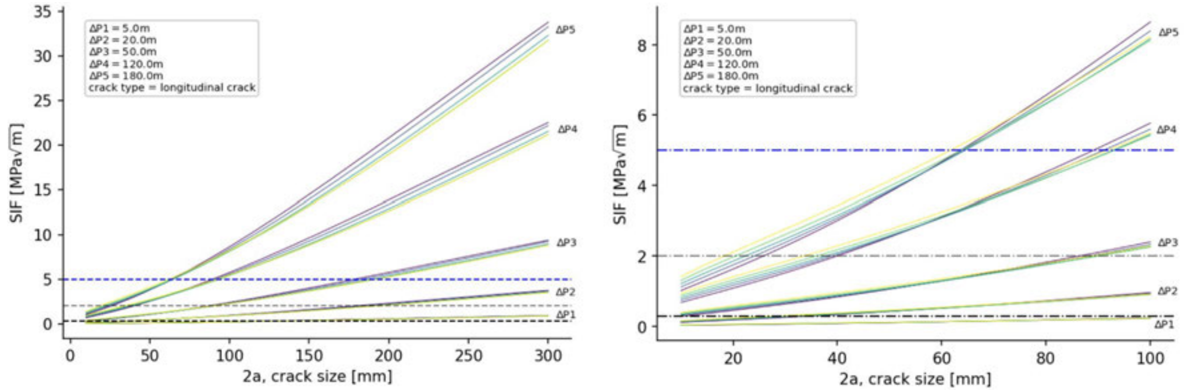


Figure 10) Stress intensity factor estimations from different water pressure magnitudes and crack length

Figure 11 shows an example of fatigue crack propagation analysis for a PVC pipe with OD 168 of mm. As it can be seen, the size of an existing crack plays an important role in the life of a pipe. However, the magnitude of the applied pressure fluctuations can significantly determine the life of a pipe. We can observe that the number of cycles required to fail the pipe can increase by an order of magnitude just by reducing the magnitude of pressure fluctuations from 30 mH_2O to 20 mH_2O, and even further for 10 mH_2O. In the previous section, we estimated a large presence of over 2000 cycles over 20 mH_2O. This number of cycles can be approximately translated to a remaining life of approximately 40 years for a pipe with a highly developed 70 mm crack without introducing safety factors. This life could be largely increased just by reducing the water pressure fluctuations.

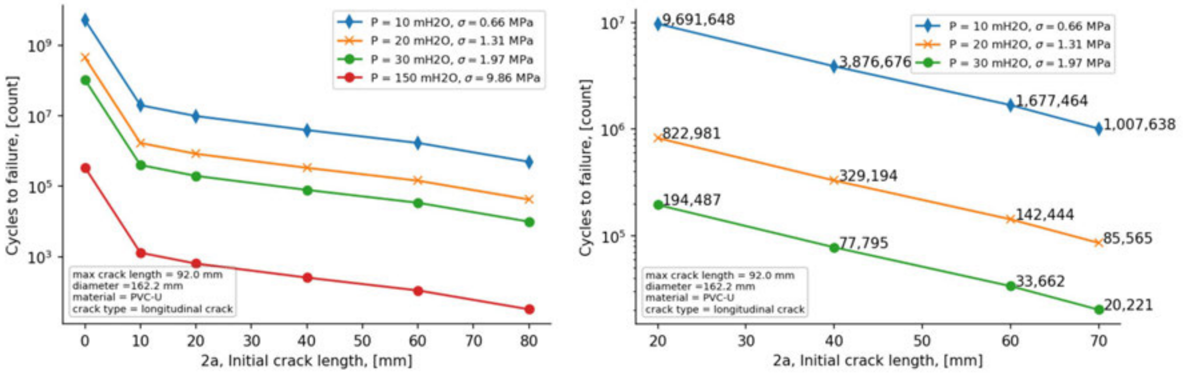


Figure 11) Fatigue life prediction for a PVC-U pipe under different constant cyclic loading conditions and initial crack sizes. a) full plot. b) Detailed view.

4 DISCUSSION

To our knowledge, the study reported here is the first study that analyses and characterises water pressure cycles over large periods using high frequency monitoring at fixed locations in a WDN. We have also quantitatively presented the implications of water pressure fluctuations in terms of fatigue crack propagation and the need to analyse their reduction with emphasis on specific materials. We evaluated the optimal sampling rate at which water pressure should be measured in order to accurately count cycles at an individual site, concluding that a sampling rate of 10 S/s and even 1 S/s can be sufficient for cycle counting. Three categories of water pressure time series were observed, our categorisation has been based on histograms for cycle counting generated by

CC BY NC SA

applying the rainflow cycle counting algorithm, updating the categorisation provided by Rezaei [32].

The proposed fracture mechanics assessment shows guidance on the magnitude of the stress intensity factor that can be generated on pipes for different crack sizes. The chart provided in Figure 10 serves as a basic tool for the understanding of the stress intensity factor applied to a pipe given its current crack length and an estimated level of water pressure cycle loading. This value of stress intensity factor can be compared to the minimum stress intensity threshold for a specific material in order to assess a preliminary estimation if a pipe can be affected by fatigue.

From the measurements and categorisations of pressure profiles obtained in the analysed WDN, it seems that pipes with high stress intensity threshold above 5 MPa√m might not suffer from pure fatigue failure. This claim must be further investigated and validated, since corrosion in cast iron can dramatically reduce the required minimum value of water pressure fluctuations to produce corrosion fatigue. Example of a reported case of a most likely failure due to corrosion fatigue was reported in [29]. In addition, previous studies that have correlated the effect of water pressure daily range as a proxy for water pressure fluctuations have shown positive correlations for reducing pressure range and reduction of pipe breaks, including pipes of high stress intensity threshold such as cast iron pipes [35], [36].

On the other hand, for pipes with lower stress intensity factor, the measured water pressure fluctuations can be significant, depending on the level of crack propagation that the pipes could have in any section of their length. Figure 11 provides an example calculation of fatigue crack propagation life for a 168 mm PVC-U pipe. This figure clearly remarks the importance of reducing water pressure fluctuations for an extended fatigue life. A highly deteriorated pipe could stop its fatigue crack growth by reducing the level of pressure fluctuations in 10 mH_2O. Thus, the importance of the application of pressure control techniques and also methodologies for pressure surge reduction as the ones indicated in [37] .

The literature has offered some estimations for fatigue crack propagation assessment of pipes such as in [33]. However, only rough estimations for cyclic loading were utilised. This study shows that water pressure fluctuations up to 20 mH_2O, can be generated constantly over a period of one year, reaching almost 2000, or in the order of 5 to 6 daily fluctuations. This number can be used as a reference for the regular pressure fluctuations observed in a network. Nevertheless, this study was limited to the characteristics of the analysed sector of the WDN, and it might be highly possible to encounter higher pressure profiles. An investigation in due course is being proposed to investigate the magnitude and propagation of water pressure fluctuations in close proximity to transmission mains connected to pump stations.

The methodology for the assessment of fatigue in this work relies on the presence of cracks in pipes, this is a limitation in the level of certainty of the results. A number of reasons can produce cracks in a underground pipe depending on the material. For cast iron pipes, corrosion is a concern for crack initiation and propagation which can be exacerbated by corrosion fatigue. For the case of plastic pipes such as PVC and PE pipes, poor installation conditions, and the presence of rocks and roots can produce initiation of cracks. More statistical understanding of the presence of cracks in pipes would allow the use of fracture mechanics together with risk analysis methodologies to provide results with confidence intervals in order to utilise this information in pipe assessment and rehabilitation programmes. This study has provided an initial overview of the effect of cycling loading on pipes from a mechanistic view. However, this study must be followed by detailed studies on specific materials in order to provide case specific guidelines for pipe failure prevention.

5 CONCLUSIONS

This paper has provided an analysis of time series of water pressure obtained from high speed pressure transducers over periods of one year at different locations in a water distribution network. We performed analyses in order to understand the potential effects of water pressure fluctuations on fatigue failure on pipes and how to utilise data from high speed pressure transducers for such purposes. Our analysis show that a sampling rate of 10 S/s is an optimal rate for the sole purpose of counting cycles of water pressure. Nevertheless, 1 S/s can be enough for applications in which cycles over 5 m are of interest. We have observed different categories of water pressure profiles an proposed a categorisation based on the histograms for cycle counting generated in water pressure profiles over one year of monitoring. The results of the number of cycles observed at the worst locations in this study when compared to fracture mechanics fatigue crack propagation analysis for an example PVC pipe have shown that fatigue crack propagation is a potential concern in pipes with low stress intensity threshold, and we have presented results that indicate that reduction of 10 m of water pressure can potentially increase in one order of magnitude the fatigue life of a pipe. Therefore confirming the importance of fatigue failure and the analysis of high speed data for materials with low stress intensity threshold.

6 ACKNOWLEDGMENTS

This work was funded by the National Agency for Research and Development (ANID)/ Scholarship Program/DOCTORADO BECAS CHILE/2020–72210314; and EPSRC EP/P004229/1 (Dynamically Adaptive and Resilient Water Supply Networks for a Sustainable Future).

7 APPENDIX

7.1 Stress intensity factors in pipes

The stress intensity factor is a concept of LEFM, which basically provides an indication of the stress intensification that is produced in a material near the tip of a crack. This value can be expressed as a function of the applied load in the material, the geometry of the material, the type of crack, and the crack length as follows:

$$K = Y \cdot \sigma \cdot \sqrt{\pi \cdot a}$$

Where Y is a geometric function that accounts for the geometry of the material and the crack configuration, σ is the applied stress on the material and a is half the crack length ($2a$).

There exists several expressions for estimating the stress intensity factor according to the geometry of the body surrounding a crack and the crack configuration. In this research we have calculated the stress intensity factor utilising the empirical solution for a longitudinal through crack in a cylinder presented in [38].

8 REFERENCES

[1] S. Folkman, “Water Main Break Rates In the USA and Canada: A Comprehensive Study,” *Mechanical and Aerospace Engineering Faculty Publications*, vol. Paper 174, 2018, Accessed: May 15, 2022. [Online]. Available: https://digitalcommons.usu.edu/mae_facpub/174

[2] D. Wilson, Y. Filion, and I. Moore, “State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains,” *Urban Water Journal*, vol. 14, no. 2. Taylor and Francis Ltd., pp. 173–184, Feb. 07, 2017. doi: 10.1080/1573062X.2015.1080848.

[3] C. Konstantinou and I. Stoianov, “A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks,” *Urban Water Journal*, vol. 17, no. 6, pp. 534–548, Jul. 2020, doi: 10.1080/1573062X.2020.1800758.

[4] H. Rezaei, B. Ryan, and I. Stoianov, "Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks," in *Procedia Engineering*, 2015, vol. 119, no. 1, pp. 253–262. doi: 10.1016/j.proeng.2015.08.883.

[5] I. Stoianov, C. Maksimovic, and N. Graham, "Field Validation of the Application of Hydraulic Transients for Leak Detection in Transmission Pipelines," in *Advances in Water Supply Management*, 2003, pp. 86–97.

[6] D. Misiunas, J. Vítkovský, G. Olsson, A. Simpson, M. Asce, and M. Lambert, "Pipeline Break Detection Using Pressure Transient Monitoring," *Journal of Water Resources Planning and Management* , vol. 131, no. 4, pp. 316–325, 2005, doi: 10.1061/ASCE0733-94962005131:4316.

[7] R. W. Gullick, M. W. Lechevallier, R. C. Svindland, and M. J. Friedman, "Occurrence of transient low and negative pressures in distribution systems," *Journal - American Water Works Association*, vol. 96, no. 11. American Water Works Association, pp. 52–66, 2004. doi: 10.1002/j.1551-8833.2004.tb10741.x.

[8] K. K. Fleming, R. W. Gullick, J. P. Dugandzic, and M. W. Lechevallier, "Susceptibility of Potable Water Distribution Systems to Negative Pressure Transients," 2005.

[9] G. Ebacher *et al.*, "Transient Modeling of a Full-Scale Distribution System: Comparison with Field Data," *Journal of Water Resources Planning and Management*, vol. 137, no. 2, 2011, doi: 10.1061/(ASCE)WR.1943-5452.

[10] I. Stoianov, L. Nachman, A. Whittle, S. Madden, and R. Kling, "Sensor networks for monitoring water supply and sewer systems: Lessons from Boston," in *Water Distribution Systems Analysis Symposium*, 2006, pp. 1–17. doi: 10.1061/40941(247)100.

[11] A. J. Whittle, M. Allen, A. Preis, and M. Iqbal, "Sensor Networks for Monitoring and Control of Water Distribution Systems," 2013.

[12] A. Hoskins and I. Stoianov, "InfraSense: A distributed system for the continuous analysis of hydraulic transients," in *Procedia Engineering*, 2014, vol. 70, pp. 823–832. doi: 10.1016/j.proeng.2014.02.090.

[13] S. Rathnayaka, B. Shannon, P. Rajeev, and J. Kodikara, "Monitoring of pressure transients in water supply networks," *Water Resources Management*, vol. 30, no. 2, pp. 471–485, Nov. 2015, doi: 10.1007/s11269-015-1172-y.

[14] D. Starczewska, R. Collins, and J. Boxall, "Occurrence of transients in water distribution networks," in *Procedia Engineering*, 2015, vol. 119, pp. 1473–1482. doi: 10.1016/j.proeng.2016.01.001.

[15] A. John Hoskins, "Monitoring the Hydraulic Dynamics in Water Distribution Systems," Imperial College London, 2015.

[16] P. Paris, "The growth of cracks due to variations in load," Lehigh University, 1962.

[17] P. Paris and F. Erdogan, "A Critical Analysis of Crack Propagation Laws," *Journal of Basic Engineering*, vol. 85, no. 4, pp. 528–533, 1963, [Online]. Available: http://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/85/4/528/5763569/528_1.pdf

[18] P. C. Paris, M. P. Gomez, and W. E. Anderson, "A rational analytic theory of fatigue," *The Trend in Engineering*, vol. 13, pp. 9–14, 1961.

[19] Y.-W. Cheng and J. J. Broz, "Cycle-counting methods for fatigue analysis with random load histories : A Fortran User's Guide," Gaithersburg, MD, 1986. doi: 10.6028/NBS.IR.86-3055.

[20] ASTM Committee E-8 on Fatigue and Fracture, "Standard Practices for Cycle Counting in Fatigue Analysis," *ASTM International*. doi: 10.1520/E1049-85R17.

[21] I. Rychlik, "Note on cycle counts in irregular loads," *Fatigue Fruct. Engng Muter. Struct*, vol. 16, no. 4, pp. 377–390, 1993, doi: 10.1111/j.1460-2695.1993.tb00094.x.

[22] H. Anzai and T. Endo, "On-site indication of fatigue damage under complex loading," *International Journal of Fatigue*, vol. 1, no. 1, pp. 49–57, 1979, doi: 10.1016/0142-1123(79)90044-6.

[23] N. E. Dowling, "Fatigue Failure Predictions for Complicated Stress-Strain Histories," *Journal of Materials, JMLSA*, vol. 7, no. 1, pp. 71–87, 1972.

[24] T. L. Anderson, *Fracture Mechanics : Fundamentals and Applications*, Fourth edition. Boca Raton: CRC Press, Taylor & Francis Group, 2017.

[25] J. A. Bannantine, J. J. Comer, and J. L. Handrock, *Fundamentals of Metal Fatigue Analysis*. Upper Saddle River, NJ: Prentice Hall, 1990.

[26] K. Walker, “The Effect of Stress Ratio During Crack Propagation and Fatigue for 2024-T3 and 7075-T6 Aluminum,” in *Effects of Environment and Complex Load History on Fatigue Life*, 1968, pp. 1–14.

[27] R. G. Forman and S. R. Mettu, “NASA Technical Memorandum 102165 Behavior of Surface and Corner Cracks Subjected to Tensile and Bending Loads in Ti-6A1-4V Alloy,” 1990.

[28] W. Elber, “Fatigue crack closure under cyclic tension,” *Engineering Fracture Mechanics*, vol. 2, no. 1, pp. 37–44, 1970, doi: 10.1016/0013-7944(70)90028-7.

[29] M. J. Cullin, T. H. Petersen, and A. Paris, “Corrosion Fatigue Failure of a Gray Cast Iron Water Main,” *Journal of Pipeline Systems Engineering and Practice*, vol. 6, no. 2, p. 05014003, May 2015, doi: 10.1061/(asce)ps.1949-1204.0000188.

[30] British Standards Institution, “Unplasticized polyvinyl chloride (PVC-U) pressure pipes for cold potable water,” *BS 3505:1986*. British Standards Institution, London, 1986. Accessed: May 13, 2022. [Online]. Available: https://bsol.bsigroup.com/PdfViewer/Viewer?pid=000000000000163172

[31] N. Samat, R. Burford, A. Whittle, and M. Hoffman, “The effects of frequency on fatigue threshold and crack propagation rate in modified and unmodified polyvinyl chloride,” *Polymer Engineering and Science*, vol. 49, no. 7, pp. 1299–1310, Jul. 2009, doi: 10.1002/pen.21369.

[32] H. Rezaei, “Impact of pressure fluctuations on pipe failures in water distribution networks,” Imperial College London, 2017.

[33] B. Rajani and Y. Kleiner, “Fatigue failure of large-diameter cast iron mains,” *Water Distribution Systems Analysis 2010*, pp. 1146–1159, 2010.

[34] H. Mohebbi, D. A. Jesson, M. J. Mulheron, and P. A. Smith, “The fracture and fatigue properties of cast irons used for trunk mains in the water industry,” *Materials Science and Engineering A*, vol. 527, no. 21–22, pp. 5915–5923, Aug. 2010, doi: 10.1016/j.msea.2010.05.071.

[35] C. Jara-Arriagada and I. Stoianov, “Pipe breaks and estimating the impact of pressure control in water supply networks,” *Reliability Engineering and System Safety*, vol. 210, Jun. 2021, doi: 10.1016/j.ress.2021.107525.

[36] Martínez-Codina, M. Castillo, D. González-Zeas, and L. Garrote, “Pressure as a predictor of occurrence of pipe breaks in water distribution networks,” *Urban Water Journal*, vol. 13, no. 7, pp. 676–686, Oct. 2016, doi: 10.1080/1573062X.2015.1024687.

[37] P. F. Boulos, B. W. Karney, D. J. Wood, and S. Lingireddy, “Hydraulic transient guidelines for protecting water distribution systems,” *Journal / American Water Works Association*, vol. 97, no. 5. American Water Works Association, pp. 111–124, 2005. doi: 10.1002/j.1551-8833.2005.tb10892.x.

[38] J. C. Newman, “Fracture Analysis of Surface and Through Cracks in Cylindrical Pressure Vessels,” NASA Technical Note, 1976.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14815

A NONLINEAR MODEL PREDICTIVE CONTROL FRAMEWORK FOR DYNAMIC WATER NETWORK OPTIMIZATION

Ernesto Arandia[1] and James Uber[2]

[1]Xylem Digital Solutions, Viareggio, LU, Italy
[2]Xylem Digital Solutions, Cincinnati, OH, USA

[1]ernesto.arandia@xylem.com, [2]james.uber@xylem.com

Abstract

This paper proposes the application of nonlinear model predictive control (NMPC) strategies within an integrating framework for planning and scheduling real-time water distribution system operations. Model predictive control (MPC) refers to a class of algorithms that make explicit use of a process model to optimize the future predicted behavior of a system. Originally developed to address the control needs of power plants and petroleum refineries, over the past 30 years it has successfully been used in a very wide range of applications in industry. In addition to the predictive model, MPC consists of a performance metric reflecting the control actions, an optimization algorithm that computes a sequence of future control signals that minimizes the performance index subject to a given set of constraints, and a moving horizon strategy, according to which only the first time point of the optimal control sequence is applied online. The predictive models are generally intended to represent the behavior of complex nonlinear dynamical systems and often consist of linear models that circumvent the stability and performance challenges associated with their nonlinear counterparts. Recent developments on methods to model and solve large-scale nonlinear optimization problems lead to a reconsideration of these formulations, in particular, through development of efficient large-scale barrier methods for nonlinear programming (NLP). As a result, it is now realistic to solve NLPs on the order of a million variables. These developments are leveraged here to develop an NMPC dynamic real-time optimization strategy that combines a nonlinear dynamical model of the water network with a large-scale NLP optimization algorithm, and a moving horizon strategy to determine how to exert control on the system at the pump facility and network levels. We consider a recent water utility case study to illustrate these concepts.

Keywords
Dynamic water network optimization, nonlinear model predictive control, operation optimization.

1 INTRODUCTION

Dynamic optimization problems involving real-world water networks present several challenges, including the large scale of the networks, the nonlinearity of the relationships among physical quantities, the discrete nature of operation decisions such as pump selection, and the fact that state-of-the-art optimization solvers do not directly handle integrals or differential equations.

The approach proposed in this paper consists of a combined nonlinear model predictive control (NMPC) strategy designed to address these challenges. The scale problem is addressed through a combination of net- work skeletonization and the use of a large-scale NLP solver that can handle millions of variables efficiently. The nonlinearities are explicitly expressed through hydraulic relationships for all the types of network components involved. The discrete decisions are handled by decomposing the optimization problem into a primary or master problem that solves the network first, and a secondary problem that solves the pump facilities using results from the primary problem. The optimization dynamics is captured through a differential-algebraic formulation in the model constraints.

Previous studies dealing with water network operation optimization consist mostly of academic analyses and very few industrial-scale undertakings. Their methodologies are widely based on the application of heuristic methods that do not address the challenges outlined above. One exception is [2], that describes the development and implementation of a large-scale NLP based pump scheduling system for a city-wide water distribution system. Other examples of NMPC studies designed for large-scale systems are found in the chemical engineering domain, e.g., [1] present a large-scale optimization framework for NLP applied to control chemical processes in real time.

The goal of this paper is to describe the proposed NMPC methodology and the proof-of-concept (PoC) that demonstrates the feasibility and optimization capabilities of the proposed solution. Since increasing energy cost efficiency is often one of the key optimization goals in operations and is also straightforward to use when contrasting optimized scenarios with the status quo, the PoC and this paper focus on energy and pressure optimization first. A near future development will explicitly incorporate water quality in addition to energy and pressure optimization.

The paper is organized as follows. Section 2 describes the NMPC technology from the perspectives of the supply network and the pump facility. Section 3 discusses the proof-of-concept (PoC) established to demonstrate the approach. Section 4 contrasts the actual and optimized operations with focus on energy optimization. Finally, Section 5 presents a summary and discusses future work.

2 METHODOLOGY

Under the proposed approach, the water supply network is a dynamic system that is controlled through optimization over a number of discrete time periods. The water network optimization problem consists thus of (1) a number of dynamic stages, (2) state, control, and algebraic variables, and (3) optimization constraints in the form of differential and algebraic equations (DAEs). The solution of this dynamic optimization problem is formally the time evolution of the network state over a moving (rolling) horizon. This solution is imposed on the facility optimization problem which deals with the selection of pumps to run and their relative speeds, in case they have variable frequency drives (VFDs), over the specified optimization horizon. The approach is general in that any number of pumps are considered and that any of them may have VFDs. The only restriction is, however, that the pumps are arranged in parallel.

Below, we briefly discuss the pump facility abstraction that is introduced to allow formulating the network optimization problem as an NLP. Then we formulate and describe the primary problem in more detail. A discussion of the pump facility problem is beyond the scope of this paper.

2.1 Pump facility abstraction

As illustrated in Fig. 1, under the introduced abstraction the pump facility consists of an upstream node (reservoir), one or more pumps arranged in parallel, one or more throttle control valves, and one or more downstream nodes. The control valves may already exist as shown in the diagram of Fig. 1 that represents a real pump facility in the system described in Section 3. Otherwise, the control valves are added to the model while ensuring the network configuration downstream of the facility is not altered. The purpose of these valves is to remove the concerns about the pump switching behavior by keeping the pumps running continuously while adjusting the control valve settings during optimization.

For consistency, the effective head gains through the pump facility are measured between the valves' downstream nodes (C and D in Fig. 1) and the upstream reservoir. The facility discharge flows are computed as well for each valve using the valves' head loss relationships (see Eq. 5 and 6). Moreover, the electrical power consumption of the facility is calculated per valve using the head gains and flows under the assumption that there is a constant global efficiency for the pump facility. This global efficiency is estimated from the calibrated head and efficiency curves for the pumps. The power consumption for the whole facility is computed by summing through the control valve powers (see Eq. 11).

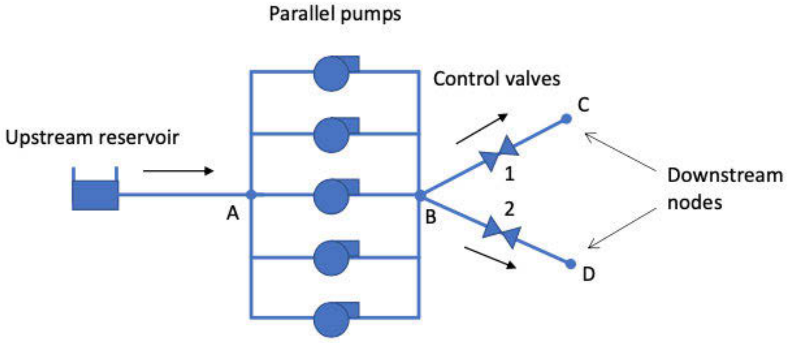


Figure 1. Diagram of the pump facility abstraction.

2.2 Network optimization problem

The main characteristics of the NMPC water network optimization model are listed below:

- The network state variables consist of hydraulic heads at non-fixed head nodes. More specifically, the state variables are partitioned into algebraic variables, i.e., hydraulic heads at junctions, and differential variables, i.e., the hydraulic heads at tanks.
- The dynamics of tank water volume is obtained through differential equations that impose flow balance constraints at the tank nodes.
- Control at the pump facilities and throughout the network is exerted through control valves where the decision variables consist of the valves' opening fraction. The pump facility control valves allow for removing the binary decisions associated with pump selection, at the network solution stage.
- Check valves are handled through a simple non-negativity constraint.
- Head loss and gain expressions at network edges consist of three types of continuous functions, namely, the Hazen-Williams equations for pipes, a power-law formulation for pumps, and the head loss equation for throttle control valves (TCVs).
- Mass balance is imposed at all junctions following the so-called node method formulation in the form of algebraic constraints.
- Appropriate upper and lower bounds are imposed to the control, state and other (auxiliary) variables.
- The objective function consists of the combined energy use and energy demand costs over the rolling operation horizon.

The following sets, parameters, and decision variables are introduced to formulate the water network optimization problem.

Sets:

T : the set of time periods

E_P : the set of pipes

E_U : the set of pumps

E_V : the set of valves

E^*_V : the set of pump facility valves

V_J : the set of junctions

V_R : the set of reservoirs

V_T : the set of tanks

F : the set of pump facilities subject to demand charges

E : the set of network links

T_P : the set of peak-hour time periods

T_O : the set of off-peak-hour time periods

Parameters:

: specific weight of water

Δ_T : length of each time period

$c_{j,t}$: electricity cost for facility valve j in time period t

η_j : efficiency associated with facility valve j, assumed constant

$D_{i,t}$: demand of junction i in period t

H_{Ri} : hydraulic head of reservoir I, assumed constant

I_{ij} : incidence of node i and link j

R_{pj} : resistance coefficient of pipe j

R_{vj} : resistance coefficient of valve j

a_{lj} : $l \in \{1, 2, 3\}$, head curve coefficient l of pump j

α : Hazen-Williams exponent

ε : small value set to 10−8

e_i : elevation of node i

$L_{i,min}$: minimum water level in tank i

$L_{i,max}$: maximum water level in tank i

$p_{i,min}$: minimum pressure head of junction i

$p_{i,max}$: maximum pressure head of junction i

A_i : section area of tank I, assumed cylindrical

s_k : demand charge on peak hours for pump facility k

w_k : demand charge on off-peak hours for pump facility k

$\bar{P}_k$: maximum to-date power peak during peak hours for pump facility k

$\bar{W}_k$: maximum to-date power peak during off-peak hours for pump facility k

Decision variables:

$H_{Ji,t}$: hydraulic head of junction i in time period t,

$H_{Ti,t}$: hydraulic head of tank i in time period t,

$h_{j,t}$: hydraulic head loss through link j in time period t,

$x_{j,t}$: fractional opening of control valve j in time period t,

S_k : highest 30-min interval power to date at pump facility k during peak hours,

W_k : highest 30-min interval power to date at pump facility k during off-peak hours.

The primary optimization problem is formulated below:

$$\min \quad \sum_{t\in T}\sum_{j\in E_V^*}\frac{\gamma\Delta_T c_{j,t}}{\eta_j}q_{j,t}h_{j,t}+\sum_{k\in F}\left(s_kS_k+w_kW_k\right), \tag{1}$$

$$\text{s.t.} \quad \sum_j I_{ij}q_{j,t}-D_{i,t}=0 \quad \forall i\in V_J,\forall t\in T, \tag{2}$$

$$q_{j,t}=-R_{p_j}^{-\frac{1}{\alpha+1}}h_{j,t}(h_{j,t}^2+\epsilon)^{-\frac{\alpha}{2(\alpha+1)}} \quad \forall j\in E_P,\forall t\in T, \tag{3}$$

$$q_{j,t}=a_{2j}^{-\frac{1}{a_{3j}}}\left(a_{1j}-h_{j,t}\right)^{\frac{1}{a_{3j}}} \quad \forall j\in E_U,\forall t\in T, \tag{4}$$

$$q_{j,t}=-R_{v_j}^{-\frac{1}{2}}x_{j,t}h_{j,t}(h_{j,t}^2+\epsilon)^{-\frac{1}{4}} \quad \forall j\in E_V,\forall t\in T, \tag{5}$$

$$h_{j,t}=\sum_{i\in V_J}I_{ij}H_{J_{i,t}}+\sum_{i\in V_T}I_{ij}H_{T_{i,t}}+\sum_{i\in V_R}I_{ij}H_{R_i} \quad \forall j\in E,\forall t\in T, \tag{6}$$

$$h_{j,t}\geq 0 \quad \forall j\in E_U,\forall t\in T, \tag{7}$$

$$H'_{T_{i,t}}-\frac{1}{A_i}\sum_j I_{ij}q_{j,t}=0 \quad \forall i\in V_T,\forall t\in T, \tag{8}$$

$$e_i+L_{i,\min}\leq H_{T_{i,t}}\leq e_i+L_{i,\max} \quad \forall i\in V_T,\forall t\in T, \tag{9}$$

$$e_i+p_{i,\min}\leq H_{J_{i,t}}\leq e_i+p_{i,\max} \quad \forall i\in V_J,\forall t\in T, \tag{10}$$

$$P_{k,t}=\sum_{j\in E_{V_k}^*}\frac{\gamma\Delta_T}{\eta_k}q_{j,t}h_{j,t} \quad \forall k\in F,\forall t\in T, \tag{11}$$

$$S_k\geq P_{k,t} \quad \forall k\in F,\forall t\in T_P, \tag{12}$$

$$S_k\geq \overline{P}_k \quad \forall k\in F, \tag{13}$$

$$W_k\geq P_{k,t} \quad \forall k\in F,\forall t\in T_O, \tag{14}$$

$$W_k\geq \overline{W}_k \quad \forall k\in F. \tag{15}$$

The objective function (1) minimizes the energy cost including energy use and demand charges. Since pump facilities may have multiple downstream control valves, the energy use cost consists of the sum of electricity costs through all facility control valves and all time periods, where the same cost $c_{j,t}$ and efficiency η_j are assigned to the valves that belong to the same facility. The energy demand costs consist of electricity charges associated with power peaks across the facilities that are subject to electricity demand charges.

Constraints (2) model the flow balance at the network junctions, where the flows $q_{j,t}$ are determined through constraints (3) for pipes, (4) for pumps, and (5) for valves. Note that, flow reversals through pipes and valves are handled by using a hyperbolic approximation of the absolute value of the link flow [3]. The head loss $h_{j,t}$ through link j in time period t is set by constraints (6) using the incidence relationship I_{ij} and the hydraulic heads of the incident nodes i. Constraints (7) prevent flow reversal through pumps. We introduce the differential constraints (8) to account for the tank flow dynamics; note that the formulation is for cylindrical tanks. Constraints (9) and (10) are used to maintain the tank levels and junction pressure heads within admissible ranges. Constraints (11) are used to determine the power use of facility k in time period t for the facilities subject to demand charges. Finally, constraints (12) to (15) are auxiliary constraints used to determine the power peaks throughout the optimization horizon.

2.3 Architecture of the Water Network Optimizer

This section describes the high-level water network optimization software architecture illustrated in Fig. 2. The water network optimizer (WNO) runs on a rolling-horizon basis following a configurable time interval (e.g., hourly). At the start of the process, real-time digital twin boundary conditions, such as tank levels, are obtained and the demand forecaster is activated to provide demand forecasts over the incoming planning horizon. As a next step, the optimization model is created (parameters, variables, constraints, discretization, optimization objective, etc.) using the boundary and demand data. Finally, the model is solved using IPOPT [4] to generate an optimized operational plan that includes control valve settings and provisional pump facility head-flow operation points. These results are used by the pump facility optimizer to generate pump and valve schedules. The subsequent steps repeat the following sequence: (1) update water demand data (from the digital twin services), (2) update the initial boundary conditions for the incoming operational plan, (3) forecast demands, (4) update the optimization model, and (5) solve the optimization formulation to deliver a new operation plan.

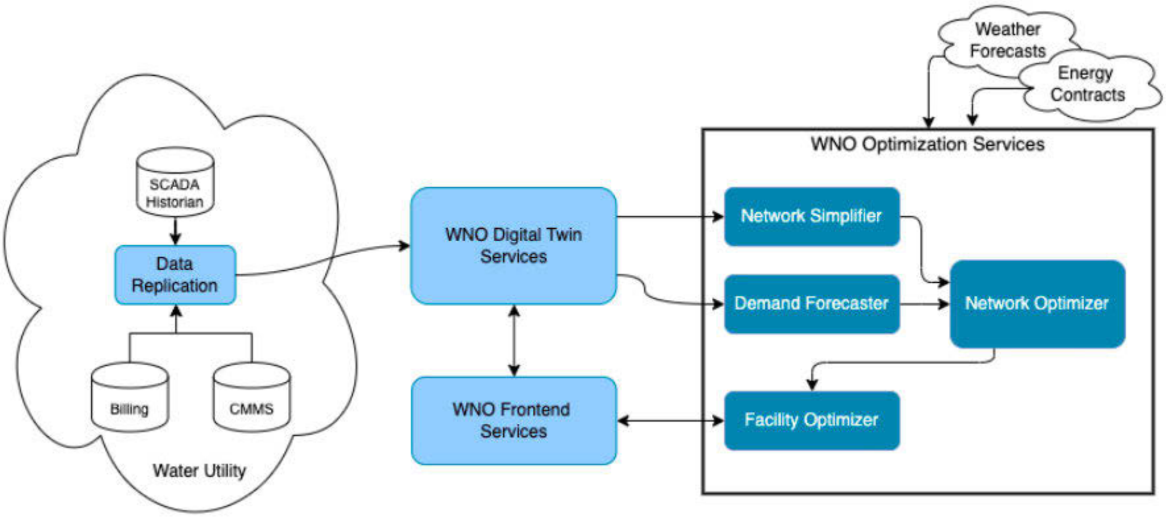


Figure 2. Diagram of the water network optimizer architecture.

3 PROOF OF CONCEPT

The Hillsborough Water Resources Department (HWRD) South-Central system is used as the testing and validation supply system. A modified and anonymized representation of the HWRD network model is illustrated in Fig. 3. The network includes three pump facilities named A, B, and C where all pumps are equipped with VFDs. Facility A accounts for most (64%) of the energy

BY NC SA

consumed for pumping operations. Both A and B operate under a time-of-use and demand-charge energy schedule while facility C operates under a fixed-rate schedule. In addition, the HWRD network is instrumented with several pressure sensors denoted in Fig. 3 by the red squares labelled from a through n. Their data was analyzed statistically to determine appropriate minimum and maximum pressure constraints that were used during optimization.

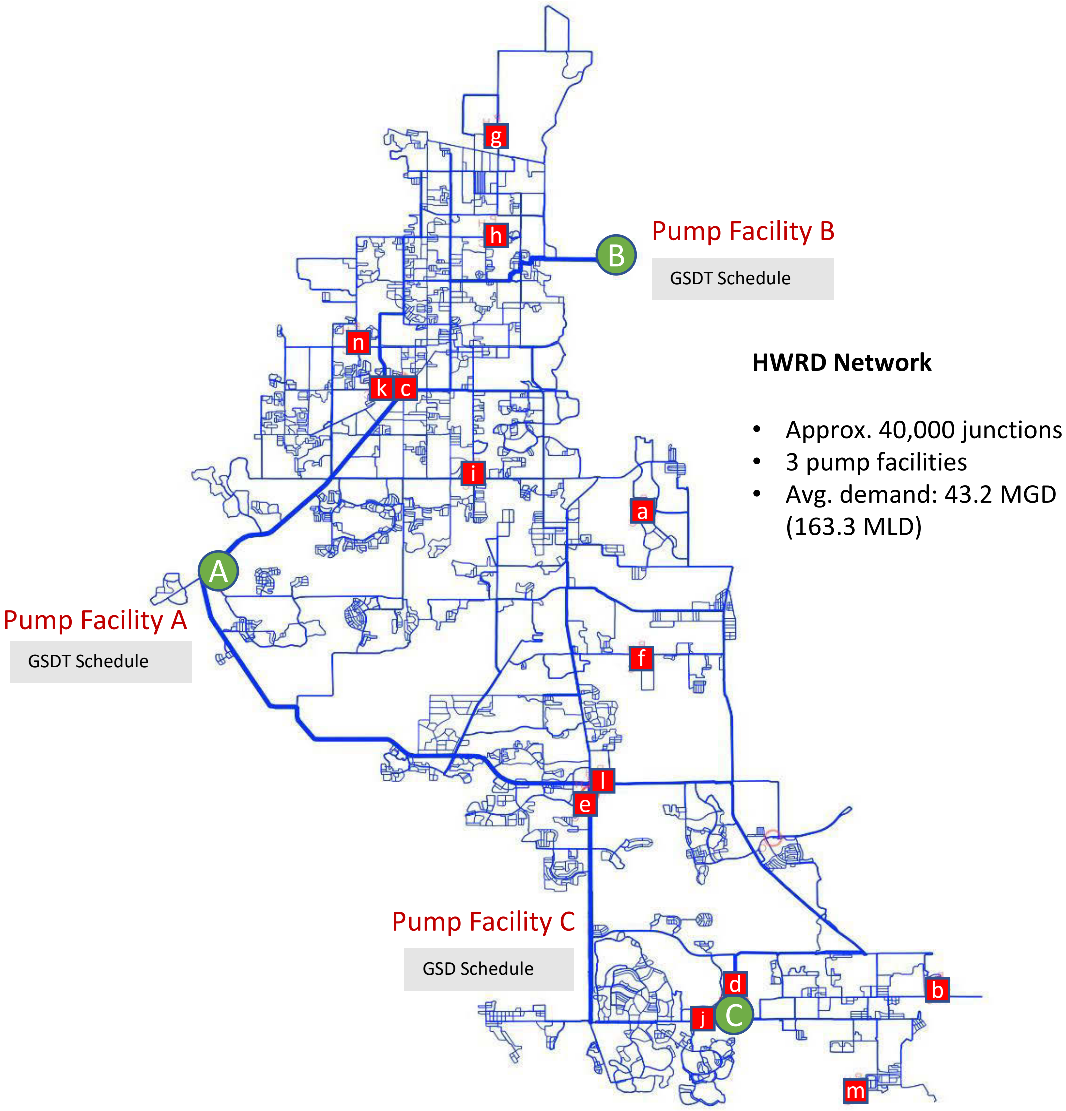

g
h
Pump Facility B
B
GSDT Schedule
HWRD Network
Approx. 40,000 junctions
3 pump facilities
Avg. demand: 43.2 MGD (163.3 MLD)
n
k
c
i
a
A
Pump Facility A
GSDT Schedule
f
l
e
Pump Facility C
GSD Schedule
d
b
j
C
m

Figure 3. The HWRD network model used in the PoC. Here, the real network has been modified geometrically and anonymized. The time frame of October 2020 is chosen for the PoC due to its interesting characteristics, such as water demand variability and the peak/off-peak demand charge tariff differences (for analysis of energy use and demand costs).

The time frame of October 2020 is chosen for the PoC due to its interesting characteristics, such as water demand variability and the peak/off-peak demand charge tariff differences (for analysis of energy use and demand costs).

A view of the actual network operation output in terms of energy demand and energy use cost is provided in Fig. 4. The energy demand or power input (Fig. 4a) was computed from SCADA measurements of pump facility discharge flow, suction pressure, and discharge pressure. The energy use cost (Fig. 4b) was computed from the energy demand time series by applying the appropriate energy schedule to each facility according to the energy contracts.

One noticeable feature of the time series of Fig. 4a is the presence of two power peaks during weekdays that are caused mainly by irrigation demands. Another visible feature is that power and energy use are lower during Sundays (Oct 4, 11, 18, and 25), in response to different consumption patterns during weekends with respect to weekdays.

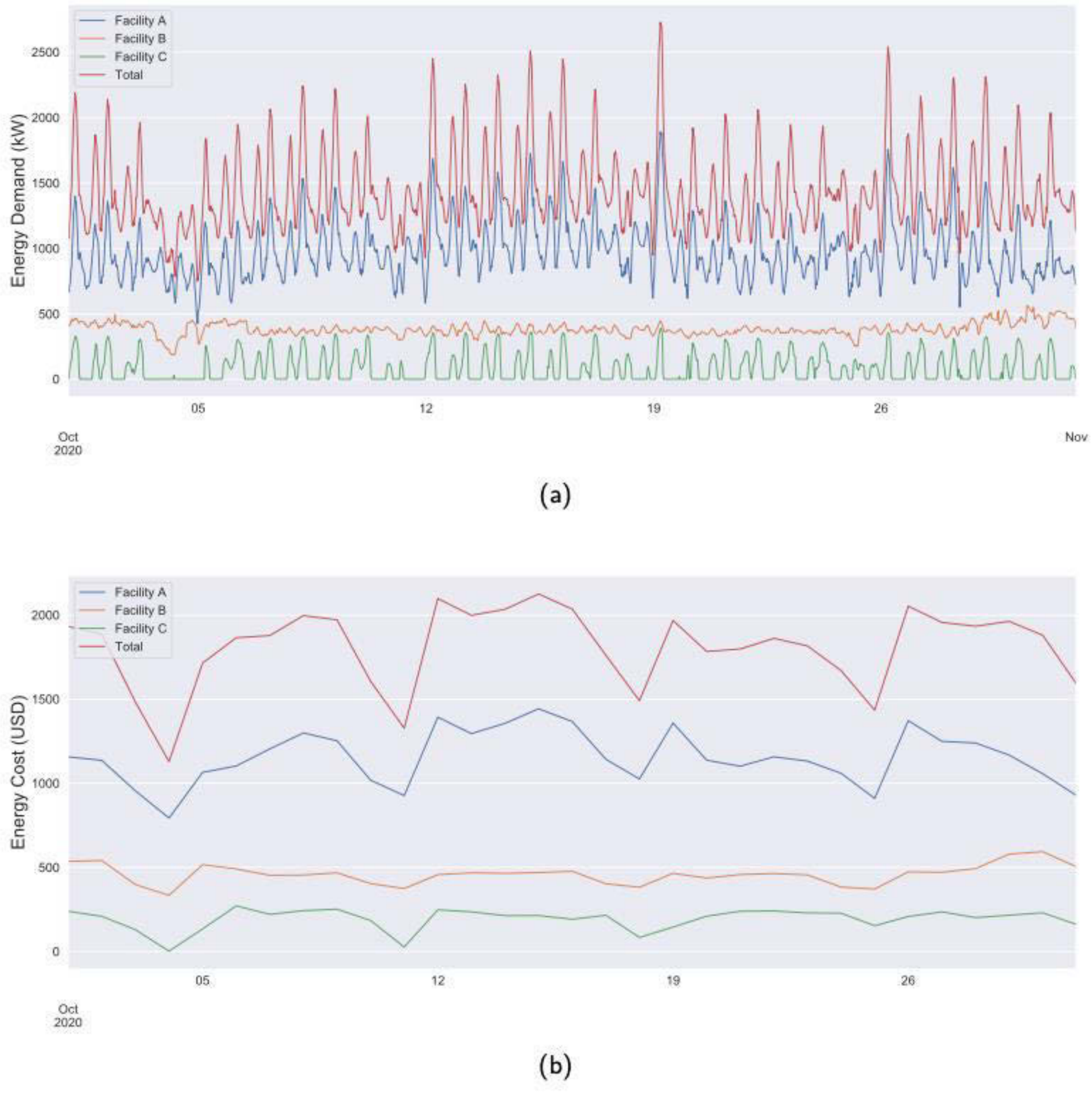


Figure 4. Actual energy demand and energy use costs for the HWRD system in the month of October 2020. (a) Instantaneous energy demand in KW per pump facility and total calculated from pressure and flow data; (b) Daily energy use cost (USD) per pump facility and total. Facilities A, B, and C account for 64, 25, and 11 percent of the total monthly energy use cost, respectively.

As part of the PoC optimization process, the distribution system was skeletonized by the network simplifier (see Fig. 2) resulting in a hydraulically equivalent network of approximately 400 nodes. Since the overall optimization process was configured to generate a 24-hour operation plan detailed at 30-min time intervals, the simplified network yielded over 21,000 decision variables after the initial optimization model construction performed by the optimizer. Most of these variables (400 x 48 = 19,200) correspond to the hydraulic heads at nodes. Moreover, the optimizer

BY NC SA

was configured to run as it would in production over a period of one month starting 2020-10-01 at 00:00 and ending 2020-11-01 at 00:00. The optimizer was setup to update the 24-hour operation plan every hour using current demand forecasts and initial conditions, such as tank levels. Other aspects of the configuration included the service-level constraints such as minimum pressures and tank levels, and the energy use and demand charge schedules. The results consist of 24 x 31 = 744 sets of optimized hydraulic states, energy costs and pump facility head-flow operation points. These results are summarized and described in the next section.

4 RESULTS

In this section we briefly discuss the computational performance and compare the optimization results with the actual operation by contrasting the energy use and costs. We also present a comparison of the levels of service before and after optimization focusing on network pressure and storage.

In terms of computational performance, Table 1 summarizes the CPU time employed by the IPOPT solver for the first and subsequent rolling-horizon iterations. While the solver took approximately 257 CPU seconds (4.28 min) for the first iteration, it took an average 10 seconds for the subsequent iterations. The total processing time including the solver CPU time, WNO service calls, etc., was 281 s (4.68 min) for the first iteration and 29.2 s on average for the subsequent iterations.

Table 1. Network Optimization Computational Performance Summary

Optimization Process Attribute	Value
Optimization output	
Number of variables	21,216
Number of equality constraints	20,971
Number of inequality constraints	1,568
Initial optimization number of iterations	3,315
Subsequent optimizations number of iterations	127
Processing time (s)	
IPOPT CPU time – first iteration	257.54
IPOPT CPU time – subsequent iterations (average)	10.08
Total processing time – first iteration	281.00
Total processing time – subsequent iterations (average)	29.20

Since the optimization objective is to minimize the sum of energy use costs and energy demand charges, it is important to describe how the network optimizer manages power use and power peaks throughout the optimization process. Fig. 5 and Table 2 present a comparison of the main features of energy demand (kW) and energy consumption (kWh) before and after optimization for pump facility A, the largest in the network. Note that we use the terms energy demand and input power as synonyms.

The energy contract for facilities A and B specifies demand charges based on the input power peak over the billing month. The power peaks are computed by averaging the facility input power time series at 30-min intervals and then selecting the maximum values both for peak and off-peak hours. The peak hours for the PoC are 12:00 Noon to 9:00 PM as specified in the utility's energy contract. Since the monthly power peaks cannot be known ahead of time, the optimizer starts from an initial guess of the power peak values and updates these estimates as the optimization progresses throughout the month. The updating process can be observed in Fig. 5 where power peaks are increasing progressively over time. By the end of the month, after optimization, the peak input power and cost for facility A were reduced by about 10 and 26% during peak and off-peak hours, respectively (see Table 2).

The energy consumption costs are related to the time-of-day charges. The optimizer aims at reducing these costs by combining the use of storage and pumps while respecting hydraulic and level of service constraints, in such a way that more pumping is performed during off-peak hours and more storage is used during peak hours. As shown in Table 2, the on-peak energy consumption and their associated costs were reduced by about 35% after optimization. On the other hand, the off-peak energy consumption and costs were reduced by approximately 23% by the optimizer.

In combination, through power peak reduction and improved time-of-use energy management, the network optimizer was able to reduce the total energy costs by approximately 23% with respect to the actual operation, for pump facility A.

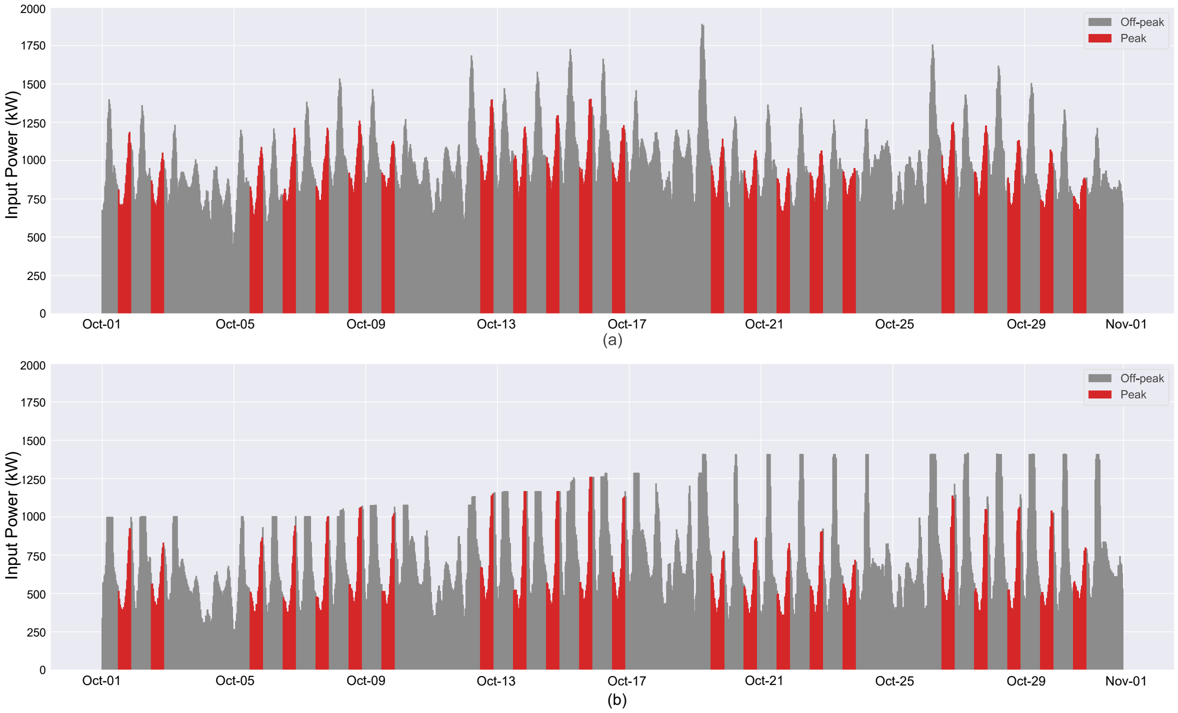


Figure 5. Comparison of input power at pump facility A: (a) Actual input power; (b) Optimized input power. The input power is expressed as a 30-min average in kW. The input power during peak and off-peak consumption hours are represented in red and gray, respectively.

Table 2. *Summary of energy use/cost, input power and demand charges at pump facility A*

Description	After Optimization	Before Optimization	Reductio After Optimization (%)
On-peak peak input power (kW)	1,253.85	1,403.09	10.64
Off-peak peak input power (kW)	1,399.71	1,892.45	26.04

On-peak demand charge ($)	10,206.34	11,421.18	10.64
Off-peak demand charge ($)	6,564.63	8,875.58	26.04
On-peak consumption (kWh)	118,134.06	181,993.83	35.09
Off-peak consumption (kWh)	417,835.20	545,354.73	23.38
On-peak consumption cost ($)	7,697.62	11,858.72	35.09
Off-peak consumption cost ($)	18,342.97	23,941.07	23.38
Total consumption cost ($)	26,040.58	35,799.79	27.26
Total demand charge ($)	16,770.96	20,296.75	17.37
Total charges ($)	42,811.54	56,096.54	23.68

A view of the total costs for the network both before and after optimization is presented in Fig. 6. These costs were computed from the hourly optimized and actual costs. The reduction of energy use costs is clearly noticeable by comparing the actual (purple-dashed line) and the optimized (red line) time series of costs. The actual demand charges were calculated by dividing the monthly demand charge by the 31 days of the analysis month. These costs are constant unlike the optimized demand charges, which are progressively increasing because of the updates that take place during optimization. A small reduction in demand charges was obtained through optimization as shown by the actual (orange dashed line) and optimized energy demand (last value of green-dash-dot line) time series.

To summarize the total daily energy cost savings after optimization, Fig. 7 presents a day-by-day sequence of percent savings obtained after optimization. These values range from 20 to 35%.

Another important aspect of the optimization process is to verify the levels of service in terms of network pressures and storage. The optimized pressures are compared with the pressures before optimization in Fig. 8. Overall, a considerable reduction of pressure was obtained throughout the network and the predetermined minimum pressure head constraint (29 m) was respected in all cases. This reduction of service pressures is responsible for much of the energy use cost savings.

Finally, the differences in water storage in the tanks are presented in Fig. 9. For visualization purposes, the optimized storage plot of Fig. 9a is showing the complete set of rolling-horizon tank levels (light colored lines). The green solid line is constructed by selecting the first time point in the sequence of optimized tank levels over the rolling operation horizon results. This line shows that the network optimizer is filling the tanks when needed but keeping the overall storage relatively low, on average, within the established minimum and maximum head constraints (14.5 and 24 m). On the other hand, the actual tank levels before optimization are shown in (Fig. 9b), where tanks are filled and drained cyclically and filled over the weekend, when energy is cheap, in preparation for the Monday demand peak in early hours. In addition, Fig. 9c shows the demands of the network to allow visualizing the opposite trends of demand versus storage, especially in the case of actual storage management.

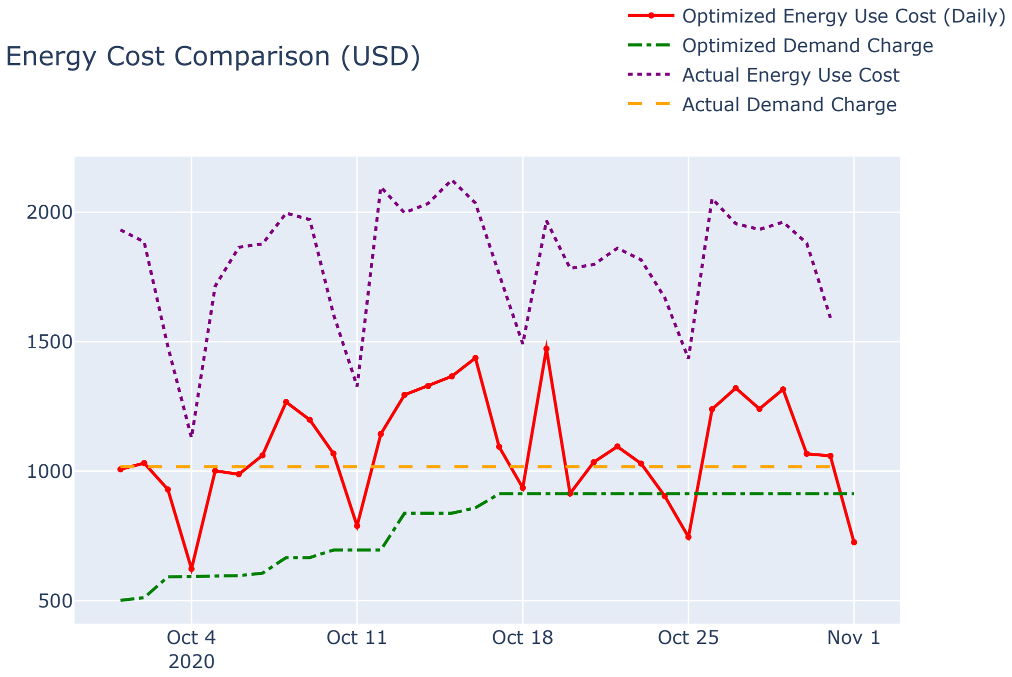


Figure 6. Comparison of total daily energy cost (USD) including use and demand charges. The red solid line corresponds to the daily energy use cost after optimization. The purple-dashed line represents the actual computed daily energy use cost. The green-dash-dot line represents the progressively increasing demand charge (apportioned by day) as computed during optimization; the last value in this series is the optimized daily demand cost. The orange-dashed line corresponds to the actual demand charge (apportioned by day).

5 SUMMARY AND CONCLUSIONS

We proposed a combined strategy designed to address challenges associated with optimizing a water network, namely, the network scale, the nonlinearity of hydraulic relationships, and the discrete nature of operation decisions. The strategy decomposes the problem in two subproblems, one that optimizes the network first, and the other that optimizes the pump facilities using results from the first subproblem. As part of the scope, we presented the NMPC methodology that represents the water supply network as an optimization problem where the state dynamics are modeled as a system of DAEs. The water network optimizer architecture includes the necessary tools that translate the optimization model from a high-level to a low-level and leverages IPOPT, a powerful large-scale NLP solver.

The proposed method was demonstrated as a rolling-horizon optimizer on a PoC where one-month worth of water supply network data was processed to obtain optimized operation policies for control valves and pump facility hydraulic head and flow operation points. The results demonstrate that it is feasible in terms of processing time (under 30 seconds) to run the NMPC optimizations to support the operation decisions of a real water network. The results also show that, for the network evaluated, considerable energy savings can be obtained on most operational days, especially when the supply system is not hydraulically stressed due to water consumption solicitation. In addition, we showed that how the network optimizer respects the level of service constraints including pressure and tank level limits.

Future work includes incorporating water quality in the optimization process, enhancing the scalability and robustness of the optimization process, and further developing and tuning the pump facility optimizer that works in coordination with the water network optimizer described here.

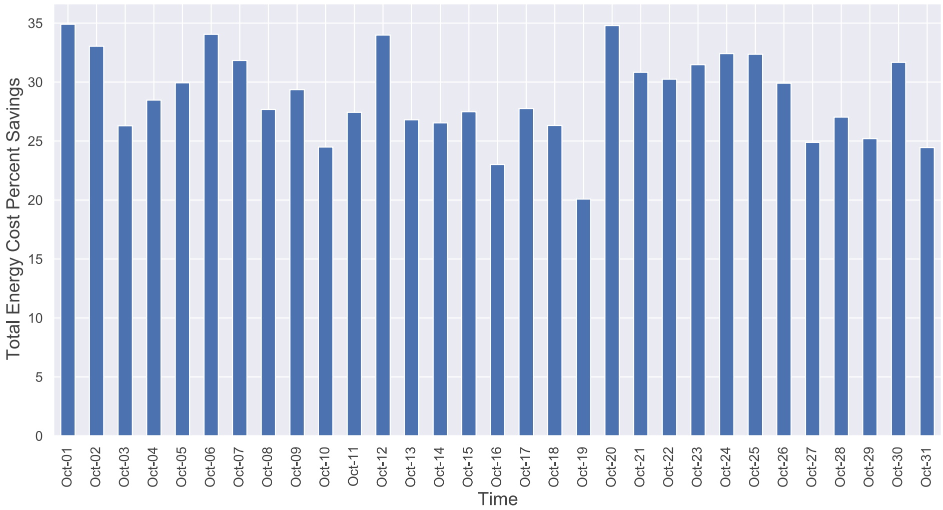


Figure 7. Daily savings of energy use cost calculated from the optimized operation schedules. The savings range from 20 to 35%.

6 ACKNOWLEDGEMENTS

We thank Dr. John McCary, Planning Section Manager and the technical staff at HWRD for providing all of the data necessary to develop the PoC presented in this paper.

7 REFERENCES

[1] Lorenz T Biegler and David M Thierry. Large-scale optimization formulations and strategies for nonlinear model predictive control. IFAC-PapersOnLine, 51(20):1–15, 2018.

[2] Jacek Blaszczyk, Krzysztof Malinowski, Alnoor Allidina, N Callaos, T Gill, and B S´anchez. Optimal pump scheduling by non-linear programming for large scale water transmission system. Proc. CCISE, pages 7–12, 2013.

[3] Ajit Gopalakrishnan and Lorenz T Biegler. Economic nonlinear model predictive control for periodic optimal operation of gas pipeline networks. Computers & Chemical Engineering, 52:90–99, 2013.

[4] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical programming, 106(1):25–57, 2006.

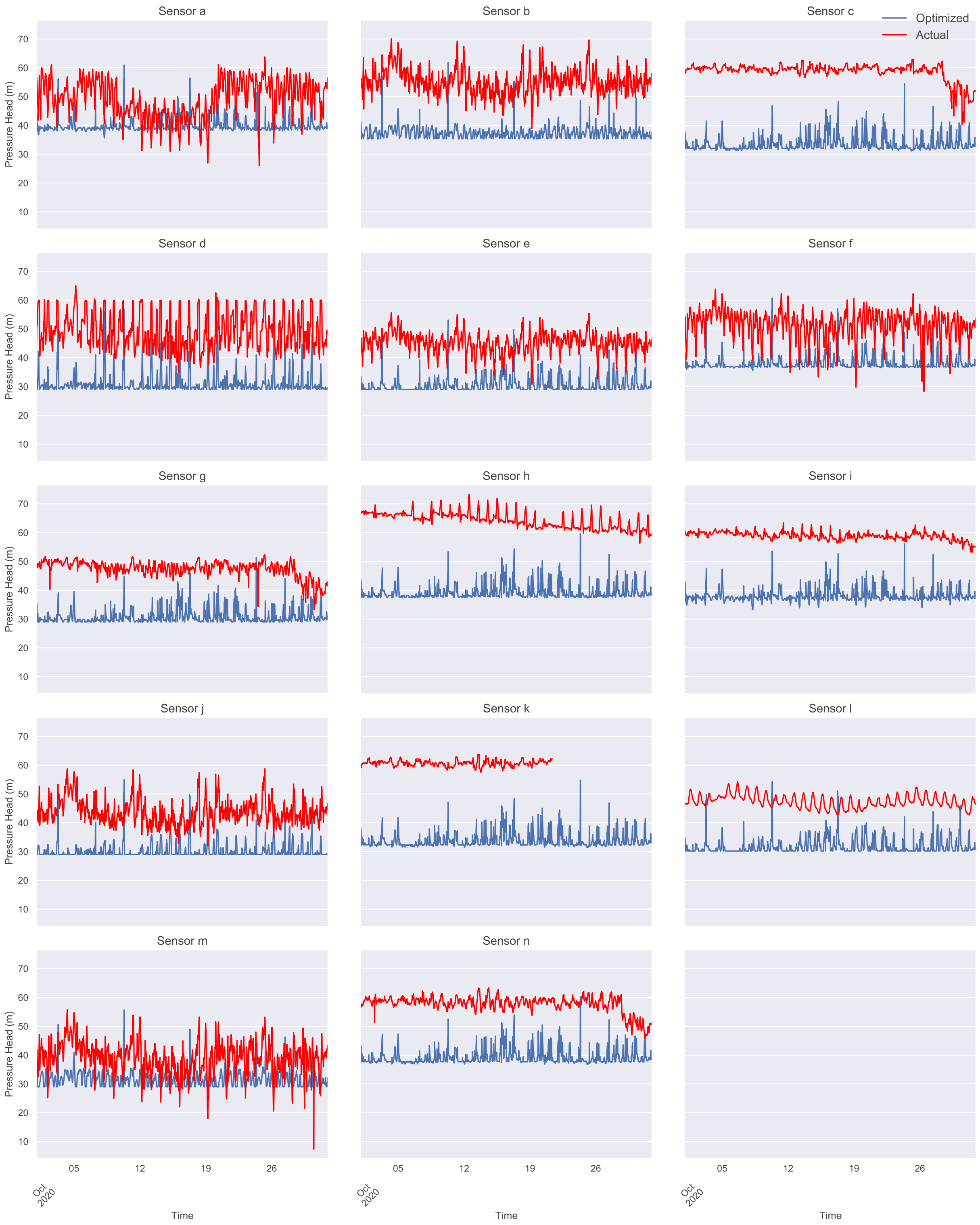


Figure 8. Comparison of optimized and actual pressure heads (m) at network junctions where pressure sensors are available. The optimization constraints at these junctions specify that pressures must be above 29 m (41 psi). The purpose of this plot is to illustrate how the levels of service in the form of pressure constraints are being applied and respected.

cc BY NC SA

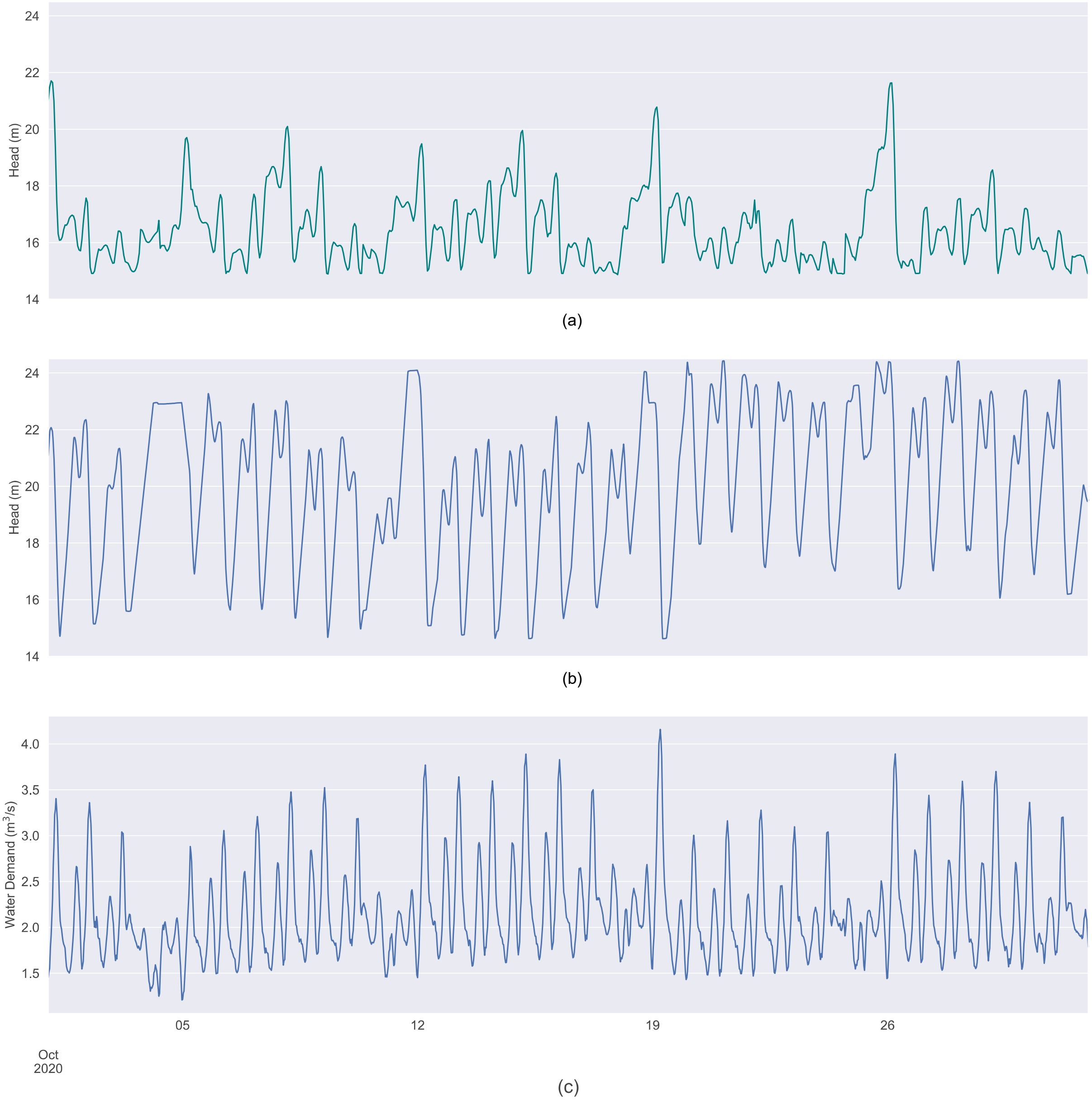


Figure 9. A comparison of actual and optimized storage management. The actual operation consists of filling the tanks during the weekend when energy is cheap in preparation for the Monday demand peak in early hours. On the other hand, the network optimizer is filling the tanks when needed but keeping the overall storage lower within limits.

BY NC SA

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14817

LEAKAGE CONTROL OF WATER DISTRIBUTION SYSTEM BY DROP-RESTORE PRESSURE BASED ON VISCOELASTIC MECHANISM

Jinliang Gao[1], Jianxun Chen[2], Wenyan Wu[3], Liqun Deng[4], Kunyi Li[5] and Yi Yuan[6]

[1,2,5,6] School of Environment, Harbin Institute of Technology, Harbin, Heilongjiang Province, China
[3] School of Engineering and the Built Environment, Birmingham City University, Birmingham, UK
[4] Shenzhen ANSO Measurement & Control Instruments Co., Ltd., Shenzhen, China

[1] *gjl@hit.edu.cn*, [2] *cjianxun877@gmail.com*, [3] *wenyan.wu@bcu.ac.uk*, [4] *380317336@qq.com*,
[5] *21b929004@stu.hit.edu.cn*, [6] *18292005003@163.com*

Abstract

As a common method to control leakage of water distribution system, pressure management has the advantages of reducing energy consumption, reducing the possibility of explosion and avoiding the aggravation of leakage. In recent years, with the popularization of plastic pipe in the world, it is necessary to study its leakage characteristic. Our research group carried out leakage experiments on high density polyethylene (HDPE) pipe, and found that the correlation curve between leakage flow and pressure did not completely coincide in the phase of pressure boost and pressure reduction. The existing FAVAD and exponential leakage models could not explain this phenomenon, which challenges the pressure management theory dominated by a single depressing-pressure process, thus it's necessary to explore pressure management strategies suitable for plastic pipes. Based on the viscoelastic properties of plastic pipe, we established the viscoelastic leakage model and proposed the leakage control method of drop-restore pressure, and verified its feasibility in practical engineering case. The main research objectives of this paper will be firstly to describe the strain response of leakage area in the process of continuous stress application with the Boltzmann superposition principle for HDPE pipe; the Voigt-Kelvin model is used to simulate the creep behavior of viscoelastic material, and a suitable leakage model for viscoelastic pipe is proposed to provide accurate expression of the leakage under the regulation of drop-restore pressure. Secondly, the viscoelastic pipe leakage model is embedded into the pressure-driven analysis model based on non-iterative method, and the pressure-driven viscoelastic leakage model is obtained. Finally, evaluating the proposed leakage model in the practical case. With the minimum of leakage flow as the objective function, the leakage control model of drop-restore pressure is established and solved by particle swarm optimization algorithm to obtain the accurate pressure regulation scheme. After applying the scheme from the optimization, the leakage rate decreases from 37.7% to 16.8% on weekday, which is great impact on leakage control.

Keywords
Pressure management, Leakage control, Pressure-driven viscoelastic leakage model, Drop-restore pressure.

1 INTRODUCTION

Pressure management is an effective method to reduce leakage and widely used by scholars and water enterprises. Global pressure reduction or local pressure reduction by using pressure reducing valves and other equipment are two commonly used methods of pressure management. The use of pressure reducing valves for pressure management is common in district metered area [1,2].

Accurate leakage model can be used to reflect the change law of leakage flow with pressure, which is helpful to estimate the benefit of leakage control measures and provide guidance for the formulation of leakage control strategies. With the constant exploration of the leakage law, the leakage model has undergone three stages: orifice outflow model, exponential model and FAVAD model. FAVAD model pays more attention to reflect the area change of leaks and improves the accuracy of leakage model. The three models all prove the positive correlation relationship between the leakage flow and pressure, which builds the foundation for leakage control through pressure reduction. However, in recent years, scholars [3,4] have found that leakage flow is not simply positively correlated with the pressure at leaks, but is affected by the pipe material and the geometry of the leaks. This challenges the conventional awareness of reducing pressure to reduce leakage.

Plastic pipe has the advantages of environmental protection, light weight and high strength, corrosion resistance, smooth inner wall without scaling, easy construction and maintenance, long service life and high cost performance. According to statistics data, from 2010 to 2012, the service length of HDPE pipes with diameter below DN600 in Shanghai accounted for 58% of the total length of commonly used pipes [5]. In recent years, the market share of plastic pipe is expanding, which also provides great impetus for the study of leakage model of plastic pipe.

In this paper, based on the leakage characteristics of HDPE pipe leaks we have studied, the concept of drop-restore pressure to control leakage is proposed. The leakage model in the process of drop-restore pressure was constructed by combining Boltzmann superposition principle and generalized Voigt-Kelvin model. Then, this model was embedded into EPANET hydraulic calculation module, and the pressure-driven leakage model based on viscoelastic theory was established to make the model simulation more close to the real operation. Combined with actual engineering cases, the water network dynamic pressure control model based on drop-restore pressure mechanism was used to calculate the pressure setting at the inlet, and verify the feasibility and effectiveness of the model based on drop-restore pressure mechanism.

2 CONCEPT OF DROP-RESTORE PRESSURE

Our research group has conducted experimental studies on three typical leaks of PE pipe: axial crack, circular leak and circumferential crack [6]. The results show that PE pipe has the following characteristics in the process of pressure boost and pressure reduction:

(1) Under the same leak shape and size, the pressure-leakage flow curve does not coincide between the pressure boost process and the reduction process;

(2) Under the same pressure, leakage flow in the pressure reduction process is higher than that in the pressure boost process;

(3) When the traditional leakage flow-pressure relationship is used to fit the experimental data, the leakage index in the pressure-boost process is generally higher than that in the pressure reduction process.

The leakage flow-pressure relationship curve of the same pressure does not coincide in the pressure boost stage and the pressure reduction stage, which means that the leakage can be reduced when the pressure operation mode is changed. The leakage can be reduced under the same operation pressure by choosing the pressure operation mode reasonably.

Taking the leakage pipe of PE pipe (DN50) with 30mm length axial crack as an example, the leakage characteristic in the process of pressure reduction and pressure boost is analysed. The experimental results are shown in Figure 1, and the conclusion could be drawn as follows:

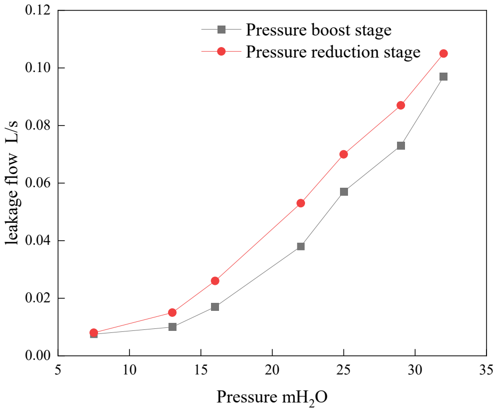


Figure 1. The relationship curve between leakage flow and pressure under pressure boost and pressure reduction stage

(1) Under the same pressure, the leakage flow in the process of pressure reduction is basically greater than that in the process of pressure boost. For example, when pressure is 21.6mH_2O, the measured leakage flow of reducing pressure process is 53.84mL/s, and the leakage flow of pressure boost is 41.57mL/s, which is 23% higher than that in the process of pressure boost.

(2) According to the exponential leakage model, the model obtained from the pressure boost process is $= 0.003h^{2.32}$, and the model obtained from the pressure reduction process is $q = 0.143h^{1.92}$. It can be seen that the leakage index of the pressure boost process is larger and more sensitive to pressure.

Based on the analysis of the above experimental result, it can be considered that when the pressure is constant, the operating pressure can be adjusted to the initial water supply pressure by lowering the pressure first and then gradually increasing the pressure, so as to achieve the purpose of reducing leakage without changing the operation pressure. Thus, the concept of drop-restore pressure leakage control is introduced, which refers to the adjustment process of first reducing the water supply pressure and then gradually increasing the water supply pressure to the original water supply pressure to achieve the ultimate goal of reducing the leakage flow.

3 LEAKAGE MODEL CONSTRUCTION BASED ON DROP-RESTORE MECHANISM

In the water supply network, the pipe material includes steel, concrete, ductile iron and high density polyethylene material. The stress-strain relationships of the first three materials are linear viscoelasticity. However, the stress-strain relationship of high density polyethylene is viscoelastic, and the strain has a certain lag after the stress application, which is characterized by pressure, time and temperature. In order to model the leakage characteristics of high density polyethylene pipe (HDPE), we used the structural equations to represent physical phenomena that assume linear-viscoelastic behaviour, and apply the Boltzmann superposition principle to describe the change of strain with stress.

3.1 Boltzmann superposition principle

Since the change of leak area is not only related to instantaneous pressure but also to historical pressure, in order to characterize the dynamic change of leak area, Boltzmann superposition principle[7] was used to quantify the change of leak area under pressure variation in the whole process.

According to Boltzmann superposition principle: when the strain is small, the continuous stress actions can be regarded as mutually independent and the strains can be simply added linearly, shown as in equation (1):

$$\varepsilon(\mathrm{t}) = \sigma(t)J_0 + \sum_{\mathrm{i}=1}^{\mathrm{m}} J(\mathrm{t}-\tau_{\mathrm{i}})\Delta\sigma_{\mathrm{i}} \tag{1}$$

The total strain generated by continuous stress application can be expressed as:

$$\varepsilon(\mathrm{t}) = \sigma(\mathrm{t})J_0 + \int_0^{\mathrm{t}} \sigma(\mathrm{t}-\alpha)\frac{\mathrm{d}J(\alpha)}{\mathrm{d}\alpha}\mathrm{d}\alpha \tag{2}$$

where J_0 is instantaneous creep compliance(m^2/N), $\sigma(t)$is the stress at time t (N), $J(\alpha)$is the creep function and τ_{i} is the time constant.
In equation (2), the total strain corresponds to leak area, and the stress corresponds to the operating pressure of the pipe network.

3.2 Voigt-Kelvin model

Creep function is introduced into equation (2). Creep occurs in viscoelastic pipes such as polymer or metal, that is, under a constant pressure, the leak area increases gradually with time. The creep behaviour of viscoelastic pipe under operating pressure was simulated by Voigt-Kelvin model[8].

Among viscoelastic models, the standard linear body model can describe the transient elasticity and hysteresis effect of viscoelastic materials due to its simple creep function expression. In this paper, the creep compliance formula of the standard linear body is introduced:

$$J(\mathrm{t}) = J_0[1+\beta(1-\exp^{-\lambda\mathrm{t}})] \tag{3}$$

Where, $J(\mathrm{t})$is the creep function at time t, β is constant which ranges from 0.1 to 0.9, and λ is constant which value is $0.5\mathrm{min}^{-1}$.

3.3 Construction of viscoelastic leakage model

The above application of Boltzmann superposition principle and the formula of standard wire creep flexibility present the strain in mathematical form, provide one method for the dynamic simulation of leak area.

The circumferential stress of viscoelastic pipe can be calculated by equation (4):

$$\sigma(\mathrm{t}) = \frac{P(\mathrm{t})(D_0-\mathrm{e})}{2\mathrm{e}} \tag{4}$$

Where P is the inner pressure of pipe(MPa), D_0 is exterior diameter of pipe(mm) and e is the thickness of pipe wall(mm).

Substituting equation (3) and equation (4) into equation (2), the total strain of viscoelastic material under ideal state can be obtained as shown in equation (5):

$$\varepsilon(\mathrm{t}) = \frac{(D_0-\mathrm{e})}{2\mathrm{e}}J_0\left[P(\mathrm{t}) + \beta\cdot\lambda\int_0^{\mathrm{t}} P(\mathrm{t}-\alpha)\exp^{-\lambda\alpha}\mathrm{d}\alpha\right] \tag{5}$$

For the actual water supply system, water pressure is in most cases discrete, so a discrete total strain formula is needed. Substituting equation (3) and equation (4) into equation (1), the follow equation (6) could be obtained.

$$\varepsilon(\mathrm{t}) = \frac{(D_0-e)}{2e}\cdot J_0\left\{P(t) + \sum_{\mathrm{i}=1}^{\mathrm{m}} (P_{\mathrm{i}}-P_{\mathrm{i}-1})[1+\beta(1-\exp^{-\lambda(\mathrm{t}-\tau_i)})]\right\} \tag{6}$$

Where τ_i is time constant, P_{i} and $P_{\mathrm{i}-1}$ （MPa） are the pipe pressure in i period and i − 1 period, respectively.

The discrete leakage area formula (equation (8)) can be obtained by combining orifice outflow equation (shown as equation (7)) and equation (6):

$$Q = A_{\mathrm{r}} C_{\mathrm{d}} \sqrt{2\mathrm{gH}} \tag{7}$$

Where Q is leakage flow($\mathrm{m^3/h}$). A_{r} is leak area($\mathrm{m^2}$), C_{d} is leakage coefficient and H is water head of leak($\mathrm{mH_2O}$):

$$A_r(\mathrm{t}) = K\frac{(D_0 - e)}{2e} \cdot J_0 \left\{P(t) + \sum_{\mathrm{i=1}}^{\mathrm{m}} (P_{\mathrm{i}} - P_{\mathrm{i-1}})[1 + \beta(1 - \exp^{-\lambda(\mathrm{t}-\tau_i)})]\right\} \tag{8}$$

Where $A_r(\mathrm{t})$ is leak area in t period.

Substitute equation (8) into the equation (7) to obtain the discrete leakage model as equation (9).

$$Q(\mathrm{t}) = C\frac{(D_0 - e)}{2e} \cdot J_0 \left\{P(t) + \sum_{\mathrm{i=1}}^{\mathrm{m}} (P_{\mathrm{i}} - P_{\mathrm{i-1}})[1 + \beta(1 - \exp^{-\lambda(\mathrm{t}-\tau_{\mathrm{i}})})]\right\} \sqrt{2gH(t)} \tag{9}$$

Where $Q(\mathrm{t})$ is the leakage flow in t moment($\mathrm{m^3/h}$), $P(t)$ is the pipe pressure(MPa), C is combined coefficient, τ_{i} is time constant.

3.4 Construction of pressure drive viscoelastic leakage model

In the water network model, the diffuser of EPANET node attribute is often used to simulate leaks, and the outflow of diffuser depends on the pressure at the junction. Considering the relationship between flow rate and pressure at diffuser junction, diffuser can realize pressure driven analysis of water supply system. The generalized equation of diffuser flow is shown in equation (10):

$$q_j^{avl} = C_d(H_j^{avl} - H_j^{min})^{\gamma}; H_j^{avl} \geq H_j^{min} \tag{10}$$

Where C_d is flow coefficient, $H_j^{avl}$and H_j^{min} are actual pressure and minimum supply pressure at junction j, respectively, and γ is one empirical value.

Based on the research of leakage control mechanism of drop-restore pressure, the research results were applied to the water supply network model. For viscoelastic pipe, it is assumed that the leakage area and strain are linearly dependent, and the pressure at the leak is equal to the operating pressure in the pipe. Based on these two assumptions, the viscoelastic leakage model (equation (9)) is embedded into the EPANET pressure-driven demand model considering the pressure boost and pressure reduction state of the pipe in the previous period.

In the existing studies, the leakage coefficient is determined by fitting the historical leakage flow and pressure data according to pressure driven demand model proposed by Wagner[9]. The circular iterative method is used to determine the leakage coefficient, and it is assumed that the leakage coefficient of every junction is equal and of a certain value, and this resolution is so imprecise. For viscoelastic materials such as HDPE, the leak area depends not only on the instantaneous pressure but also on the historical pressure. Therefore, the leak area should change dynamically with pressure, not a constant value, and the leakage coefficient is not equal at every junction.

According to equation (8) and equation (10), the expression of leakage coefficient can be obtained as follow:

$$C = \sqrt{2g} \cdot C_d \cdot K\frac{(D_0 - e)}{2e} \cdot J_0 \left\{P(t) + \sum_{\mathrm{i=1}}^{\mathrm{m}} (P_{\mathrm{i}} - P_{\mathrm{i-1}})[1 + \beta(1 - \exp^{-\lambda(\mathrm{t}-\tau_{\mathrm{i}})})]\right\} \tag{11}$$

Where K is the coefficient used to describe the relationship between leak area and strain, and the meanings of other symbols, see above.

It can be seen from equation (11) that the leakage coefficient is only a function of pressure and time, and other parameters can be determined according to the pipe characteristic. The leakage coefficient corresponds to the diffuser coefficient in EPANET. Assuming that leak occurs in every junction of the water supply network, the diffuser coefficient is assigned to every demand junction with the change of instantaneous pressure and historical pressure, and the average leakage flow at night on weekdays is checked to obtain the pressure-driven viscoelastic leakage model.

4 CASE STUDY

4.1 Brief introduction about Y network

Y network is selected as the case study. Y network has a single inlet, and flow and pressure sensors are installed before and after the inlet valves. The topology and basic information of the water supply network are clear. The schematic of Y network is shown as Figure 2.

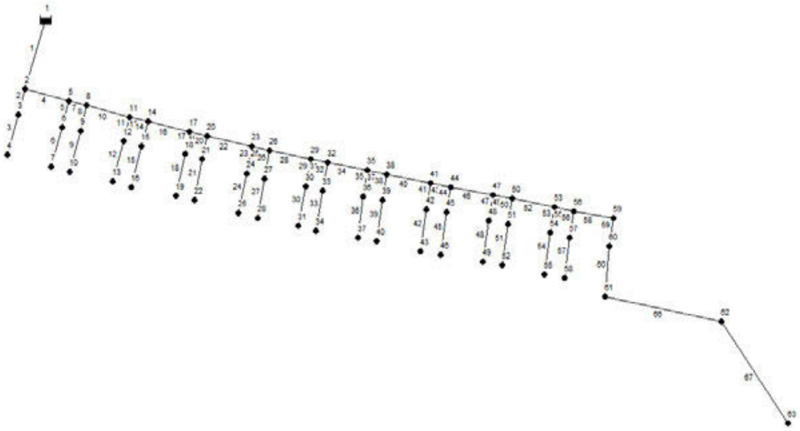


Figure 2. Schematic of Y network

The average daily inlet pressure of the network is 52.5mH$_2$O. In addition, the minimum night flow in this area are pretty high, which has enough potential for pressure reducing. The basic information of the network is summarized in Table 1.

Table 1. Basic information of Y network

Water supply Sevice area/m^2	Diameter/mm	Length/m	Pipe material
25,000	DN100	433	HDPE
	DN50	418	
Amount of costumers	Highest floor	Number of junctions	Number of pipes
210	5	62	62

Before regulating pressure, the total flow of inlet of Y network within one week was analysed, as shown in Figure 3. There are some differences in water consumption patterns during the week, and the patterns from Monday to Friday are very similar, with the peak of water consumption occurring at 8h and 22h. The pattern of water consumption on Saturday and Saturday is similar, and is significantly different from that on weekdays. The peak of water consumption occurs at 11h and 20h, respectively.

CC BY NC SA

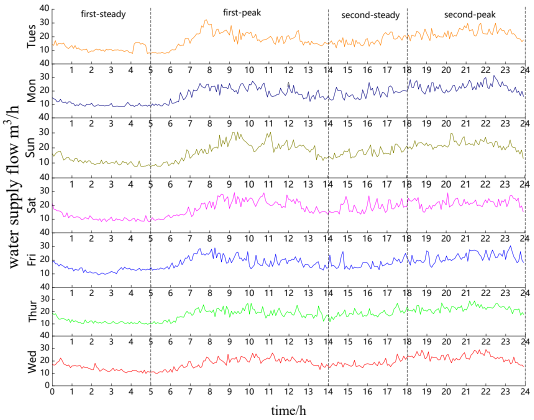


Figure 3. Analysis of weekly water consumption

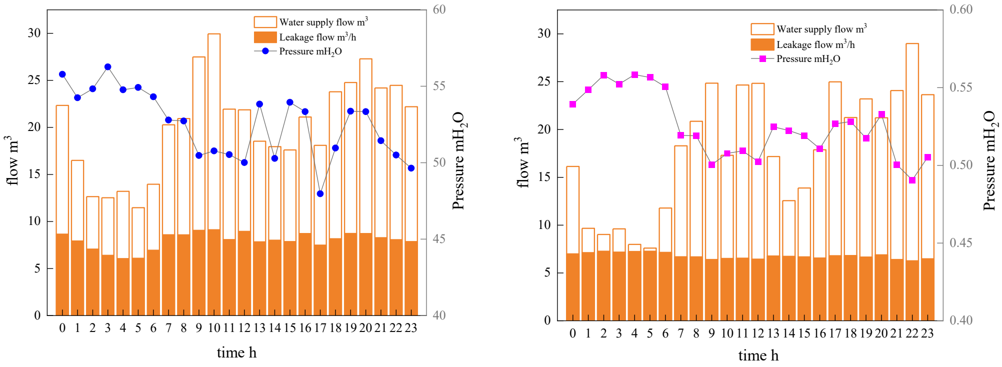
4.2 Minimum night flow analysis

Analysing the flow change of the total meter at the entrance of Y network within a week, the minimum night flow occurs during 2h to 5h at night. Using the minimum night flow during this period to estimate the leakage level. The legal water consumption at night of 1.7L/household/h is adopted, and the difference between the minimum night flow and the legal water consumption at night is used as the leakage flow.

The exponential leakage model (equation (12)) was used to quantify the relationship between inlet pressure and total leakage flow, the leakage index of 1.15 was adopted. According to the change of the inlet pressure of Y network at every hour on Friday and Saturday, the leakage flow of the two days was estimated. The results are shown in Figure 4.

$$Q_l = \alpha H^{\beta} \tag{12}$$

Where, Q_l is leakage flow (m^3/h), α is leakage coefficient, β is leakage exponent, H is the pressure at leaks(mH$_2$O), α and β could be determined by the minimum night flow and the pressure at that moment.



(a) *Friday*

(b) Saturday

Figure 4. Variation of leakage flow, total water supply flow, inlet pressure with time on Friday and Saturday

On both Fridays and Saturdays, the leakage rate was nearly 40%. Since the ground elevation of the Y network is relatively flat, considering the service floor height of the most critical node, the working pressure of water appliances and the head loss of the pipes, the water supply pressure at the entrance of the pipe network can meet the demand of water users by reaching 35mH$_2$O. The actual water supply pressure is as high as 50mH$_2$O, and the high pressure is the main reason for the high leakage rate.

4.3 Construction of leakage model

Collecting and organizing the static information and operation data of the Y network, and build a pressure-driven viscoelastic leakage model based on the water consumption data within a week and 1h as the calculation step. Since all the pipes are PE pipes, the parameter setting are as follows: $J_0 = 8.5\times10^{-3}$ (1/MPa), $\beta = 0.5$, $\lambda = 0.5$.

In order to decrease the difference between hydraulic model and actual operation of the Y network, the critical node 63 is selected as the pressure checking point. The pressure gauges was installed at 63 node (shown as Figure 5) to check the accuracy of the hydraulic model by comparing the calculated pressure with the measured pressure.

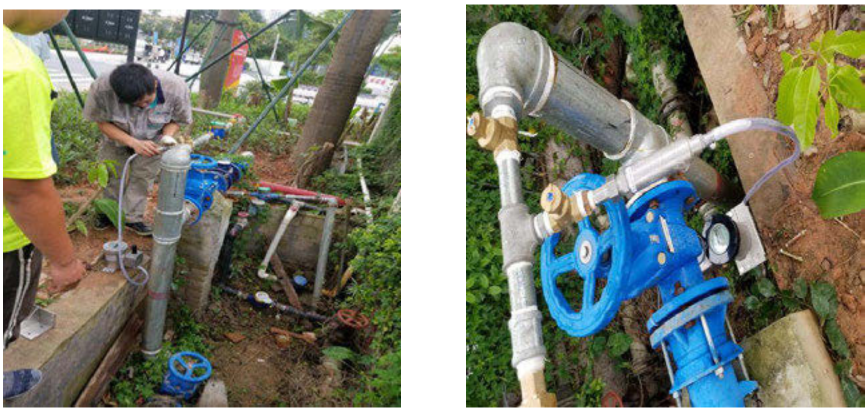
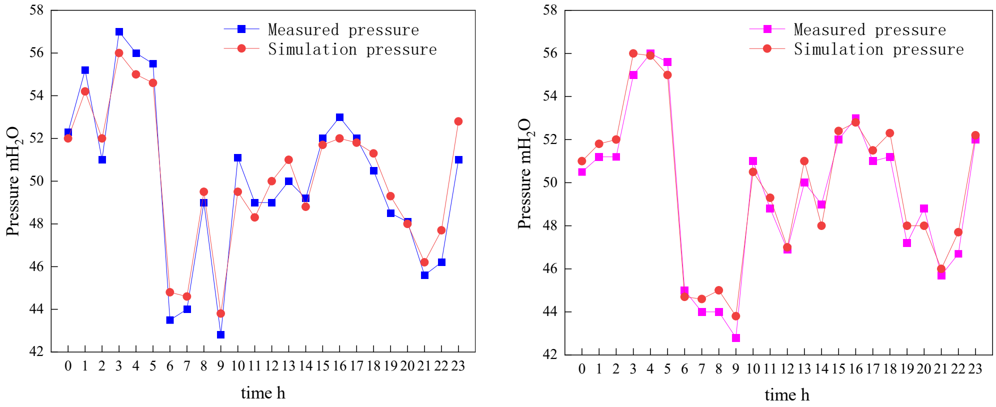


Figure 5. Pressure gauge installation at critical node

The operating data of Friday and Saturday were selected as the foundation for verification. The accuracy of the model was verified by comparing the calculated with the estimated leakage flow in Figure 4. The variation of measured and simulate calculated pressure value with time at the critical point on Friday and Saturday are shown in Figure 6.



(a) *Comparison of pressure on Friday*

(b) Comparison of pressure on Saturday

Figure 6. Comparison of measured and simulate pressure at the critical node

According to the analysis of Figure 6, the pressure fluctuation on working days is larger than that on weekend. The absolute error between the calculated and the measured pressure is within 5%. The error is very small, and the calculated total leakage volume of the whole day are 155 m^3 and 180 m^3, which are very close to the actual estimated leakage volume. So the accuracy of the established pressure-driven viscoelastic leakage model meets the application requirements and can be used for subsequent analysis.

4.4 Solution of drop-restore pressure leak model

Since the actual water supply in the Y network changes constantly and the pressure also changes hourly, in order to maximize the effect of drop-restore pressure leakage control, it is necessary to optimize the relationship between pressure regulation strategy and hourly water supply scheduling. On the premise of meeting the user's demand, find the best inlet pressure setting value.

In the optimization process, pressure setting of inlet valve every hour were set as the decision variables, the minimum leakage flow was set as the objective function (shown as equation (13)) and the minimum pressure demand at critical node was set as constraint conditions. The particle swarm optimization (PSO) algorithm was used to solve the optimal pressure regulation strategy at the entrance of the Y network and particle is the hourly inlet pressure set value, the particle fitness is the total leakage flow of the network throughout the day.

$$\min\sum_{i=1}^{n}\sum_{j=1}^{m} Q_{ij}=\min\sum_{i=1}^{n}\sum_{j=1}^{m} C\frac{(D_0\text{-}e)}{2e}\cdot J_0\left\{P_j+\sum_{t=1}^{j}(P_t\text{-}P_{t\text{-}1})[1+\beta(1\text{-}\exp^{\text{-}\lambda(t\text{-}\tau_t)})]\right\}\sqrt{2gH_j} \qquad (13)$$

Where, n is the number of junction, m is the total period to simulate, Q_{ij} is the leakage flow of j period at i junction (m^3/h), P_j is the pressure at j period (MPa), C is the coefficient, τ_i is time constant at j period, D_0 is the exterior diameter of pipe (mm), e is the thickness of pipe wall, J_o is the instantaneous creep compliance (m^2/N).

Due to the difference in water consumption patterns between weekdays and weekends of the Y network, the field experiment was conducted on Friday and Saturday of the next week in order to verify the leakage control effect of drop-restore pressure regulation. The pressure regulation strategy adopts the fixed outlet pressure control and the time-modulated pressure control, and the operation mode introduces the drop-restore pressure mode.

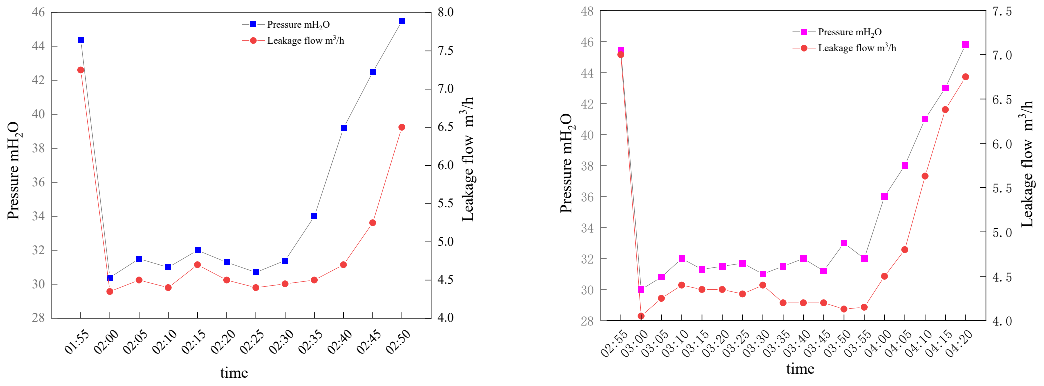
According to the analysis of the water consumption pattern of the Y network, there is still a small amount of water is used at night on weekdays, which is difficult to reflect the actual leakage situation. From 2:00 to 5:00 in the morning on Friday, the drop-restore pressure method is used for pressure regulation, and the outlet pressure is fixed for the rest of the period. For Saturday, the optimal pressure value of each time period is obtained according to the solution of the drop-restore pressure leakage control model. Considering that the inlet valve of Y network is an ordinary valve, which is difficult to be adjusted frequently and accurately, detailed pressure regulation strategy is formulated according to the pressure comfort required by historical records, as shown in Table2.

Table 2 Pressure regulation scheme on Friday and Saturday

Date	time	Pressure setting/mH_2O	time	Pressure setting/mH_2O
Friday	0:00	45.00	4:30-4:59	45.00
	0:00-1:59	45.00	5:00-5:05	35.00
	2:00	30.00	5:05-5:29	35.00
	2:00-2:29	30.00	5:30-5:40	40.00
	2:30	30.00	5:40-14:29	40.00
	2:30-2:49	45.00	14:30-14:40	30.00
	2:50-3:29	45.00	14:40-15:00	30.00
	3:30	45.00	15:00-15:30	40.00
	3:30-3:59	30.00	15:30-17:00	40.00
	4:00	30.00	17:00-17:10	45.00
	4:00-4:30	30.00	17:10-23:59	45.00
Saturday	0:00-1:59	45.00	14:30-17:29	35.00
	2:00-5:29	35.00	17:30-11:59	45.00
	5:30-14:29	45.00		

4.5 Effectiveness verification of leak control scheme

The effect of drop-restore pressure leakage control method is verified according to the filed measured results obtained after the pressure regulation strategy was carried out. During the period of 2:00-5:00 on Friday morning, the user's water consumption is relatively small and could be neglected, and the total water supply flow is approximately regarded as the leakage flow. Figure 7 shows the changes of pressure and leakage flow after the drop-restore pressure strategy is carried out at 2:00-3:00(Figure 7(a)) and 3:00-4:00(Figure 7(b)) respectively.



(a) 2:00-3:00 *(b) 3:00-4:00*

Figure 7. Variation of pressure and leakage flow with time

It can be seen intuitively from Figure 7 that the drop-restore pressure can rapidly and greatly reduce the leakage flow. When the water pressure slowly rises to 38m, the leakage flow still maintains a low value, and then increases rapidly. This verifies the operation mechanism of the drop-restore pressure: due to the lag effect in viscoelastic properties, the change of the leak area occurs behind the change of water pressure. Combined with Figure 7(b), in the process of water pressure rising slowly, the hysteresis effect is not obvious. The reason is that when a certain status is continued for too long after the pressure reduction, the control effect of the drop-restore pressure control strategy will be reduced.

Figure 8 shows the relationship between inlet pressure and leakage flow during the process of pressure reduction from 45mH$_2$O to 30mH$_2$O and then quickly returning to 45mH$_2$O: under the same condition, the relationship curves in the process of pressure-reduction and pressure boost are not coincidence. And the leakage in the pressure reduction stage is always greater than that in the pressure boost stage. The viscoelasticity theory is used to explain the phenomenon: when the operating pressure changes, the change of the leak area always lags behind the pressure, so under the same conditions, the corresponding leak area of pressure reduction is always larger than that of the pressure boost, that is, the leakage flow in the pressure reduction stage is more than that in the pressure boost stage under the same pressure. This abnormal phenomenon is reproduced in the actual network again, and it is verified that in the actual water supply network, the lag effect caused by the viscoelastic properties of the polymer pipe is still obvious.

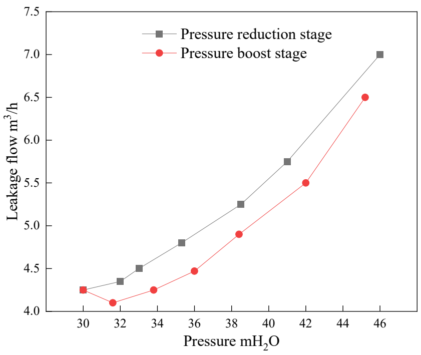
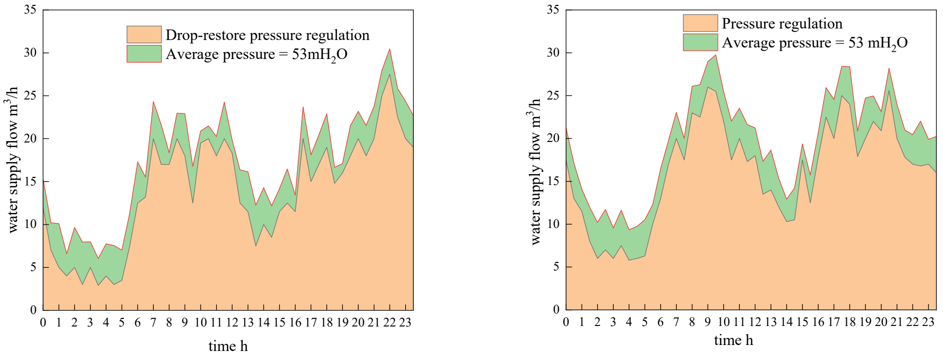


Figure 8. Relationship between pressure and leakage flow under drop-restore pressure operation

Since the small fluctuation of daily water consumption in the Y network, the operation data of Friday and Saturday and the measured data after the implementation of the pressure control strategy represent the water consumption rule of the pipe network, which are used to evaluate the effect of the pressure regulation strategy. Figure 9 shows the changes of the inlet flowmeter before and after the implementation of the pressure regulation scheme on Friday and Saturday.



(a) Friday *(b) Saturday*

Figure 9. Effect verification of pressure regulation

Comparing the changes in water supply flow before and after the pressure regulation, it can be clearly seen that the total water supply flow has decreased significantly, indicating that the drop-restore pressure regulation has a significant effect on leakage level of Y network. And the most obvious reduction in the water supply flow is between 2:00 and 5:00 in the morning. The benefit evaluation after pressure regulation is shown in Table 3.

Table 3. Benefit evaluation after pressure regulation

Stage	Item	Weekday	Weekend
Before pressure regulation (average pressure = 53.00mH$_2$O)	Daily water consumption (m^3)	431.45	485.09
	Daily leakage volume (m^3)	162.58	191.72
	Daily leakage rate (%)	37.7	39.5
After pressure regulation	Daily water consumption (m^3)	323.45	389.09
	Daily leakage volume (m^3)	54.48	95.72
	Daily leakage rate (%)	16.8	24.6
Decrement of daily leakage rate (%)		20.9	14.9
Estimation of monthly water saving (m^3)		2760	

After the pressure regulation, the leakage flow of the Y network has been significantly reduced. The leakage rate on weekdays is reduced from 37.7% to 16.8%, reduced by 20.9% and the weekend leakage rate reduced from 39.5% to 24.6%, reduced by 14.9% and the leakage rate of weekday is further reduced compared with the weekend. The simulation results of the analysis model and the measured pressure at the critical node show that when the inlet pressure is set to 40.00mH$_2$O, the pressure at the critical node can still meet the user's water demand during the peak water consumption period. After adopting the pressure regulation scheme, it is estimated that the monthly water-saving amount can reach 2,760m^3, and it is conservatively estimated that the annual saving of water supply enterprises is more than 100,000 yuan, and the economic benefits brought by pressure management are considerable.

5 CONCLUSION

In this paper, the relationship between the leakage and pressure of polymer pipes is quantitatively described by the Boltzmann superposition principle and the Voigt-Kelvin model, and a pressure-driven viscoelastic leakage model is constructed. Based on the basic information of the Y network, a hydraulic model based on the pressure-driven viscoelastic leakage model is constructed, and the optimal pressure regulation strategy is solved by the particle swarm algorithm with minimizing the leakage amount as the objective function. After the implementation of the pressure regulation strategy, it is estimated that the annual cost of water supply can be saved by 100,000 yuan, which proves that the proposed method of drop-restore pressure leakage control has widely application prospects.

In this paper, the analysis process of the Y network can be used as a reference for formulating pressure control strategies for other networks containing plastic. Drop-restore pressure leakage control is suitable for water supply areas with high water supply pressure (such as close to the water plant) and high leakage rate; Before carrying out pressure management, it is necessary to analyse the basic information of the field network and the pattern of water use to determine whether it is suitable for pressure management, and then conduct dynamic simulation through the hydraulic model to scientifically guide the field pressure control scientifically.

6 REFERENCES

[1] Prescott S, Ulanicki B, Renshaw J. Dynamic Behavior of Water Networks Controlled by Pressure Reducing Valves[C]. CCWI2005-Water Management for the 21st Century, Centre for Water Systems, University of Exeter, Devon, UK, 2005:239–244.

[2] Ulanicki B, Bounds P L M, Rance J P, et al. Open and Closed Loop Pressure Control for Leakage Reduction[J]. Urban Water.2000,2 (2):105-114.

[3] Ferrante M, Massari C, Brunone B, et al. Experimental evidence of hysteresis in the head-discharge relationship for a leak in a polyethylene pipe[J]. Journal of Hydraulic Engineering, 2010, 137(7): 775-780.

[4] Massari C, Ferrante M, Brunone B, et al. Is the leak head–discharge relationship in polyethylene pipes a bijective function?[J]. Journal of Hydraulic Research, 2012, 50(4):409-417.

[5] Jiang Liyue. Shallow discussion on the application of PE pipe in urban water supply network[J]. Shanxi Architecture, 2011,37(30):137-138.

[6] Zheng Chengzhi. Research on leakage discharge analysis model of water distribution network based on algorithm for signal processing[D].Harbin Institute of Technology,2015.

[7] GOOCH J W. Boltzmann superposition principle [M]. Boltzmann superposition principle, 2007.

[8] KAUL S. Chapter 2 - Viscoelastic modeling—passive vibration isolators [M]//KAUL S. Modeling and Analysis of Passive Vibration Isolation Systems. Elsevier. 2021: 27-61.

[9] Wagner J M, Shamir U, Marks D H. Water distribution reliability - simulation methods[J]. Journal of Water Resources Planning and Management-Asce, 1988, 114(3): 276-294.

7 ACKNOWLEDGMENTS

This study is jointly supported by the EU Horizon 2020 Marie Skłodowska-Curie Actions-ITN-IOT4Win under Grant 765921, National Natural Science Foundation of China (51778178, 51978203) and Natural Science Foundation of Heilongjiang Province of China (No.LH2019E044).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14818

STUDY ON IRON DISPERSION LAW AND CONTROL MEASURES OF DEAD-END BRANCH PIPE

Jianxun Chen[1], Jinliang Gao[2], Wenyan Wu[3], Guanghui Wang[4], Qiang Ding[5],

Shihua Qi[6] and Bo Li[7]

[1,2,7] School of Environment, Harbin Institute of Technology, Harbin, China
[3] School of Engineering and the Built Environment, Birmingham City University, Birmingham, UK
[4,5] Capital eco-pro group, Beijing, China
[6] Heilongjiang College of Construction, Harbin, China

[1] *cjianxun877@gmail.com*, [2] *gjl@hit.edu.cn*, [3] *wenyan.wu@bcu.ac.uk*,
[4,5] *18292005003@163.com*, [6] *qishihua@126.com*, [7] *2931907043@qq.com*

Abstract

The paper aims to investigate the dispersion law and mechanism of total iron from dead-end metal branch pipe to main pipe under the impact of hydraulic disturbance. A full scale experimental water distribution network including a transport main pipe as well as six dead-end branch was set up and computational-fluid-dynamics (CFD) simulation are adopted to study the dispersion phenomenon. The cavity flow theory is adopted to clarify the inner dispersion mechanism. Firstly, the CFD model is examined through the comparison of results between the two experimental methods under the same condition. Secondly, performing the effect of the Reynolds number in main pipe on the total dispersion concentration of iron. Thirdly, the characteristics of interaction between flow field and the inner mechanism linked to streamline map and concentration distribution map are analysed based on the concept of cavity flow. The future research content is briefly proposed.

Keywords

Dead-end branch pipe, iron dispersion, full scale distribution setup, computational-fluid-dynamics (CFD), cavity flow theory.

1 INTRODUCTION

The "red water" problem caused by the release of iron in the metal pipes has attracted much attention from users and water supply enterprises. Iron in cast iron pipes is released into water in the form of ions and compounds, which is the main reason for the phenomenon of "red water" [1]. Many scholars have carried out detailed research on the iron release model from different angles, and have achieved certain research results[2-4]. Most of them focus on static experiments, iron release and movement in the single and small diameter pipe. Full-scale model and the interaction influence between the pipe connections are scarce.

The existing water supply network is a complex network composed of looped and branched networks. The looped networks can not only ensure the reliability of water supply, but also improve the liquidity and avoid the generation of stagnant water. The branched pipes is mainly responsible for transferring and connecting users or water facilities. In terms of fluidity, the risk of water quality deterioration caused by branch pipes is significantly higher than that of looped pipes.

There are a large number of dead-end branch pipes in the water network due to the installation of valves, fire hydrants and other accessory equipment, and the water age in those pipes is too long and the dissolved oxygen concentration is significantly reduced, which will aggravate the iron release in the branch pipe wall. Under the condition of hydraulic disturbance and iron concentration gradient, soluble iron in the branch pipe will be released into the adjacent main pipe, which will seriously affect the water quality of the whole water supply system.

The branch pipe and the main pipe in the water supply network are usually connected by a tee fitting, forming a structure similar to a cavity at the connection. Cavity is a classical simplified physical model, which contains a large number of complex hydrodynamic problems, such as flow stability, vortex structure in cavity, coupling mechanism of flow and noise [5]. In water supply network, the connection between the main pipe and the dead-end branch pipe can be regarded as a typical cavity structure, which involves the shear layer instability of the cavity front, the generation and development of vortex, flow separation and other related physical processes. When the water in the main pipe flows through the cavity structure, the shear flow at the mouth of the cavity (that is, the connection between the main pipe and the branch pipe) interacts with the flow in the cavity (the water in the branch pipe), and the shear layer interacts with the cavity wall. Severe pressure, velocity and other pulsations are generated, causing a series of complex flow phenomenon in the cavity[6]. The theory of cavity flow plays an important role in studying the mechanism of iron dispersion in dead-end branch pipes.

Based on cavity flow theory, this manuscript analyses the influence factors and mechanism of iron dispersion from dead-end branch pipe to main pipe by constructing a full-scale water supply pipeline experiment platform and using experimental device and computational fluid dynamics numerical analysis method. Combining with cavity flow theory, some suggestions are put forward for the control of iron dispersion process in branch pipe.

2 MATERIAL AND METHOD

2.1 The Construction of experimental device

The full-scale water distribution system experimental device is shown in Figure1. It was constructed with acrylonitrile butadiene styrene(ABS) pipe and the main diameter are DN200(the main pipe) and DN100(the dead-end branch pipe). The length of main pipe is 30m and branch pipe is 6m. The other main components including two isolated stainless steel water tanks with the same volume of 2.25m³, a centrifugal pump with $Q_{max}=50m^3\cdot h^{-1}$ and $H=32mH_2O$ controlled by a frequency converter, an ultrasonic flowmeter with measuring range from 0 to 100 $m^3\cdot h^{-1}$, a pressure gauge with range from 0 to 0.6 MPa, nine butterfly valves(three of which are 100mm diameter and the rest are 200mm diameter). The device is arranged in a looped shape and all components are connected via flanges in order to facilitate modification and maintenance .

The function of the two tanks are water supply to all device and receive water from device, respectively. The water used during the experimental process is directly from the water tap in the laboratory. In order to control the length of dead-end branch pipe, the branch pipe is divided into six equal parts which are connected with butterfly valve and the length of all the parts is 1m. The length of branch pipe is depended on those valves' status. There are two ball valves are installed at both ends of one branch pipe (shown as in Fig1C) and one is used to exhaust air in the pipe and the other is used to fill the branch pipe with mixing water from the dosing tank placed near the branch pipe. The data of pump speed, flow and pressure are collected by automatic control system, and the collection frequency is 60HZ.

In the course of the experiment, the Reynold number of main pipe is controlled by adjusting frequency converter of pump. The diameter-length ratio(DLR) could be only controlled by open or close the valve in the branch pipe because the diameter of branch pipe is fixed. The concentration of iron solution is determined by controlled the ratio of iron quality and water added into dosing tank. After the completely mixing, the solution is pumped into the branch pipe through the beneath ball valve. When a set of experiments have completed, the full pipe must be flushed until the water is as clear as it was before the experiment. Through the above measure, the effect of residual iron on the experimental results can be reduced as much as possible. During the experiment, water samples were taken by the sampling ball valve at the front of the main pipe and at the branch pipe, and the iron concentration in the water samples were measured by ICP-MS (Inductively Coupled Plasma Mass Spectrometry).

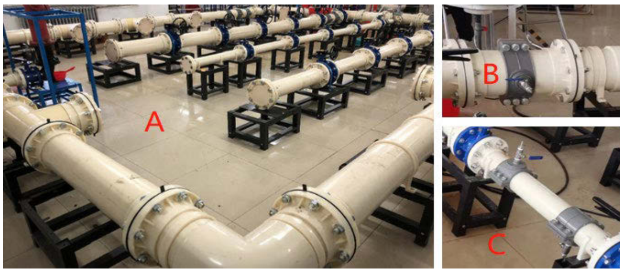


Figure 1. Full-scale water distribution system experimental device

2.2 Simulation scheme setting

Three basic methods are utilised in this experimental research including device experiment, theoretical analysis based on model hypothesis and numerical simulation by solving mathematical equations to study the law of fluid motion. Computational fluid dynamics (CFD) is widely used in fluid research because it can reduce excessive dependence on experimental devices, reduce experimental costs and improve experimental efficiency, significantly. It has been used to the study of leakage characteristics and water quality mixing in junction of water supply network[7,8]. In this paper, the iron dispersion process in dead-end branch pipe is analysed by utilising the advantage of computational fluid dynamics in numerical solution and flow field analysis.

The iron dispersion phenomenon of dead-end branch follows the three conservation laws of mass, energy and momentum. In addition, as the dispersion process is involved, the component transport equation should also be considered. The above four equations can be solved simultaneously by computational fluid dynamics to obtain the distribution of iron in branch pipe and main pipe.

2.2.1 Model construction and parameter setting

In this paper, the tee fitting at the connection between the main pipe and the dead-end branch pipe is taken as the experimental object, and Creo-3D design software is used to establish the calculation model. The flow medium in the pipeline is set as water, and incompressible; A total of 9 geometric models with different branch length were established. The main pipe length, main pipe diameter and branch pipe diameter of all models were set to the same value, which pipe diameter are 1000mm, 200mm and 100mm. And the branch pipe length was set to 100mm, 300mm, 500mm, 700mm, 1000mm, 2000mm, 4000mm, 6000mm and 9000mm, respectively.

The left opening of the main pipe was taken as the inlet boundary, and the boundary condition was set as the velocity inlet. Parameters of the initial velocity, turbulence intensity and hydraulic diameter were set. The direction of velocity was perpendicular to the inlet boundary, and the inlet was defined as no pollutant concentration according to the actual situation. The right opening of the main pipe was taken as the outlet boundary, and the boundary condition was set as the pressure outlet. The inner wall of the pipe is set as non-slip static wall, and the wall is divided into the wall containing pollutants and the wall without pollutants. The initial mass fraction of iron is set on the wall containing pollutants to simulate the reaction between the pipe wall and the components in the bulk water and the continuous transport process of pollutants into the water. The branch pipe was divided into adaptive regions, and the initial concentration of pollutants in the branch pipe was set for local initialization. The model with branch pipe length of 500mm is shown in Figure 2.

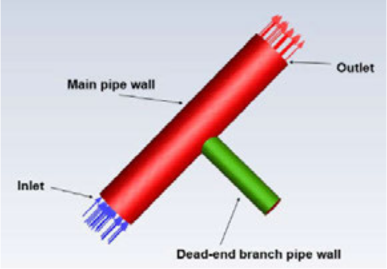


Figure2. CFD model with branch pipe length of 500mm

BY NC SA

In order to solve the changes of flow field, velocity field and concentration field with time, transient analysis was selected, large eddy simulation (LES) was used as the turbulence solution model, and WALE was selected as the sublattice stress model. At the same time, the component transport was set up to build the iron dispersion model in the dead-end branch pipe. In the setting of the transient calculation solver, SIMPLEC algorithm was adopted as the pressure-velocity coupling algorithm, and First Order Implicit was taken as the transient discrete scheme. The spatial discrete scheme was the second-order upwind scheme. After setting the time step, step number and the maximum iterative step number within the unit step size, the iterative calculation was carried out.

2.2.2 Grid partitioning and independence verification

The ICEM software was adopted to grid the whole fluid area. The main flow area, velocity inlet and pressure outlet at the fluid area are encrypted. The inlet wall is encrypted from the left to the middle, while the outlet wall is encrypted in the opposite way. A total of five grid partitioning schemes with different density were set up. Except for differences in grid size and density, the five grids were all divided on the basis of model size corresponding to Figure 2, and other parameters were set the same. Grid size and quantity are shown in Table 1. Large eddy simulation was used to carry out numerical calculation for every grid scheme. The simulated initial total iron concentration in the branch pipe was set as C0=1mol/m^3, the inlet flow rate was 0.8m/s, and the temperature is 300K. The variation of mass fraction of total iron at the main pipe outlet interface with time is shown in Figure 3.

Table 1. Grid partitioning schemes

Schemes No	1	2	3	4	5
Grid quantity	86187	134435	200320	259496	307436

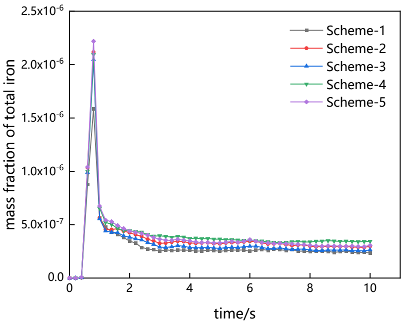


Figure 3. Variation of mass fraction of total iron with time

The variation trends of mass fraction of total iron at the outlet of different grid partitioning schemes are roughly the same, but there is a significant difference between the dispersion amount of peak and stable status for every scheme. The corresponding peak mass fraction of scheme-1 is higher than that of the other four schemes, but there is little gap with the four schemes after stability. It is speculated that the difference in the peak value is the main reason. In the stable stage of outlet mass fraction, there is no significant difference in the variation trend and the value of mass fraction between scheme-2 and scheme-5, and the number of grids are 134435 and 307436 respectively. After comprehensive consideration of calculation accuracy and efficiency, scheme-2 was selected.

2.3 Mathematical model verification

The accuracy verification of the CFD numerical model is the premise to determine whether it can be used in subsequent research. It is an effective method to prove the reliability of the model by comparing the results of the experimental device and the numerical model under the same condition. The mass fraction of total iron at the main pipe outlet was selected as the validation index. If the absolute error between the two results was within 20%, it proved that the numerical model had certain accuracy and could be used in subsequent experiments. Otherwise, the model parameters need to be adjusted continuously until the error is less than 20%. This paper mainly studied the influence of three factors on iron dispersion in the dead-end branch pipe, so it is necessary to compare the index differences between two results under three different research variants to ensure the accuracy and reliability of the numerical model.

A total of 9 schemes are designed, as shown in Table 2. Among them, the influence of diameter-length ratio on iron dispersion was analysed in Row 1-4, the influence of Reynolds number was analysed in Row 5-7, and the influence of iron concentration in branch pipe was analysed in the other two schemes.

Table 2. Schemes of mathematical model verification

Condition num	Length of branch pipe (mm)	Diameter of branch pipe (mm)	Ratio of diameter-length	Re of incoming flow	Initital iron concentration (mol/m^3)
1	1,000	100	0.100	20,000	8
2	1,000	100	0.100	28,000	8
3	1,000	100	0.100	32,000	8
4	1,000	100	0.100	20,000	1
5	1,000	100	0.100	20,000	2
6	1,000	100	0.100	20,000	4
7	2,000	100	0.050	20,000	8
8	4,000	100	0.025	20,000	8
9	6,000	100	0.017	20,000	8

2.4 Experimental scheme design

Based on the verified numerical model and its advantages in flow field analysis, several more detailed experimental scheme was designed to explore the mechanism of iron dispersion. Those schemes are shown in Table 3.

In all schemes, the diameter and length of the main pipe and the diameter of the dead-end branch pipe are kept 200mm, 1,000mm, 100mm respectively. The first nine schemes were designed to study the effect of DLR which is changed by adjusting the length of the branch pipe, the total iron concentration in the branch pipe was kept 1.0mol/m^3, and the main pipe Reynolds number was set as 200,000. The eight schemes (shown as Table 3) were designed to study the effect of main pipe Reynolds number by adjusting the flow rate of the main pipe. And the total iron concentration of the branch pipe and the length of the branch pipe are set 1.0mol/m^3 and 500mm.

Table 3. Experimental scheme design

Scheme num	Length of branch pipe/mm	Ratio of diameter-length	Concentration of total iron in branch pipe(mol/m^3)	Re of main pipe
1	500	0.200	1.0	2,000
2	500	0.200	1.0	6,000
3	500	0.200	1.0	12,000
4	500	0.200	1.0	20,000
5	500	0.200	1.0	80,000
6	500	0.200	1.0	160,000
7	500	0.200	1.0	240,000
8	500	0.200	1.0	360,000

3 RESULTS AND DISCUSSION

3.1 Model verification

The accuracy of the numerical model was studied according to the experimental scheme in Table 2. Due to the disturbance of water in the main pipe, iron dispersion in the dead-end branch pipe is a continuous process, and it takes a period of time to reach a stable status. In view of this, in the numerical model, the total iron concentration of the main pipe outlet selected the stabilized concentration; In the device experiment, water samples were taken from the downstream sampling location of the main pipe at an interval of 5 seconds, and the total iron concentration of outlet after stabilization was seen as the total iron concentration.

Figure 4 shows the total iron mass fraction and absolute error of main pipe outlet after stability obtained from numerical model and device experiment. Among the 9 experimental schemes, the total iron concentration obtained from numerical model is slightly higher than that of the device experiment. It is speculated that it is due to the disturbance of iron concentration in the dead-end branch pipe caused by the sudden opening of the dead-end branch pipe valve during the experimental operation. And the irregular shape of the fitting connection also affects the iron dispersion process. The absolute errors of the nine experiments are all less than 20%, which meet the numerical model validation conditions set in this paper, proving the applicability of parameter setting to study the effects of three factors on iron dispersion. The subsequent studies were carried out on the basis of the parameter setting.

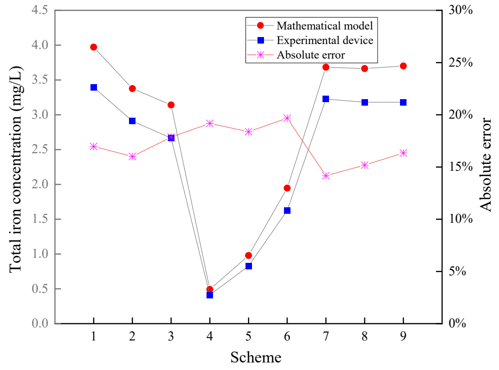
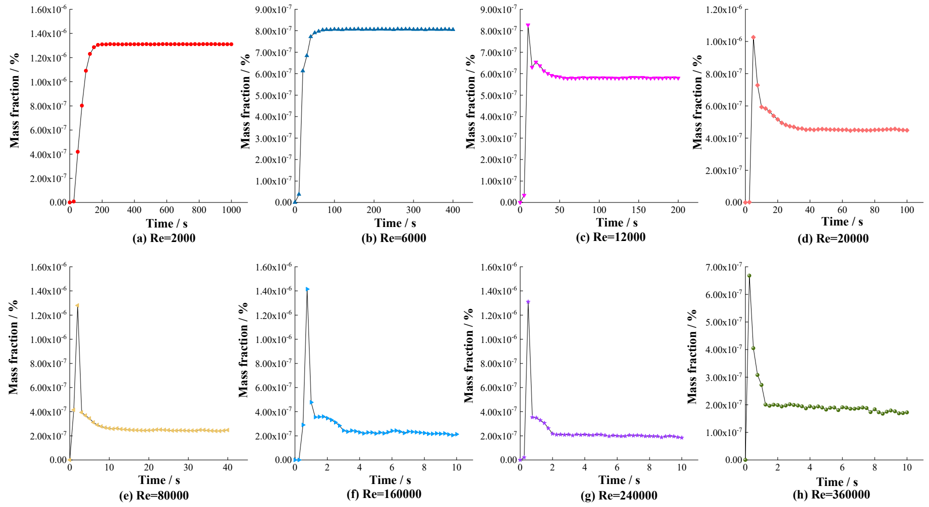


Figure 4. Comparison for numerical model and device experiment result

3.2 The effect of Reynolds number on iron dispersion

The variation rule of total iron mass fraction at the outlet with time under eight schemes obtained by large eddy simulation analysis is shown in Figure 5. Due to different flow rates, the time for dispersion to stabilization is also different. The stabilization time decreases from 200s to 2s when Reynolds number changes from 2000 to 360,000. And the larger the Reynolds number, the shorter the time required for the stabilization of the dispersion.



(a) Re=2000
(b) Re=6000
(c) Re=12000
(d) Re=20000
(e) Re=80000
(f) Re=160000
(g) Re=240000
(h) Re=360000

Figure 5. Variation of total iron mass fraction with time under different Reynolds number

Under the schemes of Re=2,000 and Re=6,000, the total iron mass fraction at the outlet increased firstly and then tended to be stable. When Re > 12,000, the fraction increased to the peak at first, then decreased gradually and maintained in a stable range. The scheme of Re=12,000 can be seen as a transition state from low Reynolds number to high Reynolds number. The main reasons for this phenomenon are as follows: when the water flow of the main pipe reaches the connection of the branch pipe under large Reynolds number, the main water containing pollutants in the branch pipe will be greatly disturbed. At this time, a large number of pollutants are included in the water of

BY NC SA

the main pipe, and this part of water reaches the outlet of the main pipe, resulting in the peak value of the mass fraction of the total iron at the outlet. In the stable stage, only the outlet mass fraction of Re = 160,000, 240,000 and 360,000 fluctuated slightly, while the rest conditions remained constant.

The variation of peak value and stable value with Reynolds number is shown in Figure 6. When Re < 160,000, the maximum value of the total iron mass fraction at main pipe outlet increases gradually with the increase of Reynolds number. When Re = 160,000, the outlet mass fraction reaches the maximum value of 1.4×10^{-6}. When Re > 160,000, the variation of maximum value of the mass fraction of total iron mass at export showed an opposite trend. When Re < 60,000, the Reynolds number has a significant effect on the outlet mass fraction. However, when Re > 60,000, the increase of Reynolds number has little effect on the outlet mass fraction at the outlet, and it always maintains around 2×10^{-7}. It can be considered that under the condition of high Reynolds number, the change of Reynolds number has little influence on the total iron mass fraction at the outlet. The stable values and maximum values corresponding are equal for Re = 2000 and Re = 6000 respectively. While the maximum values of other Reynolds number are higher than the stable values. In order to facilitate subsequent analysis, Re = 2,000 and Re = 6,000 are referred to as low Reynolds number condition, Re = 12,000 is referred to as the transition status from low Reynolds number to high Reynolds number conditions, and Re > 12,000 is referred to as high Reynolds number conditions.

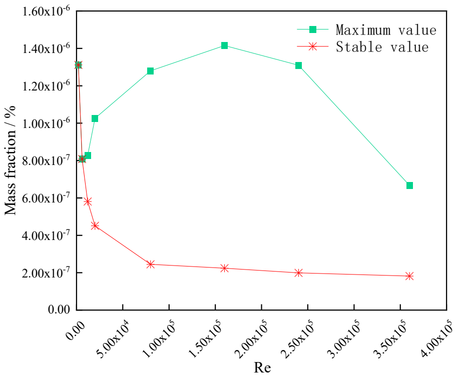


Figure 6. The maximum and stable total iron mass fraction under different Reynolds number

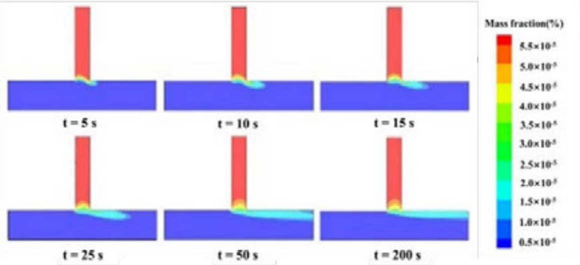
In order to explore the influence of different Reynolds numbers on iron dispersion, the maps of concentration distribution at three representative stages which including low Reynolds number stage (Re = 2,000), transition stage (Re = 12,000) and high Reynolds number stage (Re = 240,000), are selected for analysis, as shown in Figure 7.

It can be observed from Figure 7(a) that at the initial 5s of the flow, the concentration boundary at the junction of the branch pipe and the main pipe expands to the end of the branch pipe and the downstream of the main pipe, respectively. At 10s, more and more iron in the cavity diffuse from the downstream end of the cavity to the main pipe, and the concentration boundary in the branch pipe continues to expand to the end of the branch pipe. At 15s and 25s, the iron initially diffused into the main pipe continue to move to the outlet, and the flow state and concentration boundary in the dead-end branch pipe remain stable; At 50s, the initially diffused iron reached the outlet, and the mass fraction at the outlet began to gradually increase with time; At 200s, the flow state in main pipe had been stabilized, and the mass fraction at the outlet have no change.

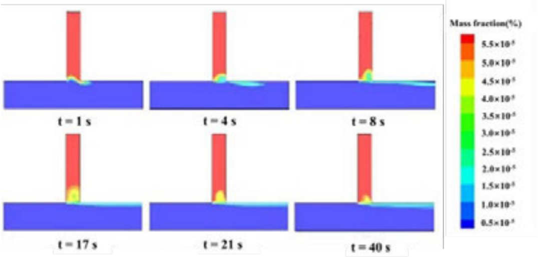
Figure 7(b) is the map of the concentration distribution in the transition stage (Re = 12,000). There are obvious differences between Re = 12,000 and Re = 2,000 in terms of the concentration peak and the dispersion state. In the initial stage, i.e. at t=1s, the main body water of the main pipe disturbs the distribution of iron in the branch pipes, which is more severe than that of the low Reynolds number, and the occurrence time is earlier. At 4s, due to the initial fluctuation taking away part of the pollutants, a vortex with concentration gradient is generated downstream of the cavity. The iron diffused in the initial stage moves to the outlet of the main pipe, and the dispersion amount of iron gradually decreases in at the later stage; At 8s, the iron generated by the strong disturbance reaches the outlet, forming a concentration peak, and the concentration boundary in the dead-end branch pipe continuously extends to the end of the branch pipe. At 17s, the low-concentration boundary at the end of the branch pipe extends to the maximum, but the iron concentration in the boundary center gradually increases, and the iron flowing out from the downstream end of the cavity tend to be stable along the pipe wall. At 21 s and 40 s, the concentration boundary in the branch pipe gradually decreases, but the concentration gradually increases. At this time, the concentration distribution change and the pollutant dispersion amount are relatively stable for both main and branch pipe. And this is called the stable dispersion stage.

Figure 7(c) shows the map of the concentration distribution under high Reynolds number (Re = 240,000). The concentration distribution in the high Reynolds number stage changes obviously. At 0.1s, a severe disturbance occurred at the connection between the branch pipe and the main pipe. Under the action of such a fluctuation, a relatively large number of iron were diffused from the back edge of the cavity, and the iron at the boundary flowed out, forming an upward convex low concentration zone in the branch pipe. At 0.2s, the low concentration band formed at the initial stage gradually became larger and extended to the end of the branch pipe, and the fraction of the band central was at a lower value. The initially diffused pollutants moved along the pipe wall to the main pipe outlet along with the main body water of the main pipe. At 0.5s, the low concentration band in the branch pipe moved along the branch pipe wall to the end of the branch pipe, and the band range continued to increase. At this time, the iron diffused in the initial stage reached the outlet of the main pipe, which corresponds to the peak concentration in Figure 6. At 0.7s and 1.6s, the amount of iron flowing out from the branch pipe to the main pipe has been stabilized. It can be clearly observed that the content of pollutants corresponding to the pipe wall of the main pipe downstream of the cavity is very small, and the mass fraction of the outlet is less than 1.5×10^{-5}. The low concentration zone in the branch pipe extended anticlockwise to the middle of the branch pipe and continued to extend. At 2.5s, the low concentration zone in the branch pipe splited into two concentration zones, and the low concentration gradient in the center of the concentration zone almost disappears. At this moment, the concentration of pollutants close to the main pipe wall remains constant and is in a stable stage.

(a) Re = 2,000



(b) Re = 12,000



(c) Re = 240,000

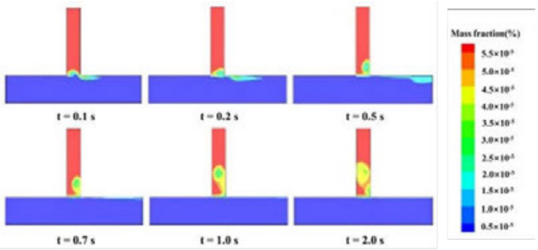


Figure 7. Concentration distribution map under different Reynolds number

3.3 Mechanism analysis

From the above analysis, DLR, iron concentration in branch pipe and incoming Reynolds number are three influencing factors for iron dispersion in dead-end branch. The main influence on the flow field characteristics is the incoming Reynolds number. Under different Reynolds numbers, the flow field characteristics of dead-end branch pipe are mainly divided into low Reynolds number stage and high Reynolds number stage. In this paper, two representative conditions are selected for flow field characteristics analysis, in which Re = 6,000 at low Reynolds number stage and Re = 160,000 at high Reynolds number stage.

3.3.1 Lower Reynold

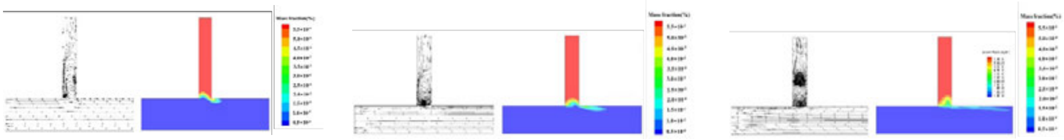
In order to analyse the relationship among the dispersion amount, the concentration distribution and the characteristics of the flow field, the instantaneous flow field analysis was carried out by combining the instantaneous streamline diagram and the corresponding concentration distribution map. Figure 8 shows the streamlines and concentration distributions at low Reynolds numbers of 2s, 6s, and 16s.

As can be seen from Figure 8 (a) that at the initial stage of flow (t=2s), when the water flow of the main pipe passes through the dead-end branch pipe, the boundary layer is lifted, and the streamline diagram is shown as an upward convex curve. It can be seen from the track and direction of the streamline that water in main pipe enters the branch pipe. The free shear layer colliders with the back wall of the branch at the middle of the branch pipe, resulting in the separation of the shear layer. A part of the separated shear layer flows back to the front edge of the cavity and forms a small vortex structure on the front edge of the cavity and the back wall of the cavity, and the other part leaves the dead-end branch from the rear edge of the cavity. Affected by the uplift of the boundary layer, the iron in the branch pipe diffuse into the main pipe.

Figure 8(b) shows the streamline diagram and concentration distribution diagram at t=6s. At this time, the large vortex developed in the branch pipe stably exists in the middle of the cavity, and the streamline state at the wall of the main pipe downstream of the branch pipe is stable. There is only one large backflow area in the branch pipe. From the concentration distribution map, it can be seen that the concentration gradient range at the front end of the branch pipe caused by the backflow area of the branch pipe gradually extends. The concentration gradient area in the main pipe gradually widens and lengthens.

The streamline and concentration distribution diagram at t=16s are shown in Figure 8(c). At this time, the cavity flow field has reached a steady state in the dead-end branch under the low Reynolds number state. There are two vortices stably in the dead end branch pipe, the vortex near the main pipe is the main vortex, and the vortex direction is counterclockwise. From the map of the concentration distribution, it can be seen that the distribution of the entire concentration gradient is highly coincident with the main vortex, indicating that the dispersion of iron is closely related to the main vortex. There is also a secondary vortex in the depth of the dead-end branch, which is driven by the main vortex and generated in a clockwise direction, so that the iron concentration in the depth of the dead-end branch pipe can be transported to the outlet of the branch. At this time, the concentration gradient range in the branch pipe continues to intensify, and the concentration gradient range in the main pipe becomes longer and narrower.

Based on the above analysis, it can be concluded that the flow field at low Reynolds number is relatively stable, the boundary layer is deeply depressed towards the end of the branch pipe. And the free shear layer collides with the back wall of the branch pipe, which affects the vortex structure of the flow field. The concentration distribution is closely related to the change of the vortex structure. The change of the streamline map also directly reflects the change of the concentration distribution map. The final concentration gradient range coincides with the height of the main vortex, and the main vortex never breaks away from the shear layer in position, indicating that the amount of iron dispersion is closely related to the main vortex.

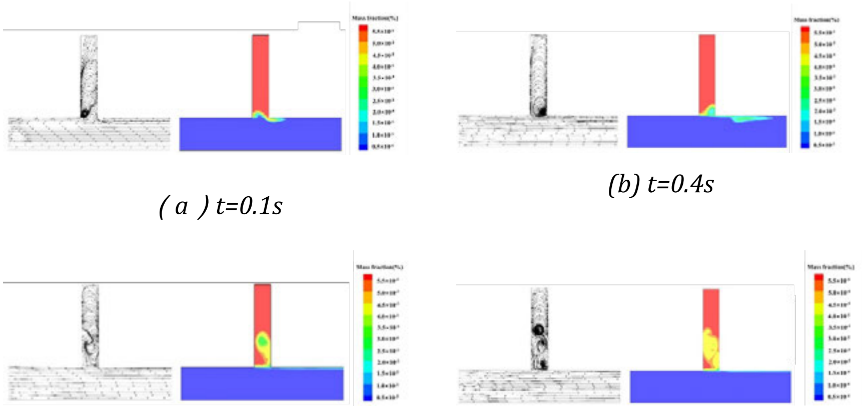


(a) *t* = 2*s* (b) *t* = 6*s* (c) *t* = 16*s*

Figure 8. streamline map and concentration distribution map under different time at lower Re

3.3.2 Higher Reynold

The instantaneous flow diagram and the corresponding concentration distribution map are used to analyse the instantaneous flow field at the stage of high Reynolds number. By comparing the flow field characteristics at low Reynolds number, the relationship between dispersion trend, concentration distribution and flow field characteristics is summarized. The streamline diagrams and concentration distribution map at the Reynolds number of 0.1s, 0.4s, 2.1s and 4s are shown as Figure 9.



(a) t=0.1s

(b) t=0.4s

(c) t=2.1s

(d) t=4s

Figure 9. Streamline map and concentration distribution map under different time at higher Re

By analysing the streamline diagram and concentration distribution map under high Reynolds number, it can be found that the flow field under high Reynolds number is not as stable as that at low Reynolds number. A significant difference occurred when the main vortex collided with the trailing edge of the cavity and moved away from the boundary to the depth of the branch. The splitting of vortices and the generation of new vortices occur during the movement. And finally form two secondary vortices and a primary vortex at the trailing edge of the cavity.

The overall flow field is characterized by the formation and development of vortexes in the shear layer, the collision and shedding of vortexes with the back edge of the cavity, the splitting of vortexes and the generation of new vortexes, and finally the three vortexes form a set of stable vortex structures.

The iron dispersion trend and concentration distribution in the dead-end branch pipe are closely related to the development process of the vortex structure, which is embodied in the form of low concentration range in the concentration distribution map.

The vortex development process includes the formation, development, collision and shedding of the low concentration range, and the final concentration range and center concentration increase with the movement and splitting of the vortex structure.

4 CONCLUSION

In this paper, the iron dispersion law of the connection between the main pipe and the dead-end branch pipe is analysed and the dispersion mechanism was studied on the basis of the concept of cavity. The mathematical model of dead-end branch iron dispersion was established by combining computational fluid dynamics numerical simulation with device experiments, and an experimental platform for dead-end branch iron dispersion was designed and built to verify the accuracy of the numerical model. The law of iron dispersion in dead-end branch pipe under different Reynolds number of inlet flow was analysed, and the iron dispersion mechanism at the connection between dead-end branch pipe and main pipe was clarified.

The numerical model was used to simulate the iron dispersion in the dead-end branch pipe under different Reynolds numbers of incoming flow. In the low Reynolds number stage, the dispersion amount first increased and then stabilized, and the dispersion amount of iron was large. The high Reynolds number stage has less dispersion after stabilization.

Combined with the theory of cavity flow, the flow field characteristics of the iron dispersion in the dead-end branch pipe are studied. At the low Reynolds number stage, the flow field characteristics were concluded as follows: the formation, development and shedding of the leading edge vortex of the cavity, which stabilized in the middle of the cavity, and finally a second-order vortex was generated along the depth of the dead-end branch pipe. The high Reynolds number stage is characterized by the formation and development of vortices in the shear layer, the collision and shedding, the splitting of vortices and the generation of new vortices, and finally the three vortices form a group of stable vortices structure. The concentration distribution is highly coincident with the vortex structure, including the formation, development, collision, shedding of the low concentration range, the final concentration range and the central concentration increasing with the movement and splitting of the vortex structure.

Future studies will combine the flow control method of the cavity to explore the dispersion mode, control mechanism and the controllable effect of iron dispersion in the blind-end branch pipe by adding the new fitting in main pipe.

5 REFERENCES

[1] SETH A, BACHMANN R, BOXALL J, et al. Characterisation of materials causing discolouration in potable water systems [J]. Water Science and Technology, 2004, 49(2): 27-32.

[2] BENSON A S, DIETRICH A M, GALLAGHER D L. Evaluation of Iron Release Models for Water Distribution Systems [J]. Critical Reviews in Environmental Science and Technology, 2012, 42(1): 44-97.

[3] Shahra Essa and Wenyan Wu (2020) Water contaminants detection using sensor placement approach in smart water networks, Journal of Ambient Intelligence and Humanized Computing (2020), https://doi.org/10.1007/s12652-020-02262-x, Springer 25 June 2020.

[4] Jinliang Gao, Shiyuan Hu, Wenyan Wu, et al. Experiment and Simulation of Ferrous Ions Diffusion at the Dead-end Branch Pipes of Water Distribution System, 1st WDSA/CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25, 2018.

[5] SHAMEKHI A, SADEGHY K. Cavity flow simulation of Carreau-Yasuda non-Newtonian fluids using PIM meshfree method [J]. Applied Mathematical Modelling, 2009, 33(11): 4131-45.

[6] DE ROECK W, RUBIO G, BAELMANS M, et al. Toward accurate hybrid prediction techniques for cavity flow noise applications [J]. International Journal for Numerical Methods in Fluids, 2009, 61(12): 1363-87.

[7] CASSA A M, VAN ZYL J E. Predicting the head-leakage slope of cracks in pipes subject to elastic deformations [J]. Journal of Water Supply Research and Technology-Aqua, 2013, 62(4): 214-23.

[8] Webb, S. W., and B. G. Van Bloemen Waanders. 2007. "High fidelity computational fluid dynamics for mixing in water distribution systems." In Proc., 8th Annual Water Distribution Systems Analysis Symp. 2006, 154. Reston, VA: ASCE

6 ACKNOWLEDGMENTS

This study is jointly supported by the EU Horizon 2020 Marie Skłodowska-Curie Actions-ITN-IOT4Win under Grant 765921, National Natural Science Foundation of China (51778178, 51978203) and Natural Science Foundation of Heilongjiang Province of China (No.LH2019E044).

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14819

ANALYSIS OF FRICTION MODELS DURING SIMULATIONS OF FILLING PROCESSES IN SINGLE PIPELINES

Oscar E. Coronado-Hernández[1], Vicente S. Fuertes-Miquel[2], Pedro L. Iglesias-Rey[3], Moshen Besharat[4], Humberto Ávila[5] and Helena M. Ramos[6]

[1]Facultad de Ingeniería, Universidad Tecnológica de Bolívar, Cartagena 131001, Colombia.
[2,3]Departamento de Ingeniería Hidráulica y Medio Ambiente, Universitat Politècnica de València, Valencia 46022, Spain.
[4] School of Engineering, Arts, Science and Technology, University of Suffolk at Suffolk New College, Ipswich IP4 1QJ, UK
[5] Institute of Hydraulic and Environmental Studies (IDEHA), Dept. of Civil and Environmental Engineering, Universidad del Norte, Barranquilla 081007, Colombia.
[6]Department of Civil Engineering and Architecture, CERIS, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, Portugal.

[1] *ocoronado@utb.edu.co*, [2] *vfuertes@upv.es*, [3] *piglesia@upv.es*, [4] *m.besharat@uos.ac.uk.*,
[5] *havila@uninorte.edu.co* [6] *hr@civil.ist.utl.pt*

Abstract

The analysis of filling processes in pressurized pipelines has been conducted using a steady friction model in the implementation of governing equations. This research is focused on the case of a filling process of a single pipeline without an air valve. Three equations were used to represent the phenomenon: (i) a rigid water column approach, which describes the water movement along the water system; (ii) a piston flow model, which assumes a perpendicular air-water interface to the main direction of the pipe; and (iii) a polytropic model for representing the thermodynamic behaviour of an entrapped air pocket. This research studies the filling processes occurrence using equations of steady and unsteady friction models, where Moody, Wood, Hazen-Williams, and Swamee-Jain equations are analysed. The analysis is applied to a case study of a single pipe of a total length of 1000 m with an internal diameter of 595 mm variation of pressure surges in the implementation of these formulations. Results confirm that there is a minimum discrepancy between steady and unsteady friction models since values of pressure surges pattern are similar.

Keywords
Filling process, unsteady, friction models, transient flow.

1 INTRODUCTION

Hydraulic installations should be designed to guarantee reliable infrastructures due to their high costs. Therefore, engineers should select pipe resistance and stiffness class to protect water pipelines of extreme absolute pressure [1].

In the last century, water hammer effects have been investigated due to pipe collapses occurrence. At the beginning, the analysis focused only on the water phase involving the operation of valve closure manoeuvring and pump's stoppages. These hydraulic events have been extensively studied both numerically and experimentally using commercial packages.

Regulations transient flow in pressurized hydraulic systems are considered in many countries around the world. In this sense, engineers and designers are implementing the developed knowledge in the field to have reliable designs [2].

In the last decades, the transient flow with entrapped air is being analysed since air pockets can produce higher extreme absolute pressure patterns compared to transient flow only for the water phase [1, 3]. This phenomenon is generated because air elasticity is higher than water and pipe elasticity. When an air pocket volume is reduced for the compression of the water phase, then the air pocket pressure is increased and vice versa [1]. Transient flow with entrapped air involves complex formulations that should be modelled to represent both hydraulic and thermodynamic behaviour of the water and air phase, respectively. Several formulations have been implemented to simulate filling processes in water pipelines. Zhou et al. (2002) [4] used an elastic water column model to simulate the behaviour of water over time considering a steady friction factor considering for filling process. Fuertes et al. (2019) [5] and Coronado-Hernández et al. (2018) [6] developed a mathematical model to predict filling and emptying operations in pressurized pipelines considering a steady friction factor. Also, filling and emptying processes have been analyzed using Computational Fluids Dynamics (CFD). The implementation of this kind of models have been used to determine the effect of backflow air in emptying operation [7], and secondary pressure peaks for uncontrolled filling operation [8].

Recently, Zhou et al. (2020) [9] introduced an unsteady friction factor (UFM) using an elastic water column model for filling operations. Coronado-Hernández et al. (2021) [10] implemented an UFM for analysing the emptying operation using a rigid water column model in the water phase.

This research focuses on the implementation of unsteady friction model for predicting filling operations using a rigid water column model. The current analysis is presented for a water pipeline without admitted air. The Brunone model [11] was included to simulate the friction factor [11]. The Moody [12], Wood [13], Hazen-Williams [14], and Swamee-Jain [15] were used to analyse a steady friction factor. The mathematical model was validated using an experimental facility with a pipe length of 3.867 m. The mathematical model using an unsteady friction factor was applied to a practical application to observe air pocket pressure, water velocity, and length of a water column variations.

2 MATHEMATICAL MODEL

2.1 Equations

- Mass oscillation equation: equation (1) represents the behaviour of a water column along of an entire pipe system.

$$\frac{dv_f}{dt}=\frac{p_0^*-p_1^*}{\rho_w L_f}+g\frac{\Delta z_1}{L_f}-gJ-\frac{R_v g A^2 v_f\left|v_f\right|}{L_f} \tag{1}$$

Where v_f = water filling velocity (m/s); p_0^* = initial pressure supplied by a tank or a pump (Pa); p_1^* = air pocket pressure (Pa); ρ_w = water density for a specific environmental temperature; Δz_1= difference elevation of ends of a pipe installation; g = gravitational acceleration (m/s^2); L_f= length of a water filling column; J = unsteady friction losses per unit length (m/m); R_v = resistance coefficient of a valve (s^2/m^5); and A= cross-sectional area of a single pipe.

The unsteady friction losses term is expressed as:

$$J=f\frac{v_f\left|v_f\right|}{2gD}+\frac{k_\delta}{g}\frac{dv_f}{dt} \tag{2}$$

Where f = friction factor; D = internal pipe diameter; and k_δ = Brunnone friction coefficient.

The left term in equation (2) represents the steady friction losses per unit length; in constant, the right term simulates the variation over time of friction losses.

The Brunnone coefficient can be computed using the formula as follows:

$$k_{\delta} = \frac{\sqrt{C^{*}}}{2} \tag{3}$$

The Vardy's shear decay coefficient (C^{*}) is calculated depending on the flow regime: for a laminar flow (Re < 2000) its value is 0.00476; and calculation for a turbulent flow (Re > 4000) is computed using equation (4).

$$C^{*} = \frac{7.41}{\mathrm{Re}^{\log(14.3/\mathrm{Re}^{0.05})}} \tag{4}$$

Where Re = Reynolds number.

By plugging equations (2), (3), and (4) into equation (1):

$$\frac{dv_f}{dt} = \frac{\frac{p_0^{*} - p_1^{*}}{\rho_w L_f} + g\frac{\Delta z_1}{L_f} - \frac{f v_f |v_f|}{2D} - \frac{R_v g A^2 v_f |v_f|}{L_f}}{1 + \omega \frac{\sqrt{\frac{7.41}{\mathrm{Re}^{\log(14.3/\mathrm{Re}^{0.05})}}}}{2}} \tag{5}$$

The coefficient (ω) is introduced to represent either SFM or UFM. A value ω = 0 implies a SFM, while a value ω = 1 represents simulation using an UFM.

- Piston-flow model: a perpendicular air-water interface was considered for simulating a filling operation, which is modelled using Equation (6).

$$\frac{dL_f}{dt} = v_f \tag{6}$$

- Polytropic model: this formulation describes how the air pocket size changes with air pocket pressure pulses. These changes are described by the polytropic law (see Equation (7)) which is applied without injected air into hydraulic installations.

$$p_1^{*} x^{k} = p_{1,0}^{*} x_0^{k} = \text{constant} \tag{7}$$

Where $\boldsymbol{x}$ = air pocket size (m) and $\boldsymbol{k}$= polytropic coefficient.

The polytropic coefficient describes a type of thermodynamic evolution. A value of 1.0 indicates an isothermal process, while a value of 1.4 simulates an adiabatic evolution. Intermediate values consider a polytropic behaviour.

2.2 Friction factor formulations

The friction factor is a dimensionless parameter, which is used in fluid dynamics to compute head loss (or pressure loss) due to friction in pipeline installations. Table 1 shows different formulations that have been used to compute the friction factor considering various boundaries of Reynolds number, relative roughness, water filling velocity, regime flow, and internal pipe diameter [16, 17]. Equations from (4) to (10) are used to evaluate the sensitivity of the filling model described in section 2.1.

Table 1. Friction factor formulations

Author	**Equation**	**Equation number**	**Application range**
Laminar flow (Re < 2000)			
Hagen-Poiseuille	$f = \frac{64}{\mathrm{Re}}$	(8)	-
Turbulent flow (Re > 4000)			
Swamee-Jain	$f = \frac{0.25}{\left[\log\left(\frac{k_s}{3.7D} + \frac{5.74}{\mathrm{Re}^{0.9}}\right)\right]^2}$	(9)	Transition zone ($10^{-6}<k_s/D<2\text{x}10^{-2}$; $3\text{x}10^{3}<\mathrm{Re}<3\text{x}10^{8}$)
Moody	$f = 0.0055\left[1 + \left(20000\frac{k_s}{D} + \frac{10^6}{\mathrm{Re}}\right)^{1/3}\right]$	(10)	$k_s/D>0.01$
Wood	$f = 0.094\left(\frac{k_s}{D}\right)^{0.225} + 0.53\left(\frac{k_s}{D}\right) + 88\left(\frac{k_s}{D}\right)^{0.44}\mathrm{Re}^{-1.62\left(\frac{k_s}{D}\right)^{0.134}}$	(11)	*Re>10000;* $10^{-5}<k_s/D<0.04$
Hazen-Williams	$f = \frac{133.89}{C_{HW}^{1.851}D^{0.017}v_f^{0.15}\mathrm{Re}^{0.15}}$	(12)	D>75 mm; $v_f<3$ m/s

Where k_s is the absolute roughness, and C_{HW} is the Hazen-Williams coefficient.

3 VALIDATION MODEL

An experimental facility was configured (see Figure 1) at the Hydraulic Lab at the University of Lisbon (Lisbon, Portugal) to check the governing equations presented in Section 2.1. The installation is composed of a hydro-pneumatic tank to supply various initial pressure, a 7.6-m-long PVC pipe, a pressure transducer located at the highest point of the installation, a manual valve (MV_1) to isolate the pipeline from the hydro-pneumatic tank, and four electro-pneumatic valves (BV_1 to BV_4). BV_1 and BV_2 remain opened during the filling phenomenon occurrence. Air pockets are injected in the highest point of system, where the pressure transducer is located. An initial air pocket size (x_0) of 0.517 m is configured for all experimental tests. The right water column remains constant during experiments (named as blocking water column). A 3.867-m-long total left branch pipe is used to produce a filling operation, which includes lengths of the inclined and vertical branch pipes (see Figure 1). The filling process begins with the opening of the BV_4. A synthetic manoeuvring of BV_4 was introduced considering an opening time of 0.2 s (given by the manufacturer) with a resistance coefficient (R_v) for a full opening of $1.7\text{x}10^5$ ms²/m⁶. Immediately, the filling water column (left water column) starts to fill the hydraulic system while a compression of the injected air pocket is provoked. Valve BV_3 is closed to produce a rapid compression of the entrapped air. The blocking water column acts as a boundary condition, producing a behaviour like a single pipeline. In this sense, the filling process of this installation is simulated using equations (1), (2), and (3). The hydro-pneumatic tank was configured using an initial gauge pressure (H_t) of 1.25 bar. Values of p_0^* were computed as the sum of H_t and atmospheric pressure.

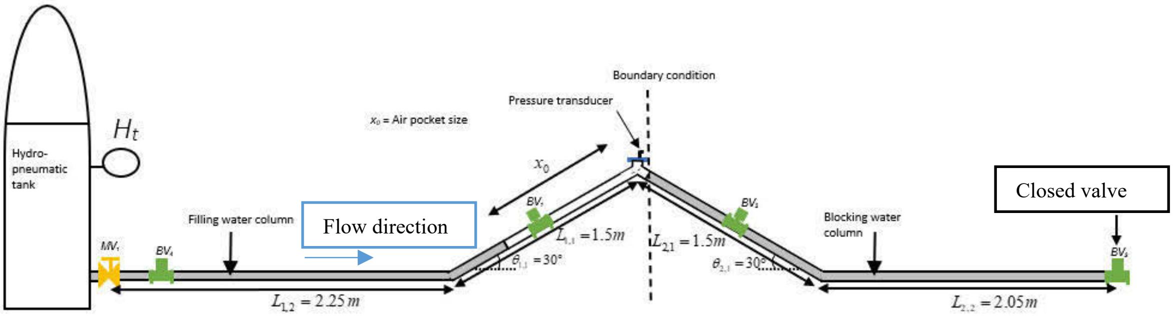

Hydro-pneumatic tank
H_t
MV_1
BV_4
Filling water column
Flow direction
x_0 = Air pocket size
x_0
BV_2
$L_{1,1} = 1.5m$
$\theta_{1,1} = 30°$
Pressure transducer
Boundary condition
$L_{2,1} = 1.5m$
BV_3
$\theta_{2,1} = 30°$
Blocking water column
Closed valve
BV_3
$L_{1,2} = 2.25m$
$L_{2,2} = 2.05m$

Figure 1. Experimental facility

Unsteady friction models are more adequate than steady friction models to represent filling operations with entrapped air since involves a variable friction factor during transient event. The Brunnone friction coefficient was used to evaluate this evolution. In addition, the Swamee-Jain equation (see equation (9)) was used to calculate the variation of the friction factor in steady flow. This equation was utilized considering that results are similar compared to the Colebrook-White formula (based on a physical formulation). Reynold numbers ranging from 0 to 2000 (laminar flow) were modelled using equation (8), while the remaining conditions of Reynold numbers were simulated using the Swamee-Jain formulation, since is suitable for values located at critical and turbulent zone flow.

The polytropic coefficient (k) was calibrated for three kinds of evolution: (i) an isothermal process (k = 1.0), (ii) an intermediate process (k = 1.2), and (iii) an adiabatic process (k = 1.4). Figure 2 presents a comparison of the mentioned polytropic coefficients the analysed run, where the best result is obtained using an isothermal evolution since the mathematical model can follow fluctuations of the average experimental. Using a polytropic coefficient of k = 1.0 (isothermal process), the peak of air pocket pressure head is 50.00 m; while for intermediate and adiabatic processes, values of 48.14 and 47.19 mm are reached, respectively. Therefore, a polytropic coefficient of k = 1.0 was selected for the analysis, which the pressure head peak is compared with the measured value of 50.42 m. Also, the mathematical model for UFM can represent oscillations of measured air pocket pressure patterns. Figure 2 shows how the mathematical model is suitable to represent air pocket pulses.

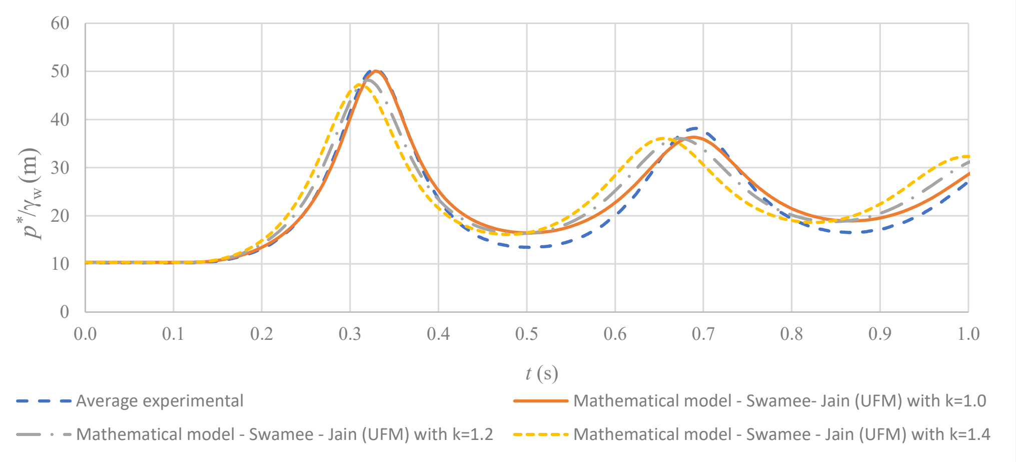
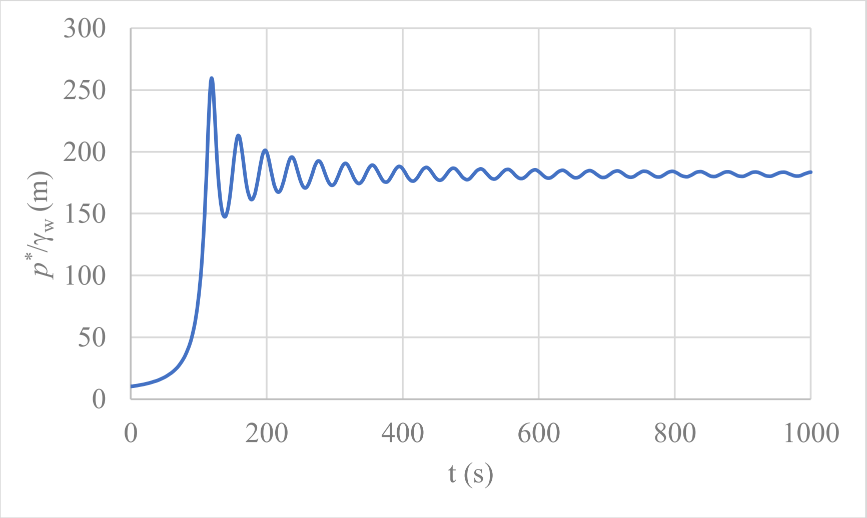


Figure 2. Comparison of values of polytropic coefficients with regards to average experimental of air pocket pressure pulses using the UFM

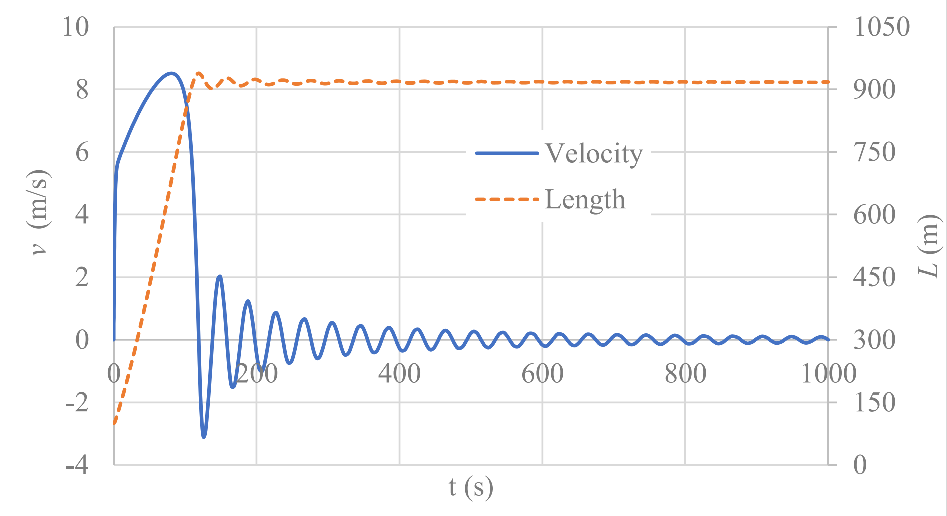
4 PRACTICAL APPLICATION

The mathematical model was performed to study the behavior of the filling operation in a single pipeline with the characteristics as follows: a total length (L_T) of 1000 m, an internal diameter (D) of 595 mm, a longitudinal slope of 10°, an absolute roughness (k_S) of 0.0015 mm, a polytropic coefficient of 1.2, an initial hydro-pneumatic absolute pressure (p_0^*) of 226387 Pa, an initial air pocket (x_0) of 900 m, and a resistance coefficient (R_v) of 12 ms^2/m^6 for a total opening condition. The numerical resolution of algebraic-differential equations (5) to (7) was carried out using the method ODE 23s in Matlab. For all simulations, the air pocket is at atmospheric conditions (p_1^*).

The UFM was performed considering the Swamee-Jain equation. Results are shown in Figure 3. According to these values, a sudden increased trend is presented at the beginning of the hydraulic event (see Figure 3a). A maximum value of air pocket absolute pressure head of 259.67 m is reached at 118.7 s, which is 11 times higher compared to the initial pressure supplied by the hydro-pneumatic pressure tank of 226387 Pa (or 23.08 m). After that, some oscillations are presented, and at the end the air pocket pressure pattern trends to be a constant around of 182 m (from 400 to 1000 s). Regarding the water velocity (see Figure 3b), it increases rapidly reaching a maximum value of 8.50 m/s at 79.7 s. When the air pocket pressure pattern attains its maximum value (at 118.7 s), then the water velocity is null. Then, some pulses are presented along of a value of 0 m/s, which indicates that the water column is reaching a rest condition. The length of the water column beings with an initial value of 100 m. After the instant opening of the discharge valve, the pipe is quickly filled by the water replacing the air column (see Figure 3b). The maximum value is obtained at the same time for the condition of the maximum value of air pocket pressure (at 118.7 s).



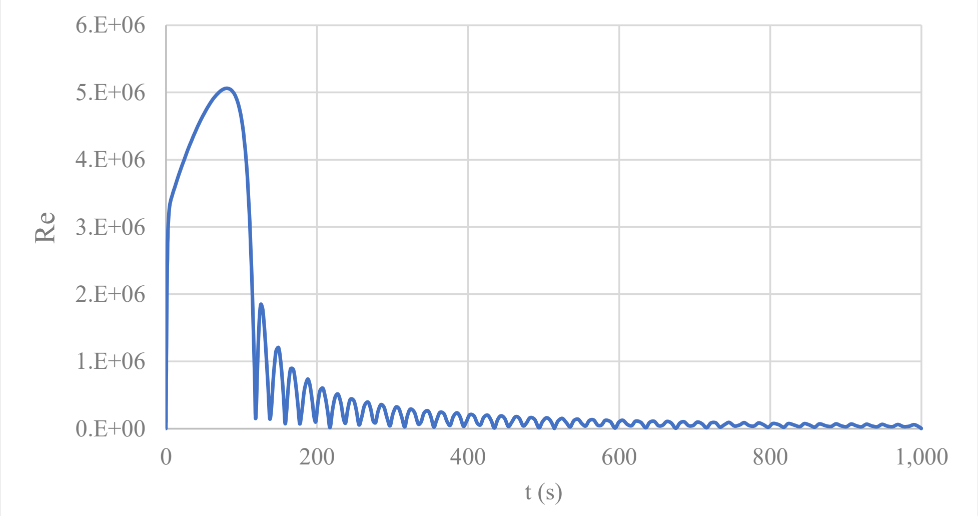
(a) Air pocket pressure



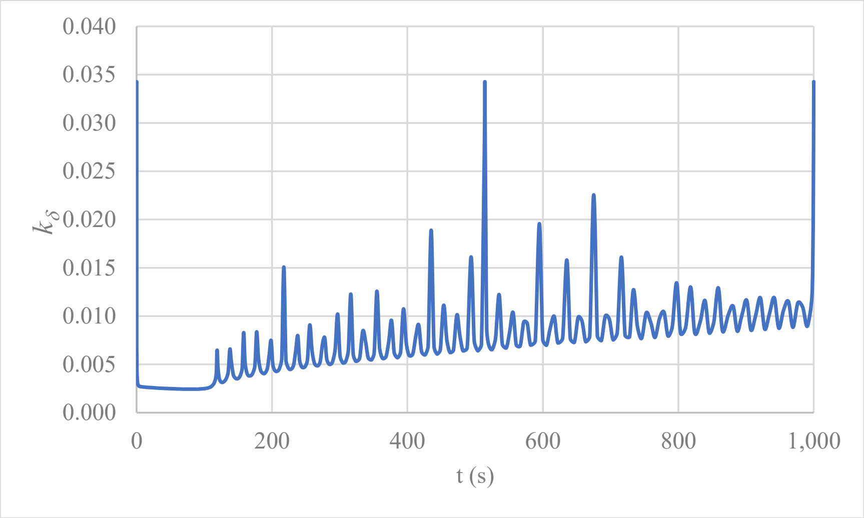
(b) Water velocity and length of the filling water column

Figure 3. Evolution of air pocket pressure, water velocity, and length of the filling water column pulses for the practical application

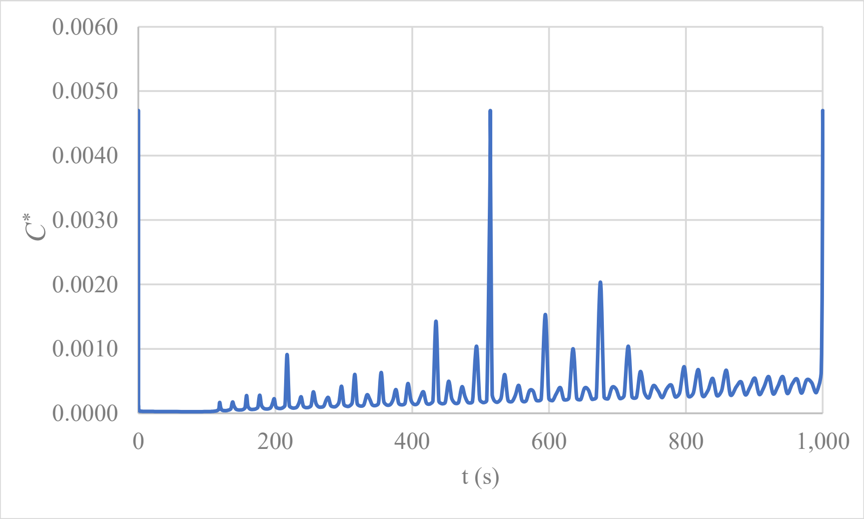
The UFMs also show the variation of the Brunnone friction and Vardy's shear decay coefficient, which depend on Reynolds number. Figure 4 shows the evolution of the Reynolds number and these coefficients. For the analysed scenario, Reynolds number varies from 0 to 5,061,794, following the same trend of the water velocity (see Figure 3b). The maximum value of Reynold number is found at 0.25 s and 0.24 s for Runs No.1 and No. 2, respectively. Brunnone friction and Vardy's shear decay coefficients have a similar trend. The higher Reynolds number, the lower Brunnone friction and Vardy's shear decay coefficients are obtained. The hydraulic transient starts with values of C^* and k_δ of 0.0047 and 0.034, respectively, which corresponds to the conditions of the laminar flow. After that, from 0 to 118.7 s, the values of these coefficients decrease since the water velocity pattern (see Figure 3b) reaches its maximum values. At the end of the transient event, an increasing trend of these coefficients are presented from 118.7 s.



(a) Reynolds Number



(b) Brunnone friction coefficient (k_δ)



(c) Vardy's shear decay coefficient (C^*)

Figure 4. Analysis of patterns of variables: (a) Reynolds number, (b) Brunnone friction coefficient, and (c) Vardy's shear decay coefficient

The SFM and UFM were compared to note the discrepancy in the determination of air pocket pressure pattern. Figure 5 shows how both models can follow the same trend during the entire transient event. In this sense, peak values of air pocket pressure head of 259.55 and 259.67 m are obtained for the SFM and UFM, respectively. Also, the peak values occurrence is obtained closely with time of 119.2 and 118.7 s for the SFM and UFM, respectively.

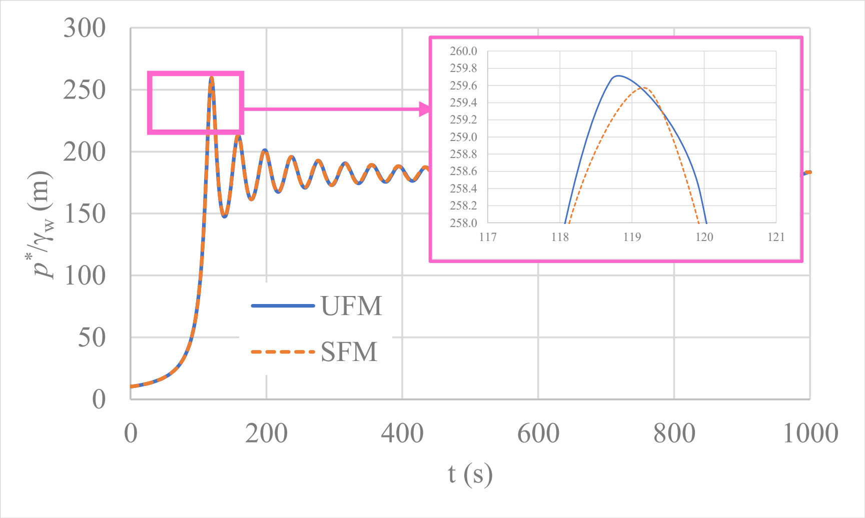
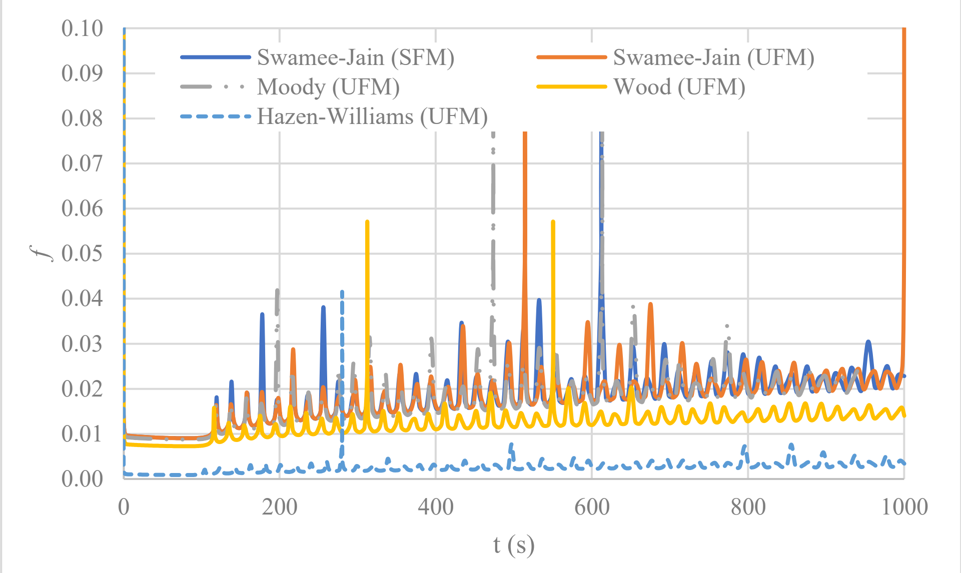


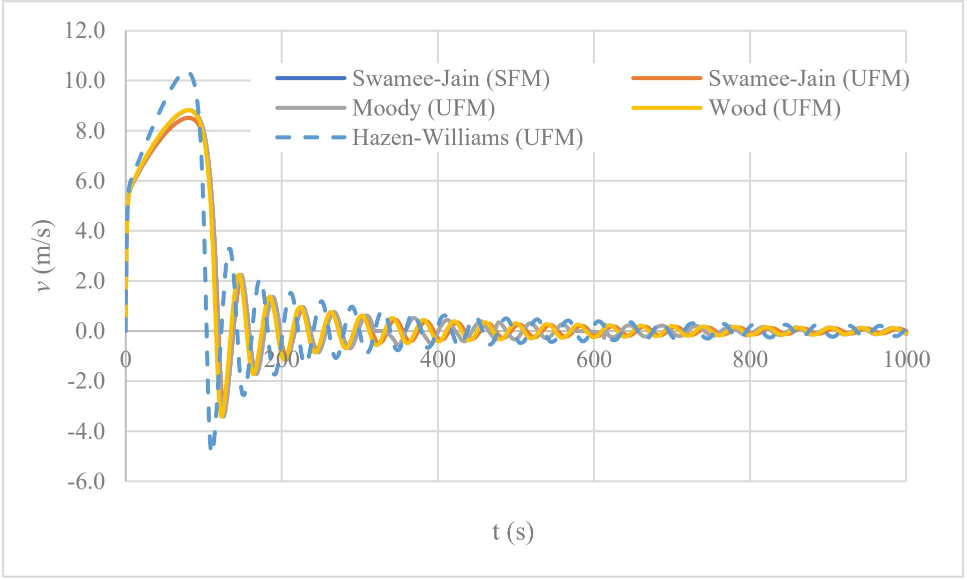
Figure 5. Comparison of air pocket pressure pulses between UFM and SFM

The analysis of the friction factor was performed considering Moody, Wood, and Hazen-Williams equations, as shown in Figure 6a. Swamee-Jain equation provided similar results using SFM and UFM. The friction factor has an increasing trend, and at the end of the hydraulic event it converges to an asymptotic value of 0.022. In addition, Moody equation also provides a good accuracy since its trend is similar compared to the results of Swamee-Jain equation. Wood equation gave lower values of friction factor than Swamee-Jain equation. The worst results were obtained using Hazen-Williams equation due to the hydraulic system exhibits water velocities higher than 3 m/s from 1.28 to 114.97 s (see Figure 6b), where the analysed empirical equation is not applicable.



(a) Friction factor

BY NC SA



(b) Water velocity

Figure 6. Analysis of friction factor formulations with UFM and SFM

5 CONCLUSIONS

A mathematical model to simulate filling processes without admitted air (or air valves) is developed considering an unsteady friction model, which is based on a rigid water column model, a piston flow model, and a polytropic model. The mathematical model was validated in an experimental facility of 7.6-m-long PVC pipe, where the analysed filling operation was performed slowly obtaining an isothermal evolution. Results confirm that the mathematical model is suitable to simulate experimental measurements of air pocket pressure. UFM provided a better calibration compared to SFM.

A practical application of a water installation is presented to note the responses of variables of Brunone model (Brunnone friction and Vardy's shear decay coefficient). Swamee-Jain equation was calibrated with measured data and used as a reference model for the friction factor. Moody, Wood and Hazen-Williams equations where evaluated for computing the friction factor. Results show that Moody equation provides a good accuracy compared to the results obtained from Swamee-Jain equation. Wood equation gave lower values of friction factor than Swamee-Jain equation. Hazen-Williams equations is not recommended for filling operations since water velocities reach values higher to 3 m/s.

Authors suggest that futures works can focus on the assessment of filling operation with air valves considering unsteady friction models.

6 REFERENCES

[1] V.S. Fuertes-Miquel, Ó.E. Coronado-Hernández, D. Mora-Meliá, and P.L. Iglesias-Rey, "Hydraulic Modeling during Filling and Emptying Processes in Pressurized Pipelines: A Literature Review", Urban Water Journal, 2019, 16(4), pp. 299-311.

[2] J. Abreu, E. Cabrera, J. Izquierdo, and J. García-Serra, "Flow Modeling in Pressurized Systems Revisited", Journal of Hydraulic Engineering, 1999, 125(11), pp. 1154-1169.

[3] L. Zhou, D. Liu, and C. Ou, "Simulation of Flow Transients in a Water Filling Pipe Containing Entrapped Air Pocket with VOF Model", Engineering Applications of Computational Fluid Mechanics, 2011, 5(1), pp. 127-140.

[4] F. Zhou, M. Hicks, and P.M. Steffler, “Transient Flow in a Rapidly Filling Horizontal Pipe Containing Trapped Air”, Journal of Hydraulic Engineering, 2002, 128(6), pp. 625-634.

[5] V.S. Fuertes-Miquel, Ó.E. Coronado-Hernández, P.L. Iglesias-Rey, and D. Mora-Meliá, “Transient Phenomena during the Emptying Process of a Single Pipe with Water-Air Interaction”, Journal of Hydraulic Research, 2019, 57(3), pp. 318-326.

[6] O.E. Coronado-Hernández, V.S. Fuertes-Miquel, M. Besharat, and H.M. Ramos, “Subatmospheric Pressure in a Water Draining Pipeline with an Air Pocket”, Urban Water Journal, 2018, 15(4), pp. 346-352.

[7] M. Besharat, O.E. Coronado-Hernández, V.S. Fuertes-Miquel, M.T. Viseu, and H.M. Ramos, “Backflow Air and Pressure Analysis in Emptying Pipeline Containing Entrapped Air Pocket”, Urban Water Journal, 2018, 15(8), pp. 769-779.

[8] A.M. Aguirre-Mendoza, D.A. Paternina-Verona, S. Oyuela, O.E. Coronado-Hernández, M. Besharat, V.S. Fuertes-Miquel, P.L. Iglesias-Rey, and H.M. Ramos, “Effects of Orifice Sizes for Uncontrolled Filling Processes in Water Pipelines”, Water, 2022, 14(6), 888.

[9] L. Zhou, Y. Cao, B. Karney, J.G. Vasconcelos, D. Liu, and P. Wang, “Unsteady friction in transient vertical-pipe flow with trapped air”, Journal of Hydraulic Research, 2020, 59(5), pp. 820-834.

[10] Ó.E. Coronado-Hernández, I. Derpich, V.S. Fuertes-Miquel, J.R. Coronado-Hernández, and G. Gatica, “Assessment of Steady and Unsteady Friction Models in the Draining Processes of Hydraulic Installations”, Water, 2021, 13(14), 1888.

[11] B. Brunone, B.W. Karney, M. Mecarelli, and M. Ferrante, “Velocity profiles and unsteady pipe friction in transient flow”, Journal of Water Resources Planning and Management, 2000, 126(4), pp. 236-244.

[12] L.F. Moody, “Friction Factors for Pipe Flow”, Transactions of the American Society of Mechanical Engineers, 1994, 66, pp. 671-684.

[13] D.J. Wood, “An Explicit Friction Factor Relationship”, American Society of Civil Engineers, 1972, pp. 383-390.

[14] Q. Travis, and L.W. Mays, “Relationship between Hazen-William and Colebrook-White Roughness Values”, Journal of Hydraulic Engineering, 2007, 133(1), pp. 1270-1273.

[15] D.K. Swamee, and A. K. Jain, “Explicit Equations for Pipe Flow Problems”, Journal of the Hydraulics Division, 1976, 102(5), pp. 657-664.

[16] J. Saldarriaga, Hidráulica de Tuberías, Alfaomega Bogotá, 2007.

[17] V. Streeter, and E. Wylie, Fluid Mechanics, Ed. McGraw-Hill, NY, 1979.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14821

INVESTIGATION OF GROUNDWATER CONSUMPTION TO COPE WITH THE INADEQUATE PIPED WATER SUPPLY IN CONTINUOUS AND INTERMITTENT SUPPLY SYSTEMS: A CASE STUDY IN BANGALORE, INDIA

B.N. Priyanka[1], Bharanidharan B.[2], Sheetal Kumar K.R.[3], M.S. Mohan Kumar[4]; V.V. Srinivas[5], Sanjana R Nibgoor[6] and Kishore Y.[7]

[1]Research Associate, Interdisciplinary Centre for Water Research (ICWaR), Indian Institute of Science (IISc), Bangalore, Karnataka, (India).
[2,3]Research Scholar, Department of Civil Engineering, IISc, Bangalore, Karnataka, (India).
[4]Formerly Professor, Department of Civil Engineering and ICWaR, IISc, Bangalore, Karnataka, (India).
[5]Professor, Department of Civil Engineering, Associate faculty ICWaR, IISc, Bangalore, Karnataka, (India).
[6,7]Project Assistant, ICWaR, IISc, Bangalore, Karnataka, (India).

[1] *E-mail: priyankabn@iisc.ac.in*; [2] *E-mail: bharanib@iisc.ac.in*; [3] *E-mail: sheetalkumar@iisc.ac.in*;
[4], *E-mail: msmk@iisc.ac.in*;[5] *E-mail: vvs@iisc.ac.in*; [6] *E-mail: sanjanan@iisc.ac.in*;
[7]*E-mail: kishorey@iisc.ac.in*

Abstract

Although the supply of piped water to the Indian cities is increasing, the demand is not always fulfilled. This gap in water demand and supply is usually bridged by using alternate sources of water, mostly groundwater. Bangalore, the capital city of Karnataka, is one of the fastest developing metropolitan cities in India and is also facing piped water supply issues. The groundwater is the main source of alternate water supply in the city. In the present study, a District Metered Area (DMA) is selected in the Bangalore South-West division; this DMA has both intermittent and continuous water supply systems. The water distribution network (WDN) of study DMA contains four inlets and five supply zones. The first is a continuous water supply system whereas the other four are intermittent systems. The impact of inequitable consumption in the study DMA is evaluated using Lorenz Curve and Gini Coefficient and the consumption of groundwater to cope with insufficient water supply is analyzed. The data used for the present investigation are from the field flowmeters, consumer meter reading, and door-to-door questionnaire survey are used for the analysis. The questionnaire survey includes RR (revenue register) number, presence of wells/borewells, horsepower (HP) of the pumps used, building type, the number of inhabitants, and the floors in each building. In the continuous supply system, a questionnaire survey was untaken for 80% of the connections, whereas in the intermittent supply system random sampling was used. The questionnaire survey analysis showed that only 27% of the consumers in the continuous supply system rely on piped water supply, whereas others used groundwater as well as piped water supply. The study illustrated the gap in groundwater consumption between supply zones within intermittent water supply systems. Reliability on groundwater was high even in continuous supply systems indicating insufficient pressures resulting in unsatisfied demands. The study indicated that just increasing the access to the piped water supply to the consumers is not sufficient, the acceptable quality with adequate pressure of water should be delivered to reduce the use of groundwater. The inferences from the study can be used to regulate groundwater extraction.

Keywords
Groundwater consumption; Intermittent water supply, Continuous water supply; Lorenz Curve and Gini Coefficient.

1 INTRODUCTION

Bangalore, the capital city of Karnataka, is facing an imbalance in the rates of water demand and supply due to rapid population growth and industrialization. About 70% of the municipal water to the city is supplied by the Bangalore Water Supply and Sewerage Board (BWSSB), which is also responsible for sewage disposal. Currently, 60% of the BWSSB water supply is met by the Cauvery River which is drawn from about 100 km against an elevation of 540 m and the rest water demand is met by groundwater pumping (Mohan Kumar et al., 2011). The gap between the water supply and demand in the city is satisfied mostly by groundwater, even at the DMA (District Metered Area) level. The groundwater consumption is not only in intermittency areas but also seen in areas with continuous water supply. To understand the reason for groundwater consumption in the 24/7 system and intermittent supply system, J.P. Nagar Phase 1 DMA in the south division of Bangalore, Karnataka, India (Figure 1)was selected.

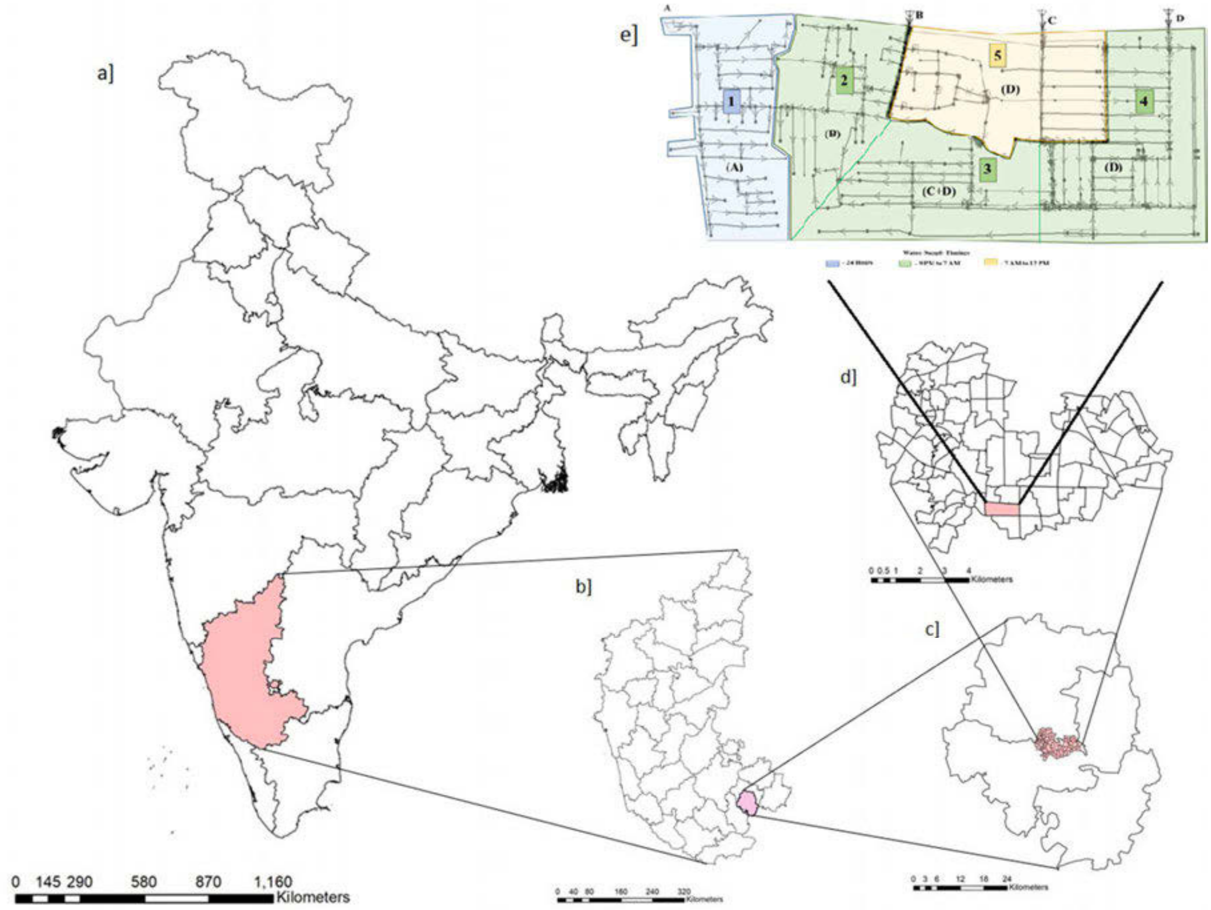


Figure 1. Geographic description of the selected DMA within a) country: India, b) state: Karnataka, c) district: Bangalore urban, and taluks: Bangalore South d) division: Southwest division 4 e) DMA: J P Nagar Phase 1 with water distribution pipe network, four inlets (named A, B, C, D), and five supply zones (supply zone 1 is continuous and zones 2, 3, 4 & 5 are intermittent).

The aerial extent of the pilot service DMA is 0.68 km² and the population is about 22158 (as per BWSSB J.P. Nagar Phase 1 DMA report, 2013). The topographic elevation of the DMA varies from 898.58 m to 918.54 m. There are 2399 properties with 295 without RR (revenue register) numbers which includes vacant plots. The total length of the pipeline in pilot DMA is 22174 m and the pipe diameters range from 100 mm to 300 mm with a weighted average diameter of 125 mm.

BY NC SA

The pipelines are of cast iron, PVC, ductile iron, and mild steel, out of which PVC is of majority with 10608 m in length and 32 years of age.

There are 4 inlets to the J.P. Nagar Phase 1 DMA (Figure 1e), which is serviced by Cauvery 2nd stage 1200 mm diameter transmission main. The supply schedule for zone 1 through inlet A (Figure 1e) is 24 hours every day. For zones 2, 3, and 4 the supply schedule is from 21:00–7:00 through inlets B, C+D, and D, respectively on alternate days, and for zone 5, the schedule is from 7:00–12:00 through inlet D on alternative days.

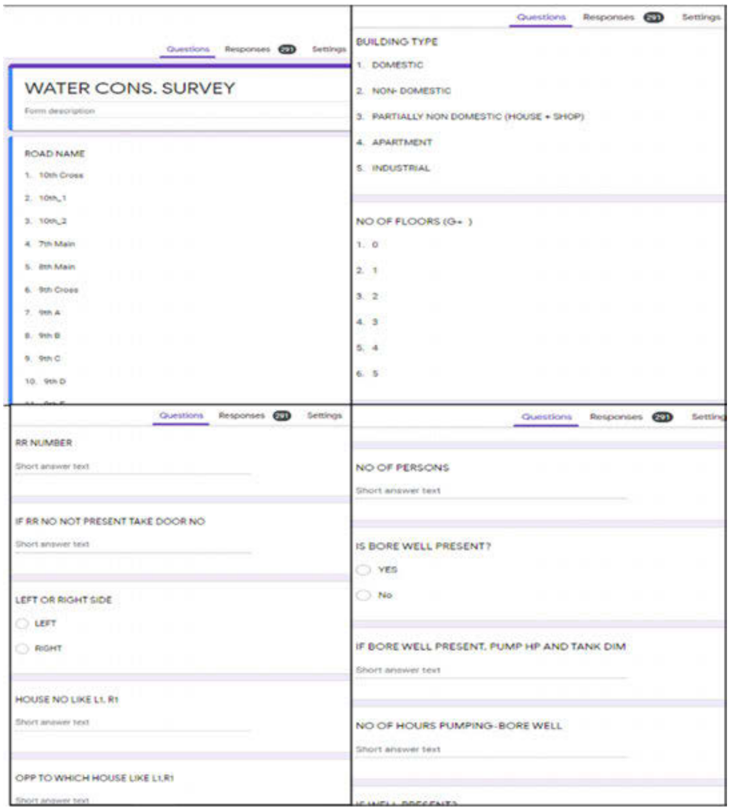
2 METHODOLOGY

2.1 Inequity Analysis

In the past several researchers have used economic indices to evaluate the inequality in water supply, consumption, supply hours, time of supply, etc., (e.g. Guragai et al., 2017; Malakar et al., 2017; Sheetal Kumar et al., 2018). For the inequity study, we have used the Lorenz curve and Gini Coefficient (GC), which are most commonly used.

The Lorenz curve is a graphical representation of the distribution of water. The deviation curve from the 45-degree equality line is called the Lorenz curve. Farther the Lorenz curve is with respect to the 45-degree perfect equality line, the more unequal is the distribution. Gini Coefficient is a value computing the ratio of the area enclosed by the line of equality and the Lorenz curve to the area under the line of equality. A higher GC value represents a more inequity and varies between 0 and 1. To plot the Lorenz curve, water consumption is considered in terms of liters per capita (lpcd), this data is converted as weighted consumption and then cumulative consumption is estimated. These cumulative consumption values are plotted on y-axis and the cumulative population on the x-axis.

2.2 Estimation of Groundwater Consumption



WATER CONS. SURVEY

Form description

ROAD NAME

1. 10th Cross
2. 10th_1
3. 10th_2
4. 7th Main
5. 8th Main
6. 9th Cross
7. 9th A
8. 9th B
9. 9th C
10. 9th D

BUILDING TYPE

1. DOMESTIC
2. NON- DOMESTIC
3. PARTIALLY NON DOMESTIC (HOUSE + SHOP)
4. APARTMENT
5. INDUSTRIAL

NO OF FLOORS (G+)

1. 0
2. 1
3. 2
4. 3
5. 4
6. 5

RR NUMBER

Short answer text

IF RR NO NOT PRESENT TAKE DOOR NO

Short answer text

LEFT OR RIGHT SIDE

○ LEFT
○ RIGHT

HOUSE NO LIKE L1, R1

Short answer text

OPP TO WHICH HOUSE LIKE L1,R1

Short answer text

NO OF PERSONS

Short answer text

IS BORE WELL PRESENT?

○ YES
○ No

IF BORE WELL PRESENT, PUMP HP AND TANK DIM

Short answer text

NO OF HOURS PUMPING-BORE WELL

Short answer text

Figure 2. Google forms papered for door-to-door questionnaire survey.

Door to door questionnaire survey was carried out to estimate groundwater consumption in J.P. Nagar phase 1 DMA. The data collected during this survey are details regarding RR number, presence of wells/borewells, horsepower (HP) of the pumps used, building type, the number of inhabitants, the floors in each building, time of pumping, and hours of pumping (Figure 2). In the continuous supply system, a questionnaire survey was taken for 80% of the connections, whereas in the intermittent supply system random sampling was used.

The groundwater consumption in each house was calculated using the following equation.

$$Q_p = 249.84 \frac{P_{hp} \times \eta}{H \times SG} \tag{1}$$

where Q_p is the pump flow rate in l/s; P_{hp} is the pump horsepower; η is the pump efficiency (assumed as 50%); H is the dynamic head in ft; SG is the specific gravity of water.

3 RESULTS AND DISCUSSION

3.1 Inequity in BWSSB Water Consumption

The data used for this analysis is collected from consumer meter readings based on RR numbers for a period of 54 months i.e., from January 2017 to June 2021. This data is analyzed streetwise in the continuous system (Figure 3) and supply zone-wise in the intermittent system (Figure 4). For the figure

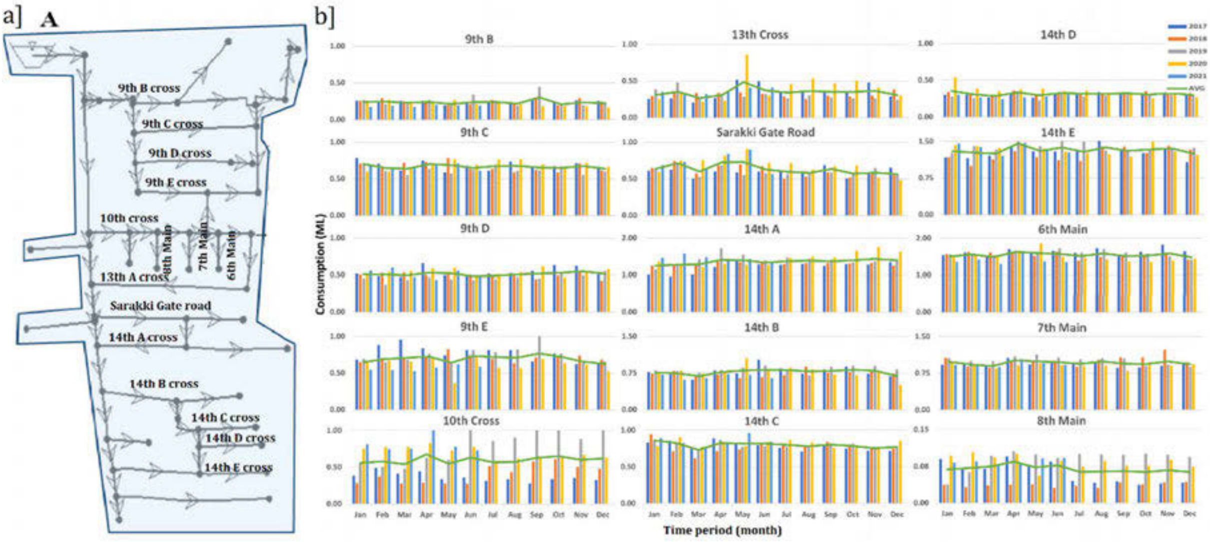


Figure 3. Continuous system a)streetwise pipe network along with flow directions, and b) streetwise monthly water consumption over 54 months i.e., from January 2017 to June 2020.

The bar graph shown in Figure 2b, describes the monthly water consumption in each street averaged over meter readings at each house. The data shown for 54 months are color-coded yearly. From the figure, it is evident that the streetwise average monthly water consumption is constant for almost all streets (except the 10th cross and 8th main), indicating water consumption in all the seasons is the same. The water consumption on the 10th cross and 8th main are increased from the year 2019, due to the replacement of pipes and reducing leakage losses. From the data, it was evident that the 8th main road consumes the least BWSSB with an average usage of 0.06–0.08 ML, whereas the consumption on the 6th main road is highest with an average usage of 1.5–1.6 ML. The 9th crossroads and the 14th crossroads, being the residential areas, are seen with an average usage of 0.5–0.6 ML and 0.7–1.0 ML, respectively. The 10th crossroad and the Sarakki Gate road, with the majority of non-domestic firms, consume around 0.6 ML of water supplied from BWSSB.

To understand the streetwise inequity, Lorenz cure and GC value are estimated for the BWSSB water consumption in the continuous supply system. From Figure 3a, it can be seen that in most of the streets the mean water consumption over 54 months is less than CPHEEO (Central Public Health and Environmental Engineering Organisation) standards of 135 lpcd (except in streets 14th A and 14th E). While in 9 streets the consumption is less than half the standard value (i.e., < 68 lpcd). This could be due to low network pressure, the use of alternate sources, and so on. The Lorenz curve and GC = 0.33 (Figure 3b) show that inequity exists. This analysis indicates due to unsatisfied demands, consumers use alternate sources, especially groundwater

a)

Street	Consumption (lpcd)	Population
8th Main	16	182
9th B	22	288
14th D	33	317
13th A Cross	41	336
9th D	44	355
9th E	59	336
9th C	61	336
Sarakki Gate Road	64	298
14th B	66	326
14th C	83	336
10th Cross	127	192
14th A	141	317
14th E	147	288
Total	903.16	3907

b)

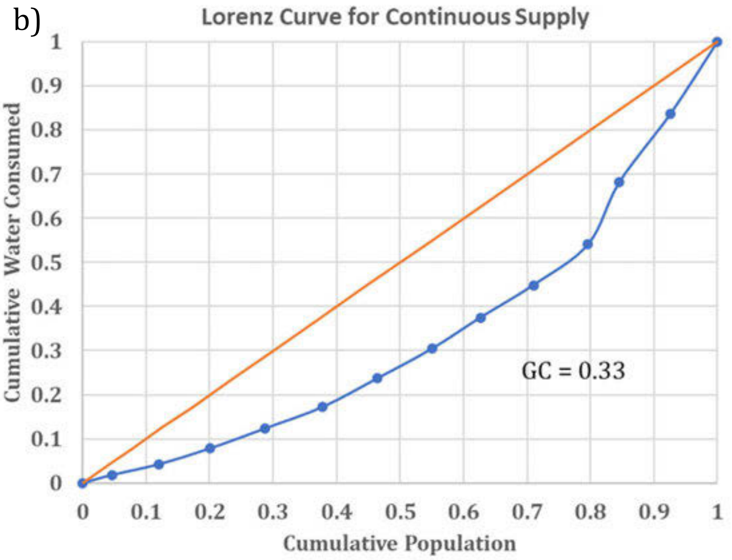


Figure 4. Continuous system a) streetwise BWSSB water consumption data, and b) Lorenz curve to evaluate inequity.

The bar graph shown in Figure 4b, describes the monthly water consumption in each supply zones over meter readings at each house. The data shown for 54 months are color-coded yearly. From the figure, it is evident that the zone-wise average monthly water consumption is constant, indicating water consumption in all the seasons is the same. Figure 4b illustrates that the water consumption in all supply zones varies from 10.1–17.8 ML and the consumption value in supply zone 4 is less whereas supply zone 5 consumes the highest.

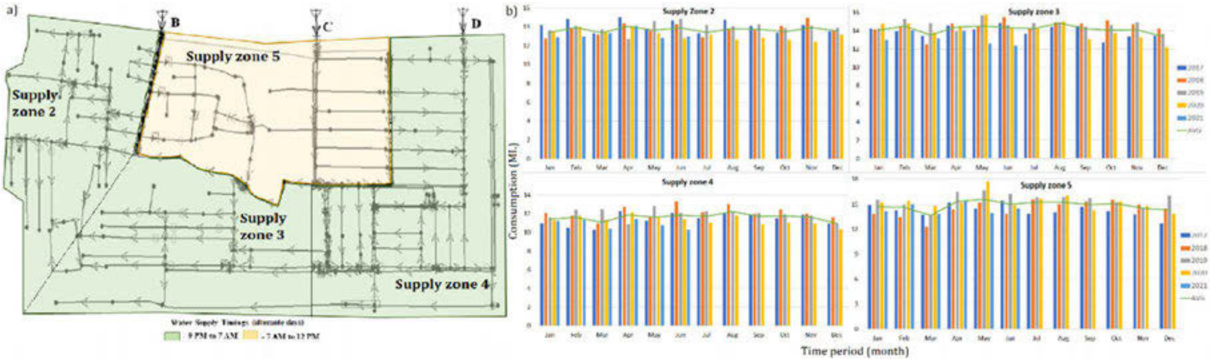


Figure 4b. Intermittent system a) supply zone-wise pipe network along with flow directions and b) supply zone-wise monthly water consumption over 54 months i.e., from January 2017 to June 2020.

To understand the supply zone inequity, Lorenz cure and GC value are estimated for the BWSSB water consumption in the intermittent supply systems. From Figure 5a, it can be seen that in all the supply zones the mean water consumption over 54 months is less than CPHEEO standards of 135 lpcd. This could be due to low network pressure, unpredictable supply times, duration of

cc BY NC SA

supply the use of alternate sources, and so on. The Lorenz curve and GC = 0.13 (Figure 5b) show that inequity exists. This analysis indicates due to unsatisfied demands, consumers use alternate sources, especially groundwater.

a)

Supply zone	Consumption (lpcd)	Population
Supply zone 5	72	6291
Supply zone 3	78	5196
Supply zone 4	88	3863
Supply zone 2	89	4412
Total	327	19762

b)

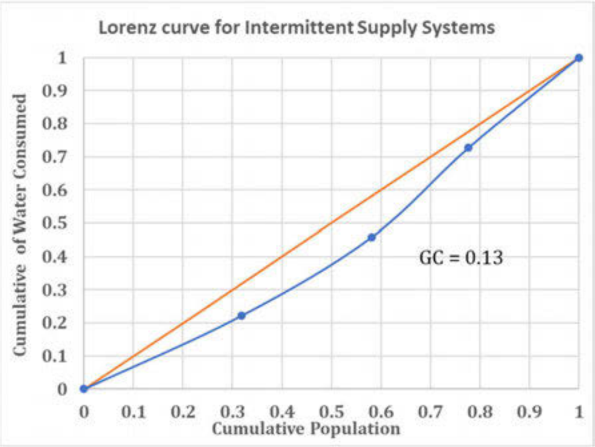


Figure 5. Intermittent systems a) supply zone-wise BWSSB water consumption data, and b) Lorenz curve to evaluate inequity.

3.2 Estimation of Groundwater Consumption

From the door-to-door questionnaire survey, it was seen that 73% of RR numbers in the continuous supply system (Figure 6a) and 84% of RR numbers in the intermittent supply system use both BWSSB and groundwater (Figure 6b). Figure 6c shows the percentage of connections using groundwater in each supply zone in the intermittent supply systems. Supply zone 5 has more number wells, whereas supply zone 4 has fewer among the four zones. The questionnaire survey was done for 80% of the connections in a continuous system whereas for the intermittent system random sampling was done at very few points.

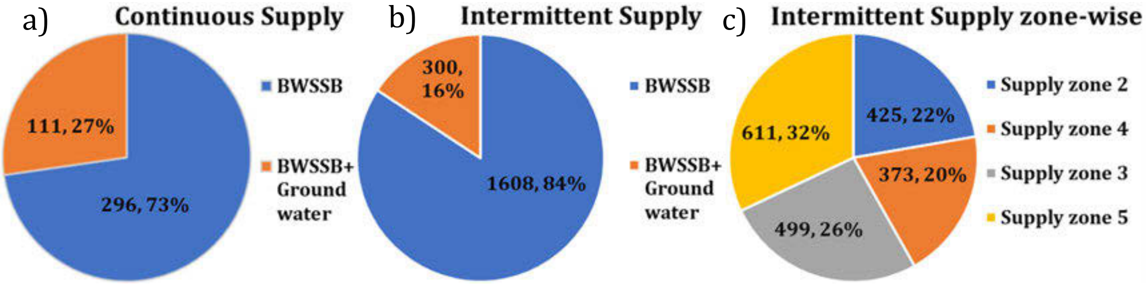


Figure 6. Pie chart showing the number of connections using groundwater a) in the continuous supply system, b) in the intermittent supply systems, and c) in the supply zones of the intermittent supply system.

As illustrated earlier in Figures 2 and 4, the consumption of BWSSB water does not change with the season, the same concept is assumed for groundwater consumption. Thus, the survey was done once i.e., in October 2021 to estimate the consumption and for the other months, the groundwater consumption is assumed to be the same. Equation (1) is used to estimate daily groundwater consumption from known pump running time, from this data monthly consumption is estimated.

From Figure 7a it is seen that the 13th A cross pumps more groundwater of greater than 1.4 ML, whereas the 7th main road consumes the least about 0.06 ML. In the intermittent system, supply zone 5 extracts groundwater of about 13 ML in a month whereas, supply zone 3 consumes groundwater of less than 4ML per month (Figure 7b).

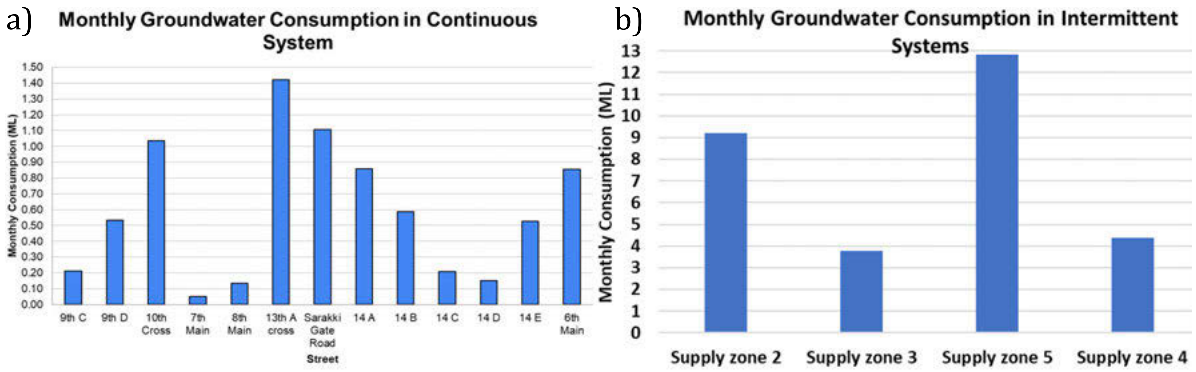


Figure 7. a) Streetwise monthly groundwater consumption in a continuous system and b) Supply zone-wise monthly groundwater consumption in the intermittent system.

The inequality in groundwater consumption is estimated using the Lorenz curve and GC for both continuous and intermittent systems. From Figure 8a, it can be seen that all the streets consume groundwater more than CPHEEO standards of 135 lpcd. While in 11 streets the consumption is more than twice the standard value (i.e., > 270 lpcd). This indicates consumers are exploiting the groundwater more than required. The GC of 0.13 indicates there exists an inequity even in groundwater consumption in a continuous system (Figure 8b).

a)

Street	Consumption (lpcd)	Population
14th E	192	32
9th E	213	106
14th A	230	116
9th D	282	117
8th Main	291	64
14th B	303	75
14th D	303	75
9th B	334	43
Sarakki Gate Road	375	142
9th C	419	47
10th Cross	453	91
14th C	495	104
13th A Cross	505	106
Total	4393	1114

b)

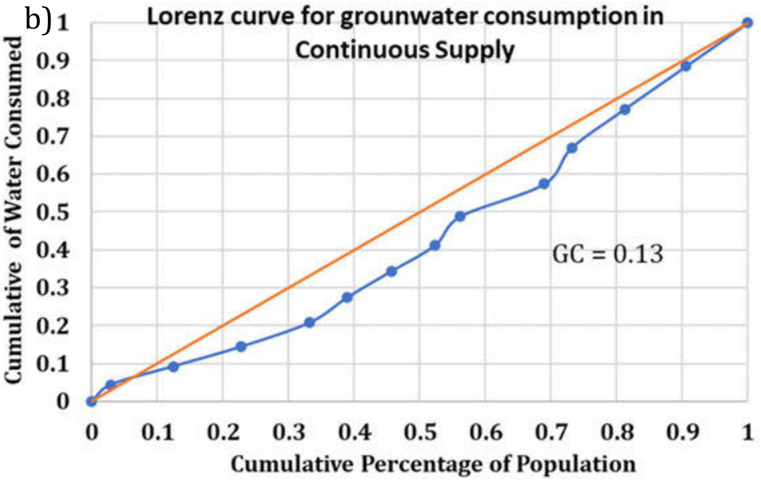


Figure 8. Continuous system a) streetwise groundwater consumption data, and b) Lorenz curve to evaluate inequity.

From Figure 9a, it can be seen that in all the supply zones the groundwater consumption is more than the CPHEEO standards of 135 lpcd. All the supply zones consume groundwater more than twice the standard value (i.e., > 270 lpcd). This indicates consumers are overexploiting the groundwater more than required. The Lorenz curve and GC = 0.3 (Figure 9b) show that inequity exists. This analysis indicates due to unsatisfied demands, consumers use groundwater water more than required.

a)

Supply zone	Consumption (lpcd)	Population
Supply zone 2	314	976
Supply zone 5	356	1200
Supply zone 3	384	328
Supply zone 4	403	364
Total	1458	2867

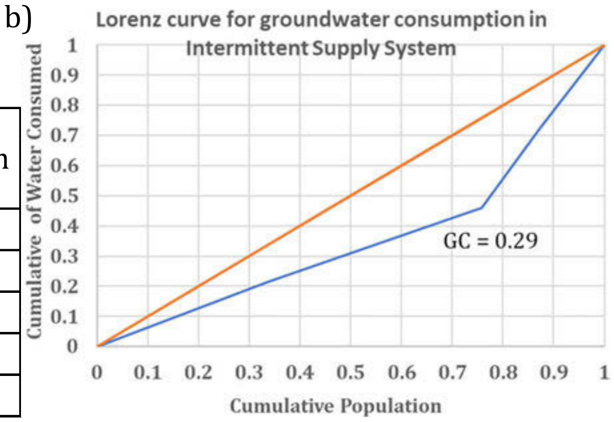


Figure 9. Intermittent system a) zone-wise groundwater consumption data, and b) Lorenz curve to evaluate inequity.

4 SUMMARY AND CONCLUSIONS

The investigation includes evaluating the demand within the DMA level. J.P. Nagar Phase 1 DMA was selected as the study DMA due to the presence of both continuous and intermittent supply systems. The consumption of BWSSB water in both continuous and intermittent for an average of 54 months is mostly less than the standard required consumption of 135 lpcd. This is mainly due to low network pressure, unpredictable supply times, duration of supply, and the use of alternate sources. Thus the inequality consumption in continuous and intermittent is evaluated using Lorenz Curve and Gini Coefficient for both BWSSB water and groundwater. To estimate the monthly groundwater consumption, a door-to-door questionnaire survey is used for the analysis. Some of the questions asked are RR number, presence of wells/borewells, horsepower (HP) of the pumps used, building type, the number of inhabitants, and the floors in each building. The survey was conducted for 80% of connections in the continuous supply system, whereas very few numbers random sampling was done for the intermittent supply system.

The results showed that inequality in both BWSSB and groundwater consumption exists in between the streets of a continuous system and in between supply zones of an intermittent supply system. Even though the water was supplied 24/7 the exists an inequity of 0.33 and the groundwater consumption was more than 190 lcpd in all the streets in a continuous system. The reliability of groundwater was very high in continuous supply systems due to insufficient pressures resulting in unsatisfied demands. In the intermittent zone, the groundwater consumption was more than twice the standard required demand as BWSSB water consumption n all the supply zones was less than 90 lpcd. This indicates just increasing the piped water supply will not be sufficient, regulation on the usage of groundwater extraction is needed.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14823

REDUCTION OF THE SEARCH SPACE FOR THE OPTIMIZATION PROBLEM OF THE DESIGN OF THE PUMPING STATION THROUGH THE AUTOMATIC IDENTIFICATION OF INFEASIBLE FLOW DISTRIBUTIONS

Jimmy H. Gutiérrez-Bahamondes[1], Bastián Valdivia-Muñoz[2], Daniel Mora-Melia[3] and Pedro Iglesias- Rey[4]

[1,2]Doctorado en Sistemas de Ingeniería. Facultad de Ingeniería, Universidad de Talca, Curicó(Chile)
[3]Dep. de Ingeniería y Gestión de la Construcción. Facultad de Ingeniería, Universidad de Talca, Curicó (Chile).
[4]Dep. Ingeniería Hidráulica y Medio Ambiente. Universitat Politècnica de València, Valencia (Spain).

[1]jgutierrezb@utalca.cl, [2] bvaldivia18@alumnos.utalca.cl, [3]damora@utalca.cl, [4] piglesia@upv.es

Abstract

The pumping station design is a critical process in water distribution networks. This set of decisions will have an immediate impact on construction costs and determine energy consumption over the entire lifetime of the system. However, the minimization of investment and operational costs at the same time is a complex problem that has been approached from different perspectives.

To achieve this goal, in recent years, it has been shown that it is possible to optimize the selection of pumps, accessories, and control systems while optimizing the flow distribution provided by the pumping stations by using the setpoint curve. However, the huge number of possible combinations and the non-linearity of the equations rule out the use of exact methods to solve the proposed mathematical model. Despite this, some metaheuristic techniques, specifically population-based evolutionary algorithms, have shown good performance against case studies in networks with a high level of simplification. Each objective function evaluation involves at least one hydraulic simulation during the analysis periods. Therefore, the computational effort grows considerably as the size of the network increases, affecting the efficiency of these algorithms and limiting their use to smaller networks. Thus, optimization of the design of pumping stations in real-size networks is a problem that has not yet been fully resolved.

To reduce the number of evaluations of the objective function during the optimization process, this work presents a new method for the reduction of the search space based on the automatic identification of infeasible flow ranges as part of the network preprocessing. The method considers the maximum capacity of the available pumps, the minimum pressure required, and the demand patterns of the network. In this way, each pumping station has different restrictions for the decision variables of the mathematical model related to the flow contributions.

From this point on, the algorithm does not waste any computational effort evaluating solutions that represent flow distributions previously classified as infeasible. Therefore, it is possible to accelerate the convergence of the algorithms while preserving the quality of the solutions obtained. This new method can be applied to any direct injection network. The amount of solution space reduction will depend on the characteristics of each network. To clarify, this work includes the analysis of one case study and a genetic algorithm was implemented to resolve the model. Finally, the results show a reduction of the solutions space of 80% for the largest network presented.

Keywords
Pumping station design, setpoint curve, metaheuristic, search space.

1 INTRODUCTION

The design of pumping stations (PS) in water distribution networks is a complex problem [1], [2]. The work of these systems directly affects the quality of life of consumers. However, its construction implies a high cost, and its operation is carried out through a large energy consumption throughout the year. In this way, the design of PSs has short- and long-term consequences for the budget of any city [3]. On the one hand, in the immediate term, it will determine the investment costs necessary for the construction of the facilities. While the operating costs will also be affected throughout the life of the project [4]. This means that efficient design could significantly reduce the costs involved in the water supply. Especially in networks where there is not enough elevation to install tanks, or the network is fed directly from groundwater [5], [6].

During the design process, pump models, accessories, and control systems must be selected [7]. Later, through network operations, the control system determines the on/off status of each pump depending on demand [4]. Therefore, the decisions involved in the design must consider the operational conditions and, consequently, cannot be optimized independently [8].

Different approaches have been proposed to optimize the efficiency of the PS design [1], [6], [8]. The authors use different criteria, like location and minimizing the leakage [9], maximizing energy production [10], and minimizing maintenance and energy costs at the same time [4] among other criteria. However, in recent times it has been shown that the determination of the optimal flow distribution can be used as an effective tool in the design of PSs [11], [12]. In particular, the research carried out by the authors de [8] implements a novel methodology based on the minimization of operational and investment costs at the same time using the optimization of flow distribution. They propose a mathematical optimization model where the decision variables are the fractions of flow provided by each PS, the model of the pumps, and the number of fixed and variable speed pump included. And to solve the model they use a pseudo genetic algorithm proposed in [13]. However, this methodology has a problem; each evaluation of a solution involves going through all the nodes to verify compliance with the minimum pressure constraints. Therefore, the algorithm loses search capacity as the network size grows. Additionally, in the case of large networks, the calculation time grows significantly due to the scarcity of feasible solutions.

This work proposes a new method to speed up the heuristic search processes of evolutionary algorithms applied to the problem posed by [8] using automatic pre-processing of the analyzed network. This method allows an empirical, systematic, and exhaustive mapping of the relationships between the possible flow contributions of the pumping stations that are being designed heads and flows in all the nodes of the network. Thus, it is possible to significantly limit the flow ranges provided by each pumping station. Specifically, it is possible to find flow rate combinations that will always give rise to an infeasible solution. The application of the methodology presented here is aimed at optimizing large networks, however, it can also be used in small networks without the need for changes.

The remainder of the paper is organized as follows: Section 2 describes the proposed methodology. Then, the developed methodology is applied to a case study, and an optimization method is implemented. Next, Section 3 provides the results, and a discussion is detailed in Section 4. Finally, the conclusions of the research can be found in Section 5.

2 MATERIALS AND METHODS

This work proposes a new method to speed up the process of searching for solutions to the problem posed in [8]. The implementation of this procedure does not imply changes in the mathematical model proposed in that investigation. For a better understanding, this model is briefly presented in the following sections.

2.1 Mathematical model

Decision variables are the key aspect of this study. They conform the solution to the problem. First, x_{ij} defines the percentage of the flow supplied from PS i (PS_i) at each time step j. Second, m_i indicates the number of fixed speed pumps and third, b_i corresponds to the identifier of the pump model to be installed in PS_i The parameters N_t and N_{ps} represent the total number of time steps and the total number of PSs, respectively; Once these values are known, it is possible to calculate the maximum flow for each PS, the number of total pumps ($N_{B,i}$), the number of VSPs.

The optimization model minimizes the sum of capital (CAPEX) and operational (OPEX) costs at the same time. Equation 1 presents the total annualized cost of the project. Where Fa is the amortization factor applying an interest rate r during Np periods.

$$\mathrm{F} = F_a \cdot CAPEX + OPEX \tag{1}$$

$$F_a = \frac{r \cdot (1+r)^{Np}}{(1+r)^{Np-1}} \tag{2}$$

The CAPEX is calculated according to Equation 3.

$$C_{CAPEX} = \mathrm{N_B} \cdot \mathrm{C_{pump}} + \mathrm{n} \cdot \mathrm{C_{inv}} + \mathrm{C_{facility}} + \mathrm{C_{control}} \tag{3}$$

Where NB is the number of pumps, C_{pump} is the purchase cost of a pump, n is the number of frequency inverters, $C_{facility}$ represents the cost of accessories including pipes, and C control is the sum of a pressure transducer, flowmeter, and programmable logic controller.

$$\mathrm{OPEX} = \sum_{\mathrm{j=1}}^{\mathrm{N_t}} \left\{ \sum_{\mathrm{i=1}}^{\mathrm{N_{ps}}} \left[\left(\sum_{\mathrm{k=1}}^{\mathrm{m_{i,j}}} \frac{\gamma \cdot \left(\mathrm{H_{o,i}} - \mathrm{A_i} \cdot \mathrm{Q^2_{i,jk}}\right)}{\left(\mathrm{E_i} - \mathrm{F_i} \cdot \mathrm{Q_{i,j,k}}\right)} + \sum_{\mathrm{k=1}}^{\mathrm{n_{j,i}}} \frac{\gamma \cdot \left(\mathrm{H_{o,i}} \cdot \alpha_{\mathrm{i,j,k}} - \mathrm{A_i} \cdot \mathrm{Q^2_{i,jk}}\right)}{\left(\frac{\mathrm{E_i}}{\alpha_{\mathrm{i,j,k}}} - \frac{\mathrm{F_i}}{\alpha^2_{\mathrm{i,j,k}}} \cdot \mathrm{Q_{i,j,k}}\right)} \right) \cdot \mathrm{p_{i,j}} \right] \Delta \mathrm{t_j} \right\} \tag{4}$$

where for each PS_i, the parameters $H_{0,i}$, Ai, E_i and F_i are the characteristic coefficients of the pump head and the performance curve and are extracted from an existing database depending on the pump model; $Q_{i,j,k}$ represents the discharge of pump k during time step j in PS i; $p_{i,j}$ is the energy cost; ϒ is the specific gravity of water; Δt_j is the discretization interval of the optimization period; and the numbers of FSPs and VSPs running at time step j are represented by $m_{i,j}$ and $n_{i,j}$, respectively. These values depend on the selected pump model and the system selected to control the operation point.

The optimization model is restricted by continuity and momentum equations and by minimum head requirements in the demand nodes. Equations 5 y 6 guarantees that the total flow supplied by the PS is equal to the flow demand.

$$x_{i,j} \geq 0 \; \forall i,j \tag{5}$$

$$\sum_{i=1}^{N_{ps}} x_{i,j} = 1 \,\forall j \tag{6}$$

2.2 The infeasibility problem

The decision variable x_{ij} determines the fraction of flow that PS_i contributes during period j. This variable can take ranges from 0 to 100 (expressed as a percentage). Where 0 means that the PS does not supply water in that period. So, on the contrary, a value of 100 indicates that it will be the only SP that operates in the period. So, there is a huge number of possible combinations, and many of them are infeasible solutions.

The main causes of infeasibility are listed below:

- The distribution of flow generates sectors of the network where it is not possible to reach the minimum required pressures.
- Some of the PSs must provide a pressure greater than the maximum head of the largest pump that exists in the available catalog.
- The sum of the flows supplied is greater than the demand.

Each evaluation of the objective function supposes an increase in the computational effort made by the algorithm. It is for this reason that this work proposes to analyze the network previously to establish minimum and maximum limits for the variable $x_{i,j}$. Unfortunately, the non-linearity of the relationships between the hydraulic variables does not allow these values to be fixed, but depends on the piezometric head of the main PS. Which supplies all the water that is not provided by the rest of the PSs. To simplify the analysis, in this paper it is assumed that there is at least one PS called PS1. Then, PS1 is always the main station. And therefore, the lower and upper limits are expressed as curves as a function of PS1.

2.3 Calculation of curve of minimum and maximum flows

The steps to find the minimum and maximum flow curves for PSs other than PS1 are described.

1) Determine the sum of the product of the base demands of the nodes according to the consumption pattern (Q_B), corresponding to the lowest consumption in all the periods of the analysis.
2) Determine the limit of the piezometric head of PS1. The minimum value is the initial piezometric head of the network (H_0). The maximum value will be the initial head plus the maximum pressure given by the pump catalog (H_{bMax}). Assign the head to the corresponding PS1 (H_{design}).
3) Select one of the pumping stations, and Q_{min} is assigned to it.
4) Set a distribution flow rate (Q_r), which corresponds to the difference between Q_B and Q_{min}.
5) Set a list of flow distribution Q_r with all possible combinations of the remaining pumping stations. These can range from 0% to 100%.
6) Initialize the hydraulic analysis, assigning and testing the possible combinations of Q_r at the remaining stations.

7) If the difference between P_{min} and $P_{Reached}$ is less than or equal to the allowed threshold (U) or the pressure of PS_i is greater than H_{Max}, then, the minimum flow rate that the station can provide is Q_{min}.

8) Select another PS in the network and go to step 3. If $Q_{mín}$ is greater than Q_B the analysis is finished. Else, update $Q_{min}= Q_{min}+1$ and go to step 4.

9) If there are no more pumping stations in the network, increase the height of PS1. The value of the increase will determine the number of points on the curve. A small increment will mean more computational effort while very large values can affect the performance of the algorithm.

10) If H_{design} is greater than H_{max} the analysis must finish, else go to step 2.

Subsequently, the analysis is carried out for the maximum flows in a similar way. First, replace in step 1 the period of lower demand for the greatest network demand. Second, in step 7, the pressure condition must be deleted.

All the points that are outside the calculated curves give rise to infeasible solutions, however, it is not possible to ensure that all the points between the curves are feasible. That is, there may still be combinations of flow distributions that can be classified according to section 2.2. These solutions must be discarded by the algorithm during the respective evolutionary process.

2.4 Decoding solution

Once the curves have been calculated they are used by the evolutionary algorithm in each evaluation of the objective function. In this new method, the value of x_{ij} represents a fraction of the difference between the highest value of the maximum flow curve and the lowest value of the minimum flow curve. Figure 1 shows how it is possible to determine the pressure of PS1 from the intersection of the flow fraction and the minimum curve. At this point, PS1 provides enough head to reach the minimum pressure in the entire network. On the left, the curves are bounded by the PS1 level and on the right, the curves extend until they reach the maximum height that a pump can provide in the catalog. Outside this range, it would be impossible to achieve a technically feasible solution.

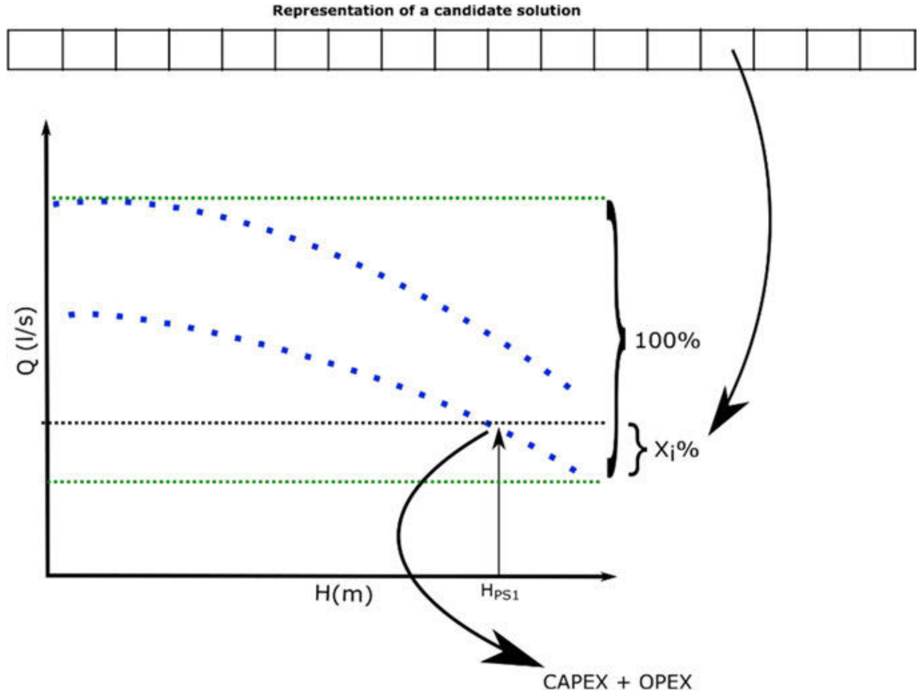


Figure 1: Decoding a solution

2.5 Case Study

To apply the methodology described above, one case study was conducted. Figure 2 shows a new WDN called Curicó Network.

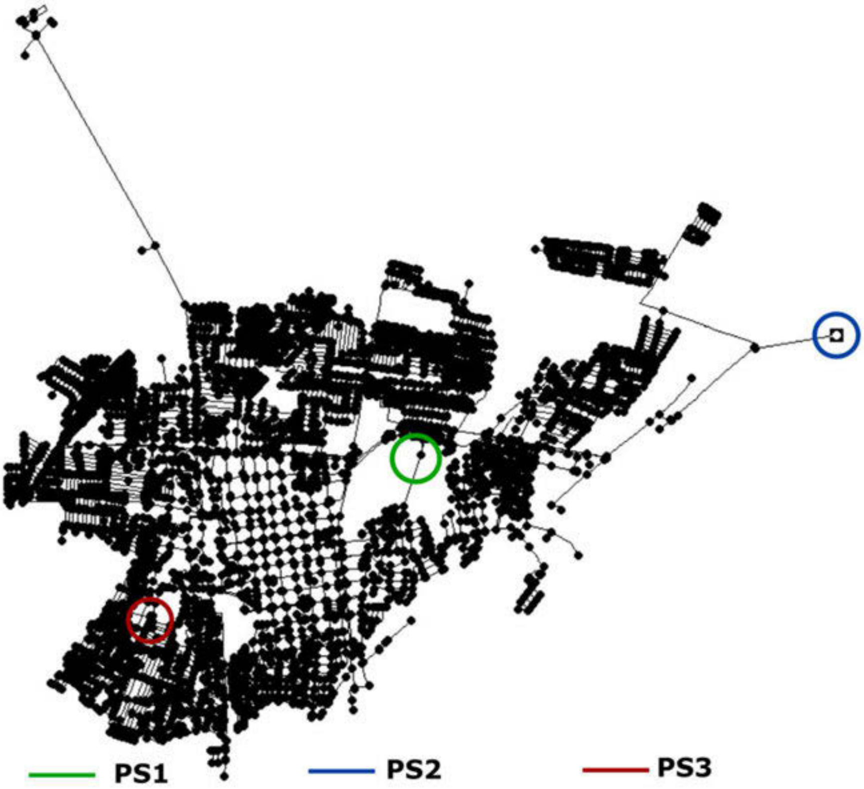


Figure 2: Red de Curicó

This network has 3 PSs (PS1, PS2 and PS3), 7630 consumption nodes, and 8359 pipes. A hydraulic analysis was carried out for one day, and a time step of one hour was considered. The demanded average flow rate is 172 L/s, the minimum pressure at the node is 15m and the roughness coefficient is 0.1. The pattern used to characterize time variation in demand is shown in Figure 3.

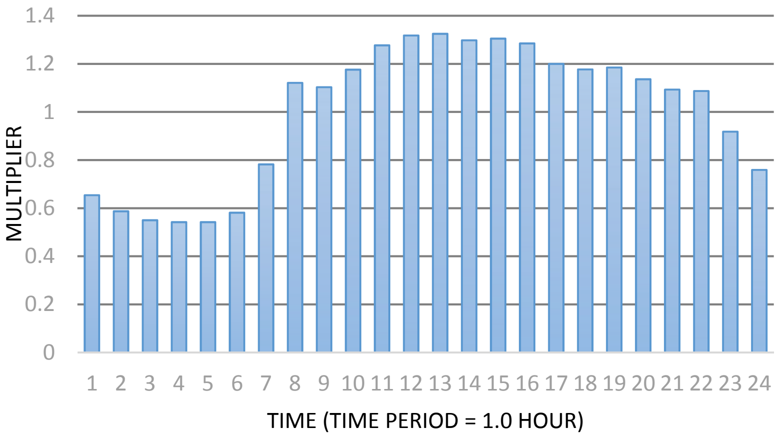


Figure 3: Demand pattern Curicó network

To calculate OPEX, Table 1 shows the hourly electricity rate used for each PS in the network. Por otro lado, todos los coeficientes necesarios para estimar CAPEX fueron obtenidos desde [8] . The maximum flow rate of the pumps in the database varied between 9 L/s and 50.7 L/s.

BY NC SA

Table 1: Electricity for the case study (€/kWh)

Time(h)	PS1	PS2	PS3
1–8	0.094	0.092	0.09
9–18	0.133	0.131	0.129
19–22	0.166	0.164	0.162
23–24	0.133	0.131	0.129

To perform the optimization process, a database with 67 pump models was used. The annualized costs of these models were calculated using an interest rate of 5% per year and a projection time of 20 years. This led to an amortization factor Fa= 7.92%.

2.6 Optimization Method

The objective of this work is to show the advantages of the application of the incorporation of the calculation of the maximum and minimum curves as a support tool for evolutionary algorithms. For this comparison to be fair, the resolution of the proposed model was carried out using the same algorithm used in [8]. Specifically, it is the pseudo genetic algorithm (PGA) developed by the authors of [13]to solve problems of an integer nature. In this way, it is possible to compare directly with the methodology proposed in avoiding unnecessary biases. In addition, the same parameters of population size (P), crossover frequency (Pc), and mutation frequency (Pm) recommended by the authors of the previous investigation were considered.

The implementation was developed according to the instructions exposed in [14] This system can conduct massive simulations and is integrated with the hydraulic network solver EPANET using the programmer's toolkit [15].

3 RESULTS

After applying the methodology proposed in section 2.3, it was possible to calculate the minimum and maximum flow curves for the PSs of the Curicó network. Figures 4 and 5 show in red the area of infeasible solutions during the period of highest demand. And in blue, is the area where there is the possibility of finding feasible solutions. In both cases about 80% of the total solutions are infeasible. It is necessary to consider that PS1 does not appear in this analysis because it is considered by the methodology as the main station. That is, it provides all the flow that has not been supplied by the rest. In fact, PS1 is not a direct part of the representation of the solutions but corresponds to the complement of the work done by PS2 and PS3.

In the case of the PS3 analysis, if PS1 has a height greater than 312m, the minimum, and maximum curves are equal to zero. That is, from that limit PS3 cannot operate. Otherwise, the supply of water to the network will exceed the demand. Unlike PS3, the PS2 curves have a regular shape. Given the topographic conditions of the network, it is possible to appreciate that PS3 is very sensitive to the contributions of the rest of the stations.

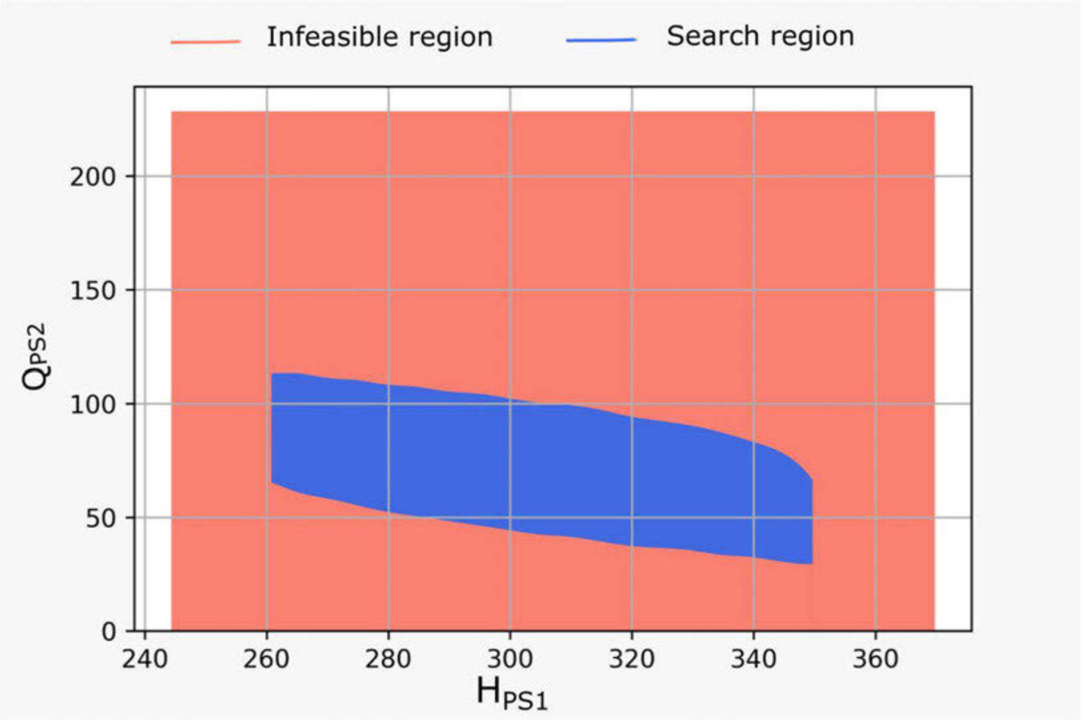


Figure 4a: Region infeasible vs search region for PS2

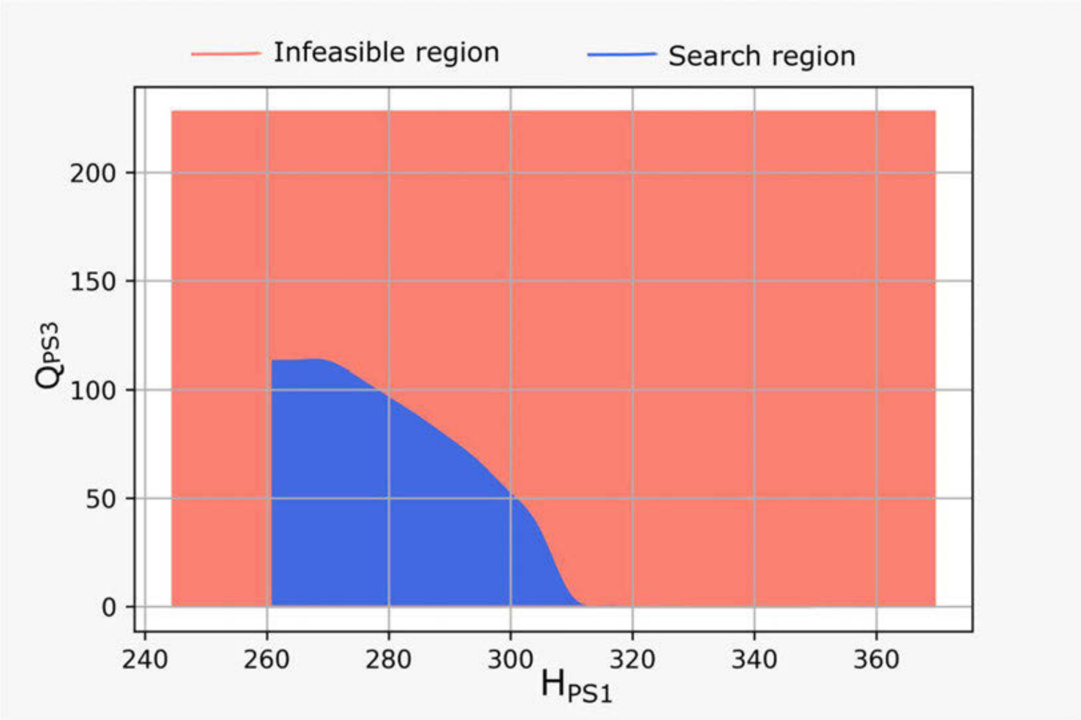


Figure 4b: Region infeasible vs search region for P3

Figure 5 shows the superposition of both curves, this analysis is important because it allows determining the influences that the behavior of one station has on the other. During the evaluation of the objective function, the PS are randomly ordered for the decoding of the solutions. In this way, it is possible to rule out infeasible solutions caused by the intersection of the two curves.

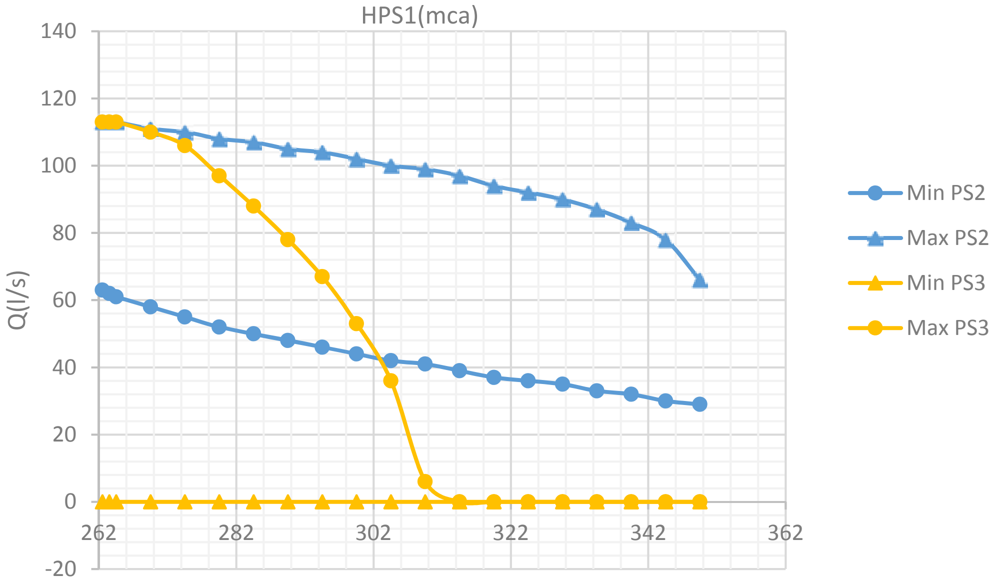


Figure 5: PS2 and PS3 curve overlay

Once the curves are calculated, the comparison of the PGA implemented in the methodology proposed in [8] vs the incorporation of the calculation to delimit the ranges of x_{ij}. Figure 6 shows the evolution of the values of the objective functions during 25,000 evaluations for the Curicó network problem.

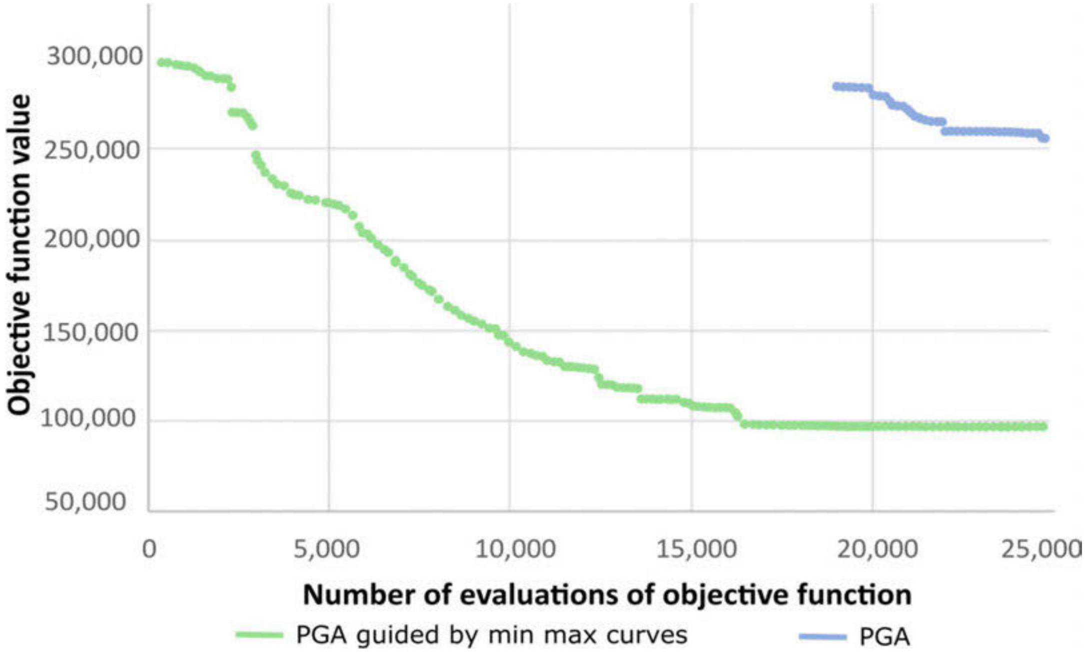


Figure 6: PGA Performance Comparison

4 DISCUSSION

The results of Figure 6 show a significant improvement in the performance of the implemented algorithm. The curve-guided algorithm finds feasible solutions very quickly and the nearly 16,000 evaluations of the objective function manage to converge. While the algorithm without the support of the minimum and maximum curves takes almost 18,000 evaluations alone to find a solution. Also, the convergence process is slower because it evaluates many infeasible solutions until a better solution is found.

On the other hand, in the work presented by [8] the calculation of the setpoint curve is used as a tool to determine the minimum energy point that must be provided by the pumps. This is an iterative procedure where each iteration of the loop must perform a computational simulation. On the other hand, in this work the setpoint curve is not used, instead, the curve of minimums calculated allows for approximate in a single iteration of the value of the height of the main PS. Therefore, this procedure is particularly advantageous in large networks.

In this case study, the results are significantly better. However, the new method presented supposes a computational effort before the implementation of an evolutionary algorithm. Therefore, it is necessary to add all the evaluations that were carried out to obtain the curves. Specifically, in the case study, these evaluations are approximately 10% of all the evaluations carried out. It is an iterative procedure that allows mapping infeasibility zones that are only performed once and can be used by any evolutionary algorithm to speed up the search for solutions. It is probable that in case studies where the number of nodes and pipes does not represent a problem, the implementation of this previous calculation will not be necessary. It could mean an increase in the number of computational simulations carried out.

5 CONCLUSIONS

The design of pumping stations in water distribution networks is a complex problem that can be optimized from different approaches. Decisions during design will affect investment costs and operational costs throughout the life of the project.

That is why the authors of [8] proposed a methodology to minimize investment and operational costs to it. And they showed that it is possible to completely design all the PSs in the network using the created model. However, this methodology is not very efficient compared to real-size networks where it is necessary to know the information of all the nodes in each evaluation of the objective function. To avoid this problem, a new methodology was proposed in this paper to speed up the search process for solutions of evolutionary algorithms compared to the mathematical model proposed above.

A systematic procedure was designed to map the infeasibility zones before running the algorithm. Then, the solutions are decoded using the result of the previous step, avoiding the evaluation of previously determined infeasible solutions. It is important to keep in mind that this procedure allows ruling out infeasibility zones, however, it does not ensure that all the solutions found within them are feasible.

The results show that the application of this new methodology can considerably reduce the number of evaluations of the objective function needed to find feasible solutions in a real size network. However, as discussed in the paper, the presented work has some limitations. In the case of small networks, the computational effort to find the infeasibility zones can generate an excess of evaluations of the objective function to achieve the same result.

Finally, experimentation with different types of networks, algorithms and parameters will be part of the future work that will allow determining the scope of the new proposal.

6 REFERENCES

[1] H. Mala-Jetmarova, N. Sultanova, and D. Savic, "Lost in optimisation of water distribution systems? a literature review of system design," Water (Basel), vol. 10, no. 3, Mar. 2018, doi: 10.3390/w10030307.

[2] T. M. Müller, P. Leise, I.-S. Lorenz, L. C. Altherr, and P. F. Pelz, "Optimization and validation of pumping system design and operation for water supply in high-rise buildings," Optimization and Engineering, vol. 22, no. 2, pp. 643–686, 2021, doi: 10.1007/s11081-020-09553-4.

[3] L. J. Blinco, A. R. Simpson, M. F. Lambert, and A. Marchi, "Comparison of pumping regimes for water distribution systems to minimize cost and greenhouse gases," Journal of Water Resources Planning And Management, vol. 142, no. 6, Jun. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000633.

[4] Y. Makaremi, A. Haghighi, and H. R. Ghafouri, "Optimization of pump scheduling program in water supply systems using a self-adaptive NSGA-II; a review of theory to real application," Water Resources Management, vol. 31, no. 4, pp. 1283–1304, Mar. 2017, doi: 10.1007/s11269-017-1577-x.

[5] H. Mala-Jetmarova, N. Sultanova, and D. Savic, "Lost in optimisation of water distribution systems? a literature review of system operation," Environmental Modelling & Software, vol. 93, pp. 209–254, Jul. 2017, doi: 10.1016/j.envsoft.2017.02.009.

[6] C. F. León-Celi, P. L. Iglesias-Rey, F. J. Martínez-Solano, and D. Savic, "Operation of multiple pumped-water sources with no storage," Journal of Water Resources Planning and Management, vol. 144, no. 9, p. 04018050, 2018, doi: 10.1061/(asce)wr.1943-5452.0000971.

[7] I. Pulido-Calvo and J. C. Gutiérrez-Estrada, "Selection and operation of pumping stations of water distribution systems," Environmental Research Journal, vol. 5, no. 3, pp. 1–20, 2011.

[8] J. H. Gutiérrez-Bahamondes, D. Mora-Meliá, P. L. Iglesias-Rey, F. J. Martínez-Solano, and Y. Salgueiro, "Pumping station design in water distribution networks considering the optimal flow distribution between sources and capital and operating costs," Water , vol. 13, no. 21, Nov. 2021, doi: 10.3390/w13213098.

[9] O. Fecarotta and A. McNabola, "Optimal location of pump as turbines (pats) in water distribution networks to recover energy and reduce leakage," Water Resources Management, vol. 31, no. 15, pp. 5043–5059, 2017, doi: 10.1007/s11269-017-1795-2.

[10] N. Nagkoulis and K. L. Katsifarakis, "Minimization of Total Pumping Cost from an Aquifer to a Water Tank, Via a Pipe Network," Water Resources Management, vol. 34, no. 13, pp. 4147–4162, Oct. 2020, doi: 10.1007/s11269-020-02661-x.

[11] C. Leon-Celi, P. L. Iglesias-Rey, and F. Javier Martinez-Solano, "Energy optimization of supplied flows from multiple pumping stations in water distributions networks," Procedia Engineering, vol. 186, pp. 93–100, 2017.

[12] C. Leon-Celi, P. L. Iglesias-Rey, F. Javier Martinez-Solano, and D. Mora-Melia, "A methodology for the optimization of flow rate injection to looped water distribution networks through multiple pumping stations," Water (Basel), vol. 8, no. 12, 2016, doi: 10.3390/w8120575.

[13] D. Mora-Melia, P. L. Iglesias-Rey, F. J. Martinez-Solano, and V. S. Fuertes-Miquel, "Design of water distribution networks using a pseudo-genetic algorithm and sensitivity of genetic operators," Water Resources Management, vol. 27, no. 12, pp. 4149–4162, Sep. 2013, doi: 10.1007/s11269-013-0400-6.

[14] J. H. Gutierrez-Bahamondes, Y. Salgueiro, S. A. Silva-Rubio, M. A. Alsina, D. Mora-Meliá, and V. S. Fuertes-Miquel, "jHawanet : An Open-Source Project for the," no. September, 2019.

[15] L. a. Rossman, "EPANET 2.0 User's manual (EPA/600/R-00/057)," 2000.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14829

NON-INTRUSIVE WATER USAGE CLASSIFICATION CONSIDERING LIMITED TRAINING DATA

Pavlos Pavlou[1], Stelios Vrachimis[1,2],
Demetrios G. Eliades[1], Marios M. Polycarpou[1,2]

[1]KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
[2] Department of Electrical and Computer Engineering, University of Cyprus, Cyprus

[1], pavlou.v.pavlos@ucy.ac.cy, [2], vrachimis.stelios@ucy.ac.cy,
[3], eldemet@ucy.ac.cy, [4], mpolycar@ucy.ac.cy

Abstract

Smart metering of domestic water consumption to continuously monitor the usage of different appliances has been shown to have an impact on people's behavior towards water conservation. However, the installation of multiple sensors to monitor each appliance currently has a high initial cost and as a result, monitoring consumption from different appliances using sensors is not cost-effective. To address this challenge, studies have focused on analyzing measurements of the total domestic consumption using Machine Learning (ML) methods, to disaggregate water usage into each appliance. Identifying which appliances are in use through ML is challenging since their operation may be overlapping, while specific appliances may operate with intermittent flow, making individual consumption events hard to distinguish. Moreover, ML approaches require large amounts of labeled input data to train their models, which are typically not available for a single household, while usage characteristics may vary in different regions. In this work, we initially propose a data model that generates synthetic time series based on regional water usage characteristics and resolution to overcome the need for a large training dataset with real labeled data. The method requires a small number of real labeled data from the studied region. Following this, we propose a new algorithm for classifying single and overlapping household water usage events, using the total domestic consumption measurements. The classification procedure is described below: 1) During the offline feature learning stage, a dataset of labeled data corresponding to water-use profile signals is analyzed to some predefined features, such as event volume, event duration, event flow peak, and event signature, to extract its statistical properties, 2) The event classification stage monitors the provided measurement time-series for events between zero-flow intervals. The identified events are then classified using Dynamic Time Wrapping and an optimization procedure that finds the best label for the observed event based on the features learned in the first stage and similarity indices. Non-classified events are processed using a variation vector technique to identify the combined events which are then split into sub-single events and classified.

Keywords
Non-intrusive water usage classification, device disambiguation

1 INTRODUCTION

The increasing water consumption due to population growth and excessive urban development is creating an unbalanced situation between water demand and supply [1]. Adding to this, the need for continuous water supply and sufficient pressure during peak times puts even more burden on the water utilities that must face these challenges [2]. Among others, water demand management practices have been proposed as a response to these problems, aiming to ensure water demand needs are met constantly while promoting water conservation [3].

New advancements in sensor technology for collecting, analyzing, and transmitting high-resolution data to both utilities and consumers, are considered important tools for water management [4]. Smart metering of domestic water consumption to continuously monitor the usage of different appliances has been shown to have an impact on people's behavior towards water conservation [5] and can be a useful tool for water utilities in managing demand during peak hours and drought periods thus eliminating the need for further investment in upgrading the water infrastructure [6].

Smart metering can be categorized into intrusive and non-intrusive metering. Intrusive metering considers the installation of a sensor in each water-consuming appliance (e.g., dishwasher, toilet, shower) while non-intrusive metering considers the installation of only one sensor on the main water supply pipe of a house thus measuring the total household consumption. Although intrusive metering offers more insight into consumer habits, the installation of multiple sensors to monitor each appliance may have a high initial cost and may be inapplicable due to practical considerations [7].

Real-time data that are available through new smart metering systems must be coupled with data analytic techniques and intelligent algorithms to play a significant role as a decision-making tool and to have an impact on water demand management and water conservation. Disaggregation algorithms process the data retrieved through non-intrusive metering and identify which water end-use appliance is active by analyzing the total water consumption signal. Identifying which appliances are in use through non-intrusive water usage classification is challenging since their operation may be overlapping while specific appliances may operate with intermittent flow making individual consumption events hard to distinguish.

Water end-use disaggregation belongs to the general spectrum of time series classification problems. Time series classification is extensively addressed using machine learning and deep learning methodologies which require large training datasets [8] as well as with pattern recognition techniques based on similarity measurements such as Dynamic Time Wrapping (DTW) [9] and Longest Common Subsequence (LCSS) [10] that generally require a reference dataset. Various studies have been conducted to address the challenge of water end-use classification using smart water metering. In a first approach (Trace Wizard and Identiflow), decision tree methods were applied for water end-use classification which required significant data [11,12]. In [13,14], the authors suggested the use of pressure sensors combined with a Bayesian approach to identify water usage events (Hydrosense). These approaches required a high initial cost for the deployment of the sensor network and did not achieve high accuracy. Non-intrusive metering combined with machine learning methods were further used to disaggregate water end-use events. The authors in [15] proposed the use of an adaptable neuro-fuzzy network to classify water end-uses achieving high accuracy, using a limited dataset of flow measurements. In more recent studies, machine learning and data analytic algorithms were developed to address the problem of water end-use disaggregation, with promising results [16–21]. Several drawbacks that were noted in these studies include the need for a large amount of historical data to train the model and the absence of disaggregation techniques for combined water events. A notable study by [22] (Autoflow) addressed the aforementioned drawbacks using a hybrid combination of Hidden Markov Models, Artificial Neural Networks, and DTW algorithms, which was further improved to avoid the need of collecting new use-data for different regional use cases [23]. The "Autoflow" model addresses the classification of single and combined water end-use events with 85.9-96.1% and 81.8-91.5% accuracy respectively. However, as stressed by the authors, more regional data are needed to improve the performance of this method.

This work has two main contributions:

- Proposes an approach for calibrating an existing synthetic time-series data generator based on regional water usage characteristics and resolution. The generated data can be used to train Machine Learning algorithms without the need of collecting real labeled data for long periods from pilot studies.

- The development of a new methodology for classifying single and overlapping household water usage events within the same dataset using non-intrusive metering. The proposed approach takes into consideration water end-use events which exhibit intermittent or non-uniform flow.

The paper is structured as follows: Section 2 describes the data models, Section 3 provides the proposed classification methodology, Section 4 presents the performance of our classification approach and in Section 5 we conclude the paper and discuss some future extensions.

2 DATA MODEL

2.1 Available usage characteristics model

In this study, we use the available usage characteristics incorporated in the *STochastic Residential water End-use Model* (STREaM) introduced by [24]. STREaM is a modeling software that generates synthetic time series of data of a household with up to 10s resolution and it was calibrated on a large dataset including observed and disaggregated water end-uses from more than 300 single-family households in nine U.S. cities [25]. Each of the water end-uses considered in the STREaM dataset (toilet, shower, faucet, clothes washer, dishwasher) is characterized by its signature (i.e., typical consumption pattern) and the probability distributions of the water event volume, the single-use durations, the number of uses per day and the time of use during the day. The number of events per day is modelled using the negative binomial and Poisson distributions, the event start time with the Kernel distribution, and the event volume and duration with two-component Gaussian Mixtures. The probability distributions are created by taking into consideration the number of house residents and the efficiency of each appliance (standard or high efficiency.

The STREaM data model requires as inputs the number of household occupants, the available water appliances with their corresponding efficiency level, the simulation time, and the data resolution. Following, it generates time series of each water end-use and their sum as the total household water consumption based on the following procedure: i) samples the number of events for each water end-use and each day of the simulation time using the Monte-Carlo method from its probability distributions, ii) samples using the Monte-Carlo method the event-usage characteristics, duration, volume and time of use form their probability distributions, iii) randomly chooses one of the available signatures of the selected water end-use, iv) scales the duration and magnitude of the signature to match the pre-selected event duration and volume, and v) positions the newly created event time-series in the total event time series of the selected water end-use according to its start time.

2.2 Model calibration using limited regional data

The data model proposed in this work extends the STREaM data model to generate synthetic data based on regional water usage characteristics. For this, we assume that we have water usage data from a limited number of households within the region. We use a 1-week dataset from a single-family house in Cyprus to update the existing signatures and generate data with up to 1s resolution. The regional dataset includes data from the following appliances: toilet, shower, faucet, clothes washer and dishwasher. We assume that these data have been correctly classified per their usage and were collected at a resolution of 1s. Finally, it is assumed that no leakages exist in the recorded data.

The drawback of having a small dataset is that we may not be able to identify the probability distribution describing event occurrence, volume, and duration. However, the characteristic signatures of events can be identified even from this small dataset which are more representative of the appliances and local usage characteristics. Thus, the main approach for the development of the data model relies on updating the existing signatures with regional signatures from the case study. In addition, during the last step of the event generation process which includes the scaling of the duration and magnitude of the selected signature, boundaries were applied to ensure that generated events comply with the consumption flow rate indicated by the regional signatures. For example, during the sampling process of the usage characteristics, our model could pick a water end-use with a short-time duration and large volume resulting in an event with an inconsistent consumption pattern compared to the regional signature.

In the following paragraphs, we describe the methodology for the creation of regional consumption patterns. Signatures from the regional labeled dataset are extracted using a hybrid combination of DTW algorithm, k-medoids clustering method evaluated based on the "Silhouette index" and an affinity search technique. We use DTW in a clustering procedure to extract water end-use signatures from the regional dataset. The partitioning algorithm k-medoids splits the time series dataset into k clusters based on the minimum distance between the points of a cluster and a specified point at the center of the cluster and can be considered faster than other clustering methods [26]. The silhouette method measures the consistency of each cluster by comparing the similarity of an object to its cluster, compared to the remaining clusters of the group [27]. Silhouette score ranges from -1 to +1, with high values indicating a better fit of the object to its predefined cluster.

Initially, the time series (events) of each water appliance are extracted from the dataset, pre-processed to remove potentially faulty sensor measurements, and normalized to avoid scale differences. A similarity matrix for each group of events is obtained using DTW followed by k-Medoids clustering. The "Silhouette index" is used to define the number of clusters per fixture and the prototype signature is generated using a similarity search technique.

Each event time-series $S_t = \{S_1, .., S_n\}$ comprised of n flow data points, is normalized to have a zero mean and standard deviation of one, thus being invariant to scale and offset, as follows:

$$\tilde{S}_t(t) = \frac{S_t(t) - m}{\sigma} \tag{1}$$

where the arithmetic mean m is given by:

$$m = \frac{\sum_{t=1}^{n} S_t}{n} \tag{2}$$

and the standard deviation is given by:

$$\sigma = \sqrt{\frac{\sum_{t=1}^{n} (S_t - m)^2}{n}} \tag{3}$$

The similarities between the time series of each group of events are calculated using the DTW method, resulting in the similarity matrix M of size AxA, where A the number of events per water-end use category, and the matrix elements are calculated as follows:

$$M_{ij} = W(\tilde{S}_i, \tilde{S}_j) \tag{4}$$

where function $W(\cdot)$ calculates the distance between points of two time-series, using the DTW method. DTW is a methodology to measure the shape similarity between two time-series with different lengths. DTW wraps the time axis to align the data points and calculates the optimal alignment between two time-series according to the following equation:

$$W(i,j) = w(i,j) + min\{W(i-1,j), W(i-1,j-1), W(i,j-1)\} \tag{5}$$

where $s = \{s_1, .., s_i, .., s_m\}$ and $t = \{t_1, .., t_j, .., t_n\}$ are the two time series with m and n data points, respectively. Distance metric w is given by $w(i,j) = |s_i - t_j|$ with the possible combinations limited to $(i-1,j), (i-1,j-1), (i,j-1)$. The accumulated DTW distance $W(m,n)$ is considered the optimal alignment between the two time-series, with initial condition $W(1,1) = w(1,1)$.

Following, the time series of each water fixture are grouped into clusters based on their similarity using the k-medoids clustering approach. Since the k-medoids method requires the number of clusters to be defined prior to clustering, the process can be carried out for a given range of clusters (e.g., 2-10 clusters). In order to define the appropriate number of clusters per water fixture, an evaluation method was simultaneously applied using the "Silhouette index".

The last step includes the extraction of the most representative signature of each cluster according to the DTW similarity results [28]. The time series with the lowest total dissimilarity M_{xy} is extracted as the main signature:

$$d_n(S_x) = \left(\frac{\sum_{y \in C_n} M_{xy}}{TC_n}\right), D_n = \min(d_n) \tag{6}$$

where:

M_{xy} : $similarity\ matrix\ bewteen\ time\ series\ S_x\ and\ S_y\ belonging\ to\ cluster\ C_n$

TC_n : $number\ of\ time\ series\ in\ each\ cluster\ C_n$

n : $number\ of\ clusters$

The extracted signature from each water end-use category can be eventually smoothened using polynomial fitting to remove measurement noise caused by the sensor and then stored in the data model. To illustrate the approach, we utilize a water-use dataset collected from a single-family household in Cyprus, in which water consumption was recorded with a 1-second resolution, and the data were labeled as toilet, shower, faucet, clothes washer, and dishwasher. Figure 1 shows the signatures extracted from each cluster of the shower category.

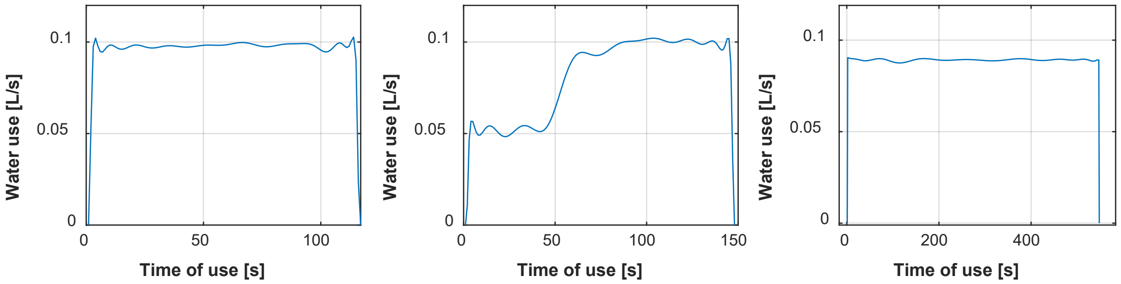


Figure 1: Signature patterns for shower water end-use

2.3 Datasets

Two synthetic datasets with a duration of 45 and 15 days respectively and 1s resolution were produced from the data model considering the following water end-uses: standard toilet, standard shower, standard faucet, high-efficiency clothes washer, and standard dishwasher. The 45-day dataset serves as the training set and the 15-day as the testing set. The training set is used to identify potential usage characteristics for each water end-use category and the testing set to evaluate the performance of the classification model described in the next section.

3 CLASSIFICATION METHODOLOGY

The water end-use classification procedure consists of two main stages.

1. In the first stage, namely *the offline feature learning stage,* the training dataset consisting of labeled data corresponding to water end-use signals is analyzed to extract the statistical properties of some predefined features including event duration, event volume, event flow peak, and event signature.
2. The *event classification stage monitors* the provided measurement time series from the test set for events between zero-flow intervals followed by the single and overlapping event classification. The classification of water end-use event relies on the DTW approach and an optimization procedure that uses similarity indices and statistical bounds extracted from the features learned in the first stage. Classification of events with an intermittent flow such as Dishwashers (DW) and Clothes washers (CW) are further processed considering a time window in the time series analysis that includes the device cycle in its entirety.

3.1 Offline Feature Learning Stage

This stage assumes the availability of inflow data of a residential household labeled according to which appliance is operating. In our case, the training dataset extracted from the data model is analyzed. The algorithm first creates event sets from labeled data by acquiring the observed time-series data with event labels and separating events, creating the set of events E. The events with the same label l are then gathered, creating the subsets of events $E_l \subset E$. The features of event duration, event volume, and event flow peak were extracted. The next step includes the calculation of the 99% confidence intervals of each feature from each water end-use. The statistical analysis showed that sets of data have a skewed distribution, thus the proposed confidence intervals were obtained by filtering out the 1% most distant data points. This was achieved by calculating the absolute distance between each data point and the arithmetic mean of the dataset. We considered that only the generated training dataset is available for the classification model and not all the data that is stored in the data model.

3.2 Event Classification Stage

Overview of the event classification process

This stage distinguishes individual events in the time-series by filtering out data points separated by a zero-flow time interval. The event classification process is applied on the extracted events and consists of the single event classification and the combined event disaggregation and categorization. A combined event includes two or more single events with overlapping operations. Single event classification is performed using a hybrid approach that includes DTW algorithm and criteria based on similarity indices using statistical bounds extracted from the features at the previous stage. Following, the combined event disaggregation takes place using initially a filtering method to split the combined event into sub-events which are then processed through the single event classification procedure. Besides the difficulty in identifying both single and overlapping events another two obstacles that were identified during the process are:

- DW and CW devices have a working cycle that exhibits intermittent flow. Classification of such events was performed using a sliding time window of measurements.
- The existence of single events with a varying flow rate that occurs in rare circumstances can be easily misclassified as combined events. To overcome this problem, a filtered variation vector technique is applied in the combined event classification procedure to identify these events.

The overall classification process is presented in Figure 2.

BY NC SA

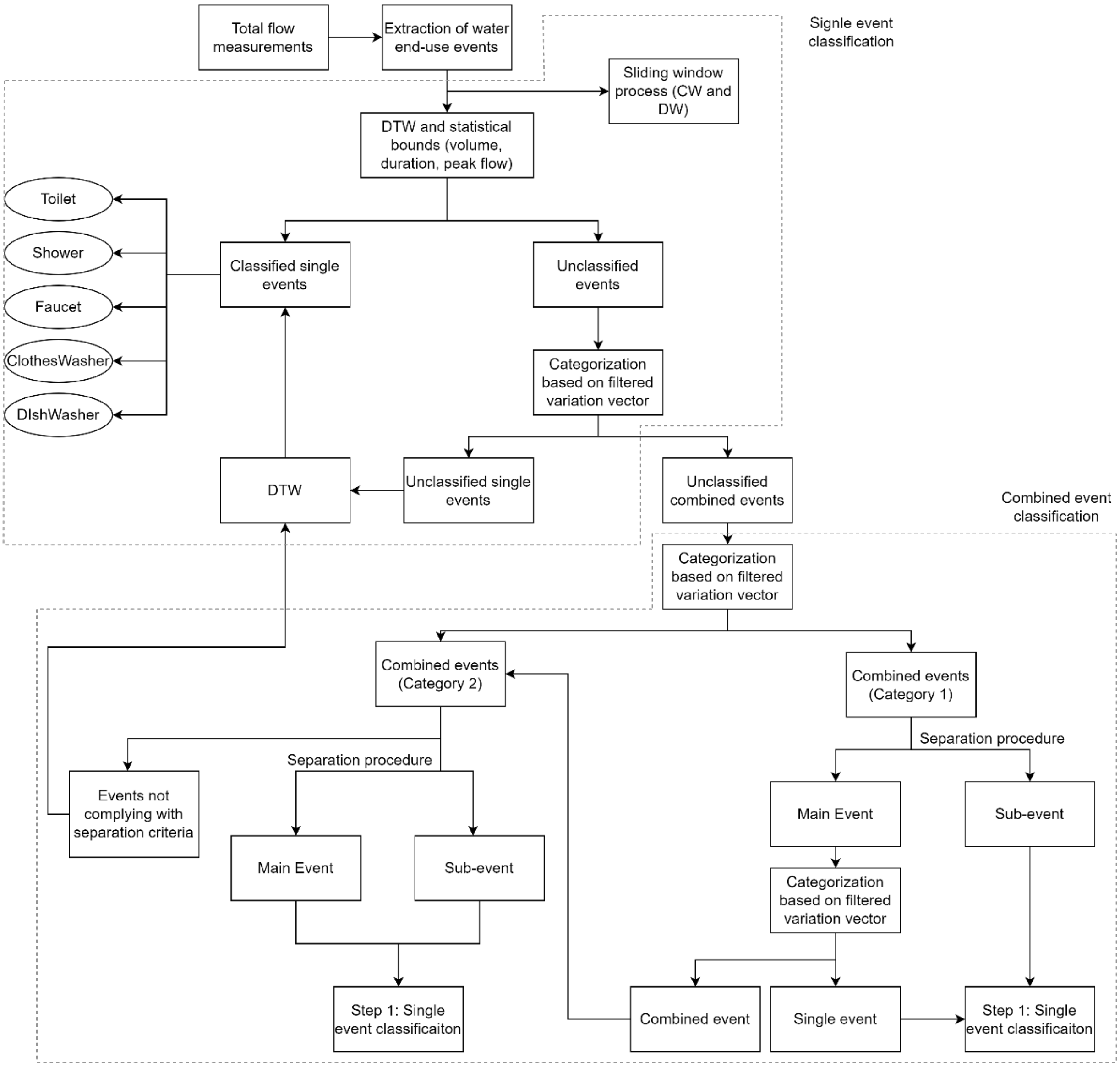

Total flow measurements
Extraction of water end-use events
Signle event classification
Sliding window process (CW and DW)
DTW and statistical bounds (volume, duration, peak flow)
Toilet
Shower
Faucet
ClothesWasher
DIshWasher
Classified single events
Unclassified events
Categorization based on filtered variation vector
DTW
Unclassified single events
Unclassified combined events
Combined event classification
Categorization based on filtered variation vector
Combined events (Category 2)
Combined events (Category 1)
Separation procedure
Events not complying with separation criteria
Separation procedure
Main Event
Sub-event
Main Event
Sub-event
Categorization based on filtered variation vector
Step 1: Single event classificaiton
Combined event
Single event
Step 1: Single event classificaiton

Figure 2: Water end-use event classification process

Single event classification

The proposed single event classification relies mostly on pattern recognition through DTW. Initially, the investigated events and labeled signatures are normalized as described in equation 1. The first task is the detection of potential time windows inside the dataset with the operation of intermittent flow devices such as DW and CW. This is achieved by applying DTW between a sliding time window with a length equal to the full cycle of operation of the selected appliance and its corresponding labeled signatures. From the Cyprus case study pilot, in Figure 3a, the signature of a DW full-cycle operation is presented with a duration of 2793 seconds which corresponds to the time window used for the classification. Bounds of maximum flow criteria are also applied in this task to avoid misclassification of DW or CW time windows.

Following, DTW is applied in all events and distinguishes them into the following categories: toilet, shower, and faucet. Classification of WM and CW single events from their full cycle of operation is performed only within the time windows specified previously. In this case, the labeled signatures of WM and CW devices are broken down forming smaller sub-patterns (Figure 3b). A similarity matrix is created between the investigated event and the available labeled signatures stored in the database. Events with signature similarity above a specific threshold are then labeled. Simultaneously, a screening procedure is performed utilizing the minimum and maximum bounds obtained from features extracted from the training dataset (volume, duration, and peak flow). Any events not complying with the criteria defined through the DTW and the water end-use feature's statistical analysis are marked as unclassified.

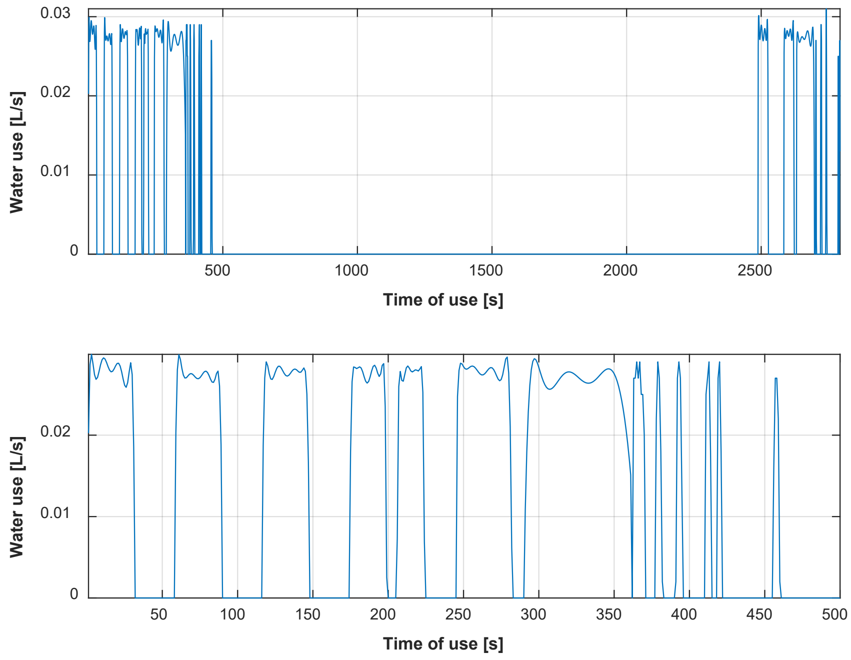


Figure 3: a) Signature of Dishwasher's full operation cycle b) Sub-single events within the Dishwasher's main signature

Unclassified events are then categorized into unclassified single and combined events. The categorization is performed using a filtering technique that detects flow rate changes within an event that exists at a specific threshold. Changes in the flow rates of an event are a good indication that another water-end use event has either been started or completed. The elements of the calculated vector are the differences between adjacent data points within an event, calculated as:

$$v_i = f_{i+1} - f_i, 1 \leq i < n \qquad (7)$$

Where $f = (f_1, f_2, .., f_i, .., f_n)$ the event flow rate points with a duration of n seconds and $v = (v_1, v_2, .., v_i, .., v_{n-1})$ the extracted vector. A threshold is then specified to neglect fluctuations within the vector that do not correspond to the use of a new water appliance. A range of thresholds calculated based on the variation between the maximum flows of labeled events from the training dataset were evaluated and the value of 0.01 L/sec was selected as it achieved the highest accuracy. Unclassified single events are selected as the events which exhibit no fluctuations in the extracted filtered variation vector. The initial and final phases of the filtered vector are ignored since they mark the starting and ending of the event (Figure 4).

The main DTW classification methodology is applied again without using statistical bounds to categorize the unclassified single events. The remaining unlabeled events are considered as combined events and their classification follows in the next step.

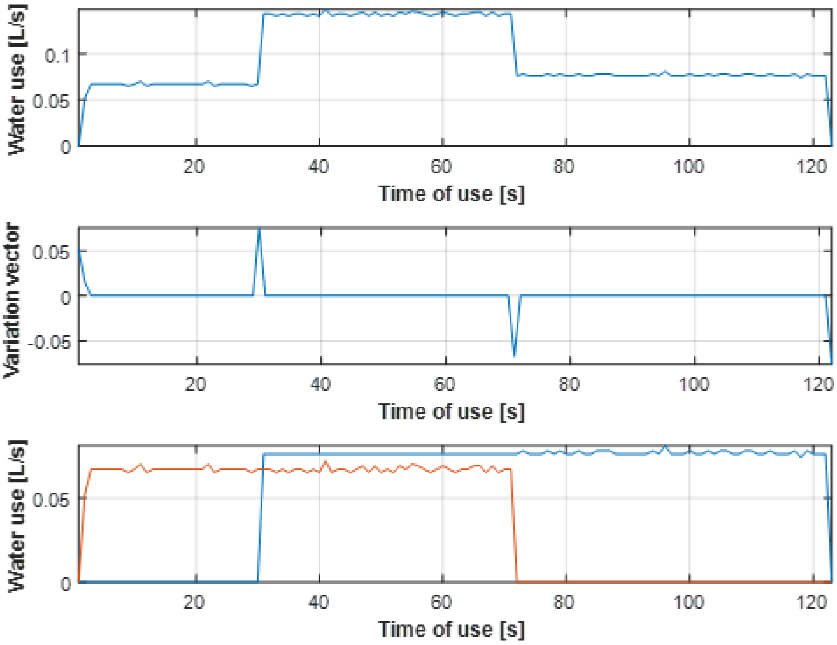


Figure 4: a) Combined event as extracted from the dataset, b) Filtered variation vector of the combined event, c) Sub-events extracted from the original event

Combined event classification

The combined event classification consists of two main tasks, the disaggregation of the combined event into single events and their classification following the approach described previously. Overlapping between events can be expressed in two different categories. The first category includes events overlapping with one sub-event a) starting and finishing before one or more other sub-events and b) starting and finishing after one or more other sub-events. The second category includes sub-events that start and finish within other sub-events.

The first step is the disaggregation of events belonging to the first category. This task is performed using an approach presented in [29], where the last flow-rate drop that corresponds to the finishing time of a combined event is compared to the last flow rate rise. If their difference is below a predefined threshold (a value of 0.005 L/sec resulted in the highest accuracy between a range of thresholds) then it is considered that a single sub-event occurred in the last phase of the combined event. The same principle applies to the starting phase. The sub-event is extracted from the initial combined event and the algorithm calculates its flow rate for the period that it was overlapping with other events. This is achieved by calculating the median flow rate during the period when only the targeted sub-event was active. An example is shown in Figure 4 with a sub-event starting and ending before the second sub-event. The remaining sub-event is evaluated again using the filtered variation vector approach and categorized as a single or combined event. If identified as a combined event, then it is included in the second category and processed as follows.

The second step includes the disaggregation of combined events included in the second category using the filtered variation vector defined previously. In this case, the algorithm searches within the filtered vector to identify the positions where a zero value is followed by a positive value and the positions where a negative value is followed by a zero value. These positions indicate the beginning and finishing of a sub-event within the combined event. The first “starting” position is matched with the first “finishing” position and the sub-event is separated from the base combined event. Events included in this category that do not meet these conditions (including at least one “starting” and “finishing” point) but they do present considerable fluctuations in their flow rate, are considered as single events and are then processed to the single event classification procedure with the use only of the DTW method. With this technique, single events with a varying flow rate that can be presented in real datasets (Figure 5), can be distiguinshed from combined events.

In the third step, the classification of the sub-events and the left-over (the remaining event after the separation process) base combined event extracted from the two previous steps takes place using the single event classification. Any events not classified are processed again through the combined event classification procedure.

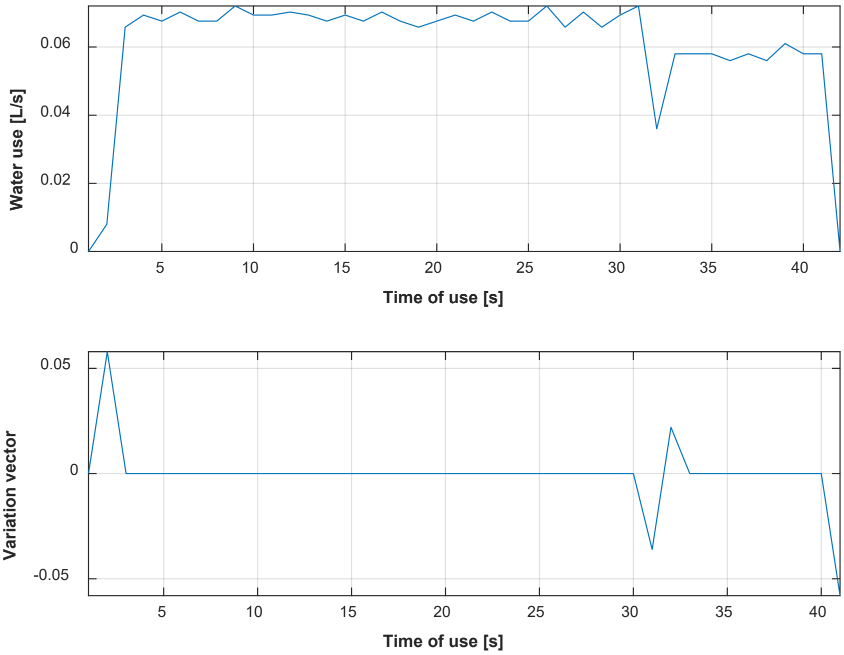


Figure 5: Example of a single event initially misclassified as a combined event. a) Original event as extracted from the dataset, b) Filtered variation vector of the event

4 RESULTS

Evaluation metrics

The macro f1-Score [30], a widely accepted metric that takes into consideration both the algorithm’s precision and recall, is used:

$$\text{Macro f1-score} = \frac{2 \times precision \times recall}{precision + recall} \tag{8}$$

Precision indicates the percentage of true positive indices among the total number of positive indices classified by the model:

$$precision = TP/(TP + FP) \tag{9}$$

and *recall* measures the amount of correctly labeled positive cases among the total number of positive cases:

$$recall = TP/(TP + FN) \tag{10}$$

TP, TN, FP, FN correspond to the number of true positives, true negative, false positive, and false negative events. The combination of the model's precision and recall makes F1-score less sensitive to imbalance classification scenarios and reaches its best value at 1 and worst score at 0. Testing accuracy is presented in terms of the number of events and consumption volume.

A confusion matrix is used to visually present the algorithm's performance by illustrating the number of correctly predicted events against the actual number of events.

Confidence intervals

The 99% confidence intervals were calculated from the statistical analysis of the three predefined features extracted from the training set (Table 1). For the DW and CW devices, the statistical analysis refers to the sub-single events that comprise a full cycle of operation. Toilet, faucet, and CW events have similar event characteristics, specifically for consumption duration and peak flow. Similarly, the calculated event volume bounds are identical as well, although CW can generate lower volume events than toilets and faucets. On the other hand, shower and DW events have more distinctive characteristics than the other categories which play a significant role in the classification process. Shower events have a longer duration, larger consumption volume, and a maximum flow higher than other categories. DW operation on the other side results in small events with low consumption and the lowest peak flow that can easily be distinguished from other appliances.

Table 1: 99% confidence intervals obtained for the water end-use features: volume, duration, peak flow

	Toilet	Shower	Faucet	CW	DW
Duration (s)	10-190	90-880	10-170	1-139	1-85
Volume (L)	0.66-9	13-90	0.43-10	0.03-11.85	0.002-2.22
Peak flow (L/s)	0.04-0.10	0.09-0.15	0.02-0.11	0.06-0.13	0.004-0.03

Classification results

The test set comprised of 1323 single and 22 combined events for a period of 15 days. The proposed approach has shown high accuracy (99%) in distinguishing the single events from the set of events while a lower F1-score of 69% was achieved for the combined event categorization although 77% of the combined events were correctly classified (Table 2). This is explained due to the existence of single events with a varying flow rate which were misclassified as combined events thus reducing the algorithm's precision. The calibration of the data model, which includes a large database of volume and duration features with regional water-end use signatures resulted in the development of a realistic dataset that included a few events with non-uniform consumption patterns. It was decided to keep these events in the dataset since they can indeed be presented in real conditions. An example is presented in Figure 5, showing a faucet event with an irregular flow trace. Although this event is considered rare, it is very realistic since it can be presented during the use of a single faucet (e.g during plate washing).

Table 2: Accuracy results in distinguishing single and combined events

	Single Events	Combined Events
Recall (%)	99.2	77.3
Precision (%)	99.6	63.0
Macro f1-score (%)	99.4	69.4

Single events

Table 3 presents the results from the classification of single events in terms of the number of events and event volume. Scoring ranges from 83% to 98% in terms of the number of events and 84% to 99% in terms of volume. Single event classification precision is also presented through the confusion matrix (Figure 6) among the percentage of misclassified events per category.

Toilet: The model demonstrates an accuracy of 84% in classifying toilet events with 87% of the total toilet events being identified. In terms of volume, we notice a total score of 90% with approximately 91% of the total water volume consumed to be correctly calculated. Toilet events were mainly distinguished from the rest of the events due to their fixed mechanical operation/signature which was identified by the DTW algorithm. A few toilet events were misclassified with faucet events as presented in the confusion matrix due to similarity between their usage characteristics.

Shower: The highest recall score in terms of the number of events and volume was achieved for the shower appliance (100%) mainly due to its distinctive consumption volume, duration, and pattern characteristics. This score indicates that all shower events were correctly classified. The precision regarding the number of events, in this case, is lower (77%) though due to misclassification with faucet events. This occurs due to the presence of a small number of shower events with a short duration. This misclassification is not considered a limitation since the algorithm precision in terms of volume is considerably high (91%). The overall score for this category reaches 87% and 95% accuracy in terms of the number of events and volume, respectively.

Faucet: An 83% accuracy was achieved for faucet event classification with an 81% recall score regarding the total number of classified faucet events and 79% recall score for their corresponding volume. The lower score in terms of volume is explained by the misclassification of some single events as combined. As previously explained, a small number of single faucet events were misclassified as combined events due to their flow trace variation. Although in small number, these events had a considerably larger volume than typical faucet events which explained the variation between the two scoring categories.

Clothes washer: The model has also been able to correctly classify most of the CW events with 91% accuracy. The few misclassified events were confused with faucet events. The high score indicates the effectiveness of applying a sliding window to detect the full operation cycle of intermittent flow devices.

Dishwasher: Regarding the DW category, the model demonstrates the highest accuracy for both scoring categories (98-99%). Approximately all DW events were identified with the corresponding algorithm precision reaching 100%. The distinctive usage characteristics of DW events obtained from the statistical analysis along with the application of DTW using sliding windows proved to be highly efficient in detecting such events.

Table 3: Single event classification accuracy in terms of number of events and volume

Number of events / Volume	Toilet	Shower	Faucet	CW	DW
Recall (%)	86.7/91.4	100/100	81.4/78.7	91.3/91.0	95.7/98.7
Precision (%)	81.6/88.6	76.9/91.1	85.3/90.4	90.1/90.5	100/100
Macro f1-score (%)	**84.1/90.0**	**87.0/95.3**	**83.3/84.2**	**90.7/90.8**	**97.8/99.4**

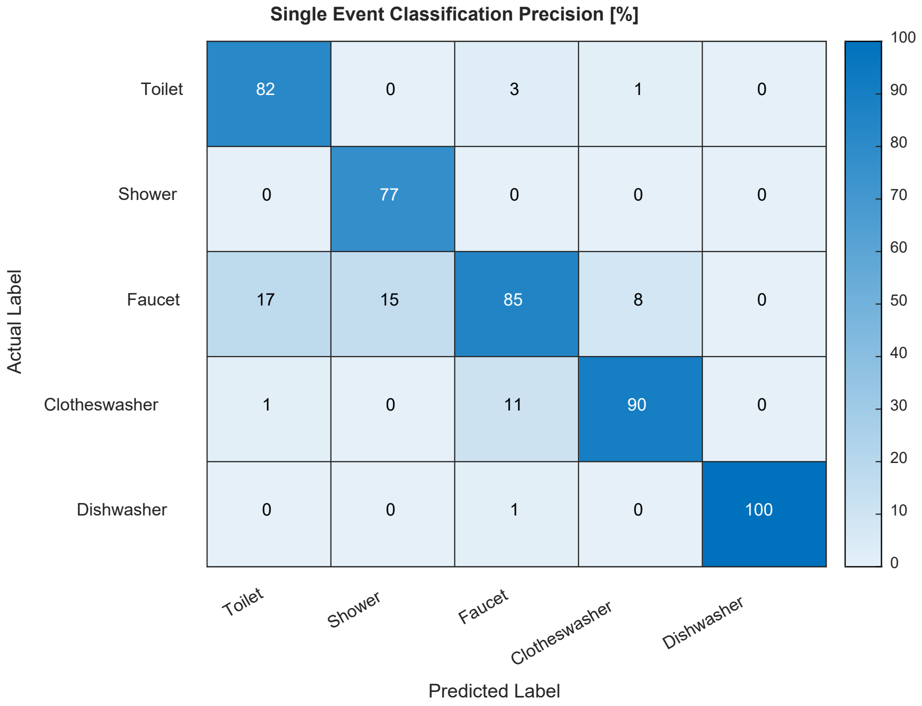


Figure 6: Confusion matrix for single event classification precision (number of events)

Combined events

As stated in Table 2, the algorithm correctly identified 17 out of 22 combined events (recall of 77%) using the proposed approach. The following approach consisting of the separation process and the classification of the extracted sub-events demonstrated an accuracy of 70%. Filtering out single events within combined events, which can be occurring completely at the same time or starting and finishing at the same time is considered a challenging task that needs to be further investigated. The extraction of sub-events under these circumstances is not always accurate, and the imbalance between the number of sub-events and single events can explain the lower classification score. Further improvements can be considered in the separation process to reach a higher precision of combined event separation and classification.

5 CONCLUSIONS AND FUTURE WORK

In this work, we initially presented an approach of extracting water end-use signatures from a limited real labeled dataset to calibrate our data model on regional water usage characteristics and resolution. The developed data model gives us the ability to use an existing large database of water end-use features from STREaM including event duration, volume, and number of events per day, and produce synthetic time series of events with regional consumption patterns. The method

BY NC SA

requires a small number of real labeled data from the target region. Following, a water end-use classification procedure is presented considering non-intrusive monitoring. The developed approach addresses the main difficulties of this challenging problem such as identifying overlapping events, devices with intermittent flow, and single events which exhibit a non-uniform consumption pattern. In the proposed hybrid approach, we use sliding windows, DTW, and confidence intervals to identify active water end-uses with accuracy ranging between 84-99% for single events and 70% for combined events. The main difficulties encountered were the identification of single events with varying flow rates and the accurate separation of combined events into sub-singe events. As shown in the results, the accurate extraction of single events from a combined event is crucial during the classification process. The applicability of this approach is further suggested to be tested in large real datasets from regions with different water usage characteristics considering also the presence of leakages.

6 ACKNOWLEDGMENTS

The work was supported by the FLOBIT Project EXCELLENCE/0918/0282 which is co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation, and the European Union Horizon 2020 program under Grant Agreement No. 739551 (KIOS CoE) and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

7 REFERENCES

[1] A. K. Biswas and C. Tortajada, “Assessing Global Water Megatrends,” in *Water Resources Development and Management*, Springer, Singapore, 2018.

[2] H. M. Ramos, A. Carravetta, and A. M. Nabola, “New Challenges in Water Systems,” *Water*, vol. 12, no. 9, Aug. 2020, p. 2340.

[3] A. Liu, D. Giurco, and P. Mukheibir, “Urban water conservation through customised water and end-use information,” *Journal of Cleaner Production*, vol. 112, Jan. 2016, pp. 3164–3175.

[4] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, “Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review,” *Environmental Modelling & Software*, vol. 72, Oct. 2015, pp. 198–214.

[5] A. Cominola *et al.*, “Long-term water conservation is fostered by smart meter-based feedback and digital user engagement,” *npj Clean Water*, vol. 4, no. 1, Dec. 2021, p. 29.

[6] M. S. Rahim, K. Anh Nguyen, R. A. Stewart, D. Giurco, and M. Blumenstein, “Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour,” in *Proc. 2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, IEEE, 2019, pp. 1–8.

[7] Y. Kim, T. Schmid, Z. M. Charbiwala, J. Friedman, and M. B. Srivastava, “NAWMS,” in *Proc. of the 6th ACM conference on Embedded network sensor systems - SenSys ’08*, 2008, p. 309.

[8] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, “The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances,” *Data Mining and Knowledge Discovery*, vol. 35, no. 2, Mar. 2021, pp. 401–449.

[9] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowledge and Information Systems*, vol. 7, no. 3, Mar. 2005, pp. 358–386.

[10] M. Vlachos, G. Kollios, and D. Gunopulos, “Discovering similar multidimensional trajectories,” in *Proc. 18th International Conference on Data Engineering*, San Jose, CA, IEEE, 2002, pp. 673–684.

[11] W. B. DeOreo, J. P. Heaney, and P. W. Mayer, “Flow trace analysis to access water use,” *Journal - American Water Works Association*, vol. 88, no. 1, Jan. 1996, pp. 79–90.

[12] R. A. Stewart, R. Willis, D. Giurco, K. Panuwatwanich, and G. Capati, “Web-based knowledge management system: linking smart metering to the future of urban water planning,” *Australian Planner*, vol. 47, no. 2, Jun. 2010, pp. 66–74.

[13] J. E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, and S. N. Patel, "HydroSense," in *Proc. of the 11th international conference on Ubiquitous computing*, New York, ACM, 2009, pp. 235–244.

[14] J. Froehlich *et al.*, "A Longitudinal Study of Pressure Sensing to Infer Real-World Water Usage Events in the Home," in *International conference on pervasive computing*, Springer, Berlin, 2011, pp. 50–69.

[15] M. A. Corona-Nakamura, R. Ruelas, B. Ojeda-Magana, and D. Andina, "Classification of domestic water consumption using an Anfis model," in *Proc.* 2008 World Automation Congress, Waikoloa, HI, IEEE, 2008, pp. 1-9.

[16] B. E. Meyer, K. Nguyen, C. D. Beal, H. E. Jacobs, and S. G. Buchberger, "Classifying Household Water Use Events into Indoor and Outdoor Use: Improving the Benefits of Basic Smart Meter Data Sets," *Journal of Water Resources Planning and Management*, vol. 147, no. 12, Dec. 2021, p. 04021079.

[17] L. Pastor-Jabaloyes, F. Arregui, and R. Cobacho, "Water End Use Disaggregation Based on Soft Computing Techniques," *Water*, vol. 10, no. 1, Jan. 2018, p. 46.

[18] R. Cardell-Oliver, J. Wang, and H. Gigney, "Smart Meter Analytics to Pinpoint Opportunities for Reducing Household Water Use," *Journal of Water Resources Planning and Management*, vol. 142, no. 6, Jun. 2016, p. 04016007.

[19] J. S. Vitter and M. E. Webber, "A non-intrusive approach for classifying residential water events using coincident electricity data," *Environmental Modelling & Software*, vol. 100, Feb. 2018, pp. 302–313.

[20] M. S. Rahim, K. A. Nguyen, R. A. Stewart, D. Giurco, and M. Blumenstein, "Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review," *Water*, vol. 12, no. 1, Jan. 2020, p. 294.

[21] F. Mazzoni, S. Alvisi, M. Franchini, M. Ferraris, and Z. Kapelan, "Automated Household Water End-Use Disaggregation through Rule-Based Methodology," *Journal of Water Resources Planning and Management*, vol. 147, no. 6, Jun. 2021, p. 04021024.

[22] K. A. Nguyen, R. A. Stewart, H. Zhang, and C. Jones, "Intelligent autonomous system for residential water end use classification: Autoflow," *Applied Soft Computing*, vol. 31, Jun. 2015, pp. 118–131.

[23] K. A. Nguyen, R. A. Stewart, H. Zhang, and O. Sahin, "An adaptive model for the autonomous monitoring and management of water end use," *Smart Water*, vol. 3, no. 1, Dec. 2018, p. 5.

[24] A. Cominola, M. Giuliani, A. Castelletti, D. E. Rosenberg, and A. M. Abdallah, "Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management," *Environmental Modelling & Software*, vol. 102, Apr. 2018, pp. 199–212.

[25] W. B. DeOreo, "Analysis of water use in new single family homes," *By Aquacraft. For Salt Lake City Corporation and US EPA*, 2011.

[26] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – A decade review," *Information Systems*, vol. 53, Oct. 2015, pp. 16–38.

[27] V. Vijayakumar and R. Bremananth, "A Study on Human Hair Analysis and Synthesis," in *Proc. International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, Sivakasi, IEEE, 2007, pp. 549–556.

[28] S. Aghabozorgi, T. Ying Wah, T. Herawan, H. A. Jalab, M. A. Shaygan, and A. Jalali, "A Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique," *The Scientific World Journal*, vol. 2014, 2014, pp. 1–12.

[29] K. A. Nguyen, R. A. Stewart, and H. Zhang, "An intelligent pattern recognition model to automate the categorisation of residential water end-use events," *Environmental Modelling & Software*, vol. 47, Sep. 2013, pp. 108–127.

[30] Haibo He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, Sep. 2009, pp. 1263–1284.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14830

WATER FOR FIREFIGHTING: A COMPARATIVE STUDY ACROSS SEVERAL CITIES

Adeshola Ilemobade[1]

[1]School of Civil and Environmental Engineering, University of the Witwatersrand, Johannesburg
Johannesburg, Gauteng, South Africa

[1]*adesola.ilemobade@wits.ac.za*

Abstract

Studies into water for firefighting are sparse in global literature. Over the past three decades, 6 published studies have been undertaken in South Africa to quantify the extent of municipal water employed to fight fires. These studies have been necessary considering the need to conserve scarce and dwindling freshwater resources while providing adequate fire protection to many South African communities. While these studies have been driven by similar objectives, their analysis have not produced results that can easily be compared in order to extract generic highlights that can aid national firefighting efforts. In addition, the recent firefighting studies postulate that the minimum fire flows in the South African National Standard (the SANS 10090) and Guideline (The Red Book) are conservative and therefore do not promote the appropriate design of water networks. This may be attributed to the fact that the fire flows in the 1st edition of the SANS 10090 were likely over-estimated for South Africa since they were compiled with the assistance of organisations from the UK, USA, Canada, New Zealand and Germany, and have not notably changed since.

This paper therefore aims to address 2 objectives. The first objective will extract as much data as is possible from each of the 6 studies and will analyse the data with the aim of comparing consistent parameters (such as fire flows). The second objective will compare results obtained from the first objective with the SANS 10090, The Red Book and available international Standards and Guidelines for firefighting. Based on the results from the second objective, this study will conclude on the appropriateness of the minimum fire flows in the SANS 10090 and The Red Book to current firefighting efforts.

Keywords
Water for firefighting, Firefighting Standards and Guidelines.

1 BACKGROUND

Few studies in the global literature have addressed the topic of water for firefighting. A cursory search using the SCOPUS search domain for *article title, keywords and abstract* using the following keywords: "*water for firefighting*", produced 787 documents, a fraction of which address flows, volumes and temporal variation of fires within communities. Including "*standards* and *guidelines*" to the above keywords narrowed the search and produced only 8 documents. When compared with the results of other keyword searches (e.g. "*optimal water networks*" – 10 217 documents, "*water meter management*" – 4 772 documents, and "*water reuse*" – 27 552 documents), the limited research in water for firefighting becomes more evident.

Firefighting infrastructure, although rarely used when compared to infrastructure that provides basic services, protects residential and non-residential communities from the devastating impacts of fires. Sadly, fires have been shown to be devastating on many levels, including economic, environmental and fatalities. To extinguish fires, several agents may be employed, with water being the most popular due to several reasons including its relative ease of availability, its relatively affordable cost, and its ability to extinguish most types of fires. Unfortunately, increasing aridity in many communities is negatively impacting the availability of freshwater resources, not only for firefighting, but more importantly, for meeting basic needs, sustaining economic development and maintaining environmental flows.

South African municipalities are legally obligated to provide water for firefighting [1] and thus, water for firefighting is given equal importance as drinking water when designing water networks. Water distribution systems are therefore designed for both peak water demand plus fire demand, and as a result, water network capacity may increase by as much as 20 % [2].

Because of the above, it is imperative to minimize freshwater extraction while meeting increasing water demand for different requirements including firefighting. This study addresses water for firefighting.

2 LITERATURE

A few South African and international studies have addressed water for firefighting. Highlights from these studies are mentioned below:

van Zyl and Haarhoff (1997) [3] investigated fire incidents spanning 1980 to 1991within a section of the City of Johannesburg (CoJ) Central Business District (CBD) in the SA Province of Gauteng and a residential area in the SA Province of Mpumalanga. Highlights were: (i) extinguishing large fires in Johannesburg over the 11-year period typically took between 30 and 90 minutes with fire flows typically ranging between 5 and 35 ℓ/s. (ii) 90 % of the large fires investigated were extinguished using fire flows of 52 ℓ/s or less and 440 kℓ of water or less; (iii) peak consumer demands were observed only at small intervals during the year, and usually did not coincide with fire events that typically took place in mid-Winter.

van Zyl, Davy and Haihambo (2011) [4] report the results of a study of water used for firefighting in the Greater City of Cape Town (CoCT) area between 1 January 2005 and 21 April 2010. The data consisted of 72 589 records separated into 10 categories. Highlights from their study include: (i) vegetation fires predominated amongst the categories investigated and occurred mostly during the summer months when climatic conditions were conducive to vegetation fires; and (ii) 98 %, 95 % and 87 % of residential, commercial and industrial fires respectively were extinguished in 2 hours or less. Industrial fires had the longest durations and required the largest volumes of water to extinguish. Residential fires had the shortest durations and required the smallest volumes of water.

Myburgh and Jacobs (2014) [5], employing a database containing 554 records with water volumes and 546 records with computed fire flows from 1 Jan 2010 to 31 Dec 2010, analyzed fire incidents within 5 towns (i.e. 3 municipalities) in the SA Province of the Western Cape (WC). Highlights were: (i) average water volume used to extinguish 77 large Category 2 (Structural and Industry) fires was 8.60 kℓ; (ii) more than 90 % of fires were extinguished using less than 10 kℓ of water; (iii) 91.4 % of the 546 fires were extinguished with non-reticulated water supplies. This means that the municipal water supply was only directly used to extinguish 47 fires or 8.6 % of the total number of fires. This was achieved because fire fighters transported water from the fire station using pre-filled tanker vehicles as the first line of defence in all cases when attending to a fire call.

Mac Bean and Ilemobade (2019) [6] analyzed about 10 years (from 01 January 2006 to 30 September 2017) of fire incident reports (3 859 records) that occurred within the CoJ. Highlights from their study were: (i) the fire flows and fire risk categories in the South African National Standard, SANS 10090 (SABS, 2018) [7] and The South African National Guideline (The Red Book; [1]) are inconsistent; (ii) The Red Book recommends fire flows that are lower than the SANS 10090 and thus, violates the SANS 10090, which is a standard that stipulates minimum acceptable values; (iii) over the study's 10-year period, 75 % of fire incidents within the CoJ were extinguished using 6.60 kℓ of water or less, 87 % of fires were extinguished using 10 kℓ of water or less and 99 % of fires were extinguished using 100 kℓ of water or less; (iv) while the frequency of fire occurrence was strongly related to climatic conditions, the volume of water used to quench fires was not a function of climatic conditions; (v) over the study's 10-year period, 99.90 % and 99.60 % of fire incidents were extinguished using average fire flows that were less than the minimum fire flows for the lowest risk categories in the SANS 10090 and The Red Book respectively; (vi) The start times of peak fires did not coincide with typical peak residential water demand periods.

Essack and Ilemobade (2022) [8], using similar but better representative criteria, re-analyzed the initial database employed by [6], and in particular, addressed 89 large fires (fires requiring more than 5 kℓ to extinguish). Highlights from their study were: (i) in the initial dataset of 4 479 records (spanning 24 February 2003 to 26 September 2017), an average of 32 incidents occurred per 30-day month with an expected increase in the average number of fire incidents during SA's dry and cold (i.e. late autumn, winter and early spring) months (May to September). The higher average number of fires during the cold months were attributed to households using unsafe appliances for heating and cooking especially in informal settlements; (ii) 99 % of fire incidents were extinguished using 100 kℓ of water or less while 87 % were extinguished using 10 kℓ or less; (iii) fire flows for 99.90 % and 99.60 % of the fire incidents were less than the minimum fire flow of the lowest SANS 10090 fire risk category and The Red Book respectively; and (iv) fire flows ranged between 0 and 3 ℓ/s. Additional highlights by [8], with specific reference to large fires were: (v) the duration of 60 % of the large fires were between 30 and 120 minutes; (vi) the average volume of water employed to extinguish the large fires was 9.63 kℓ; (vii) the fire flows employed to extinguish the 89 large fires were less than the minimum fire flows for the lowest fire risk categories in the SANS 10090 and The Red Book; and (viii) municipal water was used to supplement water conveyed in the vehicles in ~99 % out of 84 large Category 2 fires.

Thage and Ilemobade (2022) [9] analysed 3 236 fire incident reports at the Sol Plaatje Municipality (SPM), South Africa, during the period 21 July 2017 to 21 August 2020. Highlights from their study were: (i) 99.9 % of fire incidents employed less than 1 900 ℓ/min (the minimum fire flow for the SANS10090 category D1 fires), 99.5 % of incidents employed less than 1 500 ℓ/min (the fire flow specified in The Red Book for Moderate Risk 2 i.e. cluster & low-income housing and high rise flats ≤ 3 storeys) and 93.9 % of fires were extinguished using 200 ℓ/min of water or less; (ii) about 93.3 % of fires were extinguished using 7 kℓ of water or less, about 96.9 % of fires were extinguished using 10 kℓ of water or less and about 99 % of fires were extinguished using 60 kℓ of water or less; (iii) the average number of fire incidents in SPM peaked during the months of July & August (winter) and September (spring). These months experience the least rain and the highest energy (electricity, wood, paraffin) consumption for heating and cooking use; (iv) while fires in the SPM typically occurred between 11h00 and 20h00, with more than 100 fires reported per hour, the highest peak during the course of a typical day was observed at 15h00 (161 fire incidents). In addition, while the peak occurrence of fires and the residential peak water demand do not coincide in the morning, this is not the case in the afternoon as the largest number of fires (161) occur at 15h00, just before peak demands are experienced.

Davies (2000) [10], in New Zealand, found that over a 3-year period, on average 96 % of Structural fires in the study area were either extinguished with non-reticulated water supplies or with less than 10 ℓ/s of reticulated water supply.

Benfer and Scheffey (2014) [11] presents 16 fire flow methods from the USA, UK, France, Germany, the Netherlands, New Zealand and Canada. Eleven (11) addressed pre-incident building planning and at least 5 addressed on-scene firefighting. The 16 fire flow methods were applied to 2 differently sized non-residential buildings and 2 differently sized single-family residential buildings. Their study included both sprinklered and non-sprinklered calculations. Highlights from their study were: (i) minimum fire flows varied greatly across many countries; (ii) the building planning methods recommended fire flows that were higher than the on-scene methods; (iii) for residential buildings fitted with sprinklers, 12 out of the 18 methodologies required the same fire flows as non-sprinkler fitted buildings; and (iv) as can be seen in Figure 1, the range of fire flows, applied to a sample site, in all the methods are large, with the minimum fire flows in the SANS 10090 and The Red Book located close to the minimum fire flows for most ranges.

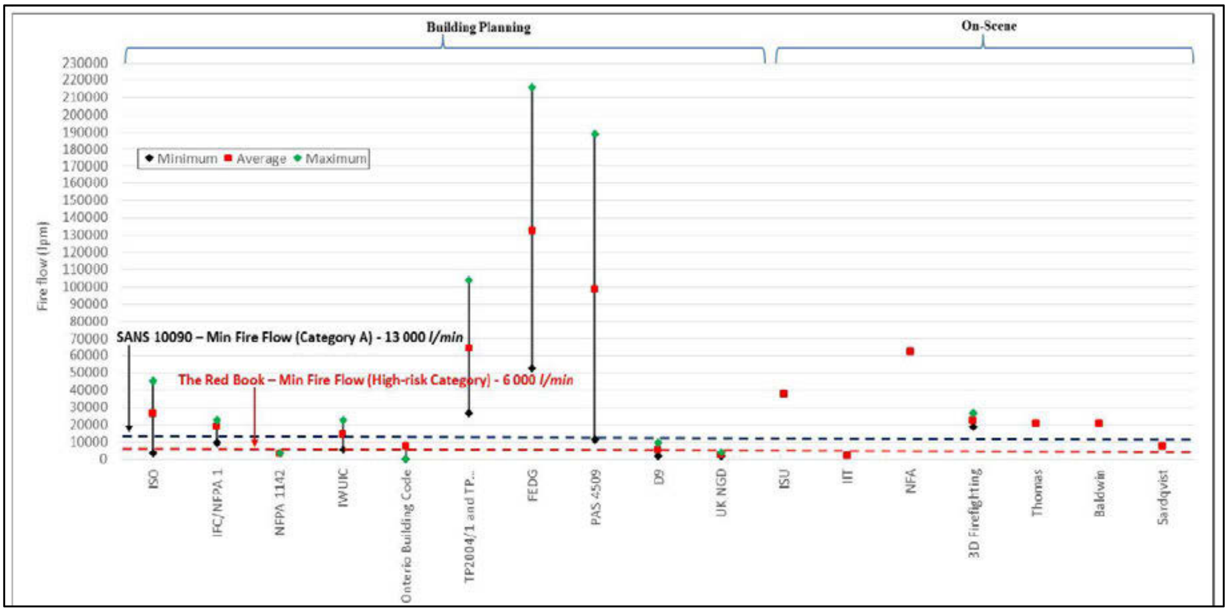


Figure 1. Benfer and Scheffey's (2014) 16 fire flows for a non-sprinklered, 4 645 m² non-residential building compared to the minimum flows in the South African Standard and Guideline

3 COMPARISON OF THE HIGHLIGHTS OF PREVIOUS STUDIES

While previous South African studies have been driven to address similar objectives, their analysis have not produced results that can easily be compared. This section therefore retrieves results published by previous studies and attempts to compare consistent parameters.

Figure 2 shows average fire flows employed to extinguish fires in each of the 3 studies plotted alongside the different minimum fire flows specified in the SANS 10090 and The Red Book. What is plainly obvious and which has been re-iterated several times in different publications ([5, 6, 8 and 9]) is that the minimum fire flows in both the SANS 10090 and The Red Book are conservative. Since these conservative flows must be included in the design of water networks, it is inevitable that larger water network component capacities will result.

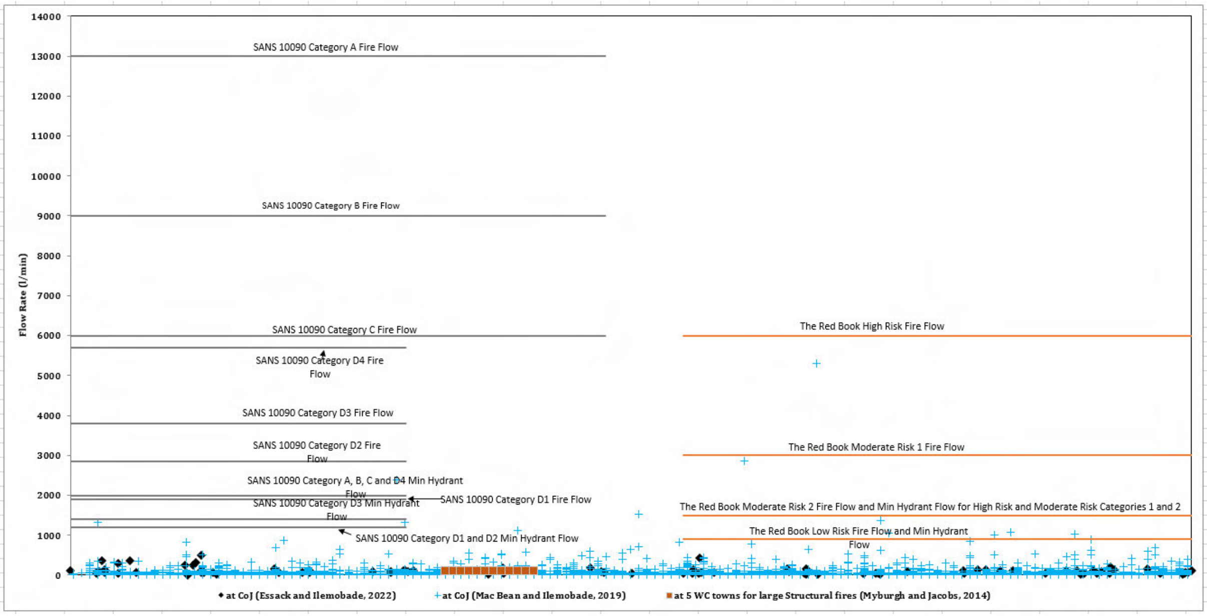


Figure 2. Fire flows from 3 studies with the SANS 10090 and The Red Book minimum fire flows superimposed

Specific highlights with regard to water volumes employed to extinguish fires were mentioned under each study's brief literature review above. Figure 3 presents the average volumes of water employed to extinguish fires for each representative month over the duration of each study. It is obvious that the average volume of water employed to extinguish fires are not correlated to the season of the year. In addition, in all studies, a significant percentage of fires (93.3% in the SPM [9], 75% in the CoJ [6], 91.4% in the 5 WC towns [5]) were (and could be solely) extinguished using certain fire engines - the Boom Pumper fire engine in the CoJ, for example, has a tank capacity of 9 kl.

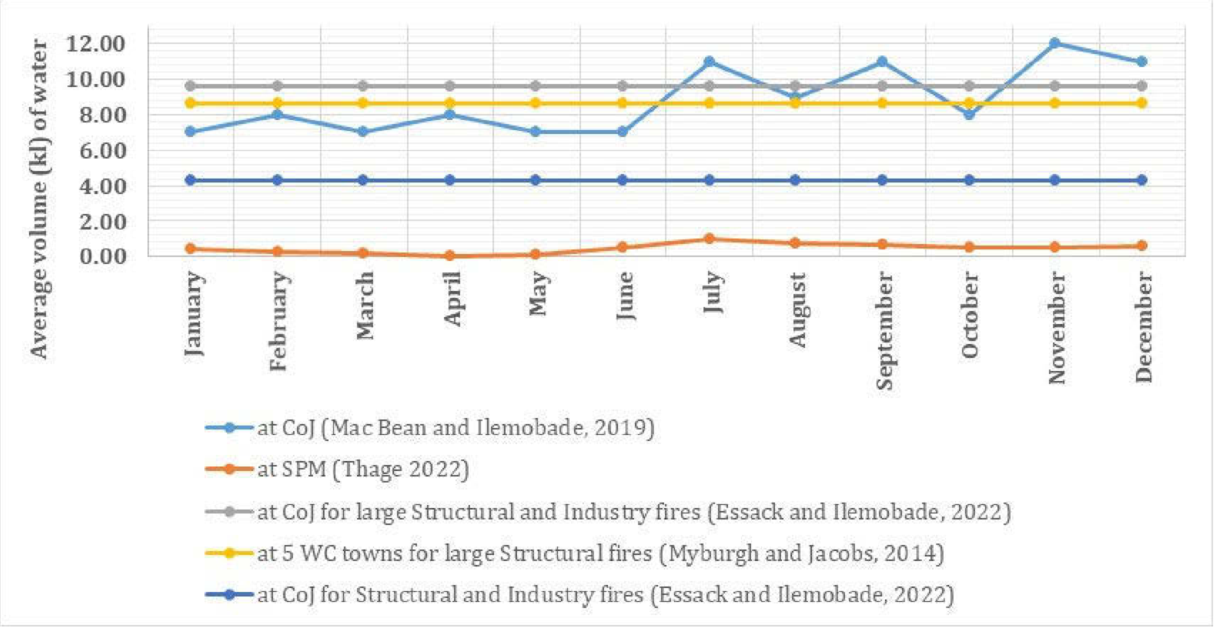


Figure 3. Average volume (kl) of water employed to extinguish fires for each representative month over the study periods

BY NC SA

Figure 4 shows that in 3 (at the CoJ [6], [8] and SPM [9]) of the 4 studies, the frequency of fire incidents increases and peaks during the dry months (April/May to September) and is the lowest during the rainy periods (December to March). In the WC [4], the rainy months are between March and November and this explains why the chart for the WC displays marginally lower frequencies between March to October when compared with frequencies at other times of the year.

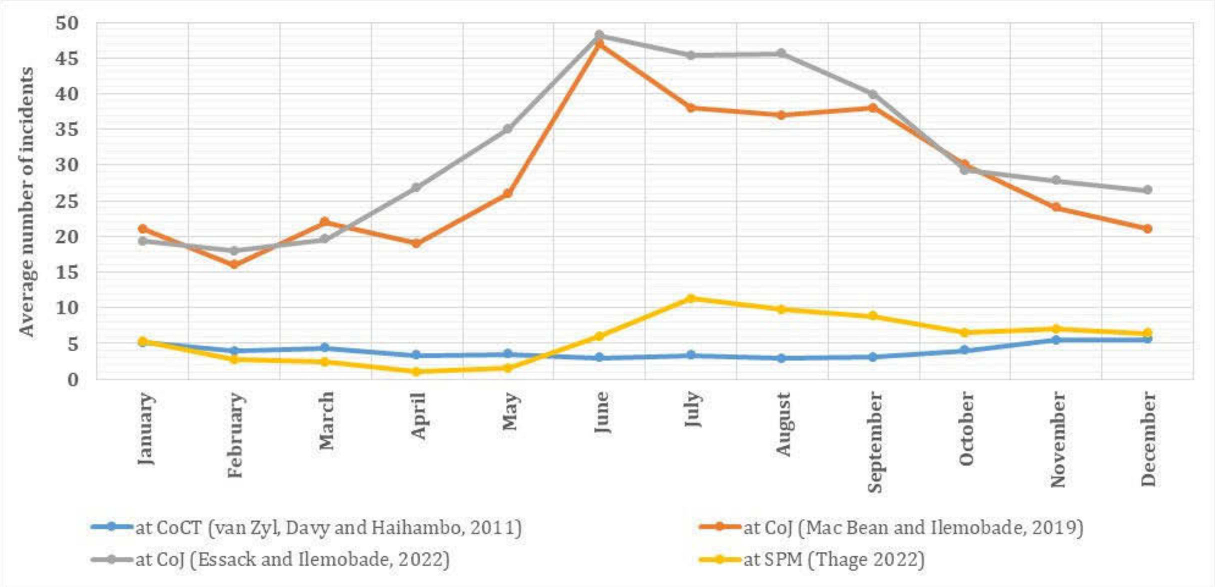


Figure 4. Average number of fire incidents for each representative month of the year over the study periods

Figure 5 shows that while the occurrence of fires and the typical residential peak water usage [12] do not coincide in the morning, this is not the case in the afternoon as the occurrence of fires in the 2 studies peak within an hour that the afternoon peak demand occurs. This may be because heating (during the dry and cold months) and cooking facilities are mostly employed in the afternoons when families are at typically at home.

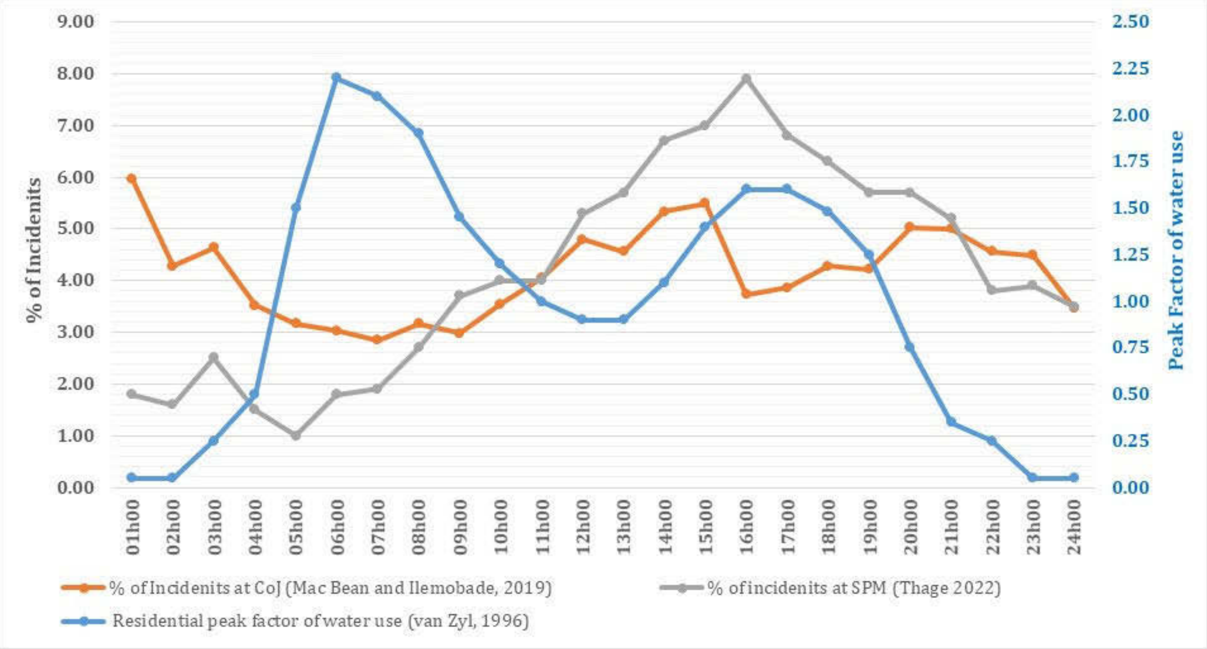


Figure 5. Daily variation of fire start times versus a typical residential peak water demand pattern

4 CONCLUSION

This paper addressed 2 objectives. The first extracted data from previous studies undertaken in South Africa and internationally, and analysed the data using consistent parameters, while the second compared the results obtained with the SANS 10090, The Red Book and available international Standards and Guidelines for firefighting. It is glaring that the fire flows in the SANS 10090 and The Red Book are conservative and while adequate fire protection must be provided to communities, the current fire flows need to be revised (reduced) in light of dire water scarcity and the need to design optimal (the least cost) water networks. The above results also provide hope to communities and local councils who cannot depend on erratic water supply to fight fires as well-equipped firefighting services (with fire engines that can convey 9 kℓ of water) may be able to extinguish at least 75 % or fires in communities.

5 REFERENCES

[1] DHS and CSIR (Department of Human Settlements, South Africa and Centre for Scientific and Industrial Research). The Neighbourhood Planning and Design Guide (The Red Book): Creating Sustainable Human Settlement. Pretoria. 2019. ISBN 978-0-6399283-2-6.

[2] Council for Scientific and Industrial Research (CSIR), Guidelines for Human Settlement Planning and Design (The Red Book). CSIR Building and Construction Technology, Pretoria, 2005.

[3] J. E. van Zyl and J. Haarhoff. “South African fire water guidelines and their impact on water supply system cost”, Journal of the South African Institution of Civil Engineering, vol. 39, no. 1, 1997, pp 16–22.

[4] J. van Zyl, M. Davy and F. Haihambo, “Water used for fire fighting in Cape Town”, Proceedings of the Computing and Control for the Water Industry (CCWI) conference, Centre for Water Systems, University of Exeter, 2011, pp 107-112.

[5] H. M. Myburgh and H. E. Jacobs. “Water for firefighting in five South African towns”, Water SA, vol. 40, no. 1, 2014, pp. 11–17. https://doi.org/10.4314/wsa.v40i1.2

[6] C. B. Mac Bean and A. A. Ilemobade. “An evaluation of the primary South African Standard and Guideline for the provision of water for firefighting”. Water SA, vol. 45, no. 4, 2019, pp. 691–699. https://doi.org/10.17159/wsa/2019.v45.i4.7551

[7] SABS (South African Bureau of Standards), SANS 10090:2018 South African National Standard. Community protection against fire. Standards South Africa (a division of SABS). Edition 3.1, 2018, Pretoria. ISBN 978-0-626-36193-8.

[8] F. Essack and A. Ilemobade, “Water for fighting large structural fires: a South African study”, Urban Water Journal, 2022. https://doi.org/10.1080/1573062X.2022.2050928

[9] Thage and Ilemobade, “Water for firefighting in Sol Plaatje municipality, Northern Cape, South Africa”, unpublished draft M.Sc. Research Report. University of the Witwatersrand, Johannesburg, 2022.

[10] S. Davies, Fire fighting water: a review of fire fighting water requirements. A New Zealand perspective. Fire Engineering Research Report 00/3, M.E. (Fire) degree at the University of Canterbury, March, 2000.

[11] M. Benfer and J. Scheffey, Evaluation of Fire Flow Methodologies. The Fire Protection Research Foundation, Quincy, Massachusetts, 2014. https://doi.org/10.1007/978-1-4939-2889-7

[12] J. E. van Zyl, “Peak factors in municipal water reticulation networks”. Proceedings of the Water Institution of South Africa Biennial Conference and Exhibition, Port Elizabeth, South Africa, 20–23 May 1996.

WDSA CCWI 2022

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14831

SUPERVISED MACHINE LEARNING MODELS FOR LEAK DETECTION IN WATER DISTRIBUTION SYSTEMS

Lochan Basnet[1], Ranji Ranjithan[2], Downey Brill[3] and Kumar Mahinthakumar[4]

[1,2,3,4]Department of Civil Engineering, North Carolina State University, Raleigh, NC, (USA)

[1]lbasnet@ncsu.edu, [2]ranji@ncsu.edu, [3]brill@ncsu.edu, [4]gmkumar@ncsu.edu

Abstract

Water distribution systems (WDSs) face a significant challenge in the form of pipe leaks. Pipe leaks can cause loss of a large amount of treated water, leading to pressure loss, increased energy costs, and contamination risks. Locating pipe leaks has been a constant challenge for water utilities and stakeholders due to the underground location of the pipes. Physical methods to detect leaks are expensive, intrusive, and heavily localized. Computational approaches provide an economical alternative to physical methods. Data-driven machine learning-based computational approaches have garnered growing interest in recent years to address the challenge of detecting pipe leaks in WDSs. While several studies have applied machine learning models for leak detection on single pipes and small test networks, their applicability to the real-world WDSs is unclear. Most of these studies simplify the leak characteristics and ignore modeling and measuring device uncertainties, which makes the scalability of their approaches questionable to real-world WDSs. Our study addresses this issue by devising four study cases that account for the realistic leak characteristics (multiple, multi-size, and randomly located leaks) and incorporating noise in the input data to account for the model- and measuring device- related uncertainties. A machine learning-based approach that uses simulated pressure as input to predict both location and size of leaks is proposed. Two different machine learning models: Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN), are trained and tested for the four study cases, and their performances are compared. The precision and recall results for the L-Town network indicate good accuracies for both the models for all study cases, with CNN generally outperforming MLP.

Keywords
Leak detection, machine learning, multilayer perceptron, convolutional neural network, hydraulic simulation, water distribution systems.

1 INTRODUCTION

Water distribution systems (WDSs) face a significant challenge in the form of pipe leaks. Pipe leaks can cause loss of large amount of treated water in WDSs leading to pressure loss and increased energy costs. Leaks can also pose risks of water contamination [1]. As reported in [2], an estimated 126 billion cubic meters of water is lost every year worldwide. With increasing demands and growing concerns about water scarcity in the face of climate change, the prevention of water losses from WDSs is crucial. Moreover, pipe leaks can grow over time and lead to breaks and bursts causing property damage and traffic disruptions. Therefore, timely detection and prevention of pipe leaks are paramount. Unlike pipe breaks, pipe leaks are tough to detect as the flow or pressure changes produced by leaks are not humanly discernable [3]. In addition, the underground location of pipes makes it even harder to detect leaks. Physical methods to detect leaks are expensive and can cause interruption to water service [4]. Computational approaches provide an economical alternative to physical methods.

Several computational approaches have been proposed for leak detection (an extensive review is provided in [5]). Machine learning approaches are one of the data-driven computational approaches that have gathered increasing interest in the last two decades in leak detection studies [6]. Machine learning methods use a large amount of data related to the hydraulic properties of WDSs such as pressure, flowrate, acoustic vibration, optics, or temperature for leak detection [7]. Pressure and flowrates are the most commonly used properties for leak detection [8].

While a good amount of research has been conducted on the application of machine learning models for leak detection in pipes [9], the question about their applicability to real-world WDSs remains unclear. One of the critical reasons for this lack of clarity concerns the scalability of the approaches considered in these studies. In [10] and [11], Convolutional Neural Network (CNN) was used for leak detection in a single pipe using simulated negative pressure wave and scalogram images of vibration signals as inputs, respectively. In [12], MLP was used with a cascade-forward back-propagation to detect leaks in a single pipe using simulated pressure data. However, analyzing leaks by isolating individual pipes in complex interconnected WDSs is not a viable solution in the field as it is difficult to isolate specific pipes. Further, the tools and resources required to collect some of these input data for individual pipes in large real-world WDSs are infeasible. Beyond single pipe analyses, several studies have considered complete or partial hydraulic systems. MLP was used in [13] to predict leaks in a simple hydraulic system using numerically obtained fluid transient waves as input. In [14], SVM was used to predict leak size and location for an isolated section of a WDS based on simulated pressure data. In [15], a model-based k-Nearest Neighbors (k-NN) classifier was used to identify leak events and locations. These studies still face the challenge of scalability as extrapolating their results and, therefore, application to the larger real-world WDSs is very challenging.

Another factor that limits the real-world application of some of the existing machine learning approaches relates to the simplifying assumptions regarding the characteristics of pipe leaks in WDSs. For example, the application of the Bayesian classifier in [16] to detect leaks assumes that there is only a single leak in the WDS, which is rarely true. In [17], unsupervised principal component analysis (PCA) was used for leak detection by assuming a single, constant size leak.

Furthermore, very few studies have considered uncertainties associated with hydraulic simulation models and imprecision of measurement devices in real-world WDSs. The parameters such as demands, pipe roughness, pipe diameters, and lengths used in the hydraulic models have associated uncertainties [18]. These uncertainties affect the accuracy of the simulated pressure and flow data. One way to account for the hydraulic model parameter uncertainties is to add noise to these parameters prior to simulation, as shown in [19] and [20]. However, such an approach is inadequate to encapsulate the uncertainties related to the imprecision of measurement devices such as pressure sensors and flow meters of the real-world WDSs.

This study proposes a machine learning-based approach to detecting and localizing leaks in WDSs, which considers multiple realistic leak scenarios and accounts for hydraulic model uncertainties and instrument imprecision. Two different machine learning models are used to predict leaks using simulated pressure measurements as input. The key contributions of this study with respect to previous approaches include:

- Overcoming the unrealistic simplification about occurrence of a single leak at a time assumed by most state-of-the-art techniques [15] by generalizing to multi-leak problems.
- Considering leaks of varying sizes to represent more realistic leak scenarios.
- Accounting for the realistic nature of leak locations by considering the possibility of random leak locations anywhere within a WDS.
- Consideration of the most common and impactful hydraulic model uncertainty, i.e., demand uncertainty, as well as measuring instrument imprecision through the addition of noise to the input data.
- Simultaneous prediction of location as well as siz e of the leaks.

Even though the machine learning models are trained using simulated pressure data, they are applicable to predict leaks using real-world measurements as long as the hydraulic model is a reasonable representation of the real system. For WDSs that have abundant real-world pressure sensor measurements, these models can easily be fine-tuned and tested using the real data.

2 METHODS

2.1 General Framework

Figure 1 illustrates the general framework proposed in this study to detect and localize leaks in WDS pipes. The framework starts with a WDS hydraulic model that generates simulated operational pressure data. First, pressure data for a leak-free scenario is generated by simulating the hydraulic model using the EPANET simulator [21]. It is followed by pressure data generation for multiple different leak scenarios. Pressure differences between the leak scenarios and the leak-free scenario are then computed and stored as a pressure readings dataset. The corresponding leak scenarios are stored as a leak values dataset. Noise is then added to the pressure readings dataset when required for the case under study described in Section 2.5. The resulting pressure readings dataset and the leak values dataset are then randomly shuffled and split into training sets and testing sets; a train to test ratio of 80 to 20 is used. The training pressure dataset and the training leak values dataset are scaled and fed to machine learning models. The pressure data is considered as covariates and the leak values as responses. The models are trained and tuned, and the optimized models are selected for the prediction of leaks. Finally, the leak prediction and model evaluation are performed on the testing pressure and leak values datasets using the optimized models; predicted model outputs are compared with the corresponding true leak values.

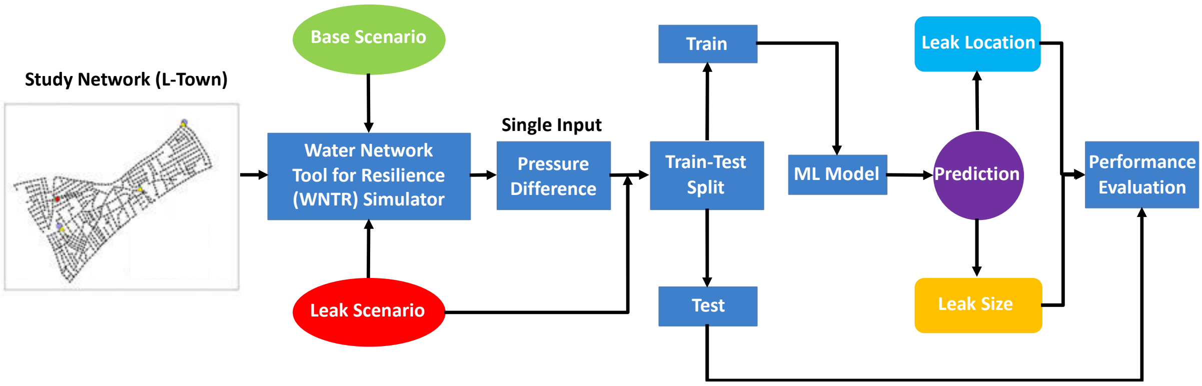


Figure 1. General framework for detecting leaks in WDS

2.2 Machine Learning Models

2.2.1 Multilayer Perceptron (MLP)

Multilayer Perceptrons (MLPs) are supervised-learning models based on deep neural networks. An MLP model consists of an input layer, an output layer, and a selected number of dense hidden layers located between the input and the output layers (Figure 2). Activation layers follow hidden layers to activate or deactivate received signals. Multiple activation functions are available to be used in these activation layers.

2.2.2 1-D Convolutional Neural Network (CNN)

2.2.2 1-D Convolutional Neural Network (CNN)

Similar to MLPs, Convolutional Neural Networks (CNNs) are also supervised-learning-based deep neural networks. The key difference between CNNs and MLPs is the presence of convolutional and pooling layers in CNNs. As shown in Figure 3, the convolutional layers produce convolved feature maps, which allow for contextual learning, and the pooling layers downsample these maps to extract abstract features from the data. The convolutional layers use kernels or filters to extract the features. A one-dimensional (1-D) CNN model uses filters that only vary in depth (i.e., one dimension). CNN models also have an input, an output layer, some dense hidden layers, and activation layers similar to the MLPs.

2.2.3 Hyperparameters and Model Tuning

Total Number of Iterations (Epochs): MLP and CNN models are trained for a number of iterations (epochs) to ensure the stability in the training process. The optimal model and its corresponding weights are determined by monitoring the training and validation errors over the entire number of epochs.

Error Function: The functions to calculate the training and validation errors are chosen based on the nature of the problem. In this study, leak detection is formulated as a regression type problem to simultaneously solve for both leak locations and sizes. Therefore, the mean squared error (MSE) function is used; mean absolute error (MAE) can be used as an alternative to MSE.

Activation Function: A trial-and-error evaluation of multiple activation functions identified the Leaky Rectified Linear Unit (L-ReLU) as a suitable activation function for this study. L-ReLU prevents the problem of vanishing gradient during forward propagation like the regular rectified linear unit (ReLU) and has an added advantage of preventing vanishing gradients during backward propagation [22].

Optimizer: The commonly used Adam optimizer is used in this study.

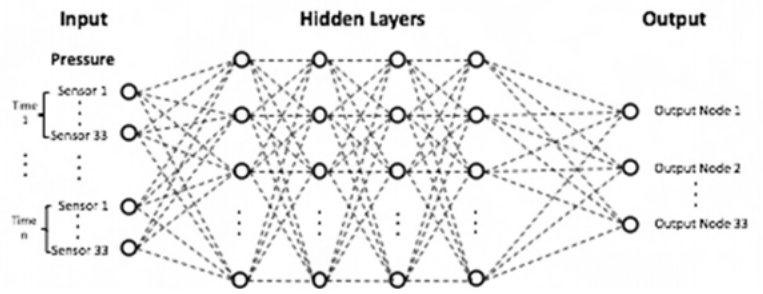

Input
Hidden Layers
Output
Pressure
Time 1
Sensor 1
Sensor 33
Time n
Sensor 1
Sensor 33
Output Node 1
Output Node 2
Output Node 33

Figure 2: Multilayer Perceptron

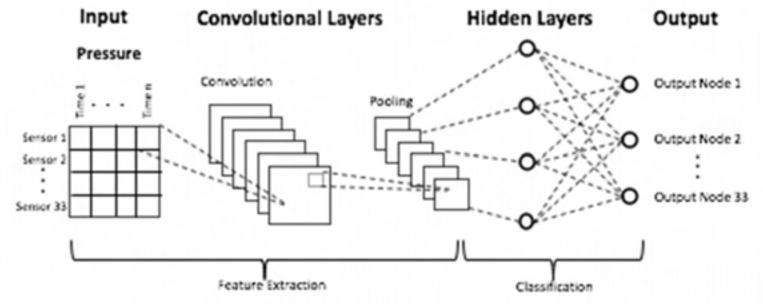

Input
Convolutional Layers
Hidden Layers
Output
Pressure
Time 1
Time n
Sensor 1
Sensor 2
Sensor 33
Convolution
Pooling
Output Node 1
Output Node 2
Output Node 33
Feature Extraction
Classification

Figure 3: Convolutional Neural Network

2.3 Study Network

In this study, the leak detection methods are applied to a standard test network called the L-Town water network (Figure 4). The L-Town network has been previously used in several modeling and simulation related researches. For example, this network was also used in the Leakage Detection and Isolation Methods (BattLeDIM 2020) [23] competition to evaluate the performances of different machine learning and computational models for leak detection. The L-Town network consists of 905 pipes and 782 junctions and is primarily a tank-regulated model network.

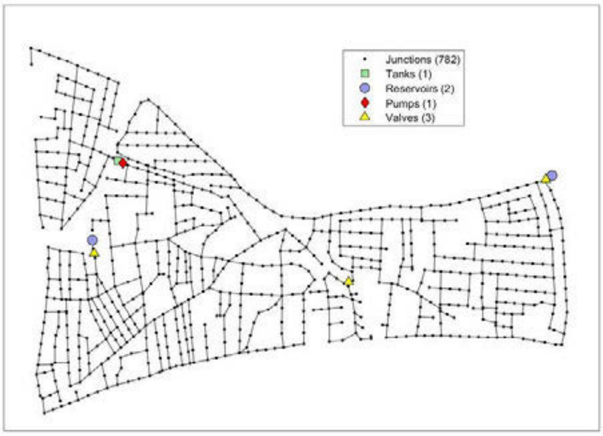
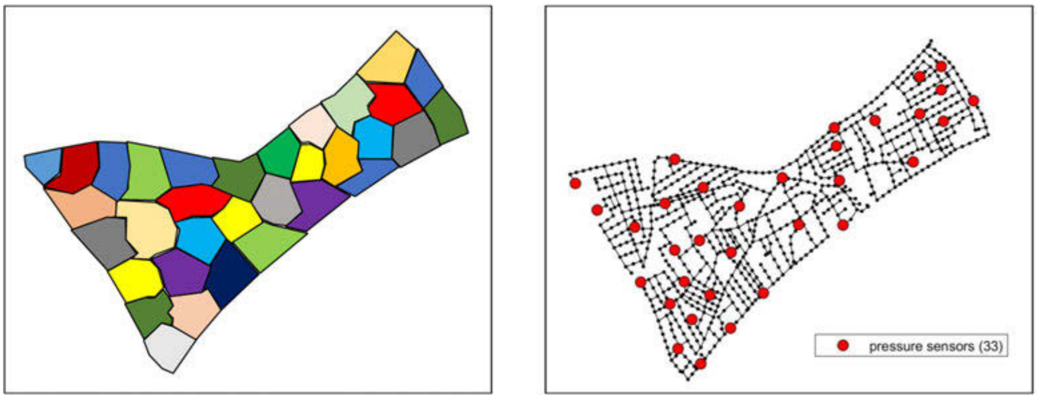


Figure 4: L-Town water network

2.4 Candidate Leak Regions

Localizing leaks to the actual pipes or junctions requires large amount of data, which is infeasible to obtain from real-world WDSs. Therefore, a lesser resolution is adopted for leak localization in this study. The entire water network is divided into several sub-areas that are considered candidate leak regions. The L-Town network is divided into 33 candidate leak regions (Figure 5a). A k-means clustering technique [24] is used to divide the network into these 33 candidate leak regions based on Euclidean distances. Leaks are modelled as emitters in EPANET and are assumed to occur at the center of each pipe. Since EPANET supports emitters only on nodes, new junction nodes are inserted at the middle of every pipe in the network using the Morph package in WNTR [25]. Candidate leak nodes representing each leak region is assumed to be at the centroid of each leak region. Centroids of leak regions are estimated using k-nearest neighbour search algorithm. For any given leak scenario, a leak located anywhere within the boundaries of a candidate leak region is defined by this region. While a hydraulic distance-based clustering measure results in more homogeneous clusters, the less homogeneous clusters obtained using Euclidean distance-based measure may pose a more significant challenge for the leak detection models. Therefore, the Euclidean distance-based clustering used here is a more conservative approach.

A pressure node is assigned to each of the 33 candidate regions to track the pressure changes due to leak/s in that region. These pressure nodes represent pressure sensors in real-world WDSs. The locations of the pressure nodes are based on the locations used in BattLeDIM 2020 and are shown in Figure 5b.



(a) (b)

Figure 5: Candidate leak regions: (a) Leak regions; (b) Pressure sensors within each region

2.5 Study Cases

Four study cases are considered in this study to represent the realistic leak characteristics, and the uncertainties in input data due to water network model inaccuracies and measuring device imprecision are considered for this study.

Case A: No-noise – Input pressure difference data is free of noise. It represents the ideal case of accurate WDS models and precise measuring devices. Leaks are assumed to occur at the centroid of each leak region.

Case B: Demand-noise – Input pressure difference data accounts for the WDS model inaccuracies. To mimic the inaccuracies in demand values in the WDS model, random Gaussian noise are added to the demands prior to simulation. Simulated pressure data are then generated using the modified WDS model. A ten percent Gaussian noise is used. Leaks are assumed to occur at the centroid of each leak region.

Case C: Mixed-noise – Input pressure difference data accounts for the WDS model inaccuracies as well as the measuring device imprecision. Unlike the *demand-noise* case, noise is added to the final pressure differences of the leak and leak free scenarios. A ten percent Gaussian noise is used. Leaks are assumed to occur at the centroid of each leak region.

Case D: Random leaks – The leaks can be located anywhere within the boundaries of the candidate leak regions instead of their centroids. No additional noise is imposed.

2.6 Data Generation

The input datasets used in this study constitute the leak scenario and the pressure difference datasets, which are generated in two sequential steps.

2.6.1 Leak Scenario Generation

The following four assumptions are considered for the generation of realistic leak scenarios for this study:

- A leak scenario must consist of at least one leak.
- A leak scenario can include a maximum of 3 leaks.
- A leak can be located in any of the 33 candidate leak regions.
- The leak size ranges from 0 to 5 as compared to the 0 to 3 range used in BattLeDIM 2020. The leak size is the discharge coefficient in the leak equation (1).

$$q = C\ p^{Y} \tag{1}$$

where q = flow rate, p = pressure, C = discharge coefficient, and Y (=0.5) = pressure exponent.

Applying the above assumptions, leak scenarios are generated using the following general procedure:

Step 1 – For a leak scenario, the total number of leaks is determined by drawing in random a number ***n*** from the set {1, 2, 3}.

Step 2 – Based on the outcome ***n*** of the previous draw, ***n*** candidate leak locations out of the 33 candidate leak locations are drawn at random.

Step 3 – For these ***n*** candidate leak locations, the leak sizes are randomly drawn from the leak size range of 0 to 5.

Step 4 – Repeat steps 1 – 3 for 100,000 times to generate 100,000 leak scenarios.

The 100,000 leak scenarios generated from the above procedure were saved as a leak scenario dataset.

2.6.2 Pressure Data Generation

Simulated pressure data are generated by the following procedure:

Step 1 – As discussed in Section 2.4, assign one pressure node each to all 33 candidate regions. The locations of the pressure nodes are based on the locations used in BattLeDIM 2020.

Step 2 – Simulate a leak-free scenario for the specified study case defined in Section 2.5 by running the base model with the EPANET simulator. Store the resulting pressure values at the 33 pressure nodes.

Step 3 – Pick a leak scenario from the leak scenario dataset and add the associated leaks to the base model. Then, run this modified model with the EPANET simulator and store the resulting pressure values at the 33 pressure nodes.

Step 4 – Repeat Step 3 for all the 100,000 leak scenarios in the dataset.

Step 5 – Compute the pressure differences between each of the 100,000 leak scenarios and the leak-free scenario. Then, combine the 100,000 pressure differences together as a pressure difference dataset.

Step 6 – Add noise to the pressure difference data depending upon the study case discussed in Section 2.5.

2.7 Model Validation and Testing

2.7.1 Train-Test Split

The input pressure and leak datasets are divided into training and test data. A training to test ratio of 80 to 20 is used to split the data. The two models are validated using the test datasets.

2.7.2 Metrics and Thresholds

The performance of the two machine learning models to predict leaks is evaluated using the two standard classification metrics: precision and recall.

$$Precision = \frac{TP}{TP + FP} \times\ 100 \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \times\ 100 \tag{3}$$

where, TP = True Positives; FP = False Positives; and FN = False Negatives.

In the context of this study, precision is the percentage of the actual leaks out of all leak predictions made by the models. Recall is the percentage of the actual leaks identified by our models out of all the leaks in the dataset.

In this study, the problem of leak detection is formulated as a regression type problem to simultaneously solve for both leak locations and sizes. To assess the model performances in terms of precision and recall, a post-processing of model outputs is required. This post-processing involves the use of thresholds to determine correct/incorrect location and size classifications. A set of nine thresholds ranging from 0.1 to 0.9 increasing incrementally by 0.1 are used. The thresholds are in the same unit as leak sizes and represent the precision of the measuring devices for real-world systems. For example, a threshold of 0.1 means that the leaks that are smaller than 0.1 in the dataset are considered as no-leaks and only the predictions that are within 0.1 units of the actual leak values are considered as correct classifications.

2.8 Software and Tools

The following software and tools were used in this study:

- EPANET Simulator 2.0 version – Hydraulic simulations are performed using EPANET simulator.
- WNTR Morph package – For splitting the network to add junction nodes at the middle of each pipe.
- MatLab 2019b version – Input data generation is done by running EPANET simulator in MatLab. Matlab is also used to generate candidate leak regions and nodes.
- Python version 3.7 – Model training, testing, and validation is done in Python.
- Tensorflow version 2.1.6 – Machine learning models are built using the Tensorflow package.

3 RESULTS AND DISCUSSION

Leak prediction performance of the MLP and CNN models are studied for the four study cases described in Section 2.5. The two models are compared by calculating precision and recall accuracies for the test dataset. Table 1 summarizes the architecture and hyperparameters for the optimal MLP and CNN models. The optimal MLP model has four dense layers: the input layer and the output layer, with 33 units each, and the two central dense layers with 64 and 128 units. The optimal CNN model consists of six layers - four dense layers and two convolutional layers. Like MLP, two out of the four dense layers are the input and the output layers, with 33 units each. The remaining two dense layers are hidden layers with 500 and 100 units, respectively. The two convolutional layers (also hidden) that follow the input layers consist of 256 and 128 filters. Figure 6 shows the trend of the training and validation mean squared errors for the *no-noise* case for the two models. The validation errors show a general decreasing trend that stops after the 100th epoch, indicating model overfitting beyond 100 epochs. The same is true for the validation errors for the other three study cases. Therefore, the required number of iterations for all model training is set to 100 epochs.

Table 1. Machine learning model details

Model	Architecture	Hidden Layers	Dense Layers	Convolutional Layers	Activation Functions	Learning Rate	Optimizer
MLP	33-64-128-33	2	4	-	LReLU	0.05	Adam
CNN	33-256-128-500-100-33	4	4	2	LReLU	0.05	Adam

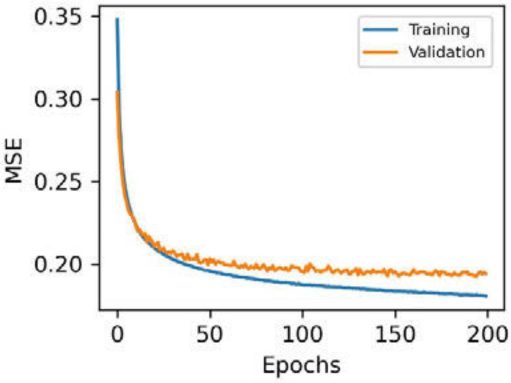
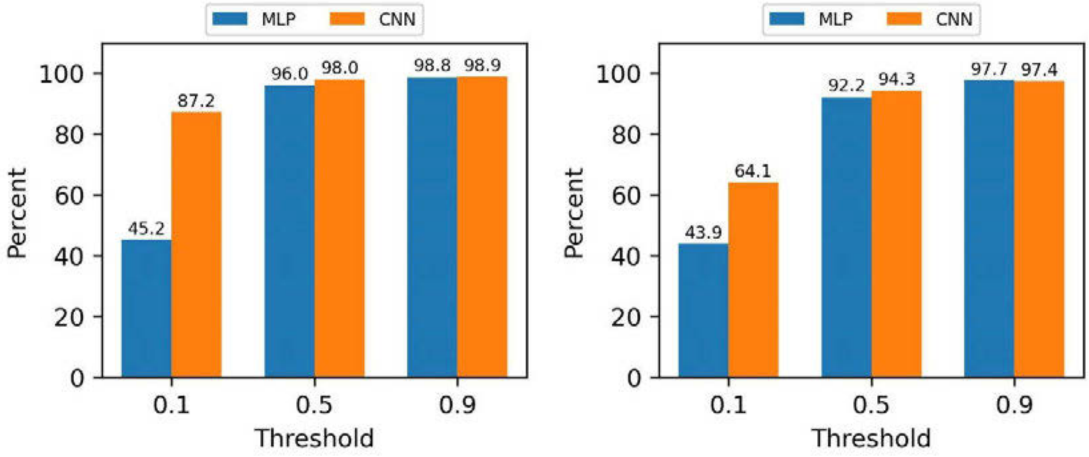


Figure 6. Training and validation error dynamics

3.1 Comparison of CNN and MLP model performance

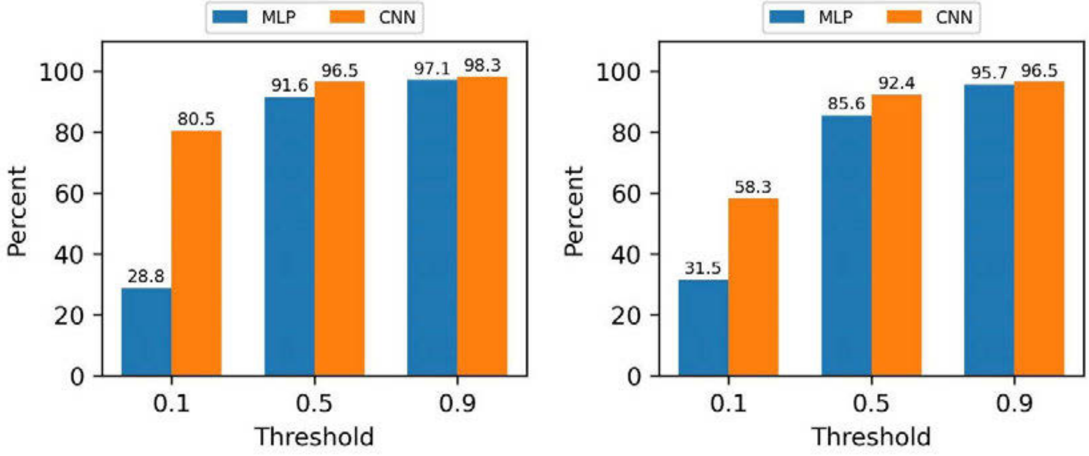
The complete model performances for the MLP and the CNN models for the four study cases are summarized in Tables 2 and 3. Figures 7 – 10 show the precision and recall for the two models at three selected thresholds (0.1, 0.5, and 0.9) for the four study cases (no-noise, demand-noise, mixed-noise, and random leaks). The results at these three thresholds are representative of all nine thresholds considered in this study, with 0.1, 0.5, and 0.9 indicating the most, the mild, and the least stringent condition, respectively. The figures show that precision is generally high (> 60%) for the CNN model at all three thresholds for all study cases except for the random leak case. Comparatively, precision for the MLP model is lower at all thresholds. The difference in precision between the two models is significantly high (> 40%) at 0.1 threshold for the no-noise, demand-noise, and mixed-noise cases. This difference, however, starts to diminish as the threshold becomes less stringent. The higher precision for the CNN model compared to the MLP model for all four study cases indicates its superiority in minimizing false leak predictions even with noise in the input data.

Similar to precision, recall for the CNN model is higher than the MLP model at the most stringent threshold (0.1) for all four study cases. However, the difference in recall of the two models at 0.1 threshold is not as high as the difference in precision. At the lesser stringent thresholds, particularly at 0.9, the difference in recall for the two models is insignificant for the no-noise, demand-noise, and mixed-noise cases. However, this difference is significant for the random leak case at all thresholds, with the CNN model outperforming the MLP model throughout. Overall, the recall results are consistent with the precision results in implying the superior performance of the CNN model over the MLP for the L-Town network.



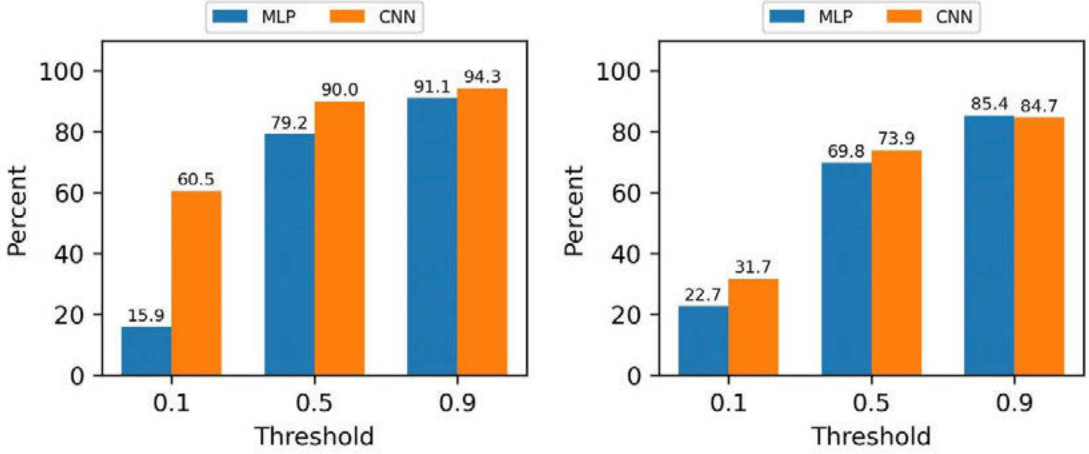
(a) *(b)*

Figure 7. Model performance for no-noise case: (a) Precision; (b) Recall.



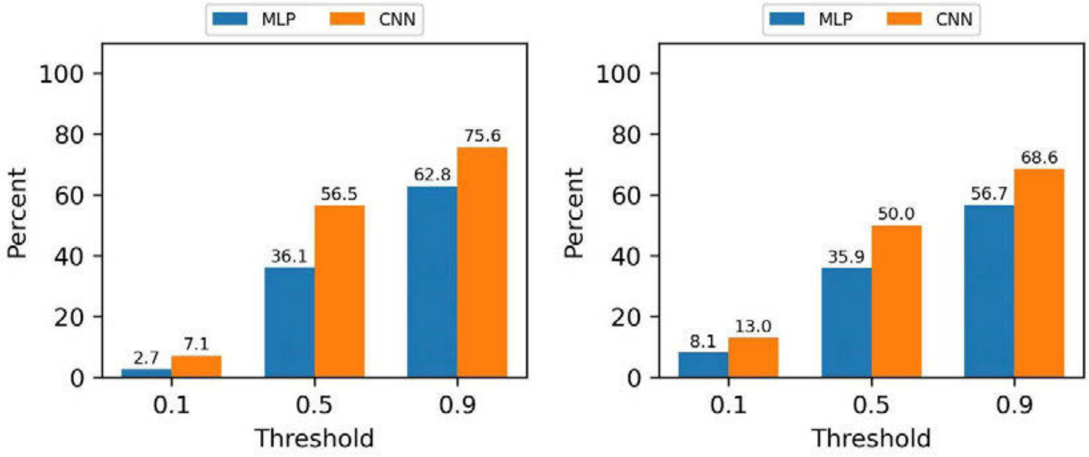
(a) *(b)*

Figure 8. Model performance for demand-noise case: (a) Precision; (b) Recall.



(a) *(b)*

Figure 9. Model performance for mixed-noise case: (a) Precision; (b) Recall.



(a) *(b)*

Figure 10. Model performance for random leaks case: (a) Precision; (b) Recall.

3.2 Problem complexity of the study cases

The precision and recall for the four study cases (with both CNN and MLP) were compared to understand the complexity of the leak detection task associated with each of the cases. Both precision and recall at all thresholds for the ideal but unrealistic *no-noise* case (Figure 7) rank highest compared to the other three study cases (Figure 8 – 10) for both MLP and CNN models. Precision and recall are comparatively high (> 40%) even at the most stringent threshold (0.1) for the *no-noise* case. These high accuracies can be attributed to the fact that the leak signatures in the input pressure difference data that are key to locating leaks are unaffected without noise. The *demand-noise* case ranks second among these four cases based on the precision and recall values. While the uncertainty in demand parameters in the hydraulic model can generate noise in the simulated pressure data, the noise is systematic. Therefore, it affects the leak signatures to a lesser degree. For the *mixed-noise* case, the 10% Gaussian noise added to the input pressure differences introduces randomness in the data that affect the leak signatures to a comparatively greater degree. Therefore, the precision and recall of the two models for the *mixed-noise* case are significantly lower compared to *no-noise* and *demand-noise* cases. The leak signatures are affected to the highest degree for the *random leaks* case. While no direct noises are added to the input data as is done for the *mixed-noise* case, the randomness in leak locations within a leak region introduces the possibility of a multitude of leak signatures for the same leak scenario, which is the most challenging for the machine learning models to learn. Therefore, the *random leaks* case ranks lowest in precision and recall. The effect of the complexity of the *mixed-noise* and the *random leaks* cases is profound at the 0.1 threshold because the artificial noise created by the randomness in the input data drowns out the changes in pressure input caused by a leak size or leak size difference of 0.1.

Table 2. Precision for the study cases

Threshold	No-noise		Demand-noise		Mixed-noise		Random leaks	
	MLP	CNN	MLP	CNN	MLP	CNN	MLP	CNN
0.1	45.2	87.2	28.8	80.5	15.9	60.5	2.7	7.1
0.2	78.5	94.4	63.7	89.7	43.2	77.0	9.1	22.5
0.3	89.4	96.3	79.9	93.5	61.6	84.0	17.7	36.9
0.4	93.9	97.4	87.4	95.4	72.6	87.7	27.1	48.0
0.5	96.0	98.0	91.6	96.5	79.2	90.0	36.1	56.5
0.6	97.3	98.5	94.0	97.2	83.7	91.5	44.2	63.0
0.7	98.0	98.7	95.5	97.7	87.0	92.6	51.2	68.3
0.8	98.5	98.8	96.5	98.1	89.3	93.5	57.3	72.3
0.9	98.8	98.9	97.1	98.3	91.1	94.3	62.8	75.6

Table 3. Recall for the study cases

Threshold	No-noise		Demand-noise		Mixed-noise		Random leaks	
	MLP	CNN	MLP	CNN	MLP	CNN	MLP	CNN
0.1	43.9	64.1	31.5	58.3	22.7	31.7	8.1	13.0
0.2	69.3	82.5	54.9	77.2	40.5	50.4	15.3	24.8
0.3	81.6	88.9	69.6	85.4	53.0	62.0	22.1	34.9
0.4	88.4	92.3	79.2	89.7	62.7	69.0	29.2	43.1
0.5	92.2	94.3	85.6	92.4	69.8	73.9	35.9	50.0
0.6	94.6	95.6	89.8	94.0	75.1	77.7	41.8	55.8
0.7	96.1	96.4	92.6	95.1	79.3	80.6	47.3	60.8
0.8	97.0	97.0	94.4	95.8	82.6	82.8	52.2	65.0
0.9	97.7	97.4	95.7	96.5	85.4	84.7	56.7	68.6

4 CONCLUSIONS

In this study, a machine learning-based approach is proposed for detecting leaks in WDSs that takes into account the characteristics of leaks present in real-world WDSs. The impact of WDS leak characteristics (varying size, multiple occurrences, and random location) and the uncertainties associated with the hydraulic model parameter and measuring devices are studied by analyzing the performance of two different machine learning models. One of the key findings of this study is that the effectiveness of the machine learning-based leak detection method is model-dependent. In this study, the CNN model is more effective than the MLP model in detecting leaks. While this result is specific to the study network (L-Town) using pressure differences as input, its implication expands beyond this study. It establishes the need to explore multiple models when developing a leak detection method. The other key finding of this study highlights the necessity of considering various types of leak scenarios that bear realistic leak characteristics to understand better the applicability of the leak detection models to real WDSs. Simplistic and unrealistic leak scenarios such as the *no-noise* case overestimate the performance of the models, as seen in this study. Models trained under such scenarios can severely underperform and be deemed useless for real WDSs. However, the high accuracies of the CNN and the MLP models for the three realistic study cases involving data noise, random leaks, and model and instrument uncertainties are proof of their potential for application to real-world leak detection problems. It is also important to point out that the locations of the pressure sensors used to generate the input data in this study are not based on hydraulic analysis and, therefore, are not optimal. Optimally located pressure nodes can further improve the accuracies of the models.

Several possibilities remain open for improving the work done in this study. Continuing the exploration of real-world leak characteristics, the addition of other types of noise can be considered for the input data. Using multiple inputs instead of a single input such as pressure is another possibility to improve leak detection accuracy. Our work in progress includes adding flow data alongside pressure data to predict leak locations and size. Finally, to understand the true potential of these leak detection models, the next step forward for this study is to apply them to a real-world WDS.

5 REFERENCES

[1] C. M. Fontanazza, V. Notaro, V. Puleo, P. Nicolosi, and G. Freni, “Contaminant intrusion through leaks in water distribution system: Experimental analysis,” Procedia Eng., vol. 119, no. 1, pp. 426–433, 2015.

[2] R. Liemberger and A. Wyatt, “Quantifying the global non-revenue water problem,” Water Sci. Technol. Water Supply, vol. 19, no. 3, pp. 831–837, 2019.

[3] J. Yu et al., “An integrated bottom-up approach for leak detection in water distribution networks based on assessing parameters of water balance model,” Water (Switzerland), vol. 13, no. 6, 2021.

[4] A. Rajeswaran, S. Narasimhan, and S. Narasimhan, “A graph partitioning algorithm for leak detection in water distribution networks,” Comput. Chem. Eng., vol. 108, pp. 11–23, 2018.

[5] Z. Hu, D. Tan, B. Chen, W. Chen, and D. Shen, “Review of model-based and data-driven approaches for leak detection and location in water distribution systems,” Water Supply, vol. 21, no. 7, pp. 3282–3306, 2021.

[6] D. Zaman, M. K. Tiwari, A. K. Gupta, and D. Sen, “A review of leakage detection strategies for pressurised pipeline in steady-state,” Eng. Fail. Anal., vol. 109, no. October 2019, 2020.

[7] I. Santos-Ruiz, J. Blesa, V. Puig, and F. R. López-Estrada, “Leak localization in water distribution networks using classifiers with cosenoidal features,” IFAC-PapersOnLine, vol. 53, no. 2, pp. 16697–16702, 2020.

[8] M. B. Abdulla and R. Herzallah, “Probabilistic multiple model neural network based leak detection system: Experimental study,” J. Loss Prev. Process Ind., vol. 36, pp. 30–38, 2015.

[9] Y. Wu and S. Liu, “A review of data-driven approaches for burst detection in water distribution systems,” Urban Water J., vol. 14, no. 9, pp. 972–983, 2017.

[10] M. Zhou, Z. Pan, Y. Liu, Q. Zhang, Y. Cai, and H. Pan, “Leak Detection and Location Based on ISLMD and CNN in a Pipeline,” IEEE Access, vol. 7, pp. 30457–30464, 2019.

[11] H. Shukla and K. Piratla, “Leakage detection in water pipelines using supervised classification of acceleration signals,” Autom. Constr., vol. 117, no. May, p. 103256, 2020.

[12] E. J. Pérez-Pérez, F. R. López-Estrada, G. Valencia-Palomo, L. Torres, V. Puig, and J. D. Mina-Antonio, “Leak diagnosis in pipelines using a combined artificial neural network approach,” Control Eng. Pract., vol. 107, no. September 2020, p. 104677, 2021.

[13] J. Bohorquez, B. Alexander, A. R. Simpson, and M. F. Lambert, “Leak Detection and Topology Identification in Pipelines Using Fluid Transients and Artificial Neural Networks,” J. Water Resour. Plan. Manag., vol. 146, no. 6, p. 04020040, 2020.

[14] J. Mashford, D. De Silva, S. Burn, and D. Marney, “Leak detection in simulated water pipe networks using SVM,” Appl. Artif. Intell., vol. 26, no. 5, pp. 429–444, 2012.

[15] A. Soldevila, J. Blesa, S. Tornil-Sin, E. Duviella, R. M. Fernandez-Canti, and V. Puig, “Leak localization in water distribution networks using a mixed model-based/data-driven approach,” Control Eng. Pract., vol. 55, pp. 162–173, 2016.

[16] A. Soldevila, R. M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig, “Leak localization in water distribution networks using Bayesian classifiers,” J. Process Control, vol. 55, pp. 1–9, 2017.

[17] M. Quinõnes-Grueiro, C. Verde, A. Prieto-Moreno, and O. Llanes-Santiago, “An unsupervised approach to leak detection and location in water distribution networks,” Int. J. Appl. Math. Comput. Sci., vol. 28, no. 2, pp. 283–295, 2018.

[18] J. Blesa and R. Pérez, “Modelling uncertainty for leak localization in Water Networks ˜ ,” vol. 51, no. 24, pp. 730–735, 2018.

[19] X. Zhou et al., “Deep learning identifies accurate burst locations in water distribution networks,” Water Res., vol. 166, p. 115058, 2019.

[20] E. G. Mohammed, E. B. Zeleke, and S. L. Abebe, “Water leakage detection and localization using hydraulic modeling and classification,” J. Hydroinformatics, vol. 23, no. 4, pp. 782–794, 2021.

[21] D. G. Eliades, M. Kyriakou, S. Vrachimis, and M. M. Polycarpou, “EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB,” in 14th Int. Conf. on Computing and Control for the Water Industry, 2016.

[22] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[23] S. G. Vrachimis et al., “BattLeDIM: Battle of the leakage detection and isolation methods.,” in Proc., 2nd Int. CCWI/WDSA Joint Conf., 2020.

[24] S. P. Lloyd, “Least Squares Quantization in PCM,” IEEE Trans. Inf. Theory, vol. 28, no. 2, pp. 129–137, 1982.

[25] K. A. Klise, M. Bynum, D. Moriarty, and R. Murray, “A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study,” Environ. Model. Softw., vol. 95, pp. 420–431, 2017.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14833

WATER DISTRIBUTION NETWORK DISRUPTIVE EVENTS. GENERATION AND EXPLOITATION OF AN INCIDENT HUB TO INCREASE THE NETWORK PREPAREDNESS

David Ayala-Cabrera[1], Jorge Francés-Chust[2], Sotudeh Hoseini-Ghafari[3], Gemma Stanton[4] and Joaquín Izquierdo[5]

[1,3,4]CWRR-School of Civil Engineering, University College Dublin, Dublin (Ireland)
[2]Business Development, Aguas de Bixquert, S.L., Xàtiva, Valencia (Spain)
[5]FluIng-IMM, Universitat Politècnica de València (Spain)

[1] david.ayala-cabrera@ucd.ie, [2] jorge@abxat.com, [3] sotudeh.hoseinighafari@ucdconnect.ie,
[4] gemma.stanton@ucdconnect.ie, [5] jizquier@upv.es.

Abstract

This paper seeks to develop increased knowledge about disruptive events in a water distribution network (WDN) through the experience acquired by previous anomalous events in the system. This work explores the various relationships between several parameters in an incident hub (specifically water loss events) in a Spanish real, small WDN. The incident hub consists of basic elements recorded during an incident (e.g. breakdown, maintenance activity, among others) and the corresponding causes that generated the incident (e.g. breakage due to excess pressure, breakage due to tree roots, etc.), as well as the management times of the incident (e.g. awareness time, isolation, and repair time). The utility collected and stored these data, which were completed with direct interviews with the system operator. Measurements were performed at pressure and flow sensors, which allowed evaluating the effects of both the incident itself and the actions taken to solve it. The records of the incidents are categorised depending on the nature of the data they contain to facilitate mapping their causes and effects. To characterise the disruptive events, a feature extraction process has been proposed using a temporal-spatial approach combined with a migration proposed that describes parameters' behaviour in the spatial dimension for a certain period of time. The characteristics obtained in the previous lessons of the incidents contained in the incident hub are compared with potential causes obtained with different control parameters. The objective is to determine the potential causal relationship of the incident that allows its characterisation. The results of this characterisation are presented and analysed in this contribution. The outcomes are promising in the sense of a clear ability to provide WDNs with key parameters that foster prediction and classification processes.

Keywords
Hub of incidents, Water distribution networks, Spatial-temporal analysis, Network preparedness, Resilience, Protection of critical infrastructures, Intelligent data analysis.

1 INTRODUCTION

Water distribution networks (WDNs) are critical infrastructures which are exposed to multiple challenges that can stop guaranteeing their correct operation concerning the satisfaction of their basic objectives, that is, ensuring the provision of an adequate quantity of water demanded, the continuity of the service and its water quality. These systems need to be prepared both to face these different challenges and to avoid disruptions or reduce their potential impacts. Utilities are currently focusing on increasing the systems' preparedness as this is seen as a powerful tool that can help them deal with these challenges. The preparedness of the system has normally been used in the context of emergencies. However, events not considered emergencies (or rarely considered as such) can cause situations that can cause or increase possible negative effects on the network.

The resilience of infrastructures normally refers to their capabilities to face certain threats. In this sense, these capabilities can be referred to as absorptive, adaptive and restorative. The absorptive capacity refers to the ability of the system to deal with abnormal events without any intervention other than its capabilities. The adaptive and restorative capabilities are more related to the interventions deployed to mitigate the effects of the anomalous event (adaptive capacity) and to fix it from a long-term perspective (restorative capacity). However, there is an additional ability of the system that many authors refer to (*e.g*. see [7]), which is the ability to learn from the events that have occurred. The events that have occurred can provide the system with information to not only face events with similar characteristics but also to face new potential challenges.

A data Hub is seen in this paper as a structured data set/digital repository (collections of data) that allows quick, digital access to information among users that may be relevant for a certain use, e.g. [1]. A data Hub with an architecture designed to centralise incident data (incident Hub) can provide the system with information on how to avoid/prepare for; how to respond to; and define a series of actions, all of these to better deal with incidents of similar characteristics and also avoid/minimise the impacts of other events with greater relevance. For example, in the field of cybersecurity, there are incident Hubs (public or private) that cover aspects such as the incident response process to improve it in the cloud service [2], cyber-attack reports on critical infrastructure [3], or offering a coordinated response to a variety of cyber security issues [4]. To increase the capacity of the system to learn about previous events, this paper presents a data Hub created from incidents in a real WDN in Spain. In addition, this paper proposes a method to extract the acquired knowledge contained in the Hub. This information can be used to establish maintenance models [5] that can minimise the impact or prevent the incident from occurring. The relationships and dependencies existing in the spatial dimension of the control parameters are contrasted with the patterns obtained from the Incident Hub to characterise the system. This ultimately increases the system resilience derived from the obtained increase in the network preparedness to cope with such potential events [6]. This characterisation also favours the elaboration of good practices in WDNs that can help in two crucial aspects: the sustainability of the system through the preservation of resources, and the mitigation of impacts due to extreme events.

2 PROPOSED FRAMEWORK

This section presents the proposed framework to increase network preparedness in WDNs through the knowledge gained from previous incidents that occurred in the system (Figure 1). This is through the generation and potential exploitation of an Incident Hub. In the proposed framework in this paper, for the incident Hub architecture, two types of cores are distinguished: i) causes/times core, and ii) effects on the system core. These cores are, initially, associated with three types of incidents: 1) water leaks, 2) maintenance or operational activity, and 3) additional demand of water. Details about the creation and categorisation of the components of the incident Hub are described in Section 3.

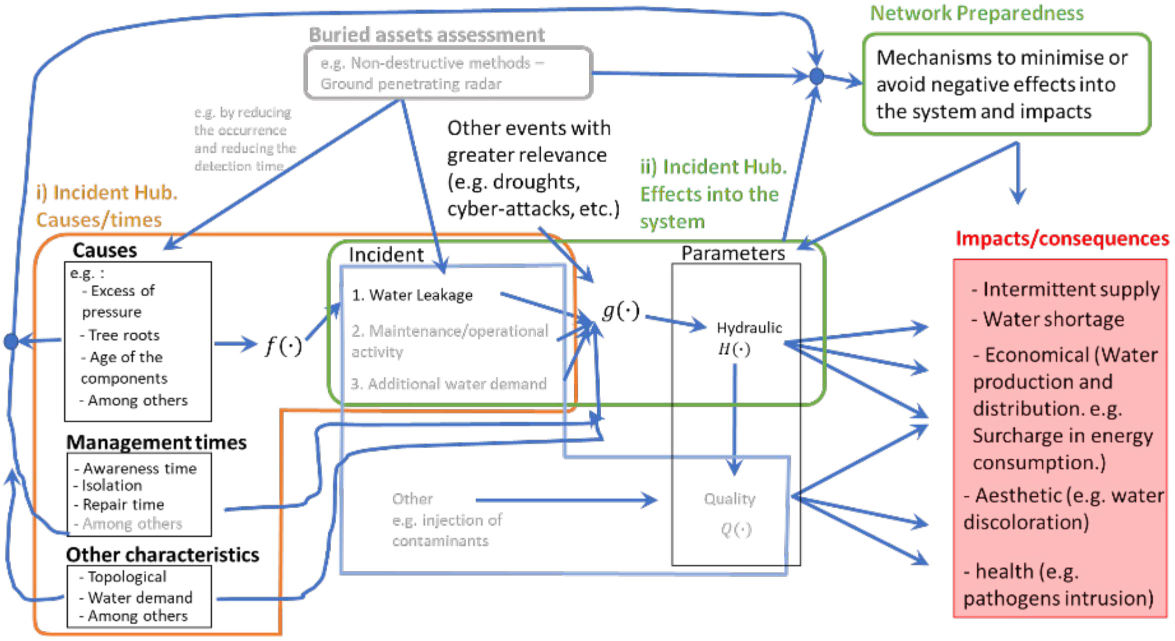

Network Preparedness
Mechanisms to minimise or avoid negative effects into the system and impacts
Buried assets assessment
e.g. Non-destructive methods – Ground penetrating radar
e.g. by reducing the occurrence and reducing the detection time
Other events with greater relevance (e.g. droughts, cyber-attacks, etc.)
ii) Incident Hub. Effects into the system
i) Incident Hub. Causes/times
Causes
e.g. :
- Excess of pressure
- Tree roots
- Age of the components
- Among others
$f(\cdot)$
Incident
1. Water Leakage
2. Maintenance/operational activity
3. Additional water demand
$g(\cdot)$
Parameters
Hydraulic
$H(\cdot)$
Quality
$Q(\cdot)$
Other
e.g. injection of contaminants
Management times
- Awareness time
- Isolation
- Repair time
- Among others
Other characteristics
- Topological
- Water demand
- Among others
Impacts/consequences
- Intermittent supply
- Water shortage
- Economical (Water production and distribution. e.g. Surcharge in energy consumption.)
- Aesthetic (e.g. water discoloration)
- health (e.g. pathogens intrusion)

Figure 1. Incident Hub. Proposed framework to increase system preparedness through knowledge gained from previous incidents

3 INCIDENT HUB. GENERATION

The network under study corresponds to a real small WDN located in Spain. Let us consider a WDN as a graph $\boldsymbol{G} = \{\boldsymbol{V}, \boldsymbol{E}\}$, where $\boldsymbol{V}$ is the set of vertices/nodes (*i.e.* tanks, reservoirs, junction nodes) with a total of $\boldsymbol{n_n}$ nodes, and $\boldsymbol{E}$ is the set of edges/links that join these nodes (*i.e.* pipes, valves, pumps) with a total of $\boldsymbol{n_l}$ links. In this case, the network under study is represented by $\boldsymbol{n_n} = \mathbf{188}$ nodes, *i.e.* 4 tanks, 2 reservoirs, and 182 junction nodes; and $\boldsymbol{n_l} = \mathbf{230}$ links, *i.e.* 212 pipes, 16 valves (it should be mentioned that only a few isolation valves whose states are closed were considered in this initial representation of the network), and 2 pumps. It is worth mentioning that the network corresponds to a schematic configuration and that its spatial coordinates do not correspond precisely to the real georeferencing.

Incident Hub. Causes/times core. To obtain relevant data about the incidents that affected in some manner the operation of the WDN (e.g. breakdown, operational or maintenance activities, water demands by users external to the network under study), a series of interviews with the operator was conducted. In the first interviews, a template was generated where information on basic aspects of the incidents was collected. The template was completed and is currently populated in an online manner by the system operator as incidents occur. Afterwards, the collected data is completed through regular interviews. The evaluation period in this paper corresponds from July 2021 to May 2022.

The aspects collected in this core of the Hub can be categorised as 1) the causes that led to the anomaly in the system, and 2) the management times of the incident. In this paper, the causes can be seen as a function $\boldsymbol{f}(\cdot)$ that characterises those factors, and the relationships among them, that can lead to a water leak incident. The main interest of this paper is to advance in the identification of these potential relationships. The effects of anomalous events can eventually be reflected, in the first instance, in hydraulic and water quality parameters, such as $\boldsymbol{H}(\cdot)$ and $\boldsymbol{Q}(\cdot)$. However, the effects of these parameters are the result of the contribution of multiple factors $\boldsymbol{g}(\cdot)$ (occurring simultaneously or sequentially). This paper considers those management elements as factors whose contribution can avoid, increase or minimise the negative effects that can be reflected in the hydraulic and quality parameters of the system (see Figure 1).

Other aspects in this core, such as topological characteristics of the system, and demand (from the own system or due to sales of water to other neighbour utilities), were located and were also collected in this Hub. They were considered as a contributor to $\boldsymbol{g}(\cdot)$. Topological aspects of the network were compiled into the hydraulic model of the network. Considering that the hydraulic model is a simplification of the real model, the incidents were allocated in the hydraulic model at the point that best fit their real position (according to the operator criteria).

Functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ jointly characterise the system in terms of a specific anomalous event and can be fed from previous experiences. This will increase the capacity of the WDN to learn and consequently avoid/minimise impacts on the system itself or other related ones.

Incident Hub. Effects on the system. Testing resilience is an effective manner to prepare the system to deal with events that may affect its proper operation. This preparation can be in terms of avoiding or minimising the negative effects that these anomalies can potentially cause. Testing resilience through simulated events, although an effective tool, has considerable limitations in terms of the selection of the event to be simulated, the number of parameters to be assessed, and in terms of the difficulty to incorporate the coexistence of more than one event simultaneously. This adds to the difficulty of obtaining a hydraulic model that can adequately capture the behaviour of the system.

The effects of an anomalous event on the system can be estimated through relationships established with its basic parameters (*i.e.* hydraulic and quality) and system characteristics (*e.g.* topological configuration). The valuable information provided by these stress tests on the system can, however, be complemented by extracting the characteristics of the effects from the basic parameters (*e.g.* measured through sensors) in events that have already occurred. In this manner, it is possible to consider a variety of other factors that are not easily incorporated into the simulation or that can normally be ignored. In this line, this work proposes the incorporation of a core for the Incident Hub that contains the effects of the anomalous event in terms of pressure and flow measured through the sensors available in the network during the period of study.

4 CATEGORISATION OF THE INCIDENTS BASED ON EXPERT KNOWLEDGE

The total number of incidents recorded during the evaluation period was 70 cases distributed in 56 water leak cases, 11 reported cases of maintenance or operational activities (it should be mentioned that not all maintenance operations were recorded), and three cases of additional water consumption events (water sold to another WDN, this last case still in progress).

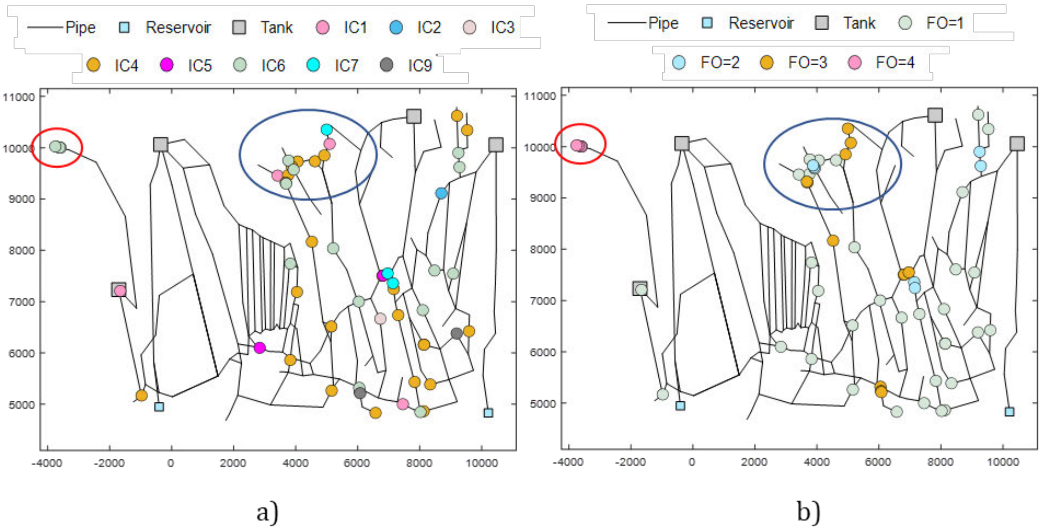
Based on the causes of each event provided by the utility's operator experience, the incidents observed during the evaluation period were classified into eleven potential incident categories (IC). The eleven selected categories were:

- **IC1. Settlement.** (4 cases) This category corresponds, among others, to the differential settlement of walls or the ground due to heavy vehicle traffic.
- **IC2. Asset fissure/crack**. (1 case) This category was assigned when no other breakage category was attributable. According to the experience of the operator, this type of case has been presented before.
- **IC3. Tree roots.** (2 cases).
- **IC4. Pressure inference.** (26 cases) This category includes all those labels provided by the operator that included pressure exclusively (high pressure, 14 cases; and pressure oscillation, 1 case), and also those that included one or more other causes besides pressure. Other causes besides pressure include 1) 2 cases - settlement (land settlement, heavy vehicles traffic); 2) 2 cases - Age of the component; 3) 5 cases - defective component;

4) 1 case - tree roots and material wear; and 5) 1 case - inappropriate pipe (low-density material).

- **IC5. Age/corrosion.** (2 cases).
- **IC6. Pipe defects**. (17 cases). This category includes inappropriate pipe wall thickness, contraction and expansion of the pipe, and poor quality of the component, among others.
- **IC7. Exposed components.** (2 cases). This category refers to incidents occurring in components of the network that are exposed to the surface.
- **IC8. Maintenance activity.** (10 cases).
- **IC9. Weather conditions/component frozen.** (2 cases).
- **IC10. Additional consumption.** (3 cases).
- **IC11. Others.** (1 case).

The spatial location of the categories that correspond to incidents that resulted in water leaks (IC1-IC7, and IC9) is presented in Figure 2a. In addition, the quantification of the frequency of occurrence (FO) of water leaks in a particular pipe is presented in Figure 2b.



a) b)

Figure 2. Incident Hub. a) Incident categories (IC), and b) frequencies of occurrence (FO) of a leak in the same pipe.

Figure 2 shows in its two insets a marked tendency of a particular zone to present water leaks (blue circle). This particular area presents diversity in incident categories (see Figure 2a) and in particular shows a high concentration of the category associated with pressure (IC4). Similarly, this zone also shows a tendency to leak water in at least two of its pipes (Figure 2b). An area that does not present diversity in terms of categories of incidents and has the highest frequency of occurrence is found at the end of the network (red circle). Other zones that present an FO of three and high diversity can be observed in Figure 2b. In general, the network shows a tendency to leak due to pressure and has its occurrences concentrated in specific pipes.

5 EXTRACTION OF KNOWLEDGE FROM PREVIOUS LESSONS - CAUSES

This section proposes a method for the extraction of knowledge from previous leakage events in WDNs and in particular the cause element of the incident Hub, *i.e.*, to increase the knowledge in

cc BY NC SA

$\boldsymbol{f}(\cdot)$. To this end, this section proposes the spatial analysis of the data collected through the use of a data migration process [7]. Two types of data are distinguished at this point, 1) data referring to the knowledge generated by the event already occurred (previous lessons), and 2) parameter data (potential cause - not exclusively hydraulic), whose potential contribution to the event under study is to be identified. On the part of the information provided by the previous events, in this document, the results obtained with the frequency of occurrence of the leak event during the sampling period are used in the first instance to conduct the migration process. Regarding the parameters to be analysed, in particular pressure, this paper proposes a clustering process based on pressure [8] as a process prior to the migration process.

MatLab's `griddata` function was used in this document to represent sparse data in a spatially homogeneous manner by constructing it in a grid. The resulting matrix has dimension $\boldsymbol{m} \times \boldsymbol{n}$. This matrix covers the maximum and minimum spatial coordinates of all vertices, $\boldsymbol{n_n}$, (network nodes). Let us denote the resulting spatial matrices as $\boldsymbol{D}_{\mathrm{pl}}$ and $\boldsymbol{D}_{\mathrm{pc}}$, for the previous lessons and from the potential cause; respectively.

Migration process. To identify the spatial relationships among the elements, this section proposes the condensation of information through its migration. This is an iterative process (based on [7]) and essentially consists of the transmission of information to a central cell, from its neighbours for a given matrix $\boldsymbol{D_p}$ ($\boldsymbol{D}_{\mathrm{pl}}$or $\boldsymbol{D}_{\mathrm{pc}}$) as $\boldsymbol{M}_{\mathrm{p},\boldsymbol{i},\boldsymbol{j}}$ = $\left(\left(\sum_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{k}=\boldsymbol{4}} \boldsymbol{D}_{\mathrm{p},\boldsymbol{cell_k}}\right) + \boldsymbol{D}_{\mathrm{p},\boldsymbol{i},\boldsymbol{j}}\right)$. A diamond-shaped configuration was proposed to transmit information from neighbours, *i.e.* $\boldsymbol{cell} = \{(\boldsymbol{i}+\boldsymbol{1},\boldsymbol{j}),(\boldsymbol{i},\boldsymbol{j}-\boldsymbol{1}),(\boldsymbol{i},\boldsymbol{j}+\boldsymbol{1}),(\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j})\}$ with $\boldsymbol{k} = \{\boldsymbol{1},\dots,\boldsymbol{4}\}$ in this paper. $\boldsymbol{D}_{\mathrm{p}}$ is updated with the information from the $\boldsymbol{M}_{\mathrm{p}}$ of the previous iteration (after $\boldsymbol{M}_{\mathbf{p}}$ has been normalised) and thus the coverage of the transferred information is increased. In this paper, three numbers of iterations were explored: 1) number of iterations when the correlation coefficient between $\boldsymbol{D}_{\mathrm{p}}$ and $\boldsymbol{M}_{\mathrm{p}}$ starts to be constant, 2) number of iterations equal to $\boldsymbol{n}$, and 3) number of iterations equal to $\boldsymbol{m}$.

This iterative process transmits information from cells with lower values to cells with higher values. To evaluate which space best characterises the event, three variants were evaluated, 1) “to-max”; it corresponds to the matrix $\boldsymbol{D}_{\mathrm{p}}$ preserved in its initial condition, 2) “to-min”; it consists of the variation of $\boldsymbol{D}_{\mathrm{p}}$ towards minima (*i.e.* $(-(\boldsymbol{D}_{\mathrm{p}} - \max(\boldsymbol{D}_{\mathrm{p}}))) + \min(\boldsymbol{D}_{\mathrm{p}})$), and 3) “to-both”, which consists of the difference for $\boldsymbol{M}_{\mathrm{p}}$ resulting from variants 1 and 2.

Previous lessons and potential causes, migration.

Figure 3 presents the migration process for the matrix of previous lessons corresponding to spatially located water leaks and their corresponding frequencies of occurrence in a particular pipeline. Variants of the migration space and also the number of iterations (50, 142 and 308) were selected, looking if any of them could represent the past event from which information is to be extracted. Considering that the WDN model is both a skeletonized and schematic configuration of the system, it can be assumed that in the analysis it is necessary to incorporate some uncertainty regarding the spatial location of the event. For this reason, in this work, it is considered that the "to-max" configuration can better represent the evaluated period when compared with "to-min" and "to-both" for the matrix of previous lessons.

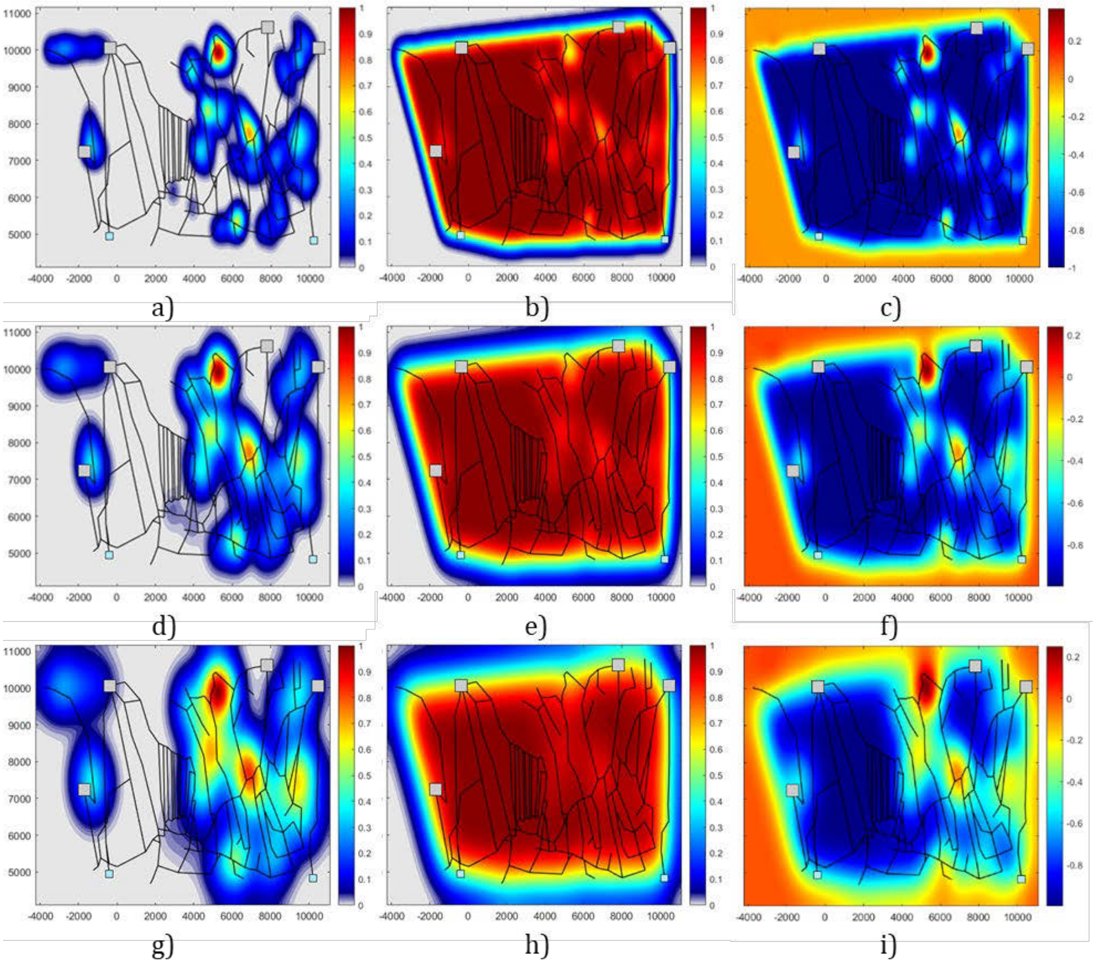


Figure 3. Previous lessons. Incidents of water leaks and frequency of their occurrences - migration. Variant: (a, d, g) to-max; (b, e, h) to-min; (c, f, i) to-both; (a-c) 50 iterations; (d-f) 142 iterations; (g-i) 308 iterations

For the analysis of the matrix of potential causes, a simulation of the hydraulic model of an arbitrary day was conducted. Each of the pressures obtained at each node was classified into clusters according to the percentage of time that each node spent in a particular range of pressures (as in [8]). The results of the conglomerate as a function of the pressure and the percentage of time that each node spent in this pressure cluster obtained are presented in Figure **4**.

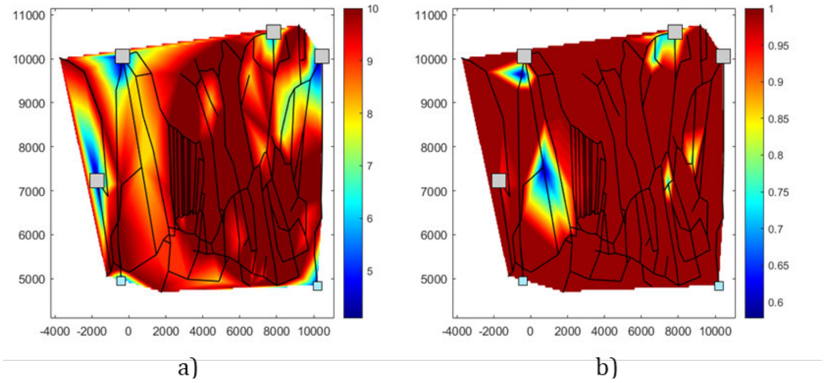


Figure 4. Potential Causes. Pressure-based clusters. a) Pressure-based cluster and b) percentage of time spent in a specific cluster

cc BY NC SA

For the matrix of potential causes (pressure-based clusters), the results in the variants of the proposed migration spaces are presented in Figure 5.

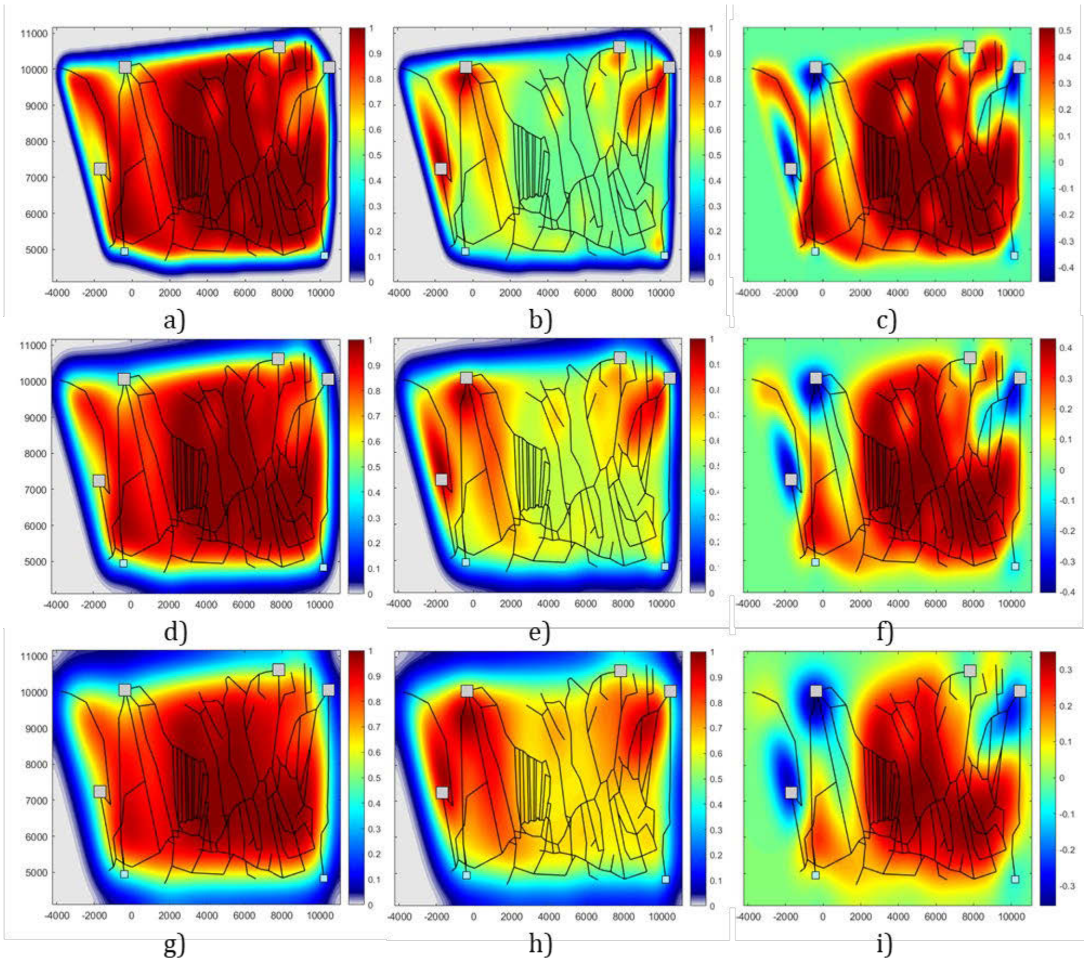


Figure 5. Potential causes. Pressure-based clusters - migration. Variant: (a, d, g) to-max; (b, e, h) to-min; (c, f, i) to-both; (a-c) 50 iterations; (d-f) 142 iterations; (g-i) 308 iterations

The results of going through the migration process for both the pressure-based cluster matrix and the percentage of time spent in a specific cluster were compared with the matrix results from previous lessons. These results are presented in Table 1 and Table 2 respectively.

Table 1. Correlation coefficients for previous lessons and pressure-based cluster

Previous lessons	Iter. 50			Iter. 142			Iter. 308		
Potential causes	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308
to-max	0.978	0.968	0.944	0.409	0.410	0.405	0.459	0.467	0.471
to-both	0.673	0.558	0.412	0.431	0.428	0.412	0.498	0.505	0.496
to-min	0.945	0.949	0.936	0.269	0.272	0.266	0.290	0.298	0.300

Table 2. Correlation coefficients for previous lessons and percentage of time spent in a specific cluster

Previous lessons	Iter. 50			Iter. 142			Iter. 308		
Potential causes	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308
to-max	0.991	0.984	0.963	0.389	0.392	0.388	0.433	0.443	0.449
to-both	0.841	0.803	0.727	0.425	0.444	0.467	0.480	0.509	0.546
to-min	0.988	0.980	0.956	0.336	0.336	0.328	0.371	0.377	0.377

Both parameters showed high correlation coefficients indicating that both the pressure and the frequency of time spent under a given pressure have a high contribution to the occurrence of the events evaluated. Likewise, correlations as high as those obtained for "to-max" with iter. 50 for the potential causes indicate that both high pressure and high times have a very big contribution in this particular network, even though the final cause labelled for the expert was a different one in some cases. Finally, both potential matrices can represent these events in a very approximate way for this system.

6 CONCLUSIONS

In this work, the generation of an Incident Hub and potential uses to increase the network preparedness and consequently its resilience has been presented. This paper presents the different causes collected in this Incident Hub. The results show that the network is prone to water leaks essentially due to high pressure in the system. In addition, this work proposes a method to extract knowledge from previous events. It is also shown how it is possible to transfer knowledge from previous events. The results are promising since the proposed methodology can be used for evaluating the contribution to the events of other parameters, as well as they can serve as bases for the training of intelligent systems.

7 REFERENCES

[1] (DRI), "Digital Repository Ireland," https://www.dri.ie/, (accessed on: Apr. 2022).

[2] SITECORE "Functional documentation," https://docs.stylelabs.com/, (accessed on: Apr. 2022).

[3] TISAFE "Incident Hub – Industrial Cybersecurity incidents Database," https://hub.tisafe.com/, (accessed on: Apr. 2022).

[4] CIH "Cyber Incident Hub," https://www.cyber-incident-hub.ch/, (accessed on: Apr. 2022).

[5] B. G. Mwanza and C. Mbohwa "The Impact of Maintenance Systems on Water Supply: A Case Study of a Utility Company," in Proc. IEOM2016. Kuala Lumpur, Malaysia. (2016).

[6] D. Ayala-Cabrera, O. Piller, J. Deuerlein, M. Herrera, "Key Performance Indicators to Enhance Water Distribution Network Resilience in Three-Stages". Water Utility Journal, vol.19, 2018, pp. 79–90.

[7] S. Hoseini-Ghafari, J. Francés-Chust, O. Piller, and D. Ayala-Cabrera, "Exploring a Spatial Dynamic Approach and Landmark Detection for Leakage/Burst event Characterisation in Water Distribution Networks," in Proc. International Environmental Modelling and Software Society (iEMSs), Brussels, Belgium, 2022. Accepted.

[8] D. Ayala-Cabrera, M. Castro-Gama, C. Quintiliani, S. Hoseini-Ghafari, "Multi-step optimization approach to rehabilitate a deteriorated and intermittent water distribution network," in Proc. 2nd International Joint Conference on Water Distribution System Analysis & Computing and Control in the Water Industry (WDSA/CCWI) - "The Battle of the Intermittent Water Supply", Valencia, Spain. Accepted.

[9] E. Lawson, S. Bunney, S. Cotterill, R. Farmani, P. Melville-Shreeve, D. Butler, "COVID-19 and the UK water sector: Exploring organizational responses through a resilience framework," Water and Environment Journal, vol. 36, no. 1, 2022, pp. 161-171.

BY NC SA

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14857

EXPERIMENTAL ANALYSIS FOR THE LOSSES ASSESSMENT IN WATER DISTRIBUTION SYSTEMS

Giovanna Darvini[1] **and Luciano Soldini**[2]
[1,2]Department ICEA Università Politecnica delle Marche, Ancona, (Italy)
[1] *g.darvini@univpm.it*, [2] *l.soldini@univpm.it*

Abstract

Water leaks assessment is a crucial aspect for the management of the water distribution systems. One strategy commonly used for the leakage reduction is to control the network pressure. Many studies have been dedicated to both the definition of the law between the leaks flow rate and the network pressure and the parameters that influencing the phenomenon. This paper presents the preliminary results of an experimental research on the water leaks evaluation conducted at the Laboratory of Hydraulics and Maritime Construction at the Università Politecnica delle Marche in Ancona, Italy. The tests were carried out varying the hydraulic operating conditions and simulating leakages from holes with different shape and size. Furthermore, different pipe materials were considered. The first results of the laboratory experiments show that the leakage from the hole increases as the pressure in the pipeline increases. The coefficients of the mathematical law for the leakage evaluation in function of the pressure were estimated. The results were compared with similar studies described in the literature.

Keywords
Water distribution systems, leakage-pressure relationship, experimental setup.

1 INTRODUCTION

The leakage reduction in water distribution systems (WDSs) is a fundamental aspect for water utilities as it is one of the macro-indicators of the technical conditions for providing the service. Leakage reduction requires the identification of the parameters that most influence the amount of volume lost and the definition of the relationships that describe the phenomenon, also in support of hydraulic modelling [1]. This issue has been addressed in recent years through both laboratory tests (e.g. [2]; [3]) and numerical modelling (e.g. [4]). The equation that generally describes the flow rate Q exiting a hole as the pressure P varies is:

$$Q = aP^b \tag{1}$$

where a and b are the coefficient and exponent respectively in the loss model. The values of the exponent b derived from experiments and field studies are between 0.5 and 2.79 (e.g. [5]). Factors influencing the value of b include the material, type of rupture, soil characteristics, flow rate and nature of the phenomenon. Recent studies have shown that the area of the leakage opening A_f (e.g. longitudinal or circular holes or cracks) increases linearly with the pressure value, for different materials and different loading conditions (e.g. [4]). The relationship linking the change in hole area to leakage flow is linear when the behaviour of the material constituting the pipe is of the elastic type. De Marchis and Milici [6] investigate the effectiveness of the suggested formulations by different authors finding that in the absence of leak area deformation, the exponent b of (1) is 0.51 and the discharge coefficient a linearly increases as the leak area grows and the slope of the linear trend is higher for circular leak than for transverse cracks. For elastic-plastic or viscoelastic behaviour, on the other hand, it is necessary to introduce other formulations. Several studies have analysed the behaviour of the leakage hole area depending on the pressure value and have proposed further modifications to introduce and quantify new aspects (e.g. [7] and [8]).

In this paper, the results of an extensive series of experimental tests in which a water leak in a pressurised system was simulated are analysed; the tests were conducted under varying system operating conditions, flow rate and pressure, and hole shape and size.

A first series of experiments was conducted on a PVC DN110 PN16 pipe by simulating leakage with the insertion of a tap on top of which metal nozzles with holes of different shapes and sizes were placed.

For a more realistic representation of leakage, the setup was recently modified to start a second series of experiments on a cast-iron pipe DN100 PN16 where a 2 mm diameter circular hole was drilled in the pipe wall.

2 EXPERIMENTAL SETUP

The experimental research for the evaluation of water loss under varying hydraulic operating conditions and hole geometry was conducted at the Laboratory of Hydraulics and Maritime Construction at the Università Politecnica delle Marche in Ancona, Italy.

2.1 Original setup

A pressurised hydraulic system consisting of DN110 PN16 PVC pipes with a total length of 20m was recently built in this facility, into which a DN100 spheroidal cast-iron pipeline with a length of 6m was inserted (see Figure 1). The circuit is fed by a 3m^3 free surface tank, while the flow rate and pressure conditions are guaranteed by a Caprari CVX321/3 5.5kW vertical multistage radial impeller pump with a nominal frequency of 50Hz. During operation the flow rate and operating head are 7.5l/s and 45.98m respectively, the minimum flow rate is 4l/s at a head of 56.93m while the maximum flow rate is 10.2l/s at a head of 32.57m. The electric pump is equipped with a three-phase inverter that allows the number of motor revolutions to be varied in terms of frequency from 35Hz to 50Hz, while a spheroidal cast-iron DN100 and PN16 flow valve with a parabolic shutter is installed at the end of the circuit, before the outlet into the discharge tank. This equipment makes it possible to regulate the operation of the system in different pressure and flow terms. The overall layout of the system is shown in Figure 1.

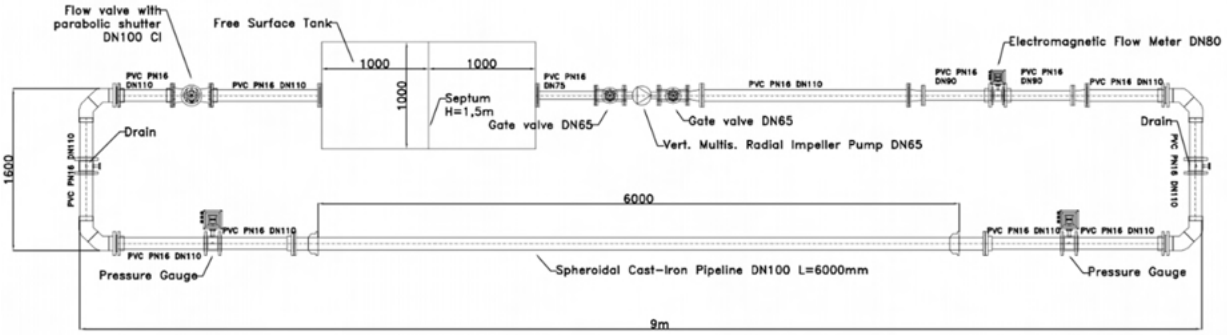

Flow valve with parabolic shutter DN100 CI
Free Surface Tank
1000
1000
1000
Septum H=1,5m
Gate valve DN65
Gate valve DN65
Vert. Multis. Radial Impeller Pump DN65
Electromagnetic Flow Meter DN80
PVC PN16 DN110
PVC PN16 DN110
PVC PN16 DN75
PVC PN16 DN110
PVC PN16 DN90
PVC PN16 DN90
PVC PN16 DN110
Drain
Drain
1600
PVC PN16 DN110
PVC PN16 DN110
6000
PVC PN16 DN110
PVC PN16 DN110
PVC PN16 DN110
PVC PN16 DN110
Pressure Gauge
Spheroidal Cast-Iron Pipeline DN100 L=6000mm
Pressure Gauge
9m

Figure 1. Layout plan of the hydraulic circuit installed at the Hydraulics and Maritime Construction Laboratory of the Università Politecnica delle Marche in Ancona

Operating conditions are measured by an electromagnetic flow meter E-H Promag 10D80 in the range (0-10.5)l/s with an accuracy of +/-0.5% of the reading and by two diaphragm pressure gauges in the range (0-1)MPa with an accuracy of +/-0.3% of the reading. There are several screwed branches in the system for the insertion of measuring instruments or drain cocks.

The simulation of the leakage was realised by inserting a ball valve and a short DN20 drain pipe closed at the end by a removable cap on which a series of metal discs with holes of different size

BY NC SA

and geometric shape were inserted into the PVC section of the circuit. The discs, 2mm thick, were drilled with a laser cutting machine with an accuracy of ± 0.1mm. Circular and rectangular holes of different sizes were used in the tests performed, the overall picture of which is shown in Figure 2 and Table 1. The dimensions of the holes were defined considering the following elements: (i) the outgoing flow rate did not exceed 50% of that circulating in the system; (ii) the maximum size of the hole was compatible with the internal diameter of the drain plug. For rectangular holes, different combinations of dimensions were also considered, but with the same area.

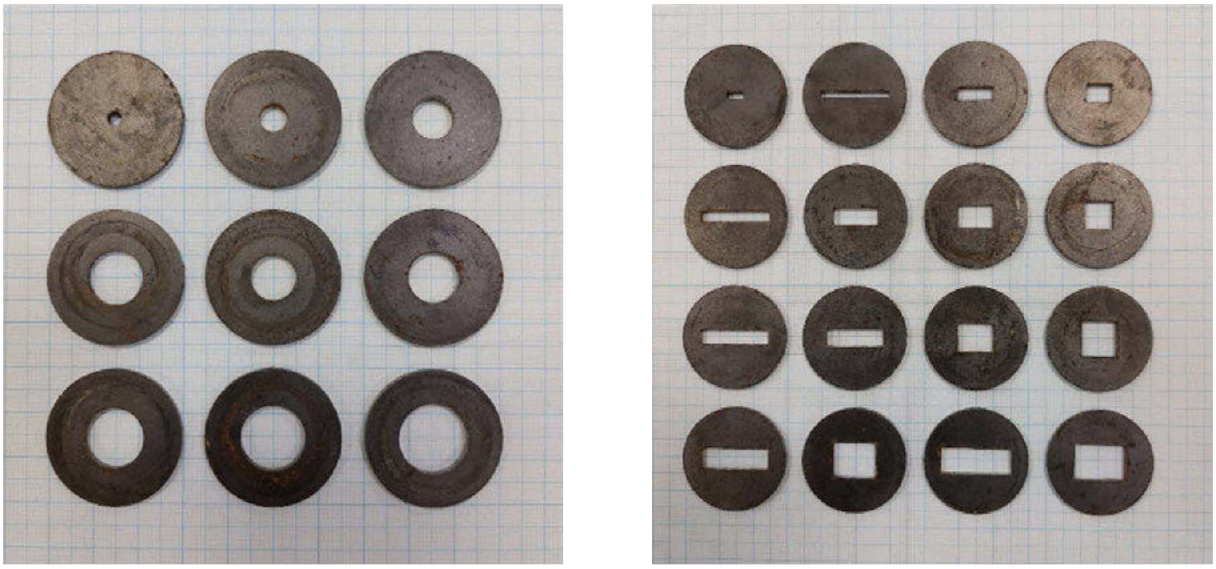


Figure 2. Nozzles used in the tests simulating leakages of different size and shape

Table 1. Characteristic dimensions of the holes simulating the leakage subject of the experimental analysis. The subscripts "f" and "i" refer to the hole and the pipe of the system respectively

Circular holes			**Rectangular holes**									
D_f	A_f	A_f/A_i	B	L	A_f	A_f/A_i	L/B	B	L	A_f	A_f/A_i	L/B
[mm]	[mm^2]	[-]	[mm]	[mm]	[mm^2]	[-]	[-]	[mm]	[mm]	[mm^2]	[-]	[-]
3	7.069	0.00096	2.00	5.00	10.00	0.00136	2.50	4.05	20.00	81.00	0.01101	4.94
6	28.274	0.00384	1.80	20.00	36.00	0.00489	11.11	5.00	16.20	81.00	0.01101	3.24
9	63.617	0.00864	3.60	10.00	36.00	0.00489	2.78	8.10	10.00	81.00	0.01101	1.23
10	78.540	0.01067	5.00	7.20	36.00	0.00489	1.44	10.00	10.00	100.00	0.01359	1.00
11	95.033	0.01291	3.20	20.00	64.00	0.00870	6.25	6.05	20.00	121.00	0.01644	3.31
12	113.097	0.01537	5.00	12.80	64.00	0.00870	2.56	10.00	12.10	121.00	0.01644	1.21
13	132.732	0.01804	6.40	10.00	64.00	0.00870	1.56	7.20	20.00	144.00	0.01957	2.78
14	153.938	0.02092	8.00	8.00	64.00	0.00870	1.00	10.00	14.40	144.00	0.01957	1.44
15	176.715	0.02401										

Tests were performed for each hole by varying the frequency of the pump using the inverter (35Hz, 40Hz, 45Hz, 50Hz) and repeating each test twice on each disc, for a total of 8 tests per hole. The test consists of opening the tap for 60s and measuring the volume exiting the bore and collected in a tank below the system. Pressure and flow rate changes during the tests are made by keeping the valve opening constant when the flow is started and varying the pump speed.

2.2 Modified setup

The objective of the hydraulic circuit modification was to reproduce in the laboratory experimental conditions as close as possible to the pipe breaks under real conditions, i.e. drilling a circular hole in the experimental pipe (see Figure 3). Given this premise, the design of a new plant part with modifications of the existing system was considered.

To have a water flow from the cast-iron pipe hole without the insertion of metal plugs and taps, a bypass pipe must be added where the water circulates normally without affecting the test pipe. The water that normally flows in the PVC pipe, called the bypass pipe, is moved to the cast-iron pipe, where the break is located 80cm above the level of the existing system. This is achieved by the use of 3 motorized valves powered by an electrical panel, 2 PVC tees, 2 PVC 90° bends and the flanged inlet and flanged cup fittings. The latter two special elements are required for the connection between cast-iron pipe (test pipe) and PVC pipes; specifically, upstream a flanged spheroidal cast-iron inlet with a nominal diameter of DN110 and PN16 was used, while downstream a flanged cup with a DN110 and PN16 spheroidal cast-iron joint was used. A view of the system layout after these modifications is illustrated in Figure 3.

The tests were carried out with four frequency values: 35Hz, 40Hz, 45Hz, 50Hz. This was possible by acting on the three-phase inverter connected to the system's electric pump, modifying the number of revolutions per second of the motor. Three tests were carried out for each frequency value in order to assess the repeatability conditions of the measurement, thus making a total of 12 tests. The parameters that were varied by varying the number of revolutions of the pump were the pressure and the operating flow rate of the system. As the opening and closure stages of the motorized valves take place in 9s, the full opening of the valves corresponds to the start for the measurement. The time interval chosen for the measure is always 60s, in order to have a sufficiently stable condition of the measurement unaffected by the transient phase of the start-up of the flow in the pipeline. Thus, the total duration of the test by considering the valve operating times of both opening and closure is 78s.

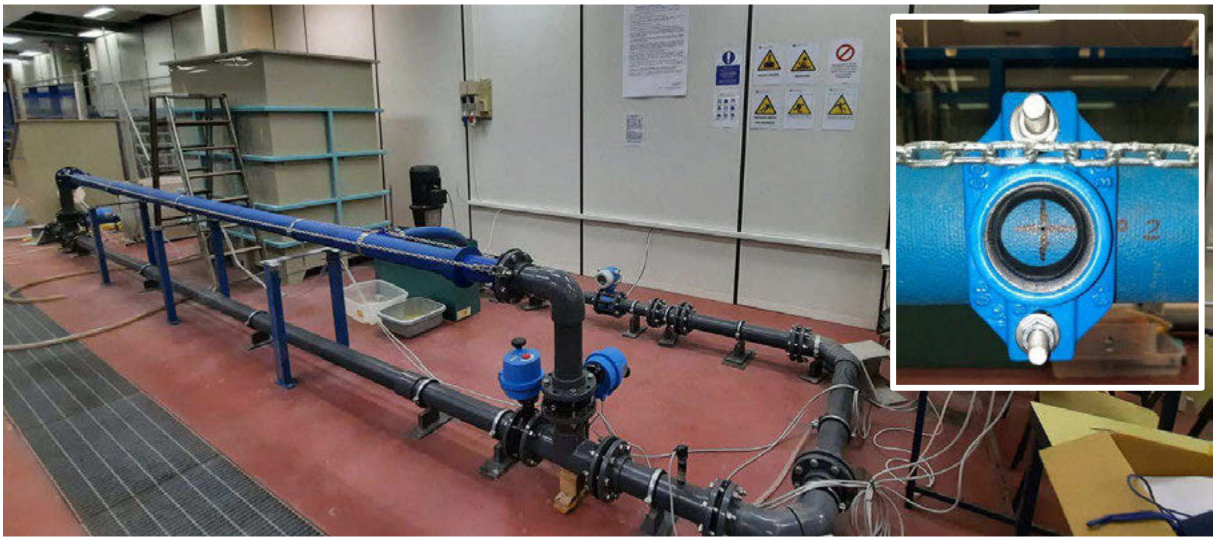


Figure 3. View of the hydraulic system after modifications; particular of the cast-iron pipe with hole used in the experiments (box on the right)

3 RESULTS

3.1 Original setup

For each test, the pressure and flow rate in the pipeline were measured, as well as the volume of water leaking out of the hole; the leakage, expressed in terms of flow rate, was obtained by dividing the volume collected by the duration of the test. The results obtained were processed to establish primarily the relationship existing between the operating pressure in the pipeline and the leakage. Secondly, an attempt was made to analyse the role played by the size and shape of the hole on the leakage itself.

The main results of the experimental tests are shown in Figure 4, where the measurement in l/s of the flow rate exiting the hole as the pressure in the pipeline varies, expressed in metres of equivalent water column, is represented for both rectangular holes (left panel) and circular holes (right panel). It can be observed that the tendency for the leakage to increase as the pressure in the duct increases is confirmed, as evidenced both under experimental (e.g. [2]) and real-life conditions. From a quantitative point of view, the operating pressures vary in the range (150÷415)kPa to which corresponds an output flow rate in the range (0.1÷2.4)l/s equivalent to a loss between 1.5% and 23% of the system flow rate.

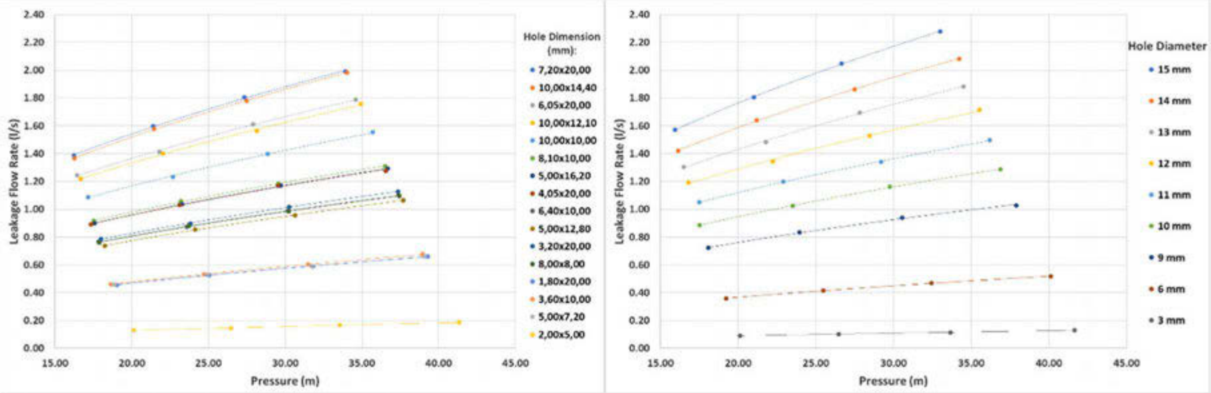


Figure 4. Variation of loss as a function of pressure for rectangular (left) and circular (right) holes.

Comparing the results obtained for rectangular and circular holes, there are no significant differences in behaviour for the two types. The greatest values in terms of measured leakage for circular holes are due to the maximum values of the circular section areas being larger than the rectangular ones. In both cases, it can be seen that the outlet flow rate has a steeper rise than the pressure as the hole area increases. Furthermore, increasing the frequency of the pump motor results in a corresponding increase in pressure, which is lower the larger the area of the hole.

As far as the influence of the hole shape on leakage is concerned, the test results do not show unambiguous behaviour. In fact, if for holes with an area greater than 100mm^2 it is observed that the leakage is greater for the holes with the largest L/B ratio, i.e. those with a more elongated shape, the same thing does not occur for holes with a smaller area where the greatest flow rate at the same area is for L/B values smaller than the maximum value.

Applying the model represented by equation (1) to the results of the experimental tests, we obtain the results shown in Figure 5 in which the average value of the coefficient a (in red) and the exponent b (in blue) are represented for each of the tests on a given hole and with an assigned frequency. The coefficient a increases as the ratio A_f/A_i increases, where A_f is the area of the hole and A_i is the internal area of the pipe, with a monomial law that deviates slightly from the linear trend line. The behaviour is essentially similar for rectangular and circular holes, the exponent of the function being 0.94 and 0.92 respectively, with a tendency for the interpolating function to overestimate the value of a as A_f/A_i increases. As regards the coefficient b, the experimental results show that the value is essentially constant for all the tests performed, with an average value of about 0.5. In this case, there is no significant difference for the two types of holes, the difference between the two coefficients being 1‰ and the standard deviation of the values of all the tests, both rectangular and circular, being of the order of 1%.

The results obtained are in good agreement with the experimental study by De Paola and Giugni [2] regarding the tests on steel pipes, under the same conditions analysed (constant pressure tests). In particular, it is observed that: (i) the exponent b of (1) is substantially coincident in the two studies for both rectangular and circular holes, with slightly higher average values for the tests on steel pipeline; (ii) the coefficient a of (1) is in agreement for values of the A_f/A_i ratio < 0.01 while for larger values the experimentation in metal pipeline presents larger values of a for both types of holes.

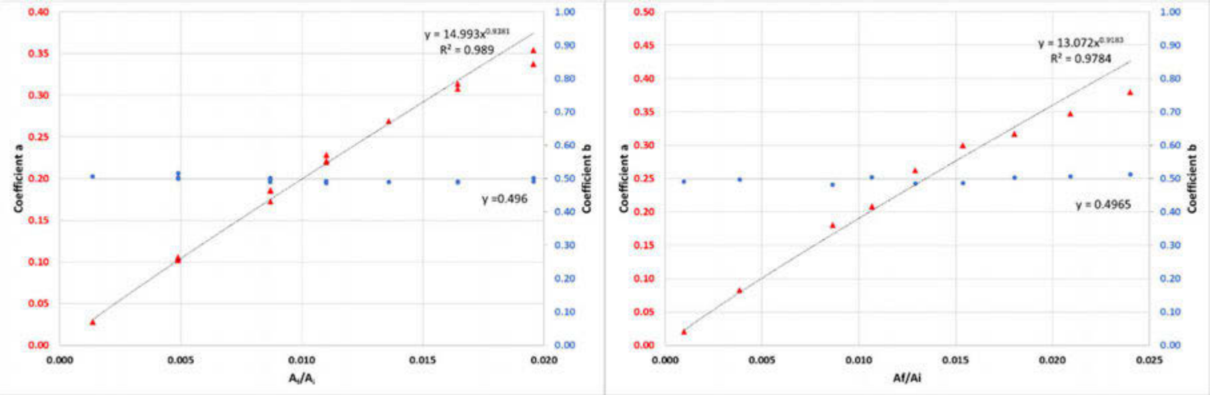


Figure 5. Values of coefficients a (in red) and b (in blue) for rectangular (left) and circular (right) holes

More tests with rectangular holes will be executed changing both the area and the shape of the hole in order to analyse the influence of some parameters as the A_f/A_i ratio or the L/B ratio on the relationship between the pressure and the leakage flow rate.

3.2 Modified setup

The preliminary results were obtained by measuring the leakages from a drilled circular hole of 2mm. They were elaborated to establish the relationship existing between the operating pressure in the pipeline and the leakage (1).

The relationship between the leakage and the average pressure is shown in Figure 6. It can be observed that the tendency for the leakage to increase as the pressure in the pipeline increases is confirmed. Furthermore, it can be seen that the increase of the frequency of the pump motor leads to a corresponding increase in the pipe pressure with an increase in the flow rate from the hole.

In order to compare the results of these preliminary tests with those obtained in the original setup for circular holes of smaller diameter, i.e. 3mm, the coefficients a and b were computed by equation (1) and by the law of the curve interpolating the experimental values of Figure 3b:

$$a = 13.072\left(A_f/A_i\right)^{0.9185} \tag{2}$$

By using the equation (1) the coefficient a obtained in the new test is 0.052 and it is larger than the value of 0.0205 obtained with the 3mm hole test of original setup; while the coefficient b of the new test is 0.422 and it is smaller than the value 0.4916 obtained in the previous tests.

By the equation (2), the coefficient a for the original setup with A_f/A_i=0.00096 is 0.023 and for the modified setup with a ratio A_f/A_i=0.0004 is 0.0099. Instead, the coefficient b is constant and equal to 0.4965.

Therefore, the value of coefficient b is smaller than results obtained in the previous experimental investigation, while the coefficient a is larger than the value obtained in a DN110 and PN16 PVC pipe with a hole with a larger diameter of 3 mm. Thus, the a-values obtained for a 2mm diameter hole deviate from the trend for larger diameter holes, where a decreases as the diameter decreases.

BY NC SA

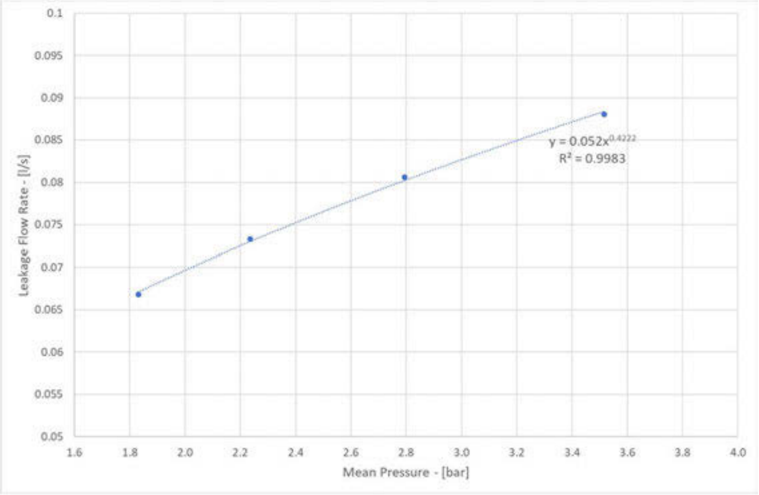


Figure 6. Average leakage as a function of the average pressure in the pipeline in the minute before the test. Equation (1) with the coefficients a and b and the associated error R^2 *is shown*

This difference could be justified by the fact that the leakage in the original setup referred to a PVC pipe, whereas in the present case the test is made on a cast-iron pipe. Moreover, in the original setup the leak was simulated by installing a junction with an interception spherical valve at the end of which there was a nozzle with the hole, while in this case the water can flow out of the hole with no constraints.

Further tests will be developed with different hole diameters and the results will be compared with the experimental values previously obtained for the PVC pipe and with tests available in the literature.

4 CONCLUSIONS

At the Hydraulics and Maritime Construction Laboratory of the Università Politecnica delle Marche, an experimental investigation was carried out in which a water leak in a pressurised system was simulated by varying the operating conditions of the system, flow rate and pressure, and the shape and size of the holes.

The first series of experiments shows that the leakage from the hole increases as the pressure in the pipeline increases. These results are in agreement with similar studies described in the literature. Comparing the results between similar experiments (e.g. [2]) it can be noted that: (i) the leakage in the metallic pipe is greater than that in the PVC one; (ii) differences in equations (1) are negligible for smaller size holes. Moreover, the preliminary results of the analysis of the influence of the hole shape on the leakage show a different behavior between holes with area greater than 100mm^2 and those with a smaller one, being the leakage greater for the holes with the largest L/B ratio in the first case and the opposite for those with smaller area. The value obtained for *b* coefficient is constant and it is very close to value of 0.5, typical of the Torricelli's law (e.g. [6]).

The experiments applied to the cast-iron pipe with a drilled hole confirmed the tendency for the leakage to increase as the pressure in the pipeline increases. The coefficients of equation (1) calculated for this set of tests are greater for the *a* coefficient and smaller for the *b* one with respect to those of the PVC pipe respectively. This preliminary result could confirm the hypothesis of De Paola and Giugni [2] that the leakage is greater in the metallic pipe.

A new set of tests will be executed to support and to generalize the model of the leakage flow rate prediction as a function of the pressure.

5 REFERENCES

[1] G. Darvini, V. Ruzza, and P. Salandin, “Performance Assessment of Water Distribution Systems Subject to Leakage and Temporal Variability of Water Demand”, Journal of Water Resources Planning and Management, 146(1), 2020.

[2] F. De Paola and M. Giugni, “Leakages and pressure relation: an experimental research”, Drink. Water Eng. Sci., 5, 2012, pp. 59-65.

[3] M. De Marchis, C.M. Fontanazza, G. Freni, V. Notaro and V. Puleo, “Experimental Evidence of Leaks in Elastic Pipes”, Water Resources Management, 30, 2016, pp. 2005-2019.

[4] A.M. Cassa, J.E. Van Zyl and R.F. Laubscher, “A numerical investigation into the effect of pressure on holes and cracks in water supply pipes”, Urban Water Journal, 7, 2010, pp. 109-120.

[5] A.M. Cassa and J.E. Van Zyl, “Predicting the leakage exponents of elastically deforming cracks in pipes”, Procedia Engineering, 70, 2014, pp. 302-310.

[6] M. De Marchis and B. Milici, “Leakage estimation in water distribution network: effect of the shape and size cracks”, Water Resources Management, 33, 2019, pp. 1167–1183.

[7] M. Franchini and L. Lanza, “Leakages in pipes: generalizing Torricelli's equation to deal with different elastic materials, diameters and orifice shape and dimensions”, Urban Water Journal, 11(8), 2014, pp. 678-695.

[8] J.E. Van Zyl, A.O. Lambert and R. Collins, “Realistic modeling of leakage and intrusion flows through leak openings in pipes”, Journal of Hydraulic Engineering, 143(9), 2017, pp. 1-7.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14869

A CRITICAL REVIEW OF LEAKAGE DETECTION STRATEGIES INCLUDING PRESSURE AND WATER QUALITY SENSOR PLACEMENT IN WATER DISTRIBUTION SYSTEMS – SOLE AND INTEGRATED APPROACHES FOR LEAKAGE AND CONTAMINATION INTRUSION

Shweta Rathi[1]

[1]Associate Professor, Medi-Caps University, AB Rd, Pigdamber, Rau, Indore, Madhya Pradesh 453331 (India)

[1] *shwetanabira12345@rediffmail.com*

Abstract

Water leakages in water distribution networks not only affects on natural water resources but also cause problems to nearby infrastructure or environment and makes water distribution networks more prone to contamination. Complexity in water distribution systems makes leakage detection and its monitoring a difficult task. Leakages in water distribution systems are caused due to damages in pipes, lack of maintenance of pressure due to uncertain demand and various operating conditions. Therefore, to manage the pressure in the water distribution systems it is necessary to identify appropriate leak location. Various studies on pressure management focused on optimal sensor placement for leak localization considering various constraints cost of sensors, demand uncertainty, damages and burst in pipes and few focused on valve location and optimization for control of pressure. The main aim of the paper is mainly to provide a comprehensive review of various studies related to leak detection, location and monitoring strategies. Also, it mainly focuses on pressure and water quality sensor placement in water distribution systems with sole and integrated approaches for leakage minimization and contamination intrusion in water distribution systems. A critical review of available methodologies for leakage detection, location, monitoring including existing simulation tools, solution approaches, and available methodologies for pressure sensor placement, technical challenges, and future research direction is presented. Till now no reviews are presented for pressure and water quality sensor placement in water distribution systems along with burst detection.

Keywords
Pressure sensor, Review, Water Distribution System, Water Quality Sensor, and Water Security.

1 INTRODUCTION

The primary aim of water distribution system is to supply adequate and safe water, efficiently and effectively at an affordable cost to all the consumers. However, due to growing challenges such as expanding populations and urbanization, climate and demographic change, infrastructure deficits and aging, inadequate maintenance of water systems, water scarcity, adverse health effects and rehabilitation costs etc. which effects on water quantity and quality in water supply systems. Further, available water sources throughout the world are becoming depleted due to industrialization and this problem is increased with rate of population are increasing specially in developing countries like India. Current issues and challenges of urban water supply systems are highlighted in Table 1. A water distribution network (WDN) is an important and major component of urban water supply systems. Water loss due to leak in pipelines is one of the major challenges in efficient

Table 1. Issues and Challenges of urban water supply including water distribution systems in developing countries like India

Issues or Challenges	Descriptions
Water Quantity	- Inconsistent and intermittent supply - High NRW/UFW - Low coverage of metering - Irrational Tariff - Low billing and collection efficiency - Demand supply gap - Water stress
Water Quality	- Lack of safe drinking water - Pollution and Contamination of drinking water - Accidental and intentional contamination - Lack of proper water treatment and distribution facilities
Technology and Management Issues	- Redundancy and Robustness in the system - Lack of technical innovations - System operation Intelligence - Adoption of innovative design and equipment - Inadequate operation, maintenance, management and monitoring of water distribution systems - Operational optimization of water distribution systems through optimal pump and valve operation – Energy efficiency in the network (Miller et al. 2013, Wakeel et. al. 2016) - Prolonger service of infrastructure like wells, pipes, tanks, pump and fitting
Structural, Institutional, Governance and Social Issues	- Data availability, reliability and transparency in the systems - Lack of existence of water policies, regulations and strategies - Shortage of skilled man power - Disconnected Agencies / lack of Coordination - Inadequate funding - Lack of monitoring - Low capacity of operational & maintenance - Lack of water conservation measures, policies and by-laws - Water efficiency devices /fixtures - Social Awareness and lack of sustainable practises

water distribution networks. Looking to the major issue of water resources availability, scarcity of water resources, high Non-Revenue Water (NRW) which is the difference between system input volume and billed authorized consumptions, may be as high as 40 to 45% in some of the cities of developing countries, of which water loss is one of its most significant parts [1, 2]. Therefore it is utmost important to manage the available resources in an efficient way as well as to minimize the leakage losses in the water distribution systems. Leakages in WDNs are due to increase in pressure and lack of operation and maintenance. Reducing leakages providing benefits in many ways as reduced operation costs and increased revenues in view of expanded existing water supply as compared to new facility of supply, better water resources efficiency, reduces pipe rehabilitation cost, minimizing infrastructure damage, preventing damage to surrounding environment and preventing adverse affect on the human health [3, 4] and ultimately helps in reducing overall operation, maintenance, management and monitoring cost.

Water losses, as part of NRW, are classified into real and apparent losses. Apparent losses are to be reported in between 1% and 9% of the total water supplied in different countries around the world. Real losses are further expressed as Infrastructure leakage index (ILI) which is the ratio between current annual real losses (CARL) and unavoidable annual real losses (UARL) [1].

The amount of leakage depends upon several factors as pipe material and age, pipe depth, pipe corrosion, pressure surges, improper handling and storage during construction, poor workmanship, operating pressure etc. In general three leakage indicators are noted as an increase of head loss and decrease of peak factor and reservoir volume [5].

Recently numbers of reviews have been provided by various studies: Puust et al. [6] provided a comprehensive of leakage management methods consisting of leakage assessment, leakage detection and leakage control methods which has been further extended by considering both real and apparent losses [7]. Other reviews assessment of water losses (Gupta and Kulat [8], Al-washali [9]), leakage detection (Li et al. [10], Datta and Sarkar[11]), condition assessment of water pipes [12]. Other focuses on transient-based methods (Colombo et al. [13]), various pipeline fault detection methods (Datta & Sarkar[11]), pressure-based methods (Abdulshaheed et al. [14]), externally and internally-based methods (Adedeji et al. [15]), data-driven methods (Wu & Liu [16]), and current and proposed intelligent methods (Chan et al. [17]), Critical Review of Steady-state leak detection and localization methods (Zaman et al. [18]), Model-based and data-driven approaches (Hu et al. [19]).

The main aim of the paper is mainly to provide a comprehensive review of various studies related to leak detection, location and monitoring strategies. It also focuses on pressure and water quality sensor placement in water distribution systems with sole and integrated approaches for leakage minimization and contamination intrusion in water distribution systems. A critical review of available methodologies for leakage detection, location, monitoring including existing simulation tools, solution approaches, and available methodologies for pressure sensor placement, technical challenges, and future research direction is presented. Till now no reviews are presented for pressure and water quality sensor placement in water distribution systems along with burst detection.

This classified is done based on Mutikanga et al. [7]. 1) Leakage assessment i.e quantifying the amount of water lost, 2) Methodology based on leakage detection and location by survey/observation or by performing signal analysis i.e. detection of leakage hotspots and leakage control methods and 3) Methodology based on leakage detection and location using modelling, calibration and optimization, and 4) Leakage monitoring and controlling methods which includes those methods which measures flows into discrete zones or DMAs, and to continuously and regularly monitor and control the flow into the DMAs, and manage the pressure in the DMAs and Night flow analysis 5) Online monitoring, leak locations and detections using pressure sensors and burst identifications. It covers methods based on water audit, District Metered Areas (DMAs), Minimum Night Flow (MNF) analysis, leakage hydraulic analysis, flow statistical analysis, acoustics, transients, pressure management, network asset management, optimization techniques, online monitoring and detection. Literature is further updated till 2022. It also covers pressure and water quality sensor placement in water distribution systems sole and integrated approaches for leakage and contamination intrusion and burst detection. This review is in progress.

2 OVERVIEW OF LEAK DETECTION METHODS

2.1 Methodology based on leakage assessment

A number of studies have been conducted previously for water audit for analysing water loss components i.e real losses in distribution systems [20]. However, this method does not focus on the apparent losses which are mainly due to meter inaccuracies, data handling errors and unauthorised consumptions. Various other techniques which are used for estimation of leakage are minimum night flow analysis (MNF analysis) and Burst and Background Estimates (BABE) [21]. BABE is a component-based approach for estimating annual losses [22]. It was observed that majority of losses are reported due to pipe-joint failure, relatively older age pipes, poor repairing and maintenance of water taps, pipe joints and shower taps, negligence of the consumer and unreliable water supply [21].

MNF analysis requires zoning and intensive field work and trained manpower. BABE is applicable only in case of regular active leakage control (ALC). The details of these methods are presented in [23]. Apart from these water and waste water balance method is also suggested by [24]. Considering the uncertainty in water distribution systems such as water demand and fluctuations, it was observed that the average of the two methods would be preferred for estimating water loss components in intermittent water supply systems. MNF methods are more suitable for the DMA scale rather than system-wide scale. Assessing the WL components by using at least two methods should improve the prioritisation, economic modelling, monitoring.

2.2 Methodology based on leakage detection and location by survey/observation/field studies or experimental studies by performing signal analysis

A leakage detection and location method includes finding location of leaks using acoustic equipments like listening devices, leak noise loggers, infrared thermography, smart-balls etc. in static and dynamic environment or using survey work or by observing pressure and flow data or by analyzing leakage-pressure relationship. Static leak detection provides early detection with minimal human interference and a dynamic leak detection system provides localization and pinpointing. Various leak detection technologies and their advantages and disadvantages are highlighted by [25]. Also, video cameras, microphones, acoustic sensors, are also helpful in finding the leaks in large-diameter pipes [26 -28]. Al-Ghamdi [29] has performed the field survey to identify the relationship between leakage rate and pressure in the selected areas of city of Makkhan, Saudi Arabia. The value of leakage exponent is identified based on network age and pipe materials.

2.3 Methodology based on leakage detection and location by theoretical/numerical method using modelling, calibration and optimization

It includes the methods based on leakage-pressure relationship to analyse the impact of pressure on leakage, modelling and optimization techniques, calibration of leakage model coefficients, infrastructure asset management, pressure management etc. considering benchmark problems and real life networks.

2.3.1 Leakage-Pressure Relationship to analyse the impact of pressure on leakage

Almandoz et al. [30] used a modelling approach for evaluation of water losses based on the physical losses in mains and service connections and the volume of water consumed but not measured by meters.

Numerous studies have worked on leakage-pressure relationship. Van Zyl [31] focused on four factors which influence leakage-pressure relationship: Leak hydraulics, pipe material behaviour and soil hydraulics and water demand. Maskit and Maskit and Ostfeld [32] proposed the methodology for calibrating leakage model coefficients for the group of pipes based on pipe age and material. The problem is solved using genetic algorithm (GA). Zyl [33] observed 4 factors for high sensitivity of leakage to pressure as pipe material behaviour, leak hydraulics, soil hydraulics and water demand along with the way of individual leak combines in a pressure management zone. Kiziloz [34] development of leakage rate prediction model using artificial neural networks (ANN) where network pressure and age is considered as a reference. Nourhan Samir et al. [35] modelled leakage as a function of pipe length, calibrating leakage coefficient to develop pressure fluctuation using PRVs. Method is applied to DMA wise.

2.3.2 Methodology based on pressure and asset management

Previous literatures have classified the leakages into three categories in WDSs as reported, unreported and background leakages. Reported leakages are emerging and visible leakages. Unreported leakages are non-surface leakages that are detectable by acoustic devices and background leakages are considered as non-surface leakages which are unreported and acoustically undetectable. Pressure control and Asset management is an effective way for minimizing leaks in WDSs, particularly helps in reducing background leakages. Several studies used for pressure management are fixed and variable area discharges [36-37]. Pressure in WDN can be controlled using varying types of control system such as fixed outlet pressure control, time-modulated pressure control Awad et al. [38] and flow modulated pressure control. Leakages could further be optimized using number of valves and their locations. Araujo et al. [39] optimised the number of valves and its location as well as valves opening adjustments for simulation in an extended period to minimise pressures and consequently leakage levels. Adedejia et al. [37] provided a critical review on pressure management strategies.

2.4 Methodology based on leakage monitoring and controlling

Leakage monitoring and controlling methods includes those methods which measures flows into discrete zones or DMAs and to continuously or regularly monitor and control the flow into the DMAs and manage the pressure using PRV at the inlet of the DMAs or by managing and analyzing the night flow and based on the excess flow determine leakages in the system. These methods are useful for prioritizing high risk zones and not identifying the location of leak.

In view of reducing NRW and reduced operation costs and increasing infrastructure life and reducing pipe bursts, water utility managers often choose to reduce excessive operating pressure or to perform service pressure regulations in WDNs. For managing the pressure in WDSs, optimization of control valves with suitable locations and settings is implemented [40-43].

Several researchers focused on Leakage reduction through Pressure Management using optimal locations of PRV [44-45].

Considering the complexity and uncertainty of water distribution systems makes it difficult for operation, maintenance and monitoring of the systems. Thus, based on the concept of graph theory a partition of WDNs is performed based on structural and connectivity analysis of the network. This partition of the network is known as cluster for topological or connectivity analysis suggested by [46]. This clustering of water distribution facilitates for various purposes as sensor placement, detection of contamination source intrusion and calibration of the model and leakage detection or pressure management studies etc. Further, clustering is one of the ways for DMA formation. Perelman et al. [47] minimizes objective function the number of open boundary valve in the pipe. In addition, pressure management is aided by installing pressure reducing valves (PRVs) at the inlet of each DMAs. The control of pressures in each DMA leads to a reduction in leakage through pipe joints and connections.

Recently several researchers work on the concept of DMAs, consisting of dividing the WDNs into small area for reducing management complexity of WDSs and provide ease in monitoring. Various studies and technical reports reported the guidelines on formation of DMAs [48-50]. Partitioning is carried out using graph theory [51-52] and DMA is established based on several factors as maximum and minimum number of consumer connections (i.e DMA size recommended for a district), the main transmission system should be kept separated from the DMAs in order to ensure a flexible and reliable water supply; the connectedness of each district to the water supply source and be independent, i.e without any connection with other DMAs; and other factors to take into account are pressure constraints at demand nodes, final leakage level target, implementation and maintenance costs [52]. Studies showed that DMAs are the cost-effective technology in case of water loss control and leakage management However, DMAs design required careful formation, in case of failure may reduce reliability, redundancy and water quality of the network. In DMAs, quantity of water leaving and entering the districts is metered. Further in case of large scale water quality contamination in WDSs, DMAs would limit the spread of contamination and minimize the extent of response actions required for the system [53]. Ulanicki et al. [54] installed PRVs at the inlet of DMAs which further helps in reduction of leakage reduction.

In past, a number of methods for formation of DMAs have been previously suggested Manual trial and error approaches (55) to automated tools integrating network analysis (56), graph theory consisting of clustering [57-58, 46]), complex networks (59-60), and heuristic methods [60-62]. The general procedure for DMA design is to identify water mains, partition the network into sub-networks, and isolate interconnecting lines using simulation-based heuristics to minimize the number of connections and dependencies between the sub-networks [46-47].

Haider, H. [63] developed a framework for intermittent water supply in order to conserve limited water resources in arid region of Saudi Arabia. Water losses are evaluated using active leakage control, passive leakage control, infrastructure asset management. Night flow analysis is performed to observe the relationship between pressure and water loss for pressure reduction.

Jadhao and Gupta [64] reduced excess pressure of DMAs during night flow. This work is further extended by Sharma et. al. [65] who considered segment identification using existing valves to identify DMAs in water network and boundary optimization is done using GA. Resilience Index is used as a performance evaluation criterion for optimal DMA configuration. Applications are shown on part of a real life network of Nagpur City.

However, few studies is carried out on topography-based partitioning method to develop DMAs in WDN. Liu et. al. [66] Compared three methods of partitioning fast greedy, random walk and Merits. Two cases are considered for complex WDN, weighted and non-weighted edges where weights are represented by demands and observed that greedy method is more effective in weighted graph. Recently, Abdulshaheed et al. [67] provided a critical review on pressure based method for monitoring leak both in steady and unsteady conditions. Gupta and Kulat [8] provided a selective review on leak management techniques in WDSs.

2.5 Methodology based on online monitoring, leak locations and detections using pressure sensors and water quality sensors and burst identifications

For better monitoring of pressure and early detection of contamination in WDSs, optimal pressure sensors and water quality sensors are placed in the systems simultaneously so as to provide effective, efficient and safe WDSs in terms of water quality and quality. Pressure sensors are helpful in order to capture abnormal pressure drop so as to locate the leak and burst in WDSs timely. Concerns about the cost of the sensors, both pressure and water quality sensors should be located at some crucial locations in the WDSs. If sufficient number of pressure sensors are placed at crucial locations in WDSs and are evenly installed, continuous pressure trend change would be helpful in location burst. A number of researchers have optimised water quality sensors in WDSs for early detection of contamination (Ostfeld and Salomons [68], Ostfeld et. al. [69], Dorini et al. [70], Weickgenannt et al. [71]. Rathi et al. [72]. Details of the study are given in Table 2. Some studies have considered uncertainty in placement of pressure sensors for leak detection [82-83].

Table 2. Online monitoring, leak locations and detections using pressure sensors and water quality sensors

Reference	Focus on	Remark	Classification
Blesa, J. (2015) [73]	Robustness in the sensor placement methodology is incorporated where only inner pressure sensors in the DMAs are considered	Robust Sensor Placement Methodology	Leak locations using Pressure Sensors
Steffelbauer et al. (2016) [74]	Sensor placement for leak localization considering demand uncertainty which leads to uncertain pressures at measurement points	Uncertainty in demand and measurement locations	Leak locations using Pressure Sensors
Cheng, Li et al. (2017) [75]	Optimal placement of pressure sensors in WDSs based on pressure sensitive matrix	• Sensor placement is optimized based on clustering analysis of pressure sensitive matrix • K-mean algorithm is used to solve optimization problems • Cluster formation to reduce the size and Complexity of the network	Leak locations using Pressure Sensors
Cao et al. (2019) [76]	Simultaneous sensor placement and pressure reduction of WDS	Both pressure sensors and PRVs optimization	Leak locations using Pressure Sensors and pressure reduction using PRV localization
Soroush, F. (2019) [77]	Optimization of no. and location of pressure sensors in WDSs using geospatial tool coupled with GA	Geo-statistical tool is used for pressure monitoring network	Geo-statistical tool coupled with an optimization algorithm
Shao et al. (2019) [78]	Leakage detection based on time series monitoring data using pressure sensor	Correlation coefficient based on time series data is considered for leak detection	Leak detection using pressure sensors
Cheng, W. et al.	Optimizing pressure sensor placement to monitor pipes	Objective is to maximize the	Monitoring of water leakage

BY NC SA

(2021) [79]	in which burst event takes place	monitoring network capability to detect leakages with pre-minimum no. of sensors	and quantification of pipe burst
Santos-Ruiz, I. et al. (2022) [80]	Optimal sensor placement for locating leaks in WDNs using information theory.	Multiple leak scenarios are generated and datasets of pressure changed are performed. Heuristic method is applied for ranking of nodes for sensor placement.	Leak locations using Pressure Sensors
Taylor et al. (2018) [81]	Equivalent orifice area to quantify pipe quality	Leakage and intrusion in intermittent water supply	Leakage and contaminant intrusion

3 SOME OBSERVATIONS AND ISSUES NEED CONSENSUS FOR FUTURE RESEARCH

- Original or Reduced Network: A real life network may involve thousands of pipe and nodes, and numbers of sensors are restricted. Considering the possibility of each and every node of original network as possible location of pressure and water quality sensors may increase computational burden, which can be significantly reduced by suitably eliminating some of the nodes from list of candidate nodes such that sensor placement accuracy is not affected. Any skeletonization method can be used.
- Dynamic water quality simulation and pressure driven approach is lacking. A cluster analysis is performed for easy applications of large and complex network problems. However, which method of cluster analysis used is a question.
- Type of leakage assessment methods: As large number of leakage assessment methods have developed in past however, quantifying the leakage assessment using any one method is inaccurate in case of uncertainties. However, the average of the two methods would be preferred for estimating water loss components in intermittent water supply systems.
- Even though hydraulic model is a valuable tool for leakage hydraulic analysis, however, few studies have used calibrated hydraulic model for DMA selection. A calibrated hydraulic model helps in the selection of the DMAs. Further, after performing DMAs, few studies have recalibrated DMA EPANET model with real data. If recalibrated model could be used it will be minimize the errors between the real and simulated network. It would be helpful for pressure management or helps in identifying crucial locations of pressure and water quality sensors in WDN.
- Installation of smart pressure gauges and automated pressure control valves: Smart pressure gauges installed at average zone points for continuous monitoring of operating pressure in pipe networks. Providing water at high pressure during night period increased water losses. Using automated pressure control valves can reduce the night pressure to some extents. It is recommended to reduce excessive operating pressure to more practical values.
- Past study showed that majority of the percentage of leakages are found to be at property connections or either in service lines or at the junction of the service lines and the property connections. Leakage Identification and repair of service connections and

storage tank in the households can reduce the leakages to some extents. Further, house connections made up of galvanized iron (GI) needs to be replaced with polyvinyl chloride (PVC) or polyethylene (PE) pipes.

- Type of solution methodology: Practical application of leakage detection and identification of appropriate location for large-scale water distribution networks is still a major challenge. Even though large no. of methodologies is developed however, few comparative works exist on leakage detection method to choose the best algorithm for generic case study.
- Even though large no. of solution methodologies are developed and applied for optimal valve location for pressure management in WDSs. However, few comparisons exist and best method in any general case study is lacking.
- As of now large number of partitioning method is suggested. However, the best method for general case is needed. Topography-based partitioning graph theory method would be more useful for establishing DMAs.
- Lacking of work in the area of leakage and contaminant intrusion simultaneously. Lack of relationship between contaminant volume intruded and system pressure, supply duration in intermittent as well as continuous supply since supply duration is important in case of flushing water quality.

4 SUMMARY AND CONCLUSIONS

The issue of water losses and reduction in efficiency of water distribution systems and security concern of water system has motivated several researchers to develop a methodology for optimal placement of both pressure and water quality sensors in WDNs for online monitoring and management of pressure and leakages which would be helpful in resolving the issues of water quantity and quality in WDSs and their adverse impacts. A critical review of available methodologies for leak detection strategies is presented in this paper with a view to raise issues requiring consensus amongst researchers. Several methods for leak localization, detection, monitoring and controlled is suggested in past. Many studies have focused on pressure and asset management. For identifying accurate leak detection, calibrated model is to be considered. Numerous studies have suggested for formation of DMA designs. However, topography-based partitioning methods for establishing DMAs are lacking. Current research work is moving towards better monitoring of pressure and early detection of contamination in WDSs. Therefore, optimal pressure sensors and water quality sensors are placed in the systems simultaneously so as to provide effective, efficient and safe WDSs in terms of water quantity and quality. Pressure sensors are helpful in order to capture abnormal pressure drop so as to locate the leak and burst in WDSs timely. Concerns about the cost of the sensors, both pressure and water quality sensors should be located at some crucial locations in the WDSs. If sufficient number of pressure sensors are placed at crucial locations in WDSs and are evenly installed, continuous pressure trend change would be helpful in location burst. The research work pertaining to these issues are highlighted for developing consensus amongst researchers for future research work on detection problems. This review is further in progress for better classification of methodologies.

5 REFERENCES

[1] A. O. Lambert, “International report: Water losses management and techniques,” *Water Sci. Technol. Water Supply*, vol. 2, 2022, 1-20.

[2] H. Cao, S. Hopfgarten, A. Ostfeld, E. Salomons and P. Li, “Simultaneous sensor placement and pressure reducing valve localization for pressure control of water distribution systems,” *Water*, 11, May 2019, 1352, doi:10.3390/w11071352.

[3] D. H. Kouchi, K. Esmaili, A. Faridhosseini, S. H. Sanaeinejad, D. Khalili and K. C. Abbaspour, "Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms," *Water*, vol. 9, no. 6, May 2017, 384; doi: 10.3390/w9060384.

[4] Q. Xu, R. Liu, Q. Chen, R. Li, "Review on water leakage control in distribution networks and the associated environmental benefits," *Journal of Environmental Sciences,* vol. 26, 2014, 955-961.

[5] G. Samudro and S. Mangkoedihardjo, "Indicators for leakage in water distribution systems," *Journal of mathematics and technology*, vol. 5, December 2010, 1-5.

[6] R. Puust, Z. Kapelan, D. A. Savic, T. Koppel, "A review of methods for leakage management in pipe networks." *Urban Water Journal*, vol. 7, no. 1, February 2010, 25–45.

[7] H. E. Mutikanga, S. K. Sharma, and K. Vairavamoorthy, "*Methods and Tools for Managing Losses in Water Distribution System*," *J. Water Resour. Plann. Manage*., Vol. 139, no. 2, March 2013, 166-174.

[8] A. Gupta, and K. D. Kulat, "A Selective Literature Review on Leak Management Techniques for Water Distribution System," *Water Resour Manage.,* Vol. 32, 2018, 3247–3269.

[9] T. AL-Washali, S. Sharma, R. Lupoja, F. AL-Nozaily , M. Haidera , M. Kennedy, "Assessment of water losses in distribution networks: Methods, applications, uncertainties, and implications in intermittent supply," *Resources*, *Conservation & Recycling,* vol. 152, 2020,104515.

[10] R. Li, H. Huang, K. Xin, & T. Tao, "A review of methods for burst/leakage detection and location in water distribution systems," *Water Science and Technology:Water Supply*, vol. 15, no. 3, June 2015, 429–441.

[11] S. Datta, & S. Sarkar, "A review on different pipeline fault detection methods," *Journal of Loss Prevention in the Process Industries*, vol. 41, March 2016, 97–106.

[12] A. Hawari, F. Alkadour, M. Elmasry, & T. Zayed, "A state of the art review on condition assessment models developed for sewer pipelines," *Engineering Applications of Artificial Intelligence,* vol. 93, August 2020, 103721.

[13] A. F. Colombo, P. Lee, & B. W. Karney, "A selective literature review of transient-based leak detection methods," *Journal of Hydro-Environment Research*, vol. 2, no. 4, April 2009, 212–227.

[14] A. Abdulshaheed , F. Mustapha, A. Ghavamian, "A pressure-based method for monitoring leaks in a pipe distribution system: A Review," *Renewable and Sustainable Energy Reviews,* vol. 69, 2017, 902–911.

[15] K. B. Adedeji, Y. Hamam, B. T. Abe, and A. M. Abu-Mahfouz, "Towards Achieving a Reliable Leakage Detection and Localization Algorithm for Application in Water Piping Networks: An Overview," Digital Object Identifier, IEEE Access, vol. 5, no. 9, September 2017, 20272-20285, 10.1109/ACCESS.2017.2752802. IEEE Access 5(9):20272 – 20285.

[16] Y. Wu, & S. Liu, "A review of data-driven approaches for burst detection in water distribution systems," *Urban Water Journal,* vol. 14, no. 9, February 2017, 972–983.

[17] T. K. Chan, C. S. Chin, & X, Zhong, "Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection," *IEEE Access*, 6, December 2018, 78846–78867.

[18] D. Zaman, M. K. Tiwari, A. K. Gupta and D. Sen, "A Review of Leakage Detection Strategies for Pressurised Pipeline in Steady State," *Journal of engineering failure analysis*, vol. 109, November 2019, 104264, 1-42.

[19] Z. Hu, B. Chen, W. Chen, D. Tan and D. Shen, "Review of model-based and data-driven approaches for leak detection and location in water distribution systems," *water supply*, vol. 21, no. 7, 2021, 3282-3306.

[20] A. Lambert, and W. Hirne, "Losses from water supply systems: Standard terminology and recommended performance measures (IWA's Blue Pages)," International Water Association, London, 2000.

[21] S. K. Bhagat, T. W. Welde, O. Tesfaye, T. M. Tung, N. Al-Ansari, "Evaluating physical and fiscal water leakage in water distribution system," *Water*, vol. 11, issue, 10, 2019. DOI 10.3390/w11102091.

[22] A. Lamber, "Accounting for losses: The bursts and background concept," *Water Environ. J*ournal, vol. 8, no. 2, 1994, 205–214.

[23] T. AL-Washali, S. Sharma, M. Kennedy, "Methods of assessment of water losses in water supply systems: a review," *Water Resour. Manage*. vol. 30, no. 14, 2016, 4985–5001.

[24] T. AL-Washali, , S. Sharma, R. Lupojad, F. AL-Nozailyc , M. Haiderac , M. Kennedya, "Assessment of water losses in distribution networks: Methods, applications, uncertainties, and implications in intermittent supply," *Resources, Conservation & Recycling,* vol. 152, 2020, 104515.

[25] S. El-Zahab and T. Zayed, "Leak detection in water distribution networks: an introductory overview," *Smart Water*, 2019, 1-23.

[26] A. N. C. Ong, and M. E. H. Rodil, “Trunk mains leak detection in Manila’s West Zone.” Proc., 7th IWA Water Loss Reduction Specialist Conf., IWA, Hague, The Netherlands, 2012.

[27] C. Stringer, R. Payton, M. Larsen, K. Laven, and D. Roy, D, “Integrated leak detection at Dallas Water utilities.” In Proc. Pipeline: Advances on Experiences with Trenchless Pipeline Projects, ASCE, Reston, VA, 2007, 1–6.

[28] Z. Y. Wu, P. Sage, and D. Turtle, “Pressure-dependent leak detection model and its application to a district water system,” *J. Water Resour. Plann. Manage.*, vol. 136, no. 1, 2010, 116–128.

[29] A. S. Al-Ghamdi, “Leakage-pressure relationship and leakage detection in intermittent water distribution systems,” *J. of water supply research and technology- AQUA*, vol. 3, 2011. 178-183.

[30] J. Almandoz, E. Cabrera, F. Arregui, E. Cabrera, & R. Cobacho, “Leakage assessment through water distribution network simulation,” *J. Water Resour. Plann. Manage*., vol. 131, no. 6, 458–466.

[31] J. E. van Zyl, and C. Clayton, “Water Distribution Systems: the Effect of Pressure on Leakage in Water Management,” January 2007.

[32] M. Maskit, A. Ostfeld, “Leakage Calibration of Water Distribution Networks,” 16th Conference on Water Distribution System Analysis, WDSA 2014, Procedia Engineering, 89, 2014, 664-671.

[33] J. E. van Zyl, “Theoretical Modeling of Pressure and Leakage in Water Distribution Systems,” 16th Conference on Water Distribution System Analysis, WDSA 2014, Procedia Engineering, 89, 2014,273 – 277.

[34] B.Kizilöz “Prediction model for the leakage rate in a water distribution system” 1-12.

[35] N. Samir, R. Kansoh, W. Elbarki, and A. Fleifle, “Pressure control for minimizing leakage in water distribution systems,” *Alexandria Engineering Journal,* May 2017, 1-12.

[36] A. M. Cassa, J. E. Van Zyl, and R. F. Laubscher “A numerical investigation into the effects of pressure on holes and cracks in water supply pipes.” *Urban Water J.*, vol. 7, no. 2, 2010, 109–120.

[37] K. B. Adedejia, Y. Hamama, T. A. Bolanle, and M. A. Adnan, “Pressure Management Strategies For Water Loss Reduction In Large-Scale Water Piping Networks: A Review,” SimHydro 2017: Choosing the right model in applied hydraulics, 14-16 June 2017, Sophia Antipolis,1-12.

[38] H. Awad, Z. Kapelan, and Savic, D. A, “Optimal setting of time modulated pressure reducing valves in water distribution networks using genetic algorithms.” *Integrating water systems*, J. Boxall and C. Maksimovic, eds., Taylor and Francis, London, 2009, 31–37.

[39] L. S. Araujo, H. Ramos, S. T. Coelho, “Pressure control for leakage minimisation in water distribution systems management,” *Water Resour. Manag*., vol. 20, 133–149.

[40] L. Reis, R. Porto, F. Chaudhry, “Optimal location of control valves in pipe networks by genetic algorithm,” *J. Water Resour. Plan. Manag.* Vol. 123, 1997, 317–326.

[41] L. Araujo, H. Ramos, S. Coelho, “Pressure control for leakage minimisation in water distribution systems management,” *Water Resour. Manag*., vol. 20, 2006, 133–149.

[42] G. Pezzinga, R. Gueli, “Discussion of Optimal Location of Control Valves in Pipe Networks by Genetic Algorithm,” *J. Water Resour. Plan. Manag.* Vol. 125, 1999, 65–67.

[43] E. Creaco, G. Pezzinga, “Embedding Linear Programming in Multi Objective Genetic Algorithms for Reducing the Size of the Search Space with Application to Leakage Minimization in Water Distribution Networks,” *Environ. Model. Softw.* Vol. 69, 2015, 308–318.

[44] Covelli, C.; Cozzolino, L.; Cimorrelli, L.; Della Morte, R.; Pianese, D. Optimal Location and Setting of PRVs in WDS for Leakage Minimization. Water Resour. Manag. 2016, 30, 1803–1817.

[45] A. Gupta, N. Bokde, K. Kulat, Z. M. Yaseen, “Nodal matrix analysis for optimal pressure-reducing valve localization in a water distribution system,” *Energies*, vol. 13, no. 8, 2020, 1878. DOI 10.3390/en13081878.

[46] L. S. Perelman, and A. Ostfeld, “Topological clustering for water distribution systems analysis,” *Environmental Modeling and Software,* vol. 26, 2011, 969-972.

[47] L. S. Perelman, M. Allen, A. Preis, M. Iqbal, A. J. Whittle, “Flexible reconfiguration of existing urban infrastructure systems,” *Environmental Science and Technology*, vol. 49, no. 22, 2015, 13378–13384. DOI 10.1021/acs.est.5b03331.

[48] M. R. Farley, “District Metering. Part 1- System Design and Installation. WRc. 1985”

[49] J. Morrison, S. Tooms, D. Rogers, “DMA Management Guidance Notes,” IWA Publication, 2007.

[50] M. Baker, “The Baker Report: Municipal water distribution system security study recommendations for science and technology investments.” Tech. rep., U.S. Department of Homeland Security, 2009.

[51] F. Scarpa, A. Lobba, G. Becciu, “Elementary DMA design of looped water distribution networks with multiple sources*,” J. Water Resour. Plan. Manag*. Vol. 142, 2016, 04016011.

[52] G. Ferrari, D. A. Savic, G. Becciu, “A graph theoretic approach and sound engineering principles for design of district metered areas.” *J. Water Resour. Plann. Manage.,* 2014, 10.1061/(ASCE)WR .1943-5452.0000424, 04014036.

[53] D. Laucelli, A. Simone, L. Berardi, O. Giustolisi, “Optimal design of district metering areas for the reduction of leakages,” *J. of Water Resources Planning and Management,* vol. 143, no. 3, 2017, 04017017. DOI 10.1061/(ASCE)WR.1943-5452.0000768.

[54] B. Ulanicki, H. A. Meguid, P. Bounds, R. Patel, “Pressure control in district metering areas with boundary and internal pressure reducing valves,” Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008, Kruger National Park, South Africa, Aug 17–20, 2008; pp 691–703.

[55] R. Murray, W. Grayman, D. Savic, R. Farmani, “Effects of DMA Redesign on Water Distribution System Performance,” Boxall, J., Maksimovic, C., Eds, Taylor & Francis Group: London, U.K., 2010;´ pp 645–650.

[56] G. Ferrari, D. Savic, D, G. A. Becciu, “Graph Theoretic Approach and Sound Engineering Principles for Design of District Metered Areas,” *J. Water Resour. Plann. Manage.* Vol. 140, 2014, 04014036.

[57] J. W. Deuerlein, “Decomposition model of a general water supply network graph,” *J. Hydraul. Eng.* Vol. 134,2008, 822–832.

[58] S. Alvisi, S.; Franchini, M. A heuristic procedure for the automatic creation of district metered areas in water distribution systems. Urban Water J. 2014, 11, 137–159.

[59] K. Diao, Y. Zhou, W. Rauch, “Automated Creation of District Metered Area Boundaries in Water Distribution Systems,” *J. Water Resour. Plann. Manage.* Vol. 139, 2013, 184–190.

[60] A. Di Nardo, M. Di Natale, G. F. Santonastaso, S. Venticinque, “An Automated Tool for Smart Water Network Partitioning,” *Water Resour. Manage.*, vol. 27, 2013, 4493–4508.

[61] A. Di Nardo, M. Di Natale, G. F. Santonastaso, V. Tzatchkov, “Alcocer-Yamanaka, V. Water Network Sectorization Based on Graph Theory and Energy Performance Indices,” *J. Water Resour. Plann. Manage.* Vol. 140, 2014, 620–639.

[62] O. Giustolisi, L. Ridolfi, “New Modularity-Based Approach to Segmentation of Water Distribution Networks,” *J. Hydraul. Eng.*, *vol. 140,* 2014, 04014049.

[63] H. Haider, I. S. Al-Salamah, Y. M. Ghazaw, R. H. Abdel-Maguid, M. S, and A. R. Ghumman, “Framework to Establish Economic Level of Leakage for Intermittent Water Supplies in Arid Environments,” J. of Water Resources Planning and Management, 2019, 1-13.

[64] R. D. Jadhao, and R. Gupta,” study of leakage reduction in water distribution network through pressure management, Journal of Indian Water work Association, April-June 2017, 124-133.

[65] A. N. Sharma, S. R. Dongre, R. Gupta, P. Pandey and N. D. Bokde, “Partioning of water distribution network into district metered area using existing valves,” Computer Modelling in Engineering and Sciences, January 2022.

[66] H. Liu, M. Zhao, C. Zhang, G. Fu, “Comparing Topological Partitioning Methods for District Metered Areas in the Water Distribution Network,” *Water*, vol. 10, 2018, 368.

[67] A. Abdulshaheed, F. Mustapha, A. Ghavamian, “A pressure-based method for monitoring leaks in a pipe distribution system: A Review,” *Renewable and Sustainable Energy Reviews*, 69 (2017) 902–911.

[68] A. Ostfeld, and E. Salomons, “Optimal layout of Early Warning Detection Stations for Water Distribution Systems Security,” *J. Water Resour Plan and Manage.* Vol. 130, no. 5, 377-385.

[69] A. Ostfeld, et al. “The battle of the water sensor networks: a design challenge for engineers and algorithms,” *J. Water Resour. Plann. Manage.* Vol. 134, 2008, 556–568.

[70] G. Dorini, P. Jonkergouw, Z. Kapelan and D. Savic,” SLOTS: Effective algorithm for sensor placement in water distribution systems,” *J. Water Resour Plan and Manage,* vol. 136, no. 6,2010, 620-628.

[71] M. Weickgenannt, Z. Kapelan, M. Blokker, D. A. Savic, “Risk based sensor placement for contaminant detection in water distribution systems,” *J. Water Resour Plan and Manage.*,vol. 136, no. 6, 2010, 629-636.

[72] S. Rathi, R. Gupta, S. Kamble, and A. Sargaonkar, “Risk based Analysis for Contamination Event Selection and Optimal Sensor Placement for Intermittent Water Distribution Network Security,” *Water Resources Management*, vol. 30, no. 8, 2017, 2671-2685. 10.1007/s11269-016-1309-7.

[73] J. Blesaa, F. Nejjaria, and R. Sarratea, "Robust sensor placement for leak location: Analysis and design," *Journal of Hydroinformatics*, September 2015, 1-33.

[74] B. David, D. F-H. Steffelbauer, "Efficient Sensor Placement for Leak Localization Considering Uncertainties," *Water Resour Manage,vol. 30,* 2016,5517–5533. DOI 10.1007/s11269-016-1504-6

[75] L. Chenga , D. Kun, , T. J. Peng, and D. Wei-xin, "Optimal placement of pressure sensors in water distribution system based on clustering analysis of pressure sensitive matrix," XVIII International Conference on Water Distribution Systems Analysis, WDSA 2016, Procedia Engineering, 186, 2017, 405 – 411.

[76] H. Cao, S. Hopfgarten, A. Ostfeld, E. Salomons, and P. Li, "Simultaneous Sensor Placement and Pressure Reducing Valve Localization for Pressure Control of Water Distribution Systems," no. 11, *Water*, May 2019, 1352, 1-18.

[77] F. Soroush, and M. J. Abedini, "Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm," 2019,1030-1047.

[78] Y. Shao, X. Li, T. Zhang, S. Chu, and X. Liu, "Time-Series-Based Leakage Detection Using Multiple Pressure Sensors in Water Distribution Systems," Sensors, 2019, 1-20.

[79] W. Cheng, Y. Chen, and G. Xu, "Optimizing Sensor Placement and Quantity for Pipe Burst Detection in a Water Distribution Network," *Journal of Water Resources Planning and Management*, 2021, ASCE, ISSN 0733-9496.

[80] I. Santos-Ruiz, F.R. López-Estrada, V. Puig, G. Valencia-Palomo, and H-R Hernández, "Pressure Sensor Placement for Leak Localization in Water Distribution Networks Using Information Theory," *Sensors* 2022, 1-13.

[81] Taylor, D, Slocum, A. H, Whittle, A. J. "Analytical Scaling relations to evaluate leakage and intrusion in intermittent water supply systems," PLOS, May 2018, 1-22.

[82] Ehan Raei, M. Ehan Shafiee, Mohammad Reza Nikoo and Emily Berglund, "Plancing an ensemble of pressure sensors for leak detection networks under measurement uncertainty," Journal of Hydroinformatics · October 2018.

[83] David B. Steffelbauer and Daniela Fuchs-Hanusch, "Efficient Sensor Placement for Leak Localization Considering Uncertainties," Water Resour Manage (2016) 30:5517–5533.

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14875

BIOFUNGUS: FUNGUS MBBR PILOT PLANT ON MURCIA ESTE WWTP

E. Mena[1], A. Gadea[2], A. Monreal-Bernal[3], S. López-García[4], V. Garre[5] and A. J. Lara-Guillén[6]

[1,2]Emuasa, 30008, Murcia, (Spain)
[3]Aquatec, Proyectos para el Sector del Agua S.A.U., 30100, Murcia, (Spain)
[4,5]Dpto. de Genética y Microbiología, Facultad de Biología, Universidad de Murcia, 30100, Murcia, (Spain)
[6]Technology Centre for Energy and the Environment (CETENMA), 30353 Murcia, (Spain)

[1]eva.mena@emuasa.es

Abstract

Concerns about energy efficiency and contaminants of emerging concern (CECs) in wastewater treatment plants (WWTP) lead development of new and alternative processes. Conventional activated sludge systems have a high energy consumption and footprint. Alternative processes are nowadays implemented to reduce them.

In this study, we present the results of Biofungus project. Influent wastewater is treated under real conditions and continuous operation in a two-step Moving Bed Biofilm Reactor (MBBR) pilot plant based on Mucor fungus. Several strains and spontaneous mutants of the Mucor fungus were investigated. Those showing improved growth performance, wastewater resistance and nitrate consumption rate were isolated in laboratory and used at the pilot plant. Moreover, the in-situ growth at the pilot plant of the fungus from spores is implemented at the plant as a parallel process. The obtained effluent water meets regulations requirements, showing a high COD and suspended solids (SS) removal (87% and 94% on average respectively) and total nitrogen removal of 35% on average.

The Biofungus pilot plant treatment works analogous to a conventional activated sludge process. One of the difficulties observed is the retention of the fungi during the process. The volume of carriers used at the MBBR is between 25-30% of the aeration tank. They allow the fungi to grow attached to them avoiding the dilution of its concentration by flotation and loss from this tank. This stage is followed by a settling tank, where the biomass is either recirculated to the aeration tank or purged to a sludge thickener tank. The process is a two-stage process with aerobic and anoxic reactors. The Mucor fungus specialized in the nitrate consumption is dosed into the anoxic reactor after a different Mucor string has consumed the DQO at the MBBR. Both stages are connected by a primary settling tank and a secondary decanter clarifies water after anoxic stage. After this secondary decanter the treated water is obtained. The retention time on the plant is between 6 and 10 hours and the treated volume is 3,6m3/day.

The *Mucor* fungus has proved able to eliminate high ammonia and nitrate concentrations in sort periods of time, resulting on consumption rates of 1,6 mgN-NO^3/h. Finally, its resistance to CECs has been evaluated under a wide range of contaminants including drugs, pesticides, herbicides, fungicides and hormones. Almost no toxicity to contaminant concentrations as high as 20 mg/l was observed. The water treatment performance has been also tested with a combination of CECs at 200µg/l and no influence has been observed after a week over the effluent water quality.

Keywords
Innovation, wastewater, Moving Bed Biofilm Reactor, fungus, denitrify, carriers.

2022, Universitat Politècnica de València
2nd WDSA/CCWI Joint Conference

1 INTRODUCTION

Activated sludge process has been the most applied biological treatment for urban wastewater since its discovery at the beginning of 20 century. Different bacterial communities are able to transform organic matter and nitrogen compounds into gas nitrogen and carbon dioxide. However, with the aim of reducing the footprint of the wastewater treatment plants (WWTP), the energy consumptions and improving the performance of conventional processes, the last two decades a lot of effort has been carried out investigating other microorganism and alternative technologies. Fungi and algae are some of the organisms' candidates as bacteria substitutes for wastewater treatment processes.

The use of filamentous fungus in wastewater treatment is still a novel approximation. Various studies have described the advantages related with the use of these organisms. Some of the benefits are: a) an improvement in nitrogen compounds elimination by the fungus or by synergistic processes with the bacteria; b) the improved flocculation and suspended solids separation; c) the reduction of produced sludge during the wastewater treatment process; d) improved performance during the dehydration stage of the sludge; e) increase the biogas production by digestion of fungal sludge; f) great resistance to contaminants of emerging concern and it assimilation and degradation.

Several fungi species have been researched in the past. Some promising results for wastewater treatment application have been obtained with *Mucor* and *Fusarium* species demonstrating a high performance for suspended solids, turbidity and chemical oxygen demand (COD) removal [1]. Although these are the most studied and known species other fungus, such as the *Aspergillus niger,* have been studied, reporting COD removal of 72% [2]. Denitrification process for wastewater treatment have been also investigated resulting in an application that worth more in-depth investigation [3].

Besides the degradation of organic matter and nitrogen compounds (ammonia and nitrate), it has been demonstrated that fungal sludge could be and excellent source of valuable subproducts such as amylase, chitin, chitosan, glucosamine, lactic acid and several antimicrobial compounds [4].

The fungus, due to its growth structure by hyphae and mycelium development (being the mycelium the aggregate of hyphae), allows the suspended solid retention at its external layer and the nutrient harvesting. Moreover, due to its higher number of gens and superior complexity as organism it has developed an improved reproductive selectivity, that translates into an enhanced flexibility and adaptation to the changing environment [5]. Another interesting aspect from fungus structure is that it produces extracellular enzymes responsible of facilitating the degradation of different refractory compounds such as phenolics, polyhydroxyalkanoates (PHAx) or pharmaceutical active compounds (PhACs), frequently present in wastewaters. This fungus peculiarity has led to several studies of complex wastewater treatment effluents or the elimination performance of contaminants of emerging concerns [6], [7], [8].

But, all these mentioned advantages sometimes do not materialized when real conditions and large scale pilots are tested. The aim of BIOFUNGUS project is demonstrating the benefits of a fungal wastewater treatment at semi-large scale in real-life operating conditions. A Moving Bed Biofilm Reactor (MBBR) pilot plant has been designed and built for testing the two-step process based on *Mucor circinelloides* fungi.

The project, ended in June 2022, demonstrates the viability of pilot plant in-situ fungus culture at large scale, enabling the continuous dosing by demand of the process. This result allows to keep the treatment performance in range of the conventional bacterial activated sludges process. Besides that, the fungus has shown good tolerance to high contaminants concentrations, such us drugs, herbicides or pesticides, with almost no inhibition of the wastewater treatment process.

2 MATERIALS AND METHODS

Several strains and spontaneous mutants of the *Mucor* where investigated. Those showing the best growth performance where used. The optimal culture conditions for the *Mucor* fungus have been identified at laboratory and later applied at the pilot plant. These strains and mutants were tested on a two-stage process. Besides, mutants specialized in assimilate nitrates have been also isolated. The isolation of this strains was carried out by growing the fungus different strains under laboratory conditions and analyzing the change in color of the culture substrate. When *Mucor* preferably consumes nitrate, the substrate pre-mixed with methyl orange changes its color from red to yellow. Finally, from the tested Petri dish, spores are isolated from the mutant *Mucor* where this change of color and preferably nitrate consume is observed (*Figure 1*).

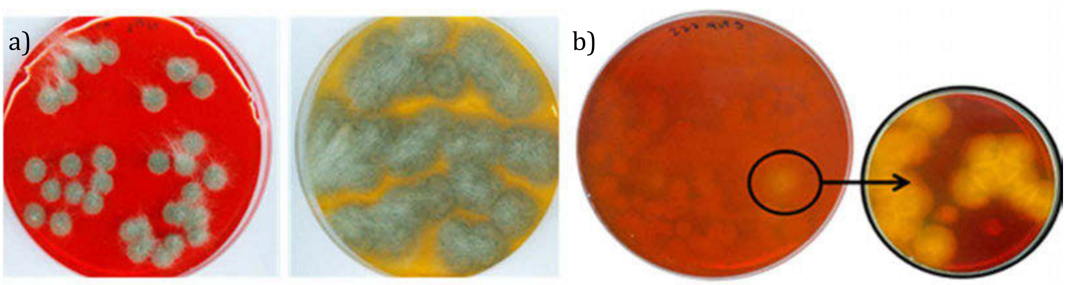


Figure 1. a) Mucor fungus culture showing the specialized nitrate consumption mutant. b) Isolation process of the nitrate specialized fungus.

By needs of the wastewater treatment process, the fungus may be preserved during long time ready to be dosed into the biological reactor (MBBR). With that aim, multiple *Mucor* candidates were kept in laboratory being periodically put in contact with aliquots of wastewater, confirming its survival and growth performance.

The pilot plant is located at Murcia Este WWTP and the influent wastewater is the same for both plants. Composites samples of the influent wastewater are collected daily automatically, whereas individual samples are collected manually every 48 from the different stages of the pilot plant (MBBR, settlers, anoxic reactor and effluent). The analysis of these sampling points was monitored during the 18 months of continuous evaluation of the pilot plant. The inlet flow was kept between 120-150 liter/hour, and the retention time between 6 and 10 hours. The average treated volume has been 3.6 m^3/day.

The control parameters during the project development consisted of: pH, conductivity, redox, temperature, chemical oxygen demand (COD), biochemical oxygen demand (BOD5), total nitrogen (TN), ammonia (NH_4^+), nitrates (NO_3^-), mixed liquor suspended solids (MLSS) and settleability (V30 and sludge volumetric index, SVI). Sporadically, it was also analyzed solid content (dried matter, DM) and volatile content (VM) from the mixed liquor and sludge. The analytical methods applied are the conventional for wastewater analysis.

The acceptance criteria used for the treated wastewater are the ones defined by the government regulatory entity to Murcia Este WWTP: (i) COD< 150 mg/l, (ii) TN<150 mg/l, (iii) NH_4^+<15 mg/l y (iv) NO_3^- <15 mg/l.

The anaerobic digestion of the sludge was tested, firstly, by biomethane potential test (BMP) using a "Automatic Methane Potential Test System (AMPTS) from Bioprocess Control" commercial equipment, and, secondly, by a laboratory scale anaerobic digestion process (10 liters reactors).

Emerging contaminants have been analyzed by an external accredited laboratory and methods using high pressure liquid chromatography (HPLC).

3 RESULTS AND DISCUSSION

3.1 Pilot plant design and operation

In a previous study [9] several mutants of the *Mucor circinelloides* were proven as good candidates for wastewater treatment, eliminating organic matter and ammonia but remaining a high concentration of nitrate in the effluent water. Biofungus project has been focused on the isolation of strains of the same *Mucor circinelloides* fungus specialized on the consumption of nitrate. The chosen fungus for the validation of the process at the pilot plant has shown nitrate consumption preference over the ammonia, though it is able to assimilate both sources of nitrogen.

For the validation of the process a two-stage pilot plant was design and built, it is presented in the *Figure 2 a)*. The plant is divided into two stages: i) a MBBR reactor (25% carrier volume), equipped with fine bubble diffusers, where the fungus responsible of the degradation of organic matter and ammonia is dosed (from now on called "aerobic fungus") and ii) a secondary anoxic reactor where the nitrate specialized *Mucor* strain is dosed (called "anoxic fungus"). The used of polymeric supports (also known as "carriers") allows the biomass (bacterial or fungal cultures) to growth attached to it and preserve a high concentration at the reactor. The plant has also decanters after the aerobic and anoxic stages, for the recovery of the biomass and its recirculation and purge. Moreover, for the in-situ growth of the fungus, two cultivation tanks were incorporated to the plant, both with pH and temperature control and nutrient solution dosing.

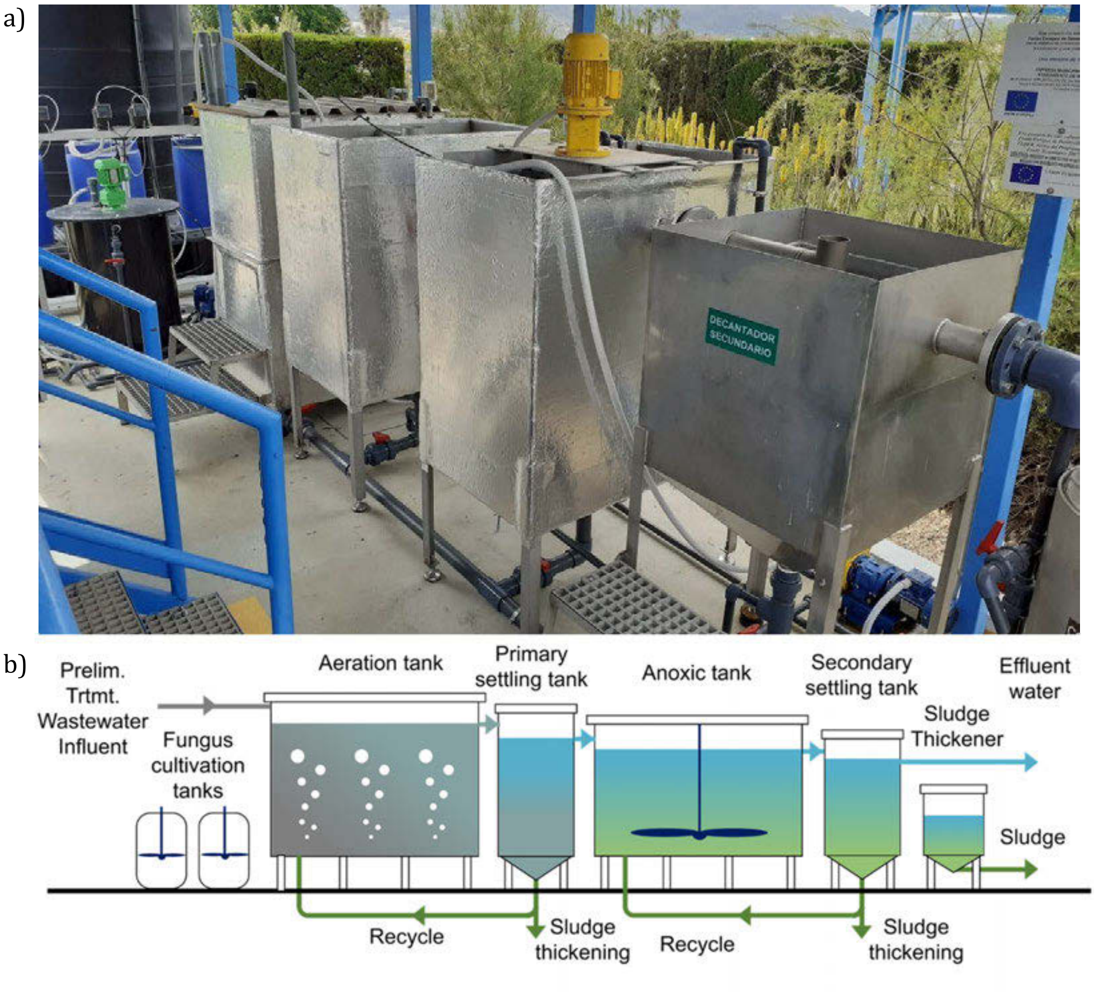

a)
b)
Prelim. Trtmt.
Wastewater Influent
Fungus cultivation tanks
Aeration tank
Primary settling tank
Anoxic tank
Secondary settling tank
Sludge Thickener
Effluent water
Sludge
Recycle
Sludge thickening
Recycle
Sludge thickening
DECANTADOR SECUNDARIO

Figure 2. a) Photography and b) schematic of the Biofungus pilot plant located at Murcia Este WWTP.

The working flow of the pilot plant *Figure 2 b)* consist of: the influent wastewater, after passing through the pre-treatment, enters the aerobic reactor where remains in contact with the fungal biomass until the organic matter and ammonia is removed. 2ppm O_2 is used as default value but it was increased up to 4-5ppm for punctual high COD periods. After the aerobic treatment, the biomass is recovered and recirculated to the aerobic reactor or purged if is required. The clarified water enters the anoxic reactor where the nitrate consumption specialized fungus is dosed at a concentration of 1-2% in volume per day and the denitrification process is carried out. Both aerobic and anoxic reactors have 1000 liters capacity. After the anoxic reactor another decanter is used for last clarification and recovery of the biomass. Primary and secondary settlers have 600 and 300 liters of capacity respectively. The purged sludge is sent to a concentration tank where water and sludge are separated. The plant is additionally equipped with a control system for timing every equipment and data logger allowing to register all sensor parameters.

Some microscopy photographs from the *Mucor* hypha structure and mycelium attached to the carrier substrate are shown in the *Figure 3 a* and *b* respectively. Image *c) shows Mucor* biofilm attached to the carrier and picture *d)* shows its growth after 48 hours.

During the project development, one of the objectives was to optimize the fungus growth conditions from spores to a ready-to-use mycelium solution and to increase its lifespan. The optimum conditions for the in-situ culture were first established at laboratory. Parameters like composition, dosing, frequency of feeding, pH and temperature were identified. Urban wastewater and residual water with high COD content from food industry were tested as culture media. *Mucor* fungus showed the same growth performance in artificial media than in presence of both food industry residual water and wastewater.

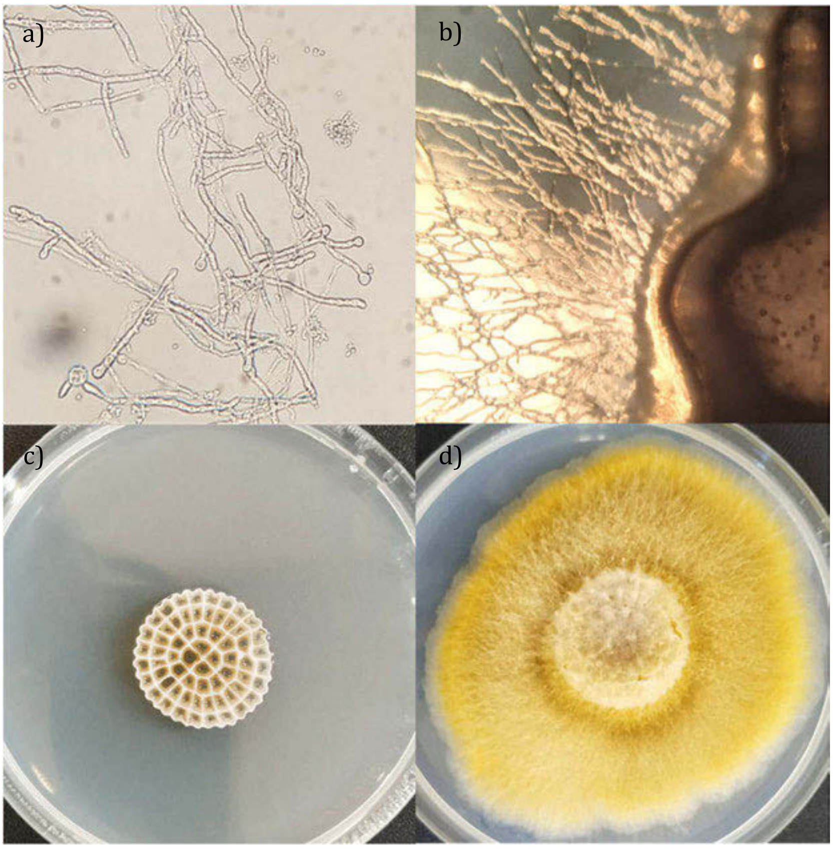


Figure 3. Microscopy images of the Mucor a) hypha structure and b) mycelium attached to the carrier. c) Pilot plant carrier with fungal biofilm and d) its growth after 48 hours.

The established conditions were later validated during the pilot plant validation and all parameters where slightly adjusted. Cultures of 100-200 liters mycelium solutions were periodically prepared for the aerobic and anoxic fungus. For 100 liters of culture media 2 kg of glucose supplemented with 0.2 kg of yeast nitrogen base (YNB) were mixed with water. The mycelium solution was ready after less than 24 hours of adding 50ml of the spore solution. Growth time mainly depends on the temperature of the tank and the age of the spores. It has been observed that preserving several weeks the spores ready at 4 or -20°C led to a delay of 12-24 additional hours. The optimum culture temperature was 26°C and pH between 4-5. In practice, the culture pH drops to 3.5-4 when the spores are mixed with the culture media and stays at 4-5 range, indicative of a good development of the fungus growth.

Among the difficulties faced during the development of the project, avoiding bacterial contamination of the fungus and nutrient solutions have been one of the priorities. One of the proposed solutions was to pasteurize the nutrients solution, increasing its temperature periodically to 60°C for 20 minutes. Although it eliminates the bacterial contamination it has a high energy cost. In practice, using solid nutrients instead of nutrients in solution avoid part of the contamination. The fungus solution last in the cultivation tank between 1 and 1.5 months before the *Mucor* culture is inhibited. The presence and concentration of the *Mucor* was confirmed by microscopy technique every 1-2 weeks.

The pilot plant was operated in continuous mode between January 2021 and June 2022. Four validation processes (stages) were carried out. Every stage lasts between two-four months and real conditions such us rainy, cold or very warm weather periods were tested, besides peaks of influent load (COD) o unprogrammed stops or failures. The results obtained during one of these stages is presented in the *Figure 4*.

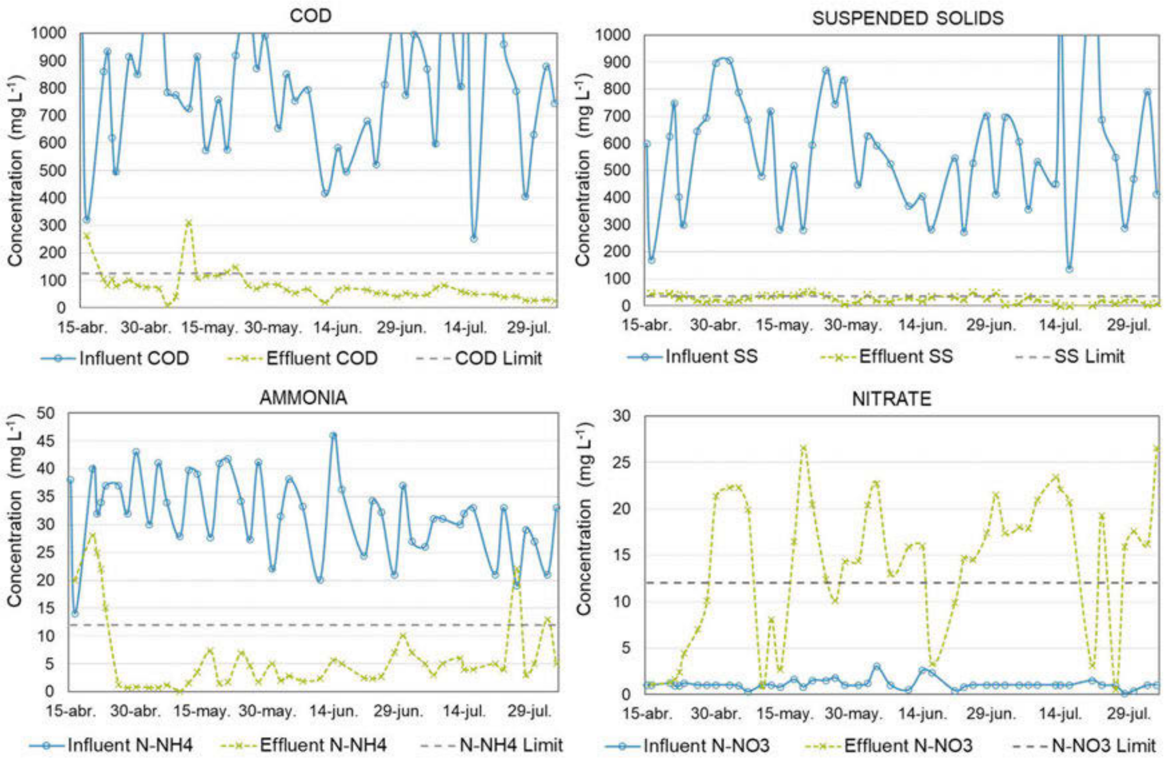


Figure 4. Influent and effluent concentrations of the main parameters monitored during the continuous operation of the Biofungus pilot plant.

Influent and effluent daily concentrations during three months of operation are shown in *Figure 4*. High performance eliminating the COD and suspended solids are observed since the first week of operation, 87% and 96% on average, respectively. On the contrary, the total nitrogen removed is in the range of 37%. Although the ammonia removal is 80% on average, the fungal sludge has very poor settleability hindering a proper mixed inside the anoxic reactor and avoiding the continuous elimination of the nitrate. The poor settleability of the fungal sludge is attributed to the nature itself of the fungus. *Mucor* fungus is an oleaginous and filamentous organism and both characteristics have negative impact on settling properties. Besides, probably related to the glucose content of the fungus solution, it has been observed that the sludge present in the anoxic reaction suffers some kind of fermentation/foaming. During several stages of the project different approaches have been tested but still the poor settleability prevent achieving a stable concentration and mixture inside the anoxic reaction. An alternative configuration for the anoxic reaction that prevent the escapement of the biomass by flotation must be investigated.

Although, with this pilot plant configuration was not possible to achieve continuous elimination of nitrate, several control experiments validating the fungus nitrate consumption were performed. In the *Figure 5* one of several control experiments is presented. The anoxic fungus characterized previously at laboratory showed its preference for the nitrate over the ammonia. Same behavior is observed when fungus is dosed at the pilot plant in presence of both compounds. Reduction rates between 10-30 mgN-NO_3/h for the nitrate and 1,5-2 mgN-NH_4/h for ammonia were obtained. First, 100% of nitrate was removed in less than an hour from the anoxic tank for initial concentration of 15mg/l. After its complete consumption, 100% ammonia was eliminated in the next 12 hours. Similar analytical results were obtained by standard laboratory test (dotted line) and by the multiparametric probe for NH_4 and NO_3 installed in the pilot plant (continuous line) (see *Figure 5*).

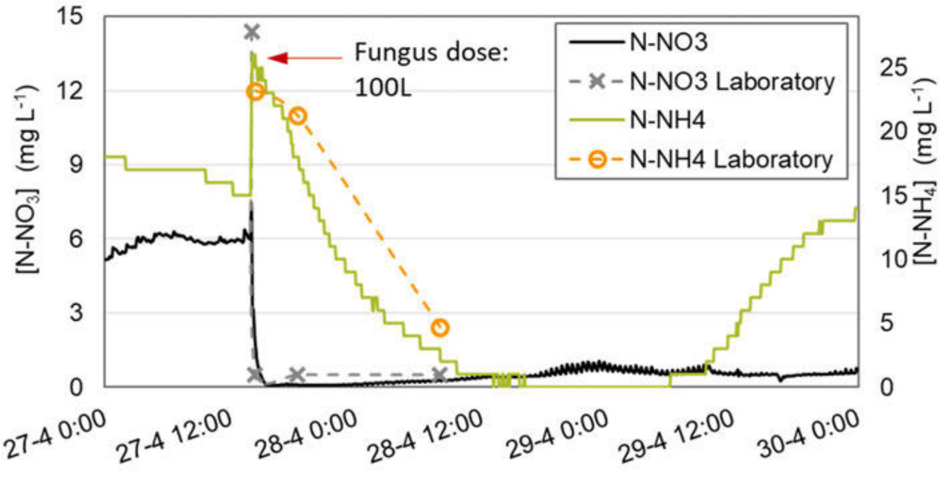


Figure 5. Control experiment validating nitrate and ammonia consumption after dosing the Mucor fungus.

Bacteria and fungus coexist in the pilot plant. The biofilm growth in the MBBR carriers, besides a high fungus content visible by microscopy (see images from *Figure 2* c and d), has the typical bacteria colonies from a conventional wastewater treatment process. This has been proven by fluorescence in-situ hybridization (FISH) technique where some fluorescent probes are attached to specific nitrifying bacteria genetic DNA sequences or chromosomes. The bacteria populations are then directly observed by fluorescence microscopy. The presence of the fungus is also confirmed for every stage of the process by adding aliquots of wastewater or sludge to Petri dishes and cultivating them for 24-48 h. This coexistence between both organisms also presents a synergistic effect for the process. The fungus provides fast response of the process and toxicity resistance whereas the conventional bacteria is still working in parallel.

During the continuous operation of the pilot plant, it has been noted the low production of obtained sludge, values between 0.3-0.6 kgDM/m^3. This low production is a great advantage of the process due to the importance of the sludge management in a conventional WWTP.

3.2 Contaminants of emerging concern: fungus resistance and concentration removal

Wastewater treatment process and fungus resistance to contaminants of emerging concern were a key point to study for the Biofungus project. Influence and interaction of contaminants for both *Mucor* strains were analyzed under laboratory conditions and pilot plant conditions. Several contaminants were chosen according to literature [10], but also, to previous influent analysis showing the presence of certain compounds in the wastewater. Three different experiments were proposed: 1) validating if the contaminants produce some evidence of toxicity over the fungus, inhibiting its normal growth. 2) to measure concentrations variations of contaminants in presence of the fungus. 3) evaluating the Biofungus pilot plant performance after the ingress of a combination of contaminants.

Different drugs, pesticides, herbicides, fungicides and hormones were added to the culture media of the *Mucor* fungus, and its growth performance was compared to a blank sample. The list of tested contaminants, all with purity between 97%-99%, are indicated in the *Figure 6 a*.

The first experiment consisted of the comparison between the mycelium growth mass after 72h in presence of the contaminant and compared to a control sample without contaminants. The test was conducted with every individual contaminant at a time, and with multiple concentrations: 20 µg/l, 200 µg/l y 20 mg/l. The results concluded that the growth performance was not clearly hindered by the presence of the contaminants. A small decrease was measured (between 2-5%) compared to the blank sample for every contaminant except for the sulfamethoxazole, that led to a mass 10% lower compared with the control sample.

For the second test, the concentration of the contaminants were measured after 72h in contact with the fungus. The difference between initial and final concentration is presented in *Figure 6 a*, and the performance of removal (by assimilation or degradation) is showed. The elimination was measured for every contaminant and, on average, 90% and 71% was removed for initial concentrations of 20 µg/l y 200 µg/l respectively. Further experiments are needed to discern whether the contaminants are assimilated or degraded by the fungus. Typical contaminants elimination processes during the wastewater treatment, such as evaporation (volatile components), oxidation reactions due to aeration or degradation by other chemical agents, are discarded.

The third and final test was performed to evaluate the Biofungus pilot plant performance when contaminants are present in the influent water. A solution of the indicated contaminants was prepared at a concentration of 200 µg/l and directly poured into the MBBR reactor. The days after the contaminants enter the plant, daily COD and ammonia analysis were conducted. As can be observed in the graph of the *Figure 6 b* the treatment performances were not affected by the contaminants, and the concentration of selected representative contaminant was zero after the third day.

These three experiments not only confirm the great resistance of the *Mucor* fungus to the selection of contaminants of emerging concern if not that it is a robust process much more resilient than conventional processes.

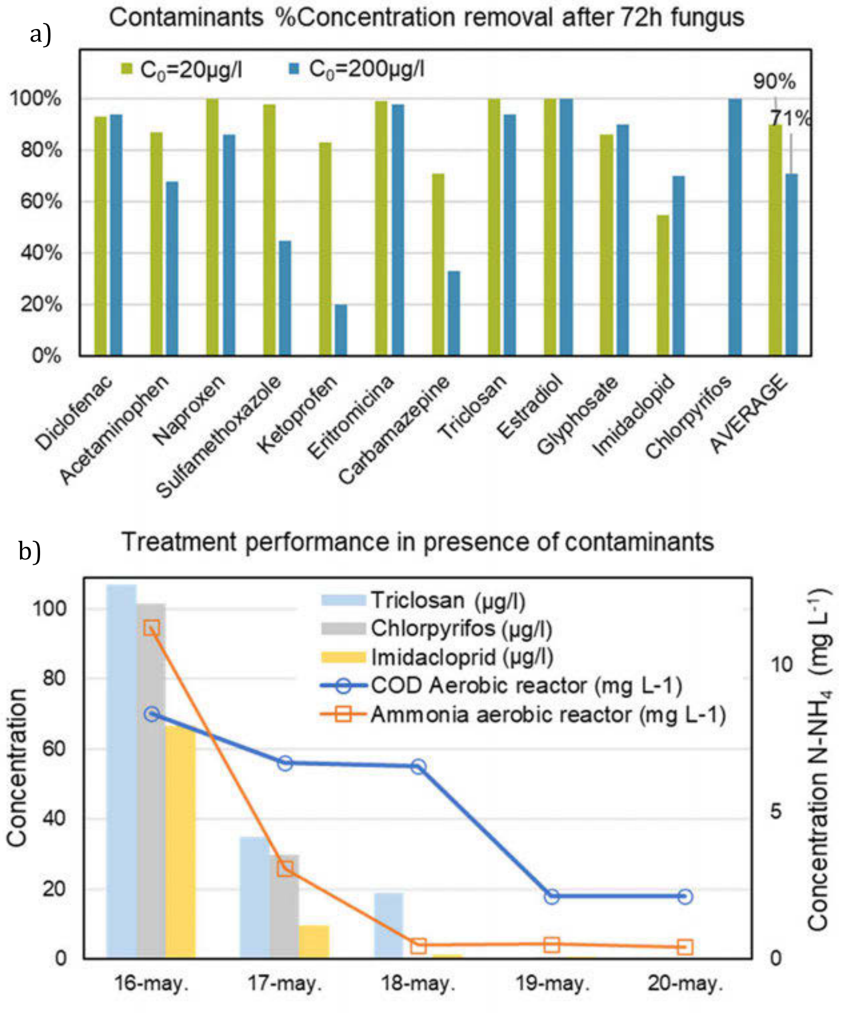


Figure 6. a) Contaminants %concentration removal after 72 hours in presence of Mucor fungus. b) Pilot plant wastewater treatment performance after adding a mixture of 13 contaminants at a concentration of 200 μg/l.

3.3 Biofungus sludge biomethane potential

Biofungus sludge was also characterized by BMP during different stages of the process. The results were in line with the sludge age of the plant. Due to the low production of sludge very few purges were needed. The biomethane production evolves accordingly with the sludge age, showing relation with the measured initial volatile content of the samples.

BMP results during the first month of operation were very promising resulting into values two-fold the ones obtained with conventional WWTP sludge. On the contrary, after 2 months of operation, biomethane productions were equal for both samples or slightly lower for the fungal sludge. The reduction in biomethane production with the sludge age was confirmed and therefore, some representative experiments of stabilized Biofungus sludge compared with the WWTP sludge are presented in *Figure 7*.

cc BY NC SA

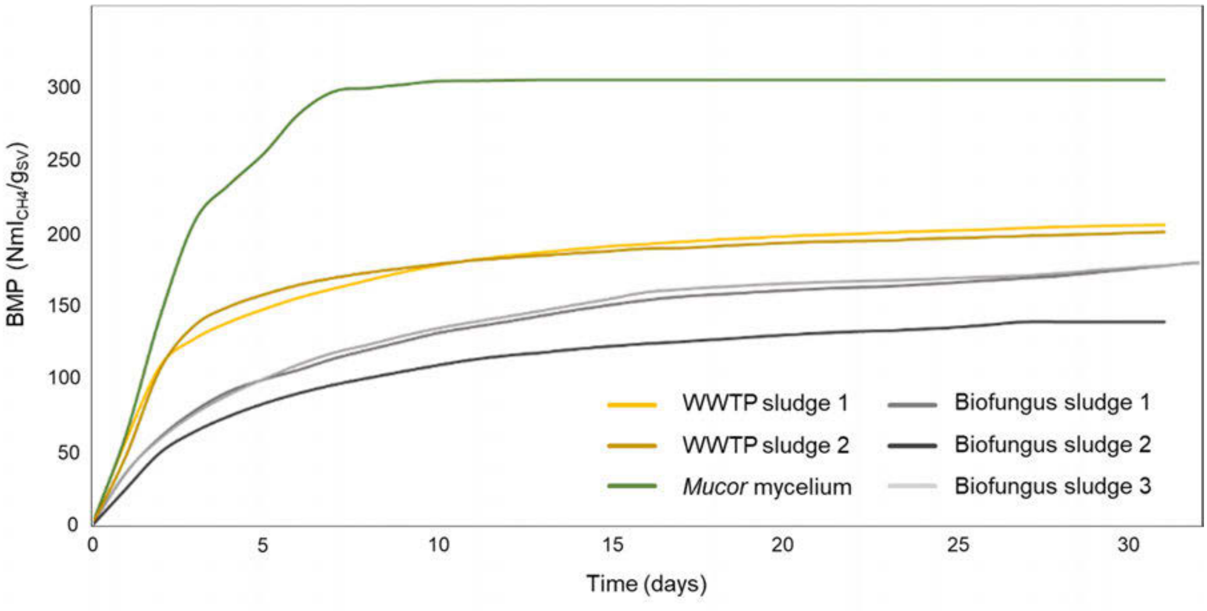


Figure 7. Average values of Biofungus pilot plant sludge samples compared with conventional WWTP sludge and the pure Mucor mycelium.

Figure 7 shows, respectively, the evolution over time of the biomethane potential of sludges produced in the Biofungus plant (from the last stage, anoxic decanter), compared with fungus mycelium and with the conventional WWTP sludge. As can be seen, BMP of mycelium is almost 50% higher than that of conventional sludge. However, the sludge generated during the daily operation using these mycelia as inoculum in the reactors has a methane production potential much lower than that of conventional sludge. It is also less that the half of the BMP of its corresponding mycelia.

Nevertheless, it is proven that Biofungus sludge is completely valid for its treatment as co-digestate by anaerobic digestion.

4 CONCLUSIONS

The development of the Biofungus project has proven that *Mucor circinelloides* is good candidate for the wastewater treatment application, not only for the aerobic treatment, achieving similar performance to that obtained in conventional treatments, but also for denitrifying process. Although, technical difficulties in the design of this plant have hindered the continuous elimination of nitrate, the elimination of 100% of ammonia and nitrate has been proved with high elimination rates.

The in-situ culture of two different strains of *Mucor* fungus has been carried out. Food industry residues have been used as culture media, being this a good initiative for residues circularity.

The high floatability observed for the fungus sludge has been related to the oleaginous and filamentous nature of this fungus. The continuous operation of the pilot plant was hampered since the beginning of the project, resulting on the escapement of the anoxic fungus from this stage and making the continuous denitrification almost impossible. No solution has been found with the proposed plant design and some alternatives considering the sludge flotability must be consider in the future.

BMP Biofungus sludge characterization was performed since the beginning of the project. High biomethane production was observe at low sludge ages, but due to the small amount of sludge produce by this process, the conventional sludge of the plant resulted on high ages and lower volatile content. The fungal sludge results, on average, on lower biomethane production compared with conventional WWTP sludge, but still can be perfectly used as co-digestate at anaerobic digestion treatment.

Finally, the *Mucor* fungus have show very high resistance to contaminant of emerging concern, included very different categories such as drug, pesticides, herbicides, fungicides or hormones, up to concentrations as high as 20mg/l. High removal efficiency of these contaminants was also measured, but still further investigation is needed to know if contaminants are assimilated o degraded. The robustness and resilience of the process was also proven in presence of contaminant after a discharge containing 200 µg/l of thirteen contaminants.

5 REFERENCES

[1] A. Fakhru'l-Razi y A. Molla, "Enhancement of bioseparation and dewaterability of domestic wastewater sludge by fungal treated dewatered sludge" J. Hazard. Mater., 147, 2012, p. 350–356.

[2] L. Coulibaly, Bioconversión de macromolécules dans un réacteur simulant un écoulement piston en régime transitoire. Cas de la bioremédiation d'eaux usées synthétique par Aspergillus niger. Thèse de doctorat, Université Catholique de Louvain, Unité de génie.

[3] R. Guest y D. Smith, "A potential new role for fungi in a wastewater MBR biological nitrogen reduction system" J. of Env. Eng. and Sci., 6, 2002, pp. 433-437.

[4] S. Sankaran y S. Khanal, "Use of filamentous fungi for wastewater treatment and production of high value fungal byproducts: A review." Env. Sci. And Tech., 40, 2010, pp 400-449.

[5] J. Bennett y L. Lasure, More gene manipulations in fungi, San Diego: Academic Press, 1991.

[6] L. Ferrando Climent, C. Cruz Morató, E. Marco Urrea, T. Vicent, S. M., R. Mozaz y B. D. S., "Non conventional biological treatment based on Trametes versicolor for the elimination of recalcitrant anticancer drugs in hospital wastewater" Chemosphere, 136, 2015, pp. 9-19.

[7] M. Shereve, A. Brockman, M. Hartleb, S. Prebihalo, F. Dorman y R. Brennan, "The White-rot fungus Trametes versicolor reduces the estrogenic activity of a mixture of emerging contaminants in wastewater treatment plant effluent" International Biodeterioration & Biodegradation, 109, 2016, pp. 132-140.

[8] C. Cruz Morató, E. Marco Urrea, T. Vicent, S. M., S. Rodríguez Mozaz, D. Barceló, D. Lucas, M. Llorca, M. Gorga y M. Petrovic, "Hospital wastewater treatment by fungal bioreactor: Removal efficiency for pharmaceuticals and endocrine disruptor" Science of the Total Environment, 493, 2014, pp. 365-376.

[9] F. Navarro-Sánchez, N. Moya, A. Gadea y E. Rodríguez, "Ignis Fungus: obtención de lodos de EDAR de alto poder calorífico mediante enriquecimiento de hongos oleaginosos" in Proc. CONAMA16, Madrid, 2016.

[10] Tran, Ngoc Han, Martin Reinhard, and Karina Yew-Hoong Gin. "Occurrence and fate of emerging contaminants in municipal wastewater treatment plants from different geographical regions-a review" Water research, 133, 2018, pp. 182-207.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14886

WATER CONSUMPTION ANALYSIS DURING NIGHT HOURS OF RESIDENTIAL CUSTOMERS

Estrada Estrada Arianne Isabella [1], Arregui De La Cruz Francisco [2], Soriano Olivares Javier [3], Ponz Carcelén Román [4] and Torres Toro David [5]

[1,2,3,4,5] Universitat Politècnica de València, Valencia, España

[1] *ariannelsabella@gmail.com,* [2] *farregui@ita.upv.es,* [3] *jasool@ita.upv.es,*
[4] *rponz@emimet.es,* [5] *datoto@emivasa.es*

Abstract

Managing and reducing water losses should be a primary concern to ensure the sustainability of a water utility. Among all the potential strategies, the design and construction of district meter areas (DMA) is probably one of the most widely used for water loss assessment and control. This is because partitioning a water network into smaller portions significantly facilitates the analysis procedures and improves the speed at which bursts and leaks are detected and located. This analysis is typically done by processing and evaluating the time series of the inflows into the DMA.

The Minimum Night Flow (MNF) represents the lowest flow into the DMA over a 24-h period. MNF typically occurs between 02:00 and 04:00 AM. During this period, most users do not intentionally use water and the inflows into the DMA are mainly composed of leakage at DMA pipes and private plumbing systems. Consequently, the analysis of the MNF allows for easy and accurate quantification of the magnitude of leakage in a particular DMA.

The main difficulty in applying this methodology appears when trying to disaggregate the night flows into its fundamental components: 1) Leakage in mains and connection pipes belonging to the distribution network 2) Leakage at customers' facilities, and 3) Intentional use of water by customers. The first two components correspond to continuous flows that, in most cases, remain constant during the night hours. The third component is inherently random and may vary in magnitude and duration.

In the proposed work, detailed data of one-year hourly readings from approximately 20,000 customers have been analyzed and disaggregated into leakage and intentional use. The aim is to improve the models that characterize the night consumption originated at the customers' facilities. Hence, the study's initial stage involves the development of algorithms that allow disaggregating the night consumption registered by the customers' water meters into a baseline consumption due to leakage and the component caused by the intentional use of water.

Probability function distributions of leakage flow rates at customers' plumbing systems have been obtained for various types of water users. Simplified probability distribution functions of intentional water use have also been developed to consider the duration of water consumption and the average flow of residential users. These probability functions allow the creation of synthetic consumption series that overlap with the baseline consumption caused by leakage inside the customers' premises.

The novelty of the proposed methodology is that the probability functions obtained have been derived from actual water consumption data of nearly 20,000 customers, monitored for one year. The night consumption model developed may enable the water utility to better estimate this parameter in those DMAs where the customers' water meters cannot be read hourly.

Keywords
Supply, Water, Probability, Leakage, Estimation.

1 INTRODUCTION

Water scarcity has been growing around the world. Activities that influence these actions involve population growth, expanding industrial activities and urban areas, climate change, and water pollution. An estimated third of the population in the world has no access to clean water. One-fifth of the world's population lives in water scarcity areas where resources are insufficient to satisfy the demands. One-fourth of the population in the world lives having not filled their water demands because of the poor management supplies of authorities. (Narmilan et. al, 2020) [1].

According to the National Statistic Institute (INE) in Spain, "During 2018, approximated 4,236 cubic hectometres (hm3) of water were supplied to public urban supply networks. Approximately Three quarters (3,188 hm3) were volumes of registered water". The data shows that a quarter of the water supplied corresponds to unregistered volumes for which users do not pay. In a high proportion, these volumes correspond to leaked volumes in the distribution networks and the internal installations of the users.

Based on these concerns, the new Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020, concerning the quality of water intended for human consumption, member states will ensure that an assessment is made of the levels of water leakage in their territory and the potential for improvement in reducing water leakage. These actions are in line with the Sustainable Development Goals of the 2030 agenda. For this purpose, the Structural Leakage Index (SLE) classification method is usually used, joint with a vast knowledge of the network behavior and a set of management strategies and methodologies that allow operating in increasingly sustainable conditions over time.

As a strategy to fit the demands, water consumption patterns have always been present in supply management. These patterns can be well-defined or very sensitive depending on the population under study. In the past, obtaining these patterns was due to the incorporation of metering technology and the construction of metering district areas. This facilitated the definition of fairer tariffs for users and improved the operation of the supply. However, the installation of these devices is concentrated in population centers. At the same time, the ends of the network have fewer meter devices and those have daily interval resolution.

Usually, the relevant information for developing future planning is unavailable to take action at the correct time. For that reason, it is common to estimate them through methods that are fundamental for resilient network management. (Avni et. al, 2015) [2]. In districts with limited information, is often available general information and approximate. In many cases, the approximations are not very precise and outdated. This consequently reduces the operational management options and limits the capacity to adapt to the global consumption of the users. Knowing more specifically about the variation of the patterns offers a window of action in detecting leaks, analyzing atypical events, and the improvement of the potential efficiency of the network.

In managing leakage reduction, it is beneficial to study hourly consumption patterns at night. The analysis of the minimum nocturnal flow allows an easy and accurate quantification of the magnitude of leakage in a given sector. This is because, at that time, the flow transported was composed of two elements. Elements are mainly due to a continuous flow that can be leakage or storage elements and random but intentional water use.

Analyzing the consumption patterns during the hours of minimum night flow in sectors with actual measurements makes it possible to extrapolate these measurements to sectors without hourly readings. These readings can be used for consumption disaggregation. In this way, to establish in magnitude and frequency the range of the function of each element, the trend of the consumption data, and the influence they have on the overall consumption of the users.

2 DATA CHARACTERIZATION

This study analyses domestic consumption during night hours in a Mediterranean City. As expected, customers decrease their water-related activities at night, and most of them spend the time resting. This typical behavior of domestic customers facilitates the analysis. Considering this, the consumption during night hours can be classified into two types: 1) one originated by leaks and other continuous consumptions and 2) other intentional uses of water like, for example, flushing toilets, drinking a glass of water, washing hands, or taking baths and showers. Most of these intentional consumptions are produced at medium flow rates, close to 0.1 L/s or 0.2 L/s, and usually happens in randomly ways with changing magnitudes and durations.

Consolidated consumption patterns, those measured by the flow meters installed in the distribution network, show how people behave as a community. They can be considered a reflection of typical human activities during night hours and may identify how specific habits and events influence domestic water consumption. A Good example of this is how human behaviors change with respect to water consumption during weekends and holidays when it is common to appreciate higher flow and larger volumes delivered to customers at night time.

The Data provided by Global Omnium that have been used in this study are composed of daily and hourly volumetric measurements during one year (2019). After data cleansing, readings of more than 20 thousand customers across five network sectors are available for analysis. Costumers in these sectors are mainly domestic, although other commercial and industrial users are also present (Figure 1a). Figure 1b shows how customers are distributed per sector. However, due to the heterogeneity of non-domestic customers and the limited sizes of the available sample, the analysis presented has been restricted to domestic customers.

Water consumption measurements were taken from the water meters located at the customer's connections. Their normal size ranges from DN13 to DN100, although domestic meters DN13 and DN15 represented more than 99% of the meters. All these meters were less than five years old and had communications capabilities allowing water consumption data at one-hour intervals. A more detailed database with installation details such as users' location, demand type, users' code, and observations is available for further analysis.

The water utility provides hourly consumption data for one complete year from all meters installed in these sectors. Initially, the dataset was composed of 499 CVS files. These files were reorganized by a module of algorithms written in VBA language to build a single simplified database in Microsoft Access. However, due to MS Access size limitations, the simplified database only includes night consumption during night hours and the total daily water usage. In the next step. The data were checked and filtered to verify hourly reading availability and calculate the daily water consumption of each customer. This calculation was restricted to users having DN13 and DN15 meters. Finally, only users with more than 300 days of full records between 1:00 and 7:00 AM were considered for the analysis.

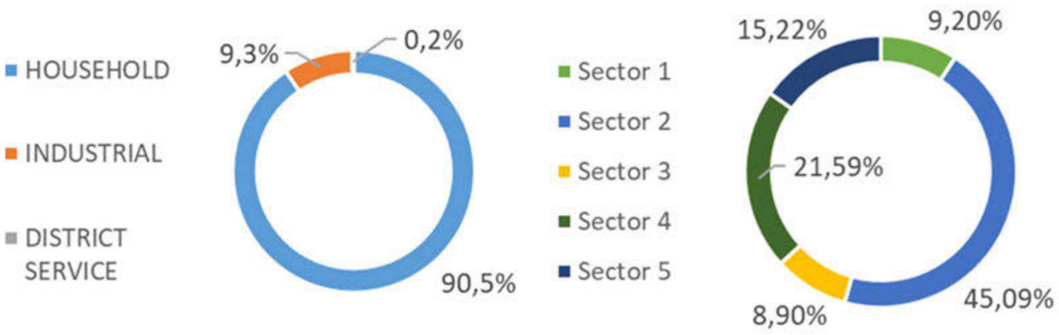


Figure 1. a) Customers distribution by users type b) Customers distribution by sectors.

At the end of this process, the database changes from 37,384 customers and 12,724,924 records to 19,145 customers and 6,883,340 records. This selection was a set point to start building up the methodology by disaggregating consumption at night.

3 ANALYSIS METHODOLOGY

The Night Flow is the lowest flow of the entire daily record. In many cases, this flow is steady, and usually, it is not modified by high consumption records. Sanitary devices' standard flows are indeed associated with those flows. For example, Figure 2 shows a night consumption that can notice a high flow during 2:00 AM and 7:00 AM. The rest of the time establishes a steady baseline of flow. Those high flows are similar to the waste of half-flush toilets.

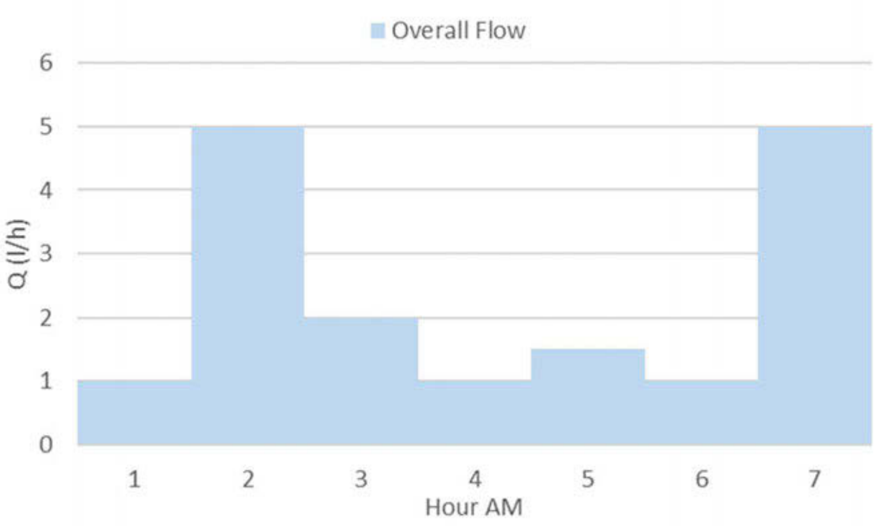


Figure 2. Example of consumption elements.

With the idea of applying a methodology to the entire database, the disaggregating of the consumption starts with evaluating the hourly measurements to identify which record is from intentional use or leakage. Therefore, it is necessary to reflect the steady behavior of leakage on statistic parameters to compare. The Median represents an excellent factor for establishing a range of evaluations. Because the high records have a low influence on the Median value, for that reason, the factor adopts the tendency of the data. The assessment compares the hourly Reading with the 10% increase in Median and, as the second conditional, The Median plus three units. These conditionals offer a range to express a minimum and maximum about the steady flow.

This first step identifies the flow caused by leakage between hours, but separating the leakage part on intentional use flow is the proper procedure to complete the disaggregating. For the second step, the PROFUGA factor was created. This factor is calculated only by the Mean of the hourly leakage consumption separated before, defining the magnitude in the daily record. The value of this parameter is removed from the intentional flow, so it notices the different flow types in every measurement. Figure 3 shows the final results applicated in the example before. This methodology can summarize the total volume of continuous/leakage flow and intentional flow over time.

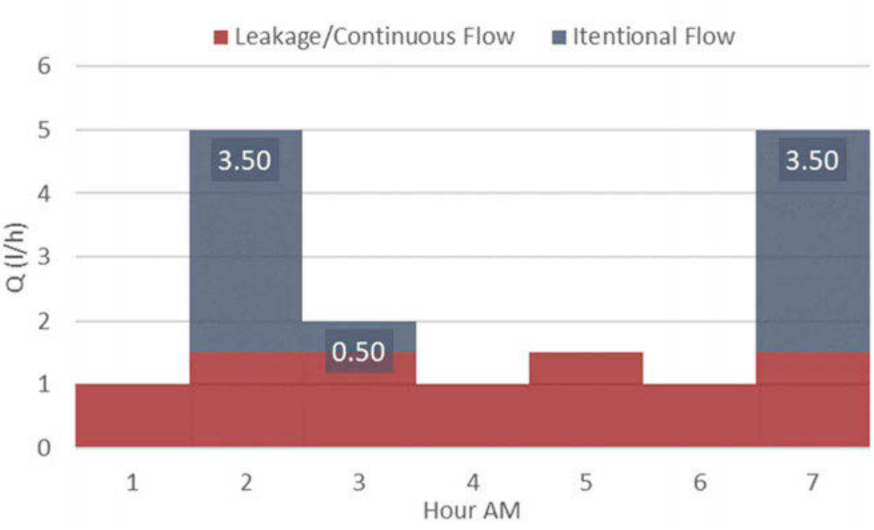


Figure 3. Example of disaggregated consumption.

Last of all, these procedures were applicated in the database by a module of algorithms written in VBA language. As a result, it is easy to see and manage the daily record decompose every hour, and make calculations as an independent data selection. This level of information gives a window of different scenarios and perspectives to manage the measurements. At this point, this work will focus on the analysis as follow in the next chapter.

4 DISCUSSION OF RESULTS

The new database comprises three information blocks: Overall consumption, leakage/continuous flow, and intentional flow. Therefore, each part must be analyzed, starting with Overall consumption. In this study, it is essential to know the tendency of the measurements, how it distributes and if there is atypical behavior.

4.1 Overall Consumption

Figures 4a and 4b show the consumption produced from 1:00 to 2:00 AM and 4:00 to 5:00 AM, respectively. A comparison between these two hours of consumption can show the contrast of values when there are the beginning of nighttime and the moment of minimum consumption. The Y-axis is the volume percentage occupied by each flow rate register compared with the total volume during 2019.

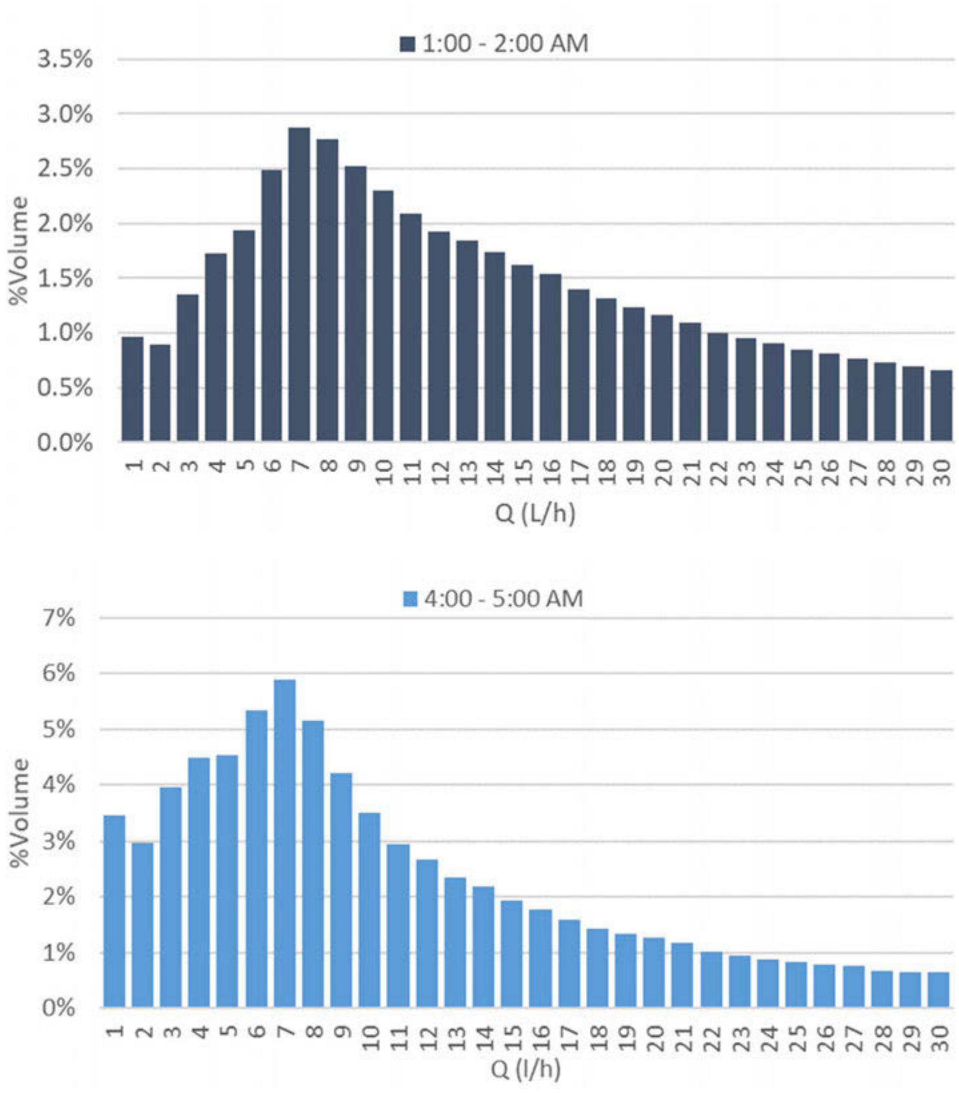


Figure 4. Distribution of flow rates in registers from a) 1:00 - 2:00 AM and b) 4:00 - 5:00 AM

At 4:00 AM, almost 6% of the total volume of his measurements is due to 7 L/h, while at 1:00 AM, only 2.9% of the total volume of his measurements is due to the same flow rate. In both cases, the flow rate with the highest volume percentage was 7 L/h. The data distribution is more concentrated at low flow rates and decreases drastically after 8 L/h. The volume percentage among the high flow rates decreases more steadily at 1:00 AM than at 4:00 AM but occupies a lower level of total volume percentage. In addition, the 4:00 AM histogram has higher volume percent records between 1 - 6 l/h due to the presence of baseline flow. At this hour is considered when it happens the minimum night flow occurs.

The following figures 5a and 5b show the consumption produced from 1:00 to 2:00 AM and from 4:00 to 5:00 AM as a rank. Also, in this figure, The secondary Y-axis is the frequency of occurrence percentage based on the total days registered at that time.

In general, low flow rate occurrence and volume are more frequent. At first seen, the highest range is not the same for both times. The rank between 31 and 70 L/h shows the maximum volume percentage at 1:00 AM, but the concurrency is also deficient. The most Frequency rate flow is in the field of 1 to 2 L/h. However, the lowest values of flow can be part of the rate error of the meters. The one that matches high values of volume and occurrence percentage in both figures is the range between 5 - 7 L/h and 8 - 12 L/h.

cc BY NC SA

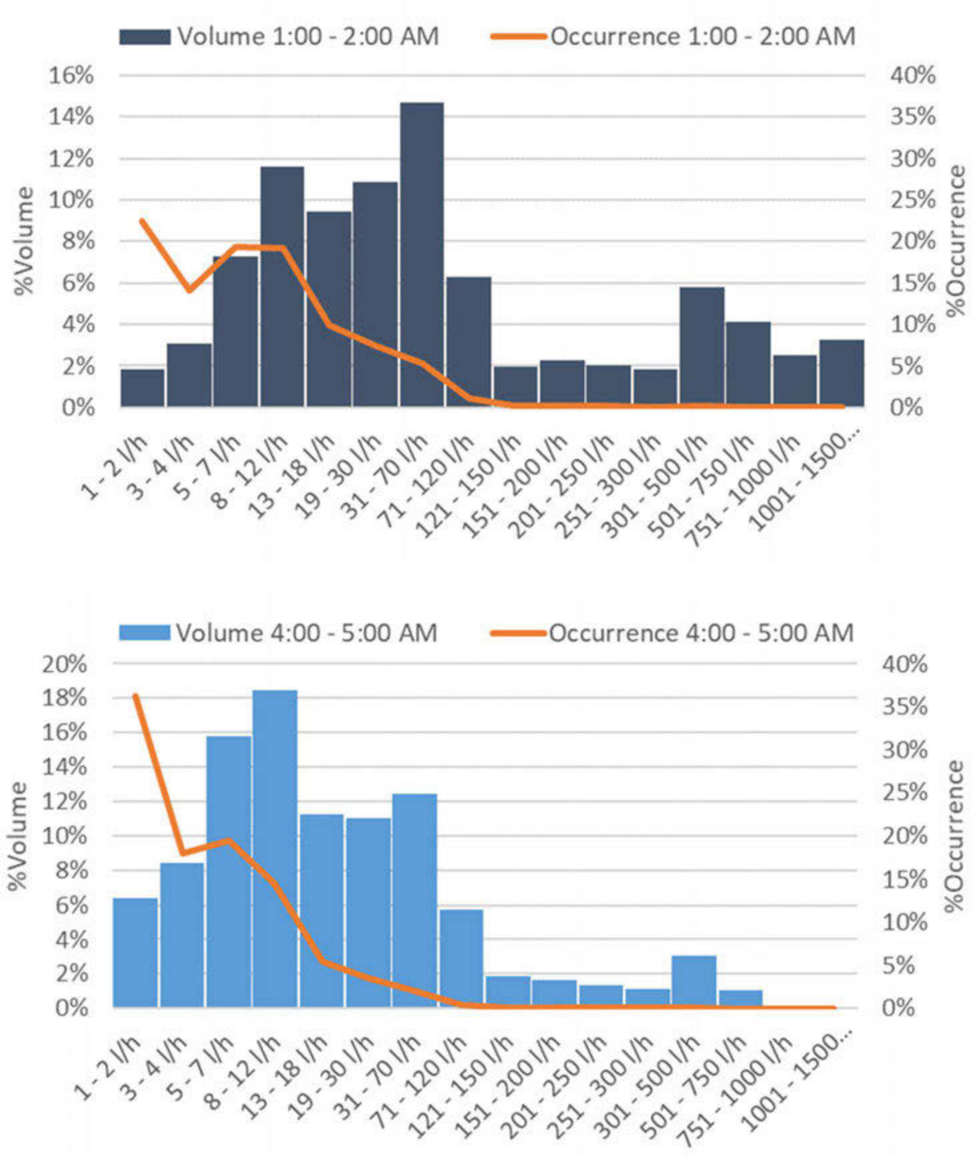


Figure 5. Distribution of flow in volume and occurrence percentages from a) 1:00 – 2:00 AM and b) 4:00 – 5:00 AM

Unifying all the hours over 4:00 AM measurements, as figure 6 shows. It was noticed a tendency in the measures except for the flow rate at 1:00 and 7:00 AM. Their values differ more in Volume and Occurrence than others and are more likely to register higher flow rates, which means they cannot be considered part of the minimum night flow. This figure could be associated with consumption patterns and reflect the influences of the values in every hour.

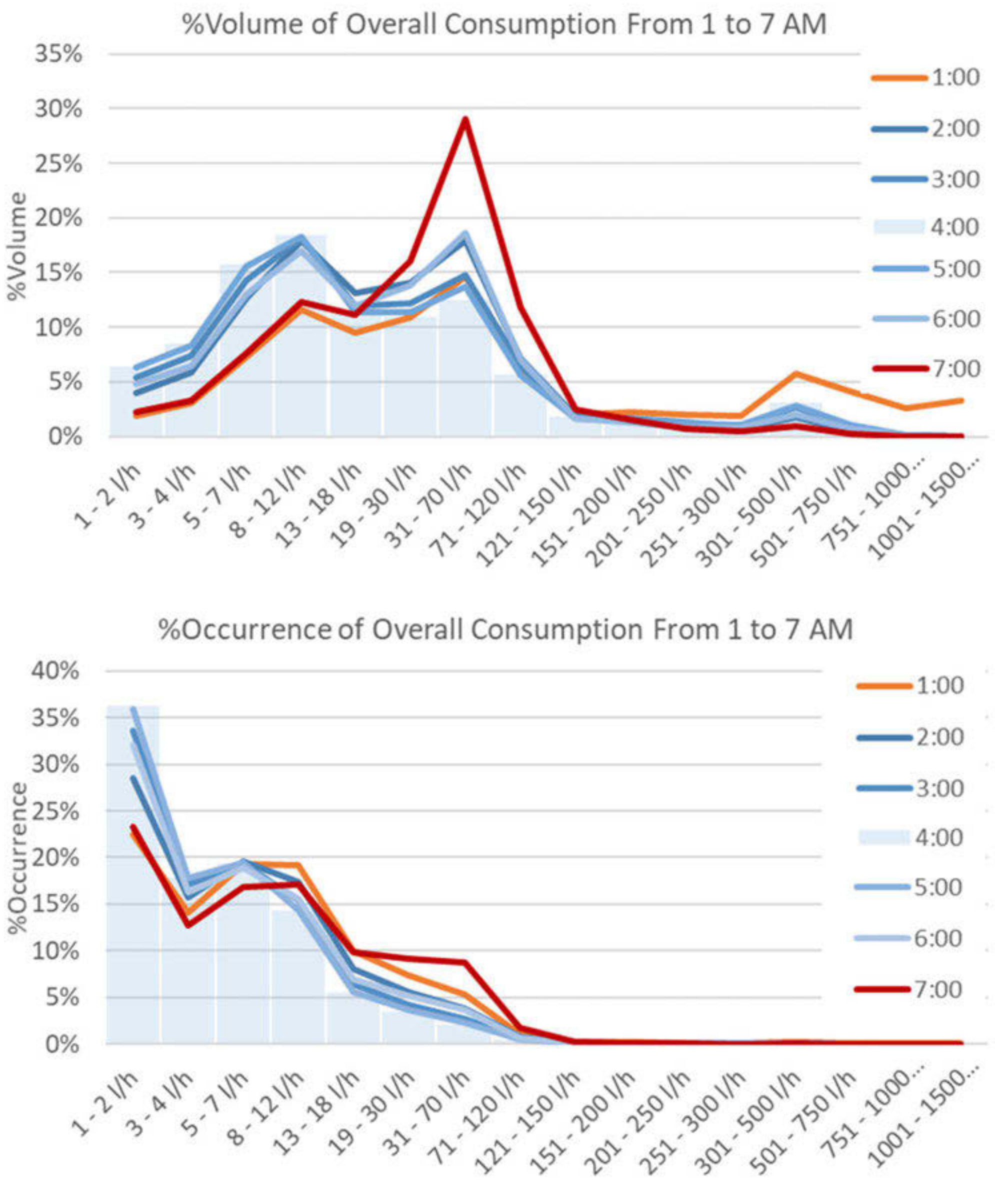


Figure 6. Distribution of Overall Consumption in a) %Volume and b) %Occurrence from 1:00 - 7:00 AM

4.2 Leakage/Continuous Flow

At the Leakage/Continuous flow data block, the flow rate was noticeable more steady. Figure 7 shows the volume and occurrence percentages of all hours compared. During the nighttime, the flow rate between hours is similar in volume and occurrence. The light blue color columns look almost the same, but dark blue color columns increase around lower ranges and then decrease at higher values. The occurrence is overlapped and higher around the lowest values. Approximately 25% to 23% of the time between 1:00 and 7:00 AM, a continuous nominal flow rate of 3 to 4 L/h occurs. The highest volume percentage is from 8 - 12 L/h, but low occurrence, as demonstrated in the figures before. Consequently, the leakage/continuous flow behavior is notoriously the same between 3:00 - 5:00 AM. Choosing any of these hours would fit the estimate of the rate flow at night.

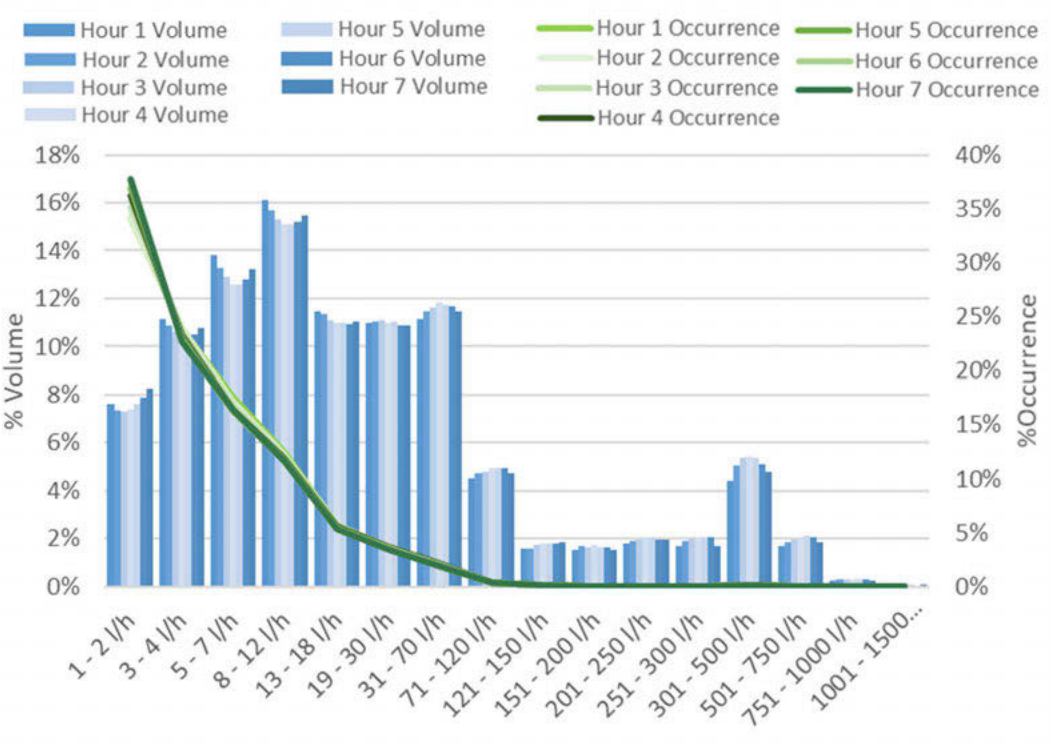


Figure 7. Distribution of Leakage/Continuous Flow by %Volume and %Occurrence from 1:00 - 7:00 AM

4.3 Intentional Flow

Intentional flow shows a solid relation to Overall consumption. Figure 8 indicates rates flow in the percentage of volume and occurrence, matching the highs and lows of Figure 6. Also, the contrast between 1:00 and 7:00 AM to the other hours clears and the tendency around 3:00 - 5:00 AM seems evident.

As a result, it is sure to say that the most common intentional consumption at night times and that occupied the most volume percentage is between 5 - 7 L/h and 8 - 12 L/h, which can be indeed associated with the action of haft-flush and full-flush toilets. Low flows as 1 - 2 L/h can be related to drinking, washing hands actions, or even an error factor of meters. All of this is reflected in Overall consumption as well. The first and last hours suffer a strong influence from other activities. Weekend consumption and morning routines increase the flow rate at these hours and separate from the night flow.

BY NC SA

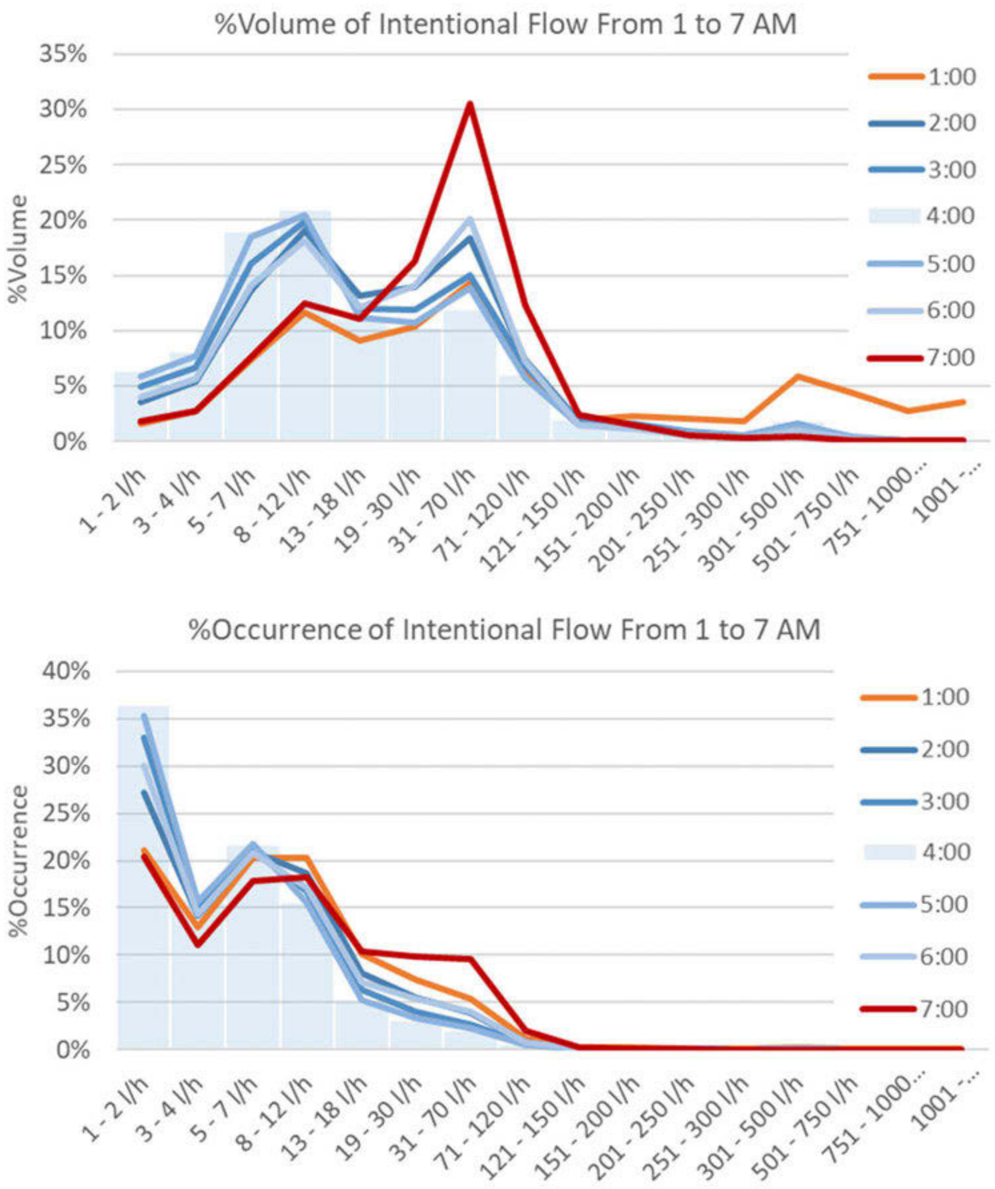


Figure 8. Distribution of Intentional Flow in a) %Volume and b) %Occurrence from 1:00 - 7:00 AM

4.4 Disaggregating Flow Rates

The disaggregating of consumption allows for defining the flow rate in three data blocks, as is shown in table 1. Also, table 2 shows filtered by sector.

Table 1. Flow estimation at Nighttime Minimums

Household Service	L/H/Customer	%TOTAL
Overall Consumption	1.50	100%
Leakage/Continuous Flow	0.53	36%
Intentional Flow	0.97	64%

cc BY NC SA

The Overall consumption at 4:00 AM is composed of 36% Leakage/Continuous flow and 64% Intentional flow. The baseline flow is steady at around 0.53 L/h/customer. This information represents a start point for managing the potential reduction of loss volume. It offers an upgraded and precise magnitude of the volume in the supply system installations and gives a reference to compare with itself and other supply systems. The partition is similar in the five sectors considered in the study, which means it could be the same in other sectors of Mediterranean City. Also, it shows through the method of minimum night flow that the leakage is hidden during the day activities while is highlighted at nighttime.

Table 2. Flow estimation at Nighttime Minimums by sectors

Sector 1	**L/H/Customer**
Overall Consumption	1.65
Leakage/Continuous Flow	0.63
Intentional Flow	1.02

Sector 2	**L/H/Customer**
Overall Consumption	1.55
Leakage/Continuous Flow	0.55
Intentional Flow	1.00

Sector 3	**L/H/Customer**
Overall Consumption	1.47
Leakage/Continuous Flow	0.47
Intentional Flow	1.00

Sector 4	**L/H/Customer**
Overall Consumption	1.40
Leakage/Continuous Flow	0.52
Intentional Flow	0.88

Sector 5	**L/H/Customer**
Overall Consumption	1.46
Leakage/Continuous Flow	0.50
Intentional Flow	0.96

The objective of decomposing consumption into sectors was to analyze the influence of each sector on overall consumption. As a result, flow rate estimation is not way different between sectors. The baseline is evident even with a significant disparity in population and commercial activities.

4.5 Heat Maps

The heat maps show by color the intensity of the average night-hour flow every day in 2019. Figure 9 shows the heat maps for Overall Consumption, Leakage/Continuous Flow, and Intentional Flow. The color blocks are organized by a timeline and flagged by month. In addition, there is the number of customers registered on every date under the month layer to observe the information's weight and the variety of population in time.

The color contrast in the first map reflects changes in customers' consumption during the day, and observing the exact position of those blocks reveals patterns in this method. In general, the lowest values are in light blue blocks around 3:00 - 5:00 AM. Those blocks temporarily mismatch at weekends, showing darker colors at 2:00 - 3:00 AM and lighter at 6:00 AM. Also, values in March, May, and August lose the continuity of the patterns because of changes in the activities. Changes due to Holidays when people enjoy time with family or going out at night and Summer Vacations when people move out of the city for weeks.

Leakages/Continuous Flow present different distributions compared to the map before. The values adopt the form of lines on this map because of the steady behavior, but they change the tone slightly at 1:00 and 7:00 AM. The darkest blocks concentrate more over January, May, September, and December. Those months are related to high amounts of population and, for that reason, demand. Besides, this method can be applied in sectors to identify atypical events and locate the leakage origin.

Intentional Flow is similar to The Overall Consumption map. As seen before, the Intentional Flow is the one that can directly reflect customer consumption. The patterns are easily observed and have the same mismatch at weekends, holidays, and summer vacations but in lower values. The maximum value at the scale is similar while different from the minimum and the 50% quantile, which is lower.

Heat maps can also open a window to analyze the time evolution at a minimum night hour because they compile all the average daily records. In this way, figure 10 shows the consumption at 4:00 AM during 2019 in the three blocks of information. The three parts of the figure have a tendency that adopts a linear function and shows the equation of itself. The tendency gives a magnitude used for future references, showing precisely the flow rate of the leakage/continuous Flow that it wanted to estimate and how Intentional Flow and Overall Consumption overlap their form.

In Figure 11, the tendency equation of Leakage/Continuous Flow has an independent term that establishes at 0.53 L/h/customer while the Intentional Flow function establishes at 0.92 L/h/customer. Besides, this tool can be used in known sectors, relate their consumption to find patterns, and apply them in sectors without remote reading. This leakage flow rate creates a value for performance expectations or to compare the efficiency of other supply systems and even sectors of the same system.

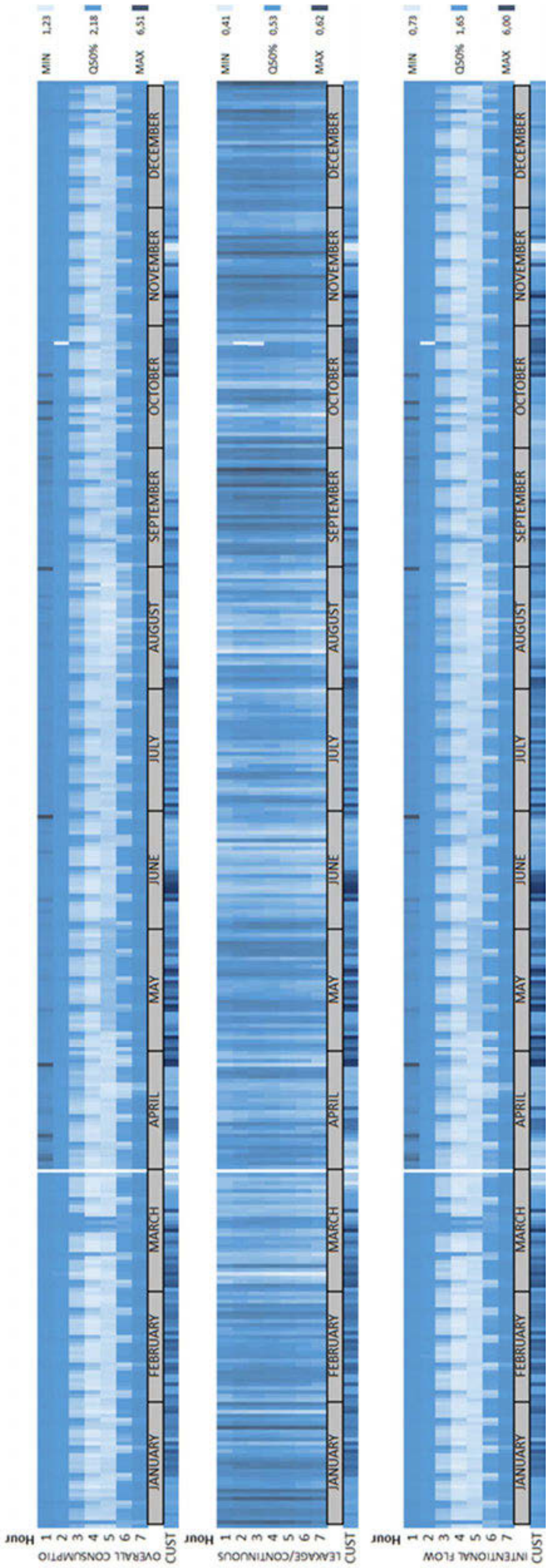


Figure 9. Heat Maps of a) Overall Consumption b) Leakage/Continuous Flow c) Intentional Flow

BY NC SA

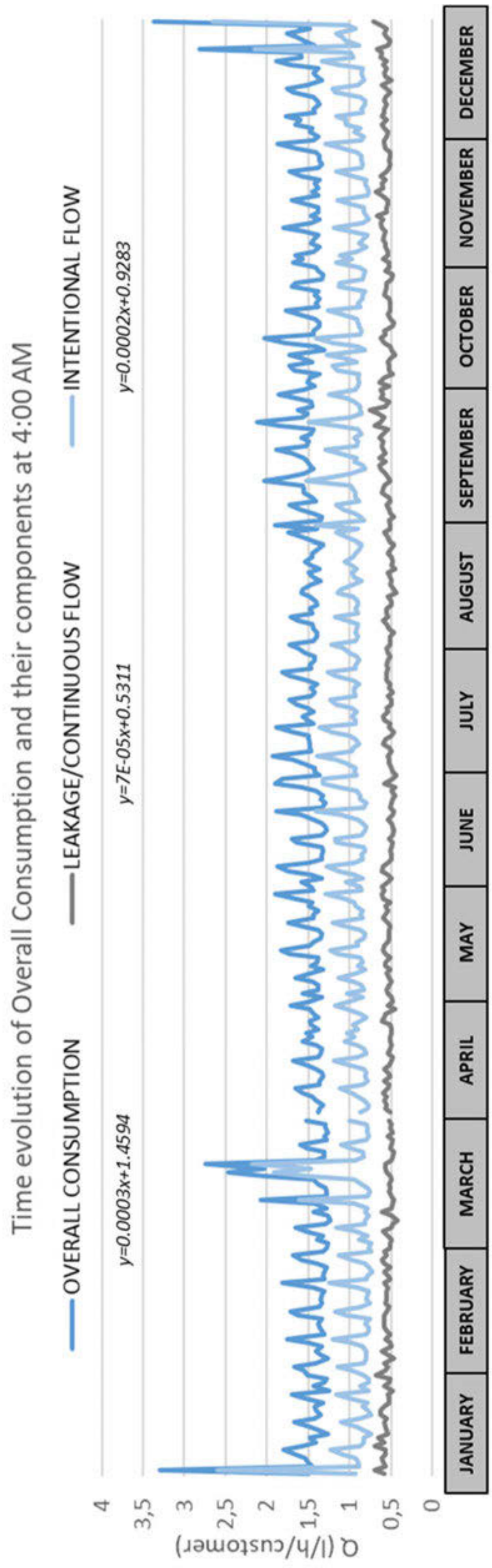


Figure 10. Time evolution of a) Overall Consumption b) Leakage/Continuous Flow c) Intentional Flow

5 CONCLUSIONS

Analyzing consumption patterns is a helpful tool for understanding the demands and maximizing the system's efficiency through management decisions. This study focuses on analyzing water consumption by disaggregation of the overall consumption during night hours of residential customers. These are the conclusions:

The method of disaggregation makes it possible to evaluate each consumption element and analyze the flow rate that handles the demands. Due to the leakage indicator PROFUGA, extracting the Leakage/Continuous Flow portion in the Intentional Flow was possible. Because of that, this estimation can use extrapolated to all-day hours and gives another point of view.

It is a fact that exists a baseline flow of leakage or continuous flow in the supply system. During night hours consumption, this baseline is a flow rate of 0.53 L/h, representing 36% of the overall flow. Analyzing the five sectors, they all keep the portion of elements in the Overall Consumption. The consumption between 3:00 – 5:00 AM offers hours alternatives to estimate this value since the behavior is the same.

The intentional flow is associated with sanitary standards device flow at that hour. It is related because the range of more percentage of volume and occurrence are 5- 7 L/h and 8-12 L/h, which correspond to flow rates of half-flush and full-flush toilets, washing hands, and drinking water at night. During these hours, the distribution is occupied mainly by low flow rates. There are registers of punctual high values but with low occurrence. The consumption at 1:00 and 7:00 AM suffers highly influenced by weekend and morning routines. They are not considerate at the estimation of minimum night-hour flow.

Heat maps provide an easy way to identify patterns and variations over time. Changes in weekends, holidays, and vacation consumption are easy to observe. It also allows us to evaluate the months with the highest baseline flows and intentional spending to take the necessary precautions. The temporal evolution shows the difference in magnitude between the components and moments of high and low consumption values. Which offers a range of demands management.

These conclusions allow applicating in other sectors of the Mediterranean City, as they were in the five sectors mentioned before. The method allows for studying the consumption of the supply systems at level details, which represents a valuable tool to create bonds and parameters for being applicated beyond the network limits. By studying temporaries series and probability functions, this study can be developed at higher levels and make a difference in management operations.

6 REFERENCES

[1] Narmilan, Amarasingam & Puvanitha, Narmilan & Niroash, Gnanachelvam & Sugirtharan, Muthucumaran & Vasssanthini, Ratnarajah. (2020). “Domestic water consumption pattern by urban households”. 10.5194/dwes-2020-32.

[2] Avni, N., Fishbain, B., and Shamir, U. (2015), “Water consumption patterns as a basis for water demand modelling”, Water Resour. Resc., 51, 8165-8181, doi:10.1002/2014WR016662.

[3] Chrysi Laspidou, Elpiniki Papageorgiou, Konstantinos Kokkinos, Sambit Sahu, Arpit Gupta, Leandros Tassiulas."Exploring Patterns in Water Consumption by Clustering". Procedia Engineering. Volume 119. 2015. Pages 1439-1446. ISSN 1877-7058.

[4] Escobar, J., Cabrera, E., Cobacho, R., (2011), “Método de los caudales minimos nocturnos: revisión de sus bases científicas, evaluación de errores potenciales y propuesta para su mejora”. Universitat Politècnica de Valencia.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.15003

DEVELOPMENT OF A TOOL FOR THE OPTIMIZATION AND REGULATION OF HYDRAULIC MICROGENERATION SYSTEMS ADAPTED TO THE DEMAND AND FLOW VARIATIONS AIMED AT THE CLEAN ENERGY RECOVERY IN WATER SUPPLY NETWORKS

Melvin Alfonso Garcia Espinal[1], Modesto Pérez-Sánchez[2], Francisco Javier Sánchez-Romero[3], P. Amparo López-Jiménez[4] and Pilar Conejos Fuertes[5]

[1,2,4,5]Hydraulic and Environmental Engineering Department, Universitat Politècnica de València, Valencia 46022, (Spain)
[3]Rural and Agrifood Engineering Department, Universitat Politècnica de València, Valencia 46022, (Spain)

[1] melgares@posgrado.upv.es, [2] mopesan1@upv.es, [3] fcosanro@agf.upv.es, [4] palopez@gmmf.upv.es; [5]pilar.conejos@idrica.com

Abstract

Incorporating energy recovery systems using renewable energies in water distribution systems is being analysed and implemented to improve sustainability, ensuring the SDG-7. Pumps working as turbines are an innovative technology, and a powerful tool to reach this energy improvement. This research develops a methodology that enables the development of an analysis tool for the optimized regulation of an energy recovery system using Simulink MATLAB.

The said methodology is capable of processing variable operation systems (VOS), which considers the flow variation through time. Also, it considered the variation of the rotational speed of the machines linked directly to the efficiency aiming to maximize the recovered energy.

Finally, the methodology was applied to a real case study defining the signal parameters to the regulation equipment in terms of numbers of pumps working as turbines (PATs) in operation, rotational speed, and the degree of opening of the pressure reducing valves installed.

The tool contemplates the energy analysis, which shows the recovery of 60% of the supplied energy in the system. This daily recovered energy was above 1898 kWh, and it contributed to an increase in efficiency and sustainability.

Keywords
Hydraulic regulation, Electronic regulation, SDG, Pump working as turbine, Water distribution network, Flow variation, sustainability, variable operation system.

1 INTRODUCTION

In recent years, energy recovery systems have been proposed to lessen the use of non-renewable energies. [1]. At the same time, microgeneration systems are the most frequently demanded[2]. The addition of these new technologies leads towards new steps for further sustainable development of the society and more conscious use of the accessible resources. As a consequence of the population growth over the years, the energy cost and consumption involving the water cycle have increased, for that reason, the efficient management of the water distribution networks has been a priority [3], [4].

Numerous studies show that sustainable indexes can be improved by using microgeneration systems. Its implementation benefits decreasing pressures in the network; this decrease ensures an upturn of the leakage indexes in the supply. [5], [6]. Also, renewable energy is increased, and the optimization of the energy generation [7].

Thus, the use of reverse mode pumps, the so-called pumps working as a turbine (PAT), is a practical implementation that has been considered over recent years [8]. The application of microhydropower in combination with other renewable systems established, such as solar or wind systems, can improve the use and production of clean energy in the water cycle[9].

Various authors have carried out PATs analyse since the 1940s, including Stepanoff [10], Childs [11], Williams [12] and Pérez-Sánchez et al. [13], who defined new expressions for the estimation of the best efficient point (BEP) of a PAT and its characteristic curve from a database with over 181 different pumps. While it is true that with the use of PATs, it is possible to recover energy, regulatory strategies are needed to maximize the use of these benefits.

Hydraulic Regulation (HR) is carried out through a hydraulic circuit with elements series and in parallel and the Electrical Regulation (ER) is by means of controlling equipment such as frequency variator drives and inverters are the most common solutions in systems where PATs are used for energy production [14]. The generated power can be supplied directly to the grid or self-consumed.

Optimization tools must support the regulation improvements in the water network. The quality of service to the user needs to be a priority and water managers have to know the main constraints of the system over time such as flow, and upstream and downstream pressure to guarantee it [15].

The need arises to provide the required pressure in a variable operating regime. Variable speed operation is an approach for controlling the discharge at the pump as a turbine inlet aiming at increasing operational efficiency. For example variations in pressure and flow rate[16], [17]

This manuscript develops an optimisation modeln Simulink MATLAB[18]. This tool will represent an upgrade in the PAT management in the water system. Regarding the issue related to PATs, their selection and simulation in different scenarios, there is no software or tool previously developed for such purposes, for which emphasis is placed on the creation of a program.

The use of empirical methods to estimate the characteristic curves of the PATS is part of the innovation in order to optimize the operation in the system is part of the innovation of this tool.

2 METHODS

In this section, the methodology in which is based the tool is proposed and defined. The analysis, optimization and regulation of energy recovery is the main objective. First, it is essential to have the system constraints (Flow and upstream and downstream pressures) that will define the operational limitations.

The main field of application, but not limited to, is water distribution supplies, where the demand changes over time, hence is a variable operation system (VOS). Regulation strategies should be designed to improve the efficiency and sustainability of the supply and at the same time, never compromise the level of service for the population.

Figure 1 represents the basic layout for the hydraulic model or an energy recovery system. The model is composed of a SCADA or Flowmeter which is used to generate the input data, recording the demanded flow and pressure setpoints. It is followed by the energy recovery group, for this case, the tool and methodology only consider PATs. Two pressure-reducing valves (PRV), one is installed in parallel to the PAT group and the second one is installed in series at the outlet of the system. Last, there are control valves installed which are used for sectioning, protection of the infrastructure.

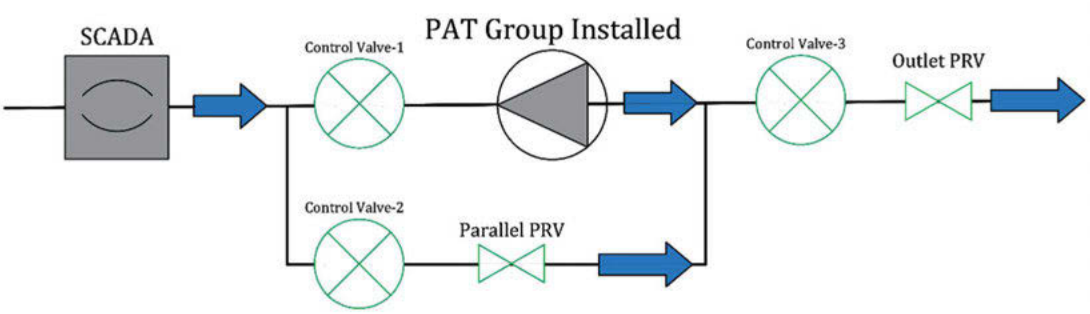


Figure 1. Hydraulic model layout.

The tool outputs the signals needed for the regulation strategies: Number of PATs working, Rotational speed and the Degree of opening of both PRV installed in the system. It is important to note that this tool have some limitations that should be addressed. PAT group should be selected beforehand, limited to simulate only 3 PAT installed in parallel, and all of the parameters should be introduced before running the simulation.

2.1 Methodology

In Figure 2, the developed methodology is shown. This is made up of three blocks: A. Model setup, B. Model simulation, and C. Model out put. Every block is composed of different sections, and those sections as well into steps that contain the different actions to be executed through the model simulation.

In the first block *(A. Model setup)* the main objective is to prepare the model prior to its simulation. All the input data should be loaded in this block for the tool to work correctly. This block has two main steps:

Step I is where all the recorded data is extracted from the database and then processed, previous being inputted into the model into a compatible format for the tool. The three main constraints that must be known are demanded flow (Qd), the upstream pressure (Pu) and downstream pressure (Pd) variation over time.

This tool can work both as an optimization tool for existing networks with energy recovery installed or as an analysis tool for proposing the implementation of energy recovery systems where the conditions allow it to. Because of that, after step I there is a decision-making block were depending on the conditions of the supply, the next step to follow will be conditioned.

If there is an existing PAT group installed, the next step is *Step II,* where all the initial configurations should be typed in the MATLAB console. Parameters are divided into six different categories according to the function they represent: 1. Parameters of the PAT group, 2. Variable-frequency drive settings, 3. Electric generator efficiency details, 4. Maximum power determination thresholds, 5. Pressure reducing valve characteristics and 6. Fluid properties.

Otherwise, if no energy recovery system is installed, *Step II.I* should be completed first. Parting from the input data, the selection process for the PAT system is made following the methodology proposed by Camilo Rosado et al. [19]. After the PAT group is selected, advance to Step II.

BY NC SA

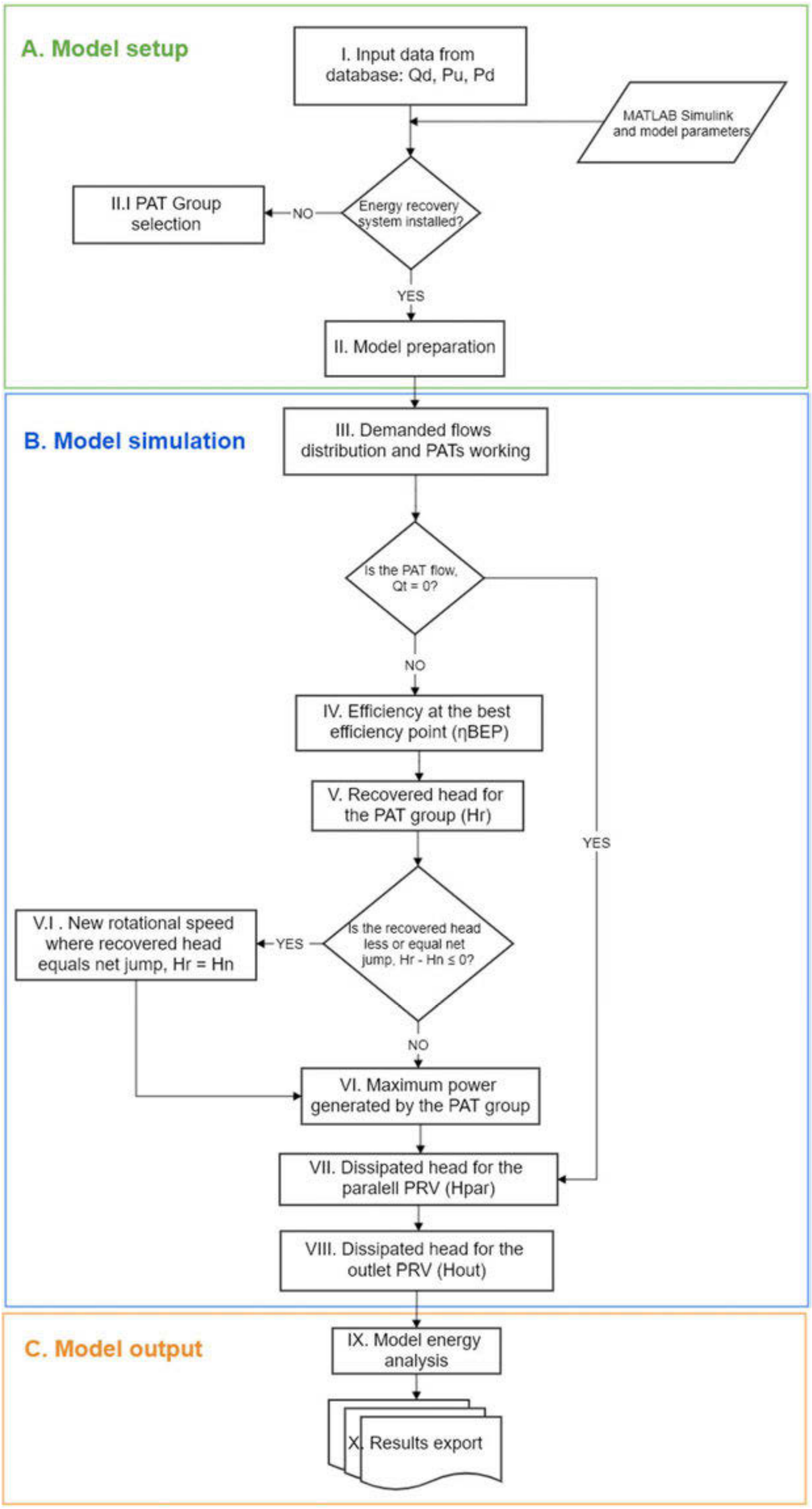

A. Model setup
I. Input data from database: Qd, Pu, Pd
MATLAB Simulink and model parameters
Energy recovery system installed?
NO
II.I PAT Group selection
YES
II. Model preparation
B. Model simulation
III. Demanded flows distribution and PATs working
Is the PAT flow, Qt = 0?
NO
YES
IV. Efficiency at the best efficiency point (ηBEP)
V. Recovered head for the PAT group (Hr)
Is the recovered head less or equal net jump, Hr - Hn ≤ 0?
YES
V.I . New rotational speed where recovered head equals net jump, Hr = Hn
NO
VI. Maximum power generated by the PAT group
VII. Dissipated head for the paralell PRV (Hpar)
VIII. Dissipated head for the outlet PRV (Hout)
C. Model output
IX. Model energy analysis
X. Results export

Figure 2. Proposed methodology for the analysis of the regulation in an energy recovery system.

The next block *(B. Model simulation)* comprehends the hydraulic model code, in which calculations for the PAT system and pressure reducing valves are carried out. It covers the central part of the methodology and is from *Step III* thru *Step VIII.*

The main goal is to optimize the recovered energy in order to increase the generated power with the given input data while fulfilling the population demand corresponding to flow and pressure.

Step III is where the flow distribution for the PATs and PRVs at any given moment is done. The outputs of this step comprehends the number of machines working (nf), as well as the flow turbinated for each one (Qt). Note that this tool layout is restricted to a maximum of three parallel PATs installed, for any instance, the flow will be the same for all the PATs in operation at that moment. In order to determine how many PATs should work based on the flow, the strategies proposed by [20] were used.

On the other hand, after this step another decision-making block is set, where based on the turbinated flow, methodology can go into one of two paths. If the flow rate is zero, advances to *Step VII.*

Step IV involves a iterative proccess where the goal is to appoximate to the maximum efficiency capable given the conditions at that time. Operating at the best efficient point (BEP) will be the ideal scenario, if not possible the closest is selected. To estimate the BEP, dimensionless curves are used, based on the research made by Plua et al. [21], represented in the equations:

$$h = -0.31070 \left(\alpha \frac{Qt}{Q_{BEPt}}\right) + 0.1958 \left(\frac{Qt}{Q_{BEPt}}\right)^2 - 0.0118 \left(\frac{Qt}{Q_{BEPt}}\right) - 0.06429\alpha^2 + 1.8489\alpha - 0.2241 \tag{1}$$

$$e = 0.8271 \left(\alpha \frac{Q_t}{Q_{BEPt}}\right) - 0.3187 \left(\frac{Qt}{Q_{BEPt}}\right)^2 - 0.1758 \left(\frac{Qt}{Q_{BEPt}}\right) - 1.035\alpha^2 + 1.1815\alpha - 0.5019 \tag{2}$$

Next, in *Step V,* the recovered head for the PATs is determined based on the rotational speed and variable-frequency drive setpoint from the previous step. Before advancing to step VI, the difference between the net jump available (Hn) and the recovered head (Hr) has to be verified to match the criteria of Hn - Hr ≤ 0. After meeting this condition, advances to Step VII.

On the contrary, if the recovered head is larger than the available jump, a new rotational speed is determined to comply with the established criteria. With a new recovered head value, proceeds to Step VI.

Step VI is the last step involving PAT calculations. The goal is to maximize the generated power by varying the α value into a limited bracket where the tool searches for the best fitting value within the range. After completing this step, a new α value is set for the PAT, and therefore, new recovered head and efficiency values and the calculations that rely to this are obtained.

Step VII and *Step VIII* are similar steps, the dissipated head and degree of opening by both the parallel and outlet PRV is calculated, respectively. This to ensure the regulation strategies in which downstream pressure setpoint is met.

The last block is *(C. Model output),* is where all the energy analysis and results are carried out. *Step IX* is where all the results are organized, classified, and then processed. And in *Step X*, the results are exported in the format of graphs and tables; this part also generates the regulation strategies table that will guarantee an increase in the recovered energy.

3 RESULTS

3.1 Case study description

This case study was implemented in a study point of the high-pressure water distribution network of the Valencia Metropolitan System (VMS) (Spain). At this point, a PRV is installed to maintain the pressure requirements downstream. The installation of energy recovery systems was contemplated, increasing sustainability and efficiency indexes.

Figure 3 represents the layout of this scenario, where point A from the water distribution network and defined as the upstream pressure point (Pu). The net available hydraulic jump is represented (Hn) and the symbol for the energy recovery system to be installed. Demanded flow is Qd, and the population demand point is when the downstream pressure (Pd) is met.

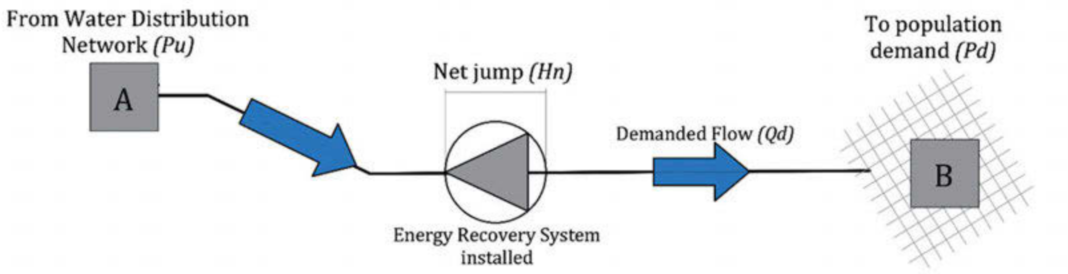

From Water Distribution Network *(Pu)*
A
Net jump *(Hn)*
Demanded Flow *(Qd)*
To population demand *(Pd)*
B
Energy Recovery System installed

Figure 3. The layout of the case study scenario.

The data is a temporal series with records of the demanded flow and upstream and downstream pressures. The data is processed for an entire day with a 10 minute time interval for each reading. Table 1 shows a fragment of the input data structure in which all the primary constraints are defined at any given time. Also, statistics calculations were performed for analyzing the data prior to selecting the PAT group and running the simulation.

Table 1. Input Data fragment and statistics calculations

Time (Δt = 10min)	Qd (L /s)	Pu (m w.c.)	Pd (m w.c.)
2:50:00	403.76	77.94	39.74
3:00:00	385.22	78.64	38.60
3:10:00	291.79	80.89	34.38
3:20:00	183.57	83.09	31.02
3:30:00	80.94	83.98	29.21
3:40:00	56.57	83.99	29.00
Maximum Value	406.16	84.90	39.74
Minimum Value	49.79	75.58	25.08
Mean Value	302.67	79.38	32.84
Standard Deviation	149.68	3.24	4.11

The PAT group selected consists of three pumps of Bombas IDEAL, model CPH 80-210 (215 mm), with a nominal rotational speed of 2900 rpm at 50 Hz. The second part of the proposed energy recovery system is the selection of the PRV. For the valve installed in parallel, the model selected is a BELGICAST Hydroblock DN250, and for the valve connected in series is a BELGICAST Hydroblock DN350. This is complemented with five control valves for management and maintenance tasks. In Figure 4, the previously described proposed regulation layout is shown with all of its components.

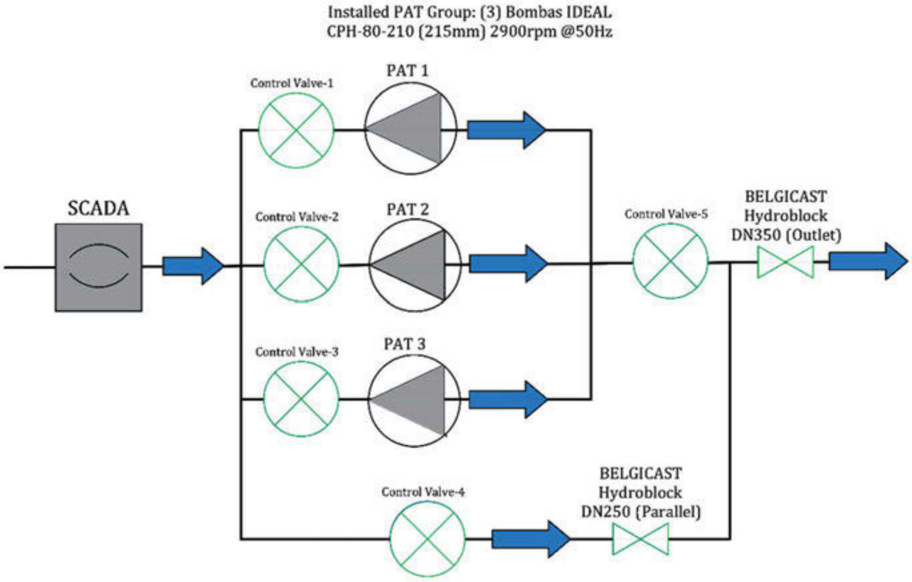


Figure 4. Proposed regulation layout

In Figure 5a is represented a turbinated flow by one PAT during the simulation. All PATs are the same model and are connected in parallel. Therefore, the PAT flow will be the same for all the working machines at that given time. The PAT operation interval was between 49 and 150 L/ s, and a mean value of 114 L/ s. On the other hand, Figure 5b shows the evolution of the number of PATs working at each time. At every moment during the simulation, at least one PAT was operating, and in the majority of the instances, three machines were used.

Figure 5c represents the temporal evolution of the heads and pressures during the simulation. Where is shown the upstream pressure (Pu), downstream pressure (Pd) and the net hydraulic jump (Hn), which is obtained by the difference between upstream pressure and downstream pressure. The recovered head for the PATs (Hr) was within the range of 28 and 54 m w.c. and a mean value of 40 m w.c. The difference between the upstream pressure and the recovered head is labelled as (Pu-Hr); the recovered head was never superior to the available net jump in any instance.

Similarly, Figure 5d shows the evolution of the dissipated heads for both pressures reducing valves. (Hpar) is for the installed in parallel and (Hout) represents the valve installed in series. During the lower demand period, the outlet PRV had a max dissipated head at 29 m w.c. The parallel valve did not work during those times because the demanded flow was above the minimum working flow for the PAT.

Figure 5e shows the graph for the accumulated and dissipated energies during the simulation. The total available accumulated energy was 3113 kWh, and the total recovered energy for the PAT group was 1898 kWh, representing 60% of the available energy and 90% of the total recoverable energy.

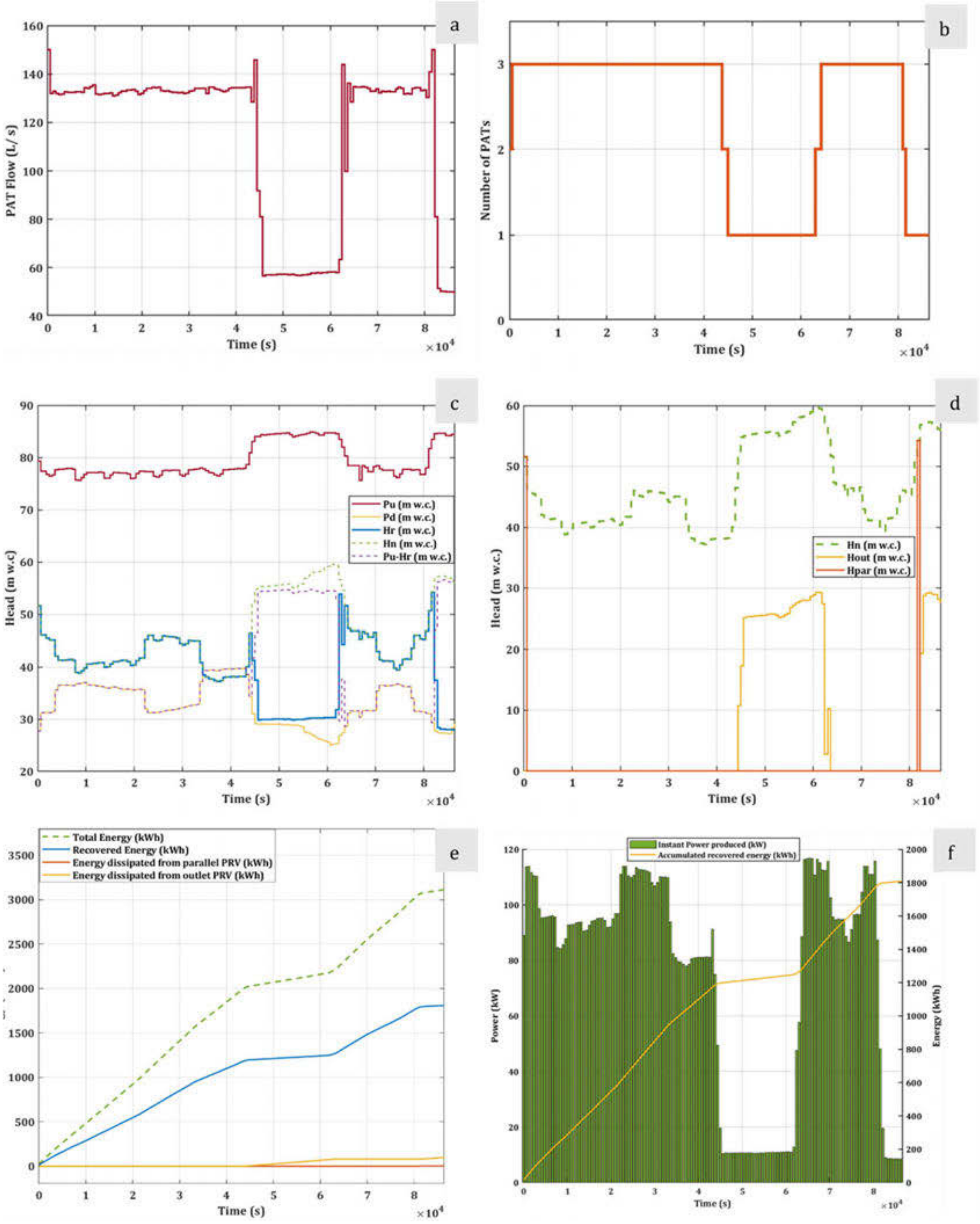


Figure 5. (a) Flow over time, (b) Number of PATs working, (c) Heads and pressures over time, (d) Dissipated heads by PRV; (e) Accumulated and dissipated energies over time; (f) Instant power produced and accumulated recovered energy.

Figure 5f shows the representation of the instant power produced for the PAT (in kW) during the simulation and the evolution of the accumulated recovered energy (kWh). Instant power oscillated between 10 and 114 kW, and average instant power of 50kW. The low demand period is where instant power is the least.

CC BY NC SA

The second section of the model results is the regulation strategies, in which both hydraulic and electronic regulation for the case study are defined in Table 2. This table is composed of five columns: 1. Time, expressed in seconds parting from 0 to 86400 in this case study, 2. Number of PAT working, 3. PAT α value, 4. Degree of opening for the parallel valve and 5. Degree of opening for the outlet valve. These setpoints ensured the maximum recovered energy at any moment during the simulation.

Table 2. Regulation strategies from 43200s to 45600s

Time (s)	Number of PAT working	PAT α value (%)	Parallel PRV Degree of opening (%)	Outlet PRV Degree of opening (%)
43200	3	0.918	0	100
43800	2	0.959	0	100
44400	2	1.001	0	28
45000	1	1.045	0	13
45600	1	0.920	0	8

The variable-frequency drive values were between 0.816 and 1.045 for the minimum and maximum values, respectively, and the average value of 0.948; this supports the PAT selection wherein most of the time during the simulation PATs worked near their nominal rotational speed.

The pressure reducing valve operation was modified to optimize the water supply. For the installed in parallel, this was closed most of the time, except for one instance. On the other hand, the outlet valve was fully open most of the time and with the lowest degree of opening of 70% during the lower demand period. With the valves, the regulation strategies are completed.

4 CONCLUSIONS

Due to the lack of tools for simulation and optimization for energy recovery systems and the need to constantly seek improvements to increase the sustainability and efficiency indexes in water supply networks, the need to create this tool arises. In this manuscript, a methodology for an optimization model developed in Simulink MATLAB is presented as well as its implementation in a case study.

The optimization strategy defined in this methodology aims to maximize the energy recovered. In the same way, the parameters that dictate the behavior of the system based on the number of PAT working, rotation speed and degree of opening of the PRV are obtained.

The tool developed, operating with the methodology outlined in this work, serves both to analyze an existing installation and seek improvement solutions and in cases where it is necessary to evaluate the proposal of an energy recovery system. The innovation of this study is the incorporation of empirical methods, where characteristic curves can be approximated.

However, the limitations of the model must be taken into account: 1. If there is no energy recovery system, it must be selected first, 2. The largest number of PATs allowed is 3, and they must be installed in parallel 3. All parameters must be entered prior to running the simulation.

This optimization tool implemented in a network point of the high-pressure water distribution network of the Valencia Metropolitan System (Spain), and as a result, improving the recovered energy. Moreover, it can be used in any water distribution system as long as the variables required for the correct model simulation are available.

This is a field of research of great interest for sustainable development since it shows that it is possible to increase the energy recovered in a water distribution network and, at the same time, improve its hydraulic and energy efficiency by regulating the water resources in a more concious way.

Table 3. Abbreviations

Parameter	Definition	Units
Qd	Demanded flow.	L/ s
Pu	Upstream pressure.	m w.c.
Pd	Downstream pressure.	m w.c.
Q_{BEPt}	Flow at the best efficiency point (BEP) of the machine in turbine mode.	L/ s
H_{BEPt}	Recovered head at the best efficiency point (BEP) of the machine in turbine mode.	m w.c.
η_{BEPt}	Efficiency at the best efficiency point (BEP) of the machines in turbine mode.	%
ns_t	Specific number in turbine mode.	m, kW
Q_{BEPp}	Flow at the best efficiency point (BEP) of the machine in pump mode.	L/ s
H_{BEPp}	Head at the best efficiency point (BEP) of the machine in pump mode.	m w.c.
η_{BEP}	Efficiency at the best efficiency point (BEP) of the machines in pump mode.	%
Q_{mint}	Minimum operating flow for the PAT.	L/ s
Q_{maxt}	Maximum operating flow for the PAT.	L/ s
N_o	Nominal rotational speed of the PAT.	rpm
n	Rotational speed of the PAT at a given moment.	rpm
α	Frequency inverter value setpoint for the rotational speed.	%
α_{min}	Minimum frequency inverter value setpoint.	%
α_{max}	Maximum frequency inverter value setpoint.	%
lb	Lower bound of the range for determining the maximum power.	%
ub	Upper bound of the range for determining the maximum power.	%
Diam	Pressure-reducing valve (PRV) diameter.	mm
Kvo	Flow coefficient for the valve while fully open.	$m^3/h/\sqrt{Pa}$
g	Gravity acceleration.	in m/s^2

Qt	Flow rate turbinated by the PAT.	L/ s
h	Ratio between recovered head and recovered head at the BEP in turbine mode.	Dimensionless
e	Ratio between efficiency and the efficiency at the BEP in turbine mode.	Dimensionless
Hr	Head recovered by the PAT	m w.c.
ηt	Efficiency of the PAT	%
Hn	Net available head. Difference between upstream and downstream pressures.	m w.c.
Hpar	Dissipated head for the parallel PRV	m w.c.
Hout	Dissipated head for the outlet PRV	m w.c.
MP	Mechanical power generated by the PAT.	kW
EP	Electrical power generated by the generator.	kW
η_{elec}	Efficiency of the electrical generator.	%

5 REFERENCES

[1] L. Corcoran, P. Coughlan, and A. McNabola, “Energy recovery potential using micro hydropower in water supply networks in the UK and Ireland,” Water Supply, vol. 13, no. 2, pp. 552–560, Mar. 2013, doi: 10.2166/ws.2013.050.

[2] L. Corcoran, A. McNabola, and P. Coughlan, “Optimization of Water Distribution Networks for Combined Hydropower Energy Recovery and Leakage Reduction,” Journal of Water Resources Planning and Management, vol. 142, no. 2, p. 4015045, Feb. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000566.

[3] A. del Borghi, L. Moreschi, and M. Gallo, “Circular economy approach to reduce water–energy–food nexus,” Current Opinion in Environmental Science & Health, vol. 13, pp. 23–28, Feb. 2020, doi: 10.1016/j.coesh.2019.10.002.

[4] Kanakoudis V, Tsitsifli S, Samaras P, Zouboulis A, and Demetriou G, “(PDF) Developing appropriate Indicators for urban water supply systems evaluation across the Mediterranean.” https://www.researchgate.net/publication/258286246_Developing_appropriate_Indicators_for_urban_water_supply_systems_evaluation_across_the_Mediterranean (accessed Feb. 21, 2022).

[5] M. Morani, A. Carravetta, G. del Giudice, A. McNabola, and O. Fecarotta, “A Comparison of Energy Recovery by PATs against Direct Variable Speed Pumping in Water Distribution Networks,” Fluids, vol. 3, no. 2, p. 41, Jun. 2018, doi: 10.3390/fluids3020041.

[6] T. Grubic, L. Varga, Y. Hu, and A. Tewari, “Micro-generation technologies and consumption of resources: A complex systems’ exploration,” Journal of Cleaner Production, vol. 247, Feb. 2020, doi: 10.1016/J.JCLEPRO.2019.119091.

[7] T. I. Reigstad and K. Uhlen, “Optimized Control of Variable Speed Hydropower for Provision of Fast Frequency Reserves,” Electric Power Systems Research, vol. 189, p. 106668, Dec. 2020, doi: 10.1016/j.epsr.2020.106668.

[8] M. Tahani, A. Kandi, M. Moghimi, and S. D. Houreh, “Rotational speed variation assessment of centrifugal pump-as-turbine as an energy utilization device under water distribution network condition,” Energy, vol. 213, p. 118502, Dec. 2020, doi: 10.1016/j.energy.2020.118502.

[9] A. Alhejji et al., “Energy Harvesting and Water Saving in Arid Regions via Solar PV Accommodation in Irrigation Canals,” Energies 2021, Vol. 14, Page 2620, vol. 14, no. 9, p. 2620, May 2021, doi: 10.3390/EN14092620.

[10] A. J. Stepanoff, Centrifugal and Axial Flow Pumps: Theory, Design, and Application. 1957.

[11] S. Childs, "Convert pumps to turbines and recover HP," Hydro Carbon Proces- Sing Pet, vol. 41, 1962.

[12] A. A. Williams, "The Turbine Performance of Centrifugal Pumps: A Comparison of Prediction Methods," Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, vol. 208, no. 1, pp. 59–66, Feb. 1994, doi: 10.1243/PIME_PROC_1994_208_009_02.

[13] M. Pérez-Sánchez, P. A. López-Jiménez, and H. M. Ramos, "Modified Affinity Laws in Hydraulic Machines towards the Best Efficiency Line," Water Resources Management, vol. 32, no. 3, pp. 829–844, Feb. 2018, doi: 10.1007/s11269-017-1841-0.

[14] A. Carravetta, "Energy Recovery in Water Systems by PATs: A Comparisons among the Different Installation Schemes," Procedia Engineering, vol. 70, pp. 275–284, Jan. 2014, doi: 10.1016/j.proeng.2014.02.031.

[15] M. Crespo Chacón, J. Rodríguez Díaz, J. García Morillo, and A. McNabola, "Pump-as-Turbine Selection Methodology for Energy Recovery in Irrigation Networks: Minimising the Payback Period," Water (Basel), vol. 11, no. 1, p. 149, Jan. 2019, doi: 10.3390/w11010149.

[16] A. Carravetta, G. del Giudice, O. Fecarotta, and H. Ramos, "PAT Design Strategy for Energy Recovery in Water Distribution Networks by Electrical Regulation," Energies (Basel), vol. 6, no. 1, pp. 411–424, Jan. 2013, doi: 10.3390/en6010411.

[17] J. Delgado, J. P. Ferreira, D. I. C. Covas, and F. Avellan, "Variable speed operation of centrifugal pumps running as turbines. Experimental investigation," Renewable Energy, vol. 142, pp. 437–450, Nov. 2019, doi: 10.1016/j.renene.2019.04.067.

[18] Inc. The MathWorks, "MATLAB - El lenguaje del cálculo técnico." 2021. [Online]. Available: https://es.mathworks.com/products/matlab.html

[19] L. Camilo Rosado, P. López-Jiménez, F.-J. Sánchez-Romero, P. Conejos Fuertes, and M. Pérez-Sánchez, "Applied Strategy to Characterize the Energy Improvement Using PATs in a Water Supply System," Water (Basel), vol. 12, no. 6, p. 1818, Jun. 2020, doi: 10.3390/w12061818.

[20] M. Pérez-Sánchez, F. J. Sánchez-Romero, H. M. Ramos, and P. A. López-Jiménez, "Improved Planning of Energy Recovery in Water Systems Using a New Analytic Approach to PAT Performance Curves," Water 2020, Vol. 12, Page 468, vol. 12, no. 2, p. 468, Feb. 2020, doi: 10.3390/W12020468.

[21] F. A. Plua, F.-J. Sánchez-Romero, V. Hidalgo, P. A. López-Jiménez, and M. Pérez-Sánchez, "New Expressions to Apply the Variation Operation Strategy in Engineering Tools Using Pumps Working as Turbines," Mathematics, vol. 9, no. 8, p. 860, Apr. 2021, doi: 10.3390/math9080860.

[22] D. Novara and A. McNabola, "A model for the extrapolation of the characteristic curves of Pumps as Turbines from a datum Best Efficiency Point," Energy Conversion and Management, vol. 174, pp. 1–7, Oct. 2018, doi: 10.1016/j.enconman.2018.07.091.

[23] A. McNabola et al., "Energy recovery in the water industry using micro-hydropower: an opportunity to improve sustainability," Water Policy, vol. 16, no. 1, pp. 168–183, Feb. 2014, doi: 10.2166/WP.2013.164.

[24] M. Pérez-Sánchez, P. A. López-Jiménez, and H. M. Ramos, "PATs Operating in Water Networks under Unsteady Flow Conditions: Control Valve Manoeuvre and Overspeed Effect," Water 2018, Vol. 10, Page 529, vol. 10, no. 4, p. 529, Apr. 2018, doi: 10.3390/W10040529.

[25] B. Capelo, M. Pérez-Sánchez, J. F. P. Fernandes, H. M. Ramos, P. A. López-Jiménez, and P. J. C. Branco, "Electrical behaviour of the pump working as turbine in off grid operation," Applied Energy, vol. 208, pp. 302–311, Dec. 2017, doi: 10.1016/j.apenergy.2017.10.039.

2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.15011

POTENTIAL REDUCTION OF LEAKAGE VOLUME BY COMBINING DYNAMIC PRESSURE MANAGEMENT AND ENERGY RECOVERY IN VALENCIA

Alejandra Guarachi Quiñones[1], F. Javier Martínez-Solano[2],
Carmen Sánchez Briones[3] and Andross Pérez Lleó[4]

[1, 2] Hydraulic and Environmental Engineering Department, Universitat Politècnica de València, Valencia (Spain)
[3, 4] Global Omnium, Valencia (Spain)

[1]aguaqui@posgrado.upv.es, [2] jmsolano@upv.es, [3]csbriones@emivasa.es, [4]anpelle@globalomnium.com

Abstract

The main objective of a drinking water distribution network is to supply all consumption points with the quantity of water demanded under sufficient pressure and quality conditions. Currently, pressure regulation in water supply networks consists of managing pressure of district metered areas (DMAs) to ensure a sufficient supply to users, being traditionally carried out by reducing excessive pressures, and therefore diminishing water losses. The most basic form of regulation is to maintain a constant pressure at the inlet of the DMAs by means of a pressure reducing valve (PRV). Another turn on the screw in regulation is dynamic sectorization, modifying the set points of the PRVs according to the time of the day to reduce pressure during off-peak hours. This modification of the set point in the PRV involves the movement of a mechanism, requiring a constant energy source.

This work proposes the incorporation of an energy recovery system in the distribution network by installing small water turbines at the inlet of the DMA in order to obtain the necessary energy to feed the dynamic regulation equipment, coming the energy required from the network itself. The case study proposed is the water distribution system of the city of Valencia, where a DMA will be selected for the analysis of the dynamic sectorization. To do that, in the first place it is required to analyse the potential energy recovery in the inlets of the DMAs taking into count average values of pressure upstream and downstream and the minimum night flow. This parameter is one of the criteria selections for the pilot DMA. A hydraulic model will be calibrated for the subsequent analysis of the pressure regulation enabling to evaluate the alternatives of dynamic pressure management: by time control, flow control and critical point control. The results from this pilot DMA will include quantifying economically both water savings and the cost of installation of the small water turbines, with the purpose of bringing the city closer to an example of climate-neutral supply.

Keywords
Water Supply Networks, Pressure Dynamic Management, Pressure Reducing Valve, Energy Recovery, Circular Energy, Climate-neutral Water Supply Network.

1 INTRODUCTION

Increasing energy efficiency is a cutting edge topic in any economic activity. The excessive use of energy cause, among others, increasing greenhouse gases (GHG) emissions and energetic dependencies. GHG emissions are related with climate change and global warming. Energetic dependencies generate geopolitical instabilities. As an example, the price of energy is dramatically increasing in the 21st century, as can be seen in Figure 1. The average price of energy in Spain is in 2022 seven times the price it had in 2000, and it is still raising. Same behaviour has been observed in the rest of Europe.

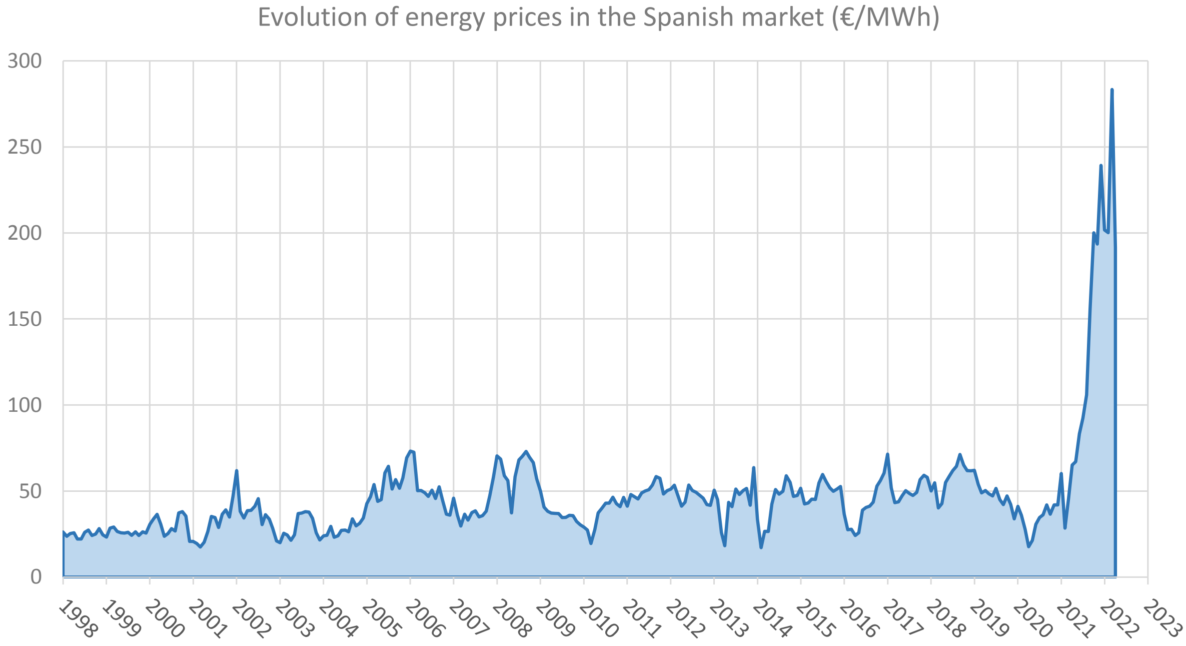


Figure 1. Evolution of the energy prices in Spain. (Source: OMIE, Spanish NEMO Nominated Electricity Market Operator)

The water industry (distribution, sewage and treatment) is one of the main energy consumers in the world. It has been proved that around 7% of energy consumption is directly related with the different water industry activities. Up to 40% of this energy is spent in the transportation and distribution of water. Therefore, energy is highly related with water. Water and energy are two limited resources used in the same utility that has been very important throughout the history but hardly considered jointly [1]. However, some authors are starting studying the water-energy nexus (WEN) [2]. Zaragoza et al. [3] created the concept Watergy to describe the mutual relation between water and energy in the water industry.

Colombo and Karney [4] described two different interactions between water and energy involving leakage. On the one hand, leakage implies additional water volume to be introduced in the water network. This waste of water supposes a loss of revenue for the industry but also may compromise water quality issues. On the other hand, the larger the flow rate, the larger the head losses in the network. It leads to additional power in pumping stations, increasing the energy consumption. Therefore, leakage reduction has an additional outcome consisting in the reduction of consumed energy.

The relation between leakage flow and pressure was defined by Germanopoulos and Jowitt [5]. As a result of this relation, Jowitt and Xu [6] started to optimize the pressure in a network in order to reduce leakage, also known as pressure management. These authors concluded that using pressure reducing valves (PRV) might significantly reduce leakage volumes. Besides, modifying the setting of the valves along the day leads to a larger reduction of leakages but has also some drawbacks. First, changes in the setting may affect operation of pumps in the network. Second, someone (meaning labour costs) or something (energy costs associated to the actuator of the valve) must change the setting. From an energy point of view, this action may be taken as a waste of energy: energy is first provided by pumps or elevated tanks and then dissipated in a valve. In an energy-scarcity scenario, this can be understood as illogical. That is why use of small turbines (pump as turbines –PAT–, miniturbines and picoturbines) are starting being used in water distribution networks (WDN). In other words, the energy we need to dissipate can be converted in electricity. That is way Iglesias-Castelló et al. [1] defines Potential Recoverable Energy (PRE).

BY NC SA

Use of small hydropower plants presents some difficulties for their implementation in WDN. First, it is difficult to find information about small turbines. Besides, operation of turbines in installations as WDN, with a highly varied conditions, is difficult to control [7]. To solve this issue, Nautiyal et al. [8] propose a setup with two valves: one in series and another in parallel. Nowadays, the use of small turbines in WDN is widely extended in networks with high differences of pressure.

To sum up, pressure management has demonstrated its efficiency to reduce water losses. Pressure management requires the use of PRV, that is, needs to dissipate excess of pressure. Small turbines may convert this excess of pressure into energy. The variation of the settings of the valves with time requires a small amount of energy. The main objective of this work consists of studying the viability of using small turbines to generate the energy needed by actuators to modify the setting of the valves. The method used will consist into three steps: a) selecting suitable turbines based on the available head drops and flow rates, b) selecting suitable district metering areas (DMA) with PRE enough to feed the actuators, and c) using a calibrated model to determine the improvement of both water and energy consumptions and evaluate the validity of the system.

2 METHODOLOGY

2.1 Installation description

Regulating turbines is a complicated task, especially in a WDN where the flow demands are variable throughout the day. As mentioned above, the setup proposed by Nautiyal et al. is an installation with a PRV in series and a control valve in parallel. These two valves control the behaviour of the turbine to maintain the design operating point. On one side, when the flow rate is larger than the operating point the turbine, following its characteristics curve, will generate a head loss greater than expected. However, control valve is opened and this excess of discharge is taken to the by-pass. On the other side, when the flow rate is less than operating point, the head loss that will produce the turbine is less than expected. In that case, PVR reduces pressure head until it reaches an operating point within turbine characteristic curve.

In the installation proposed in this work there is a main PRV and its setting will be the only one to be modified according to the pressure necessities of the WDN. On that basis, the proposed setup consists of a small turbine and PRV in parallel to the main PRV. In this case, the PRV in parallel has a protection purpose in the event of a turbine failure. Therefore, this setup can only admit design flow rate or higher. When the flow rate is larger than the operating point, the excess flow is by-passed to the main PRV.

Demand flow fluctuates throughout the day in an urban WDN. Normally there are two valleys of consumption, one in mid-afternoon and other in the early morning hours, in this last the consumption is minimum. The measured flow during these hours is known as Minimum Night Flow (MNF) and it is characterised by low user consumption and high level of leakage.

Considering the setup proposed, for the optimum operation of the installation the design flow rate shall be the MNF for the WDN. In this way, during the early morning hours all flow will only pass through the turbine and the main PRV will stay closed.

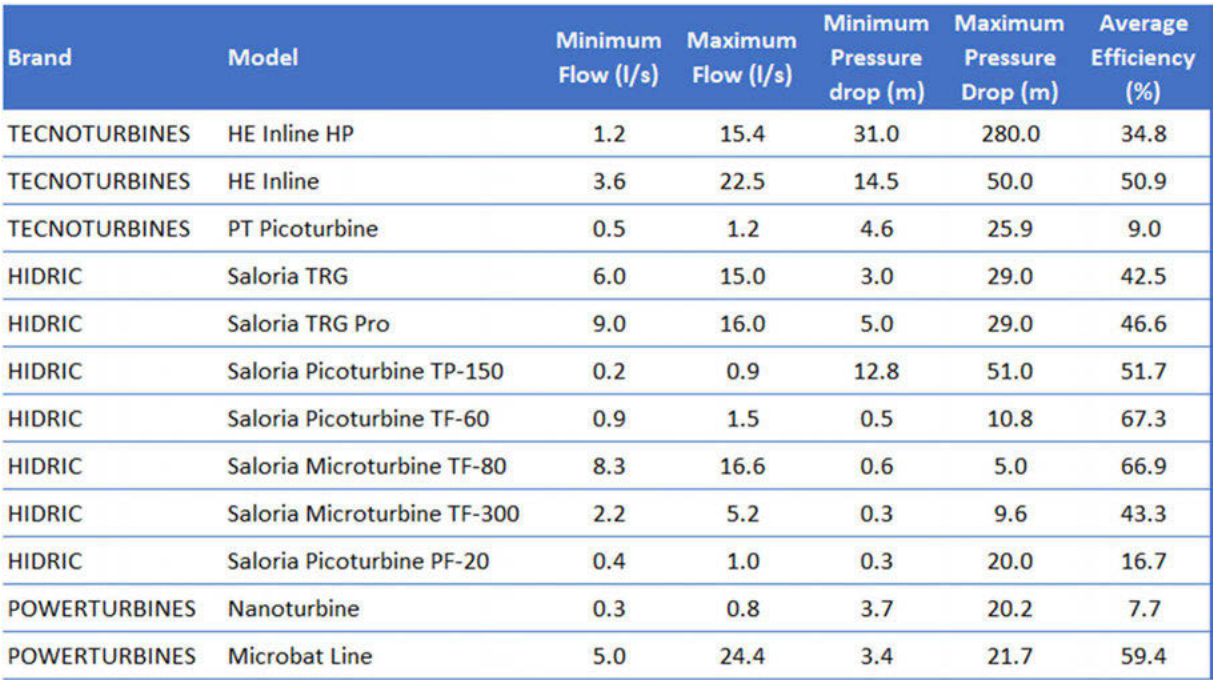
Other variable of the turbine design operating point to be determined is pressure drop. Pressure drop at the same time depends on two variables: pressure at the inlet point of the WDN and the pressure delivered to the WDN. Therefore, it will be defined at a later stage.

In fact, small turbine must generate the energy to manipulate the main PRV for that MNF and pressure drop. It is necessary 18 W in case to use a PRV controller or 50W in case to use a motorised pilot, based on the information from brand catalogues.

2.2 Selection of turbines

A wide variety of turbines for different flow and pressure drop ranges are available on the market. In this research it is analysed the turbine catalogue from different brands which are focus on recovering energy from WDN. Based on the information extracted from the operating range curves of twelve turbine models, where the recovered power in terms of flow and pressure drop are displayed, it has been possible to calculate the theoretical recovered power and hence, average efficiency for each model.

Table 1. Flow and pressure drop ranges and average efficiency for turbine models from different brands (Source: commercial catalogues of Tecnoturbines, Hidric and Powerturbines, available at the Internet).



Brand	Model	Minimum Flow (l/s)	Maximum Flow (l/s)	Minimum Pressure drop (m)	Maximum Pressure Drop (m)	Average Efficiency (%)
TECNOTURBINES	HE Inline HP	1.2	15.4	31.0	280.0	34.8
TECNOTURBINES	HE Inline	3.6	22.5	14.5	50.0	50.9
TECNOTURBINES	PT Picoturbine	0.5	1.2	4.6	25.9	9.0
HIDRIC	Saloria TRG	6.0	15.0	3.0	29.0	42.5
HIDRIC	Saloria TRG Pro	9.0	16.0	5.0	29.0	46.6
HIDRIC	Saloria Picoturbine TP-150	0.2	0.9	12.8	51.0	51.7
HIDRIC	Saloria Picoturbine TF-60	0.9	1.5	0.5	10.8	67.3
HIDRIC	Saloria Microturbine TF-80	8.3	16.6	0.6	5.0	66.9
HIDRIC	Saloria Microturbine TF-300	2.2	5.2	0.3	9.6	43.3
HIDRIC	Saloria Picoturbine PF-20	0.4	1.0	0.3	20.0	16.7
POWERTURBINES	Nanoturbine	0.3	0.8	3.7	20.2	7.7
POWERTURBINES	Microbat Line	5.0	24.4	3.4	21.7	59.4

This information is used to identify the minimum flow rate (0.2 l/s) and minimum head loss (0.3 m) for install any of these 12 turbines. A representative value of efficiency is also estimated as first quartile of average efficiency ($\eta_{Q1} = 20\%$).

2.3 Selection of DMA based on Potential Recoverable Energy (PRE)

Potential recoverable energy (PRE) is excess energy, over the minimum required, supplied in the nodes of WDN. A portion of the PRE could be recovered inside the WDN itself by installing energy recovery devices (PAT, turbines), in certain points of the network in order to reduces the excess energy supplied to all network node [1].

Valencia city WDN is formed of more than 50 DMAs of which 32 have installed at least one PRV and have available data related with it. With the available data is possible to obtain the variables that affect the turbine operation.

For a first estimation of pressure drop, it can be calculated as the difference between average upstream pressure and average downstream pressure of the PRV. In addition, there is a target MNF for each DMA set by water utility. For the DMAs with two or more inlets, it is assumed that MNF distribution among the inlets is the same as the average flow distribution during all day. In this way, it is calculated a MNF per inlet.

In energy recovery using turbines, the portion of the PRE that could be recovered is limited by turbine efficiency. Therefore, the useful power out is calculated as shows in equation (1).

$$P(W) = \gamma \cdot \Delta p \cdot Q \cdot \eta \tag{1}$$

Where P is useful recovered power in watts, γ is the specific weight of water in newton per cubic metres, Δp is the pressure drop in metres column water, $Q$is the flow rate in cubic metres per second and η is the efficiency dimensionless.

For the selection of which DMA would be most convenient to install turbines were considered the following criteria:

a) The turbine must be installed in line, that is, it must be a reaction turbine. This criterion excludes Pelton turbines.

b) Hydraulic criteria: MNF per inlet must be achieved and a minimum pressure drop for install a turbine.

c) Useful recovered power must be enough to move the actuator of the PRV. A PRV needs at least 18 W in the case the PRV uses a controller or 50 W in the case of a motorised pilot.

d) Number of inlets to the DMA. Due to the complexity of controlling a DMA with more than one entry, DMA with few entries are preferred.

2.4 Preparation of the calibrated models

Hydraulic models utilised in this work are built using QGISRed plugin, which is a free professional software to build and analyse WDN models from a shapefile database. The initial hydraulic model consists of junctions with base demand that is real information of water consumption at each service connection. The demand pattern is assumed same as input flow rate pattern. Furthermore, reservoirs in the inlets of the DMAs represents the PRV upstream pressure with total head of 1 and a head pattern that includes elevation and pressure head and PRV with setting same as average of PRV downstream pressure.

To obtain the calibrated model, it is followed the calibration methodology for pressure-dependent demand and consumption introduced by Martinez-Solano [9]. One relevant hypothesis is that non-revenue water volume consists of leakage in the WDN and therefore it is considered pressure dependent. The objective of this methodology is to calculate emitter coefficient for each junction, based on a general emitter coefficient for the entire WDN, and adjust consumption pattern coefficients for the base demand in an iterative process.

On one hand, daily balance of injected flow and revenue flow results in the estimated average leakage rate ($\overline{Q_L}$).

$$\overline{Q_L} = \overline{Q_S} - \sum_{i=1}^{N_D} \overline{Q_{BD,i}} \tag{2}$$

Where $\overline{Q_S}$ is average water supplied, $\overline{Q_{BD,i}}$ is base demand of node i within the total of all consumption nodes in the network N_D.

On the other hand, it is admitted that all junctions have the same emitter coefficient and same pressure temporal variation to calculate average leakage flow in equation (3). To calculate average pressure for the entire network, it is considered as initial hypothesis that base demand follow same pattern as injected flow rate.

$$\overline{Q_L} = \sum_{i=1}^{N_D} c_{E,i} \cdot \sqrt{\overline{p_i}} \tag{3}$$

Where $c_{E,i}$ is emitter coefficient for node i and $\overline{p_\iota}$ is average pressure in node i. In this work, it is assumed that emitter coefficient is same for all nodes in the network ($c_{E,i} = c_E$). Consequently, if equation (2) and (3) are equated, general emitter coefficient c_E is calculated as in equation (4).

$$c_E = \frac{\overline{Q_S} - \sum_{i=1}^{N_D} \overline{Q_{BD,\iota}}}{\sum_{i=1}^{N_D} \sqrt{\overline{p_\iota}}} \tag{4}$$

Average leakage rate calculate in equation (2) must coincide with the result in (3). If not, general emitter coefficient is corrected iteratively until the average pressure in junction in two successive iterations is differs less than a certain value.

Adjustment of the consumption pattern coefficients is the next step in the process. The coefficients are calculated following equation (4).

$$m_d(t) = \frac{Q_S(t) - c_E \sum_{i=1}^{N_D} \sqrt{p_i(t)}}{\sum_{i=1}^{N_D} Q_{BD,i}} \tag{4}$$

Where $m_d(t)$ is the consumption coefficient at time t, $Q_S(t)$ is water supplied at time t and $p_i(t)$ is pressure in node i at time t. For this work, leakage rate is calculated with pressure temporal variation of a representative junction of the network.

The alteration of the consumption pattern involves changes in network pressure, thus iteration continuous until the error between observed and simulated input flow rate is less than a certain value (in this work this value was assumed to be <1%).

In DMA with more than one inlet is essential that the model represents exactly the flow rate that passes in every entrance. For that, when the iteration process is concluded, valve settings are modified. At this point is worth mentioning the modification with a lower setting in PRV implies that passes less flow rate in this entrance. However, this flow is compensated by the other inlets. It happens the opposite with a higher setting in PRV.

2.5 Setpoint curve

Once the model is calibrated, it is built the setpoint curves in order to know the specific head pressure that is needed to deliver a flow rate considering the pressure in the critical node of the network, as accurately as possible. The critical node is the node with the lowest pressure and can change depending on demand variation. In a network each inlet or source of supply has a setpoint curve.

The process followed to calculate the setpoint curves is explained by Leon Celi et al [10]. Specifically, the explanation for the cases with pressure dependent consumptions both only one water source and more than one water source supply. Besides, in the case of more than one water source the flow rates to be supplied are fixed considering the data from the calibrated model.

According to the Urban Planning Regulations of the General Urban Development Plan of Valencia, the minimum pressure in the service connection must not be less than 25 m. Being conservative, the minimal pressure established as required is 27m if it is takes into account possible losses in the service connections.

The setpoint curve obtained can be approximated by the following equation (5) as a function of flow demand Q_d.

$$H^{(c)} = H_0 + R_c \cdot Q_d^n \tag{5}$$

Where H_0 is the minimum head pressure for zero flow demand, R_c and n are parameters for a power regression.

This way, the head pressure required for any flow demand in the network is easily calculated.

3 CASE STUDY

3.1 Description of the network

The network under study is Benimaclet sector which includes the area of the old Benimaclet neighbourhood in Valencia city. This network has 1,047 service connections and 10,708 clients. These clients are mainly domestic (88.1%), besides industrial (10.8%) and municipal and fire hydrants (1.1%).

The Benimaclet network has two water inlets. First one is located in the central west area and water through a 250mm pipe in Vicente Zaragoza St. The second one is in the southeast area and water through a 200mm diameter pipe in Catalunya Ave. In each inlet there is a PRV of 200mm of diameter. In addition, pressure gauges are installed both upstream and downstream of each PRV, and a water meter is installed in each of the inlet pipes.

In the same way as other areas of Valencia City, topography is practically flat with no major slopes. The difference between the highest and lowest elevation in DMA is only 6 metres. The lowest elevation in east area and the highest in west area, where is located the critical point of the DMA. The network has 23.1 km of pipelines of which 57.6% was installed before 2001, 34.4% between 2002 and 2003, and the remaining percentage from 2004.

The hydraulic model is composed of 2 reservoirs, 2 valves, 3342 pipelines 3202 nodes of which 822 are consumption nodes.

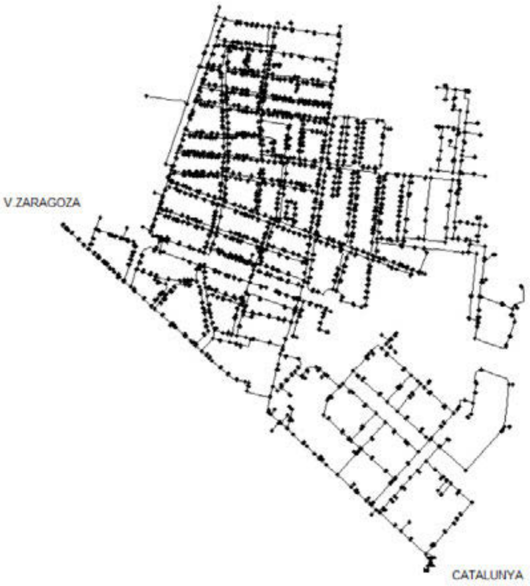


Figure 2. Hydraulic model of Benimaclet sector in Epanet [Source:]

3.2 Initial operation and pressure management

Calibrated model has revenue water in each service connection for the maximum consumption day of March 2020 with a demand multiplier to adjust average base demand for a week of August 2020, when was possible to take pressure data from 3 points inside the DMA.

Until November 2021, PRVs were fully open thus water arrived at the inlets with a pressure around 41 m was not reduced and average pressure inside network was around 40. Under these conditions, for calibrated model for a week of August 2020, average water supplied was 29.27 l/s and average revenue water was 18.52 l/s. Therefore, non-revenue water (NRW) was 10.75 l/s, which represents 36.7% of water supplied.

Considering this, from November 2021 the DMA has a pressure modulation based on fixed pressure output where PRV setting are 34 and 35 m in Vicente Zaragoza St. and Catalunya Ave., respectively. Consequently, average pressure inside the network is around 35.3 m and average pressure at the critical point is 31.2 m. Besides, this implies a reduction of 1.6% in NRW compared to the previous situation.

DMA water consumption has the common daily variations for urban networks. For instance, there is a great valley between 2 a.m. and 6 a.m. with very low the consumption compared to daytime. In factor, the average flow rate for this period is 17.13 l/s and the target MNF sets is 14.1 l/s.

3.3 Dynamic pressure management

In this work it is proposed a time-based dynamic pressure management which allows higher downstream pressure to be set for daytime and lower pressure at night when consumption decreases. For the calibrated model input flow is lower in the period between 2 a.m. and 6 a.m., when the consumption is around 55-60% of average input flow.

It is used the setpoint curves of the DMA for determining the valves setting (Figure 3). On one hand, setpoint curve for PRV in Catalunya Ave. indicates that for the range of input flow, the minimum pressure needed barely fluctuated and it is around 32 m. Thus, PRV will remain with a modulation based on fixed pressure output. On the other hand, PRV in Vicent Zaragoza St. will have a setting of 29.8 m in the night period and 30.7 m for the daytime period.

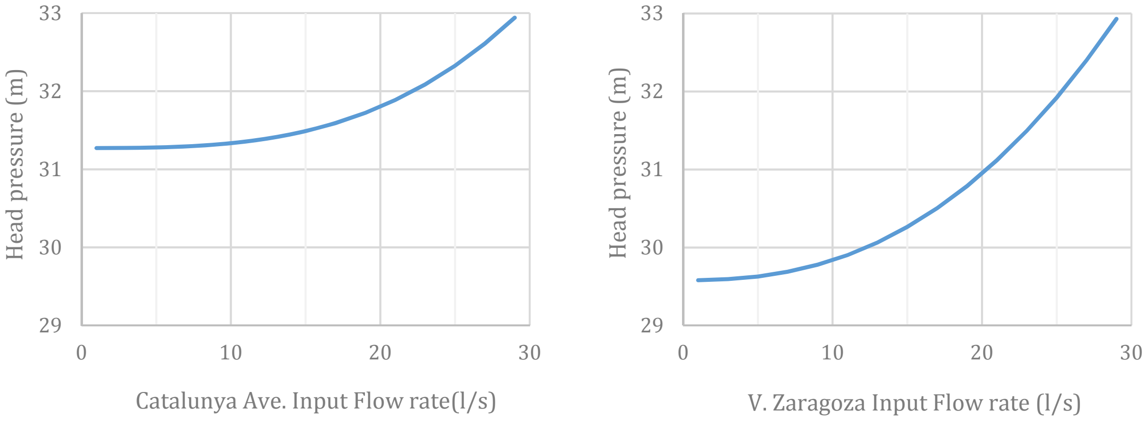


Figure 3. Setpoint curves for both sources water supply in Benimaclet network.

The proposed operation will allow pass 9.18 l/s through Vicente Zaragoza St. and 18.85 l/s trough Catalunya Ave. Besides, it will suppose a reduction of 2.8% in NRW.

In fact, as it is showed in Figure 4, it could be an improvement in volumetric efficiency as minimum required pressure in the network is lower (Figure 4). The volumetric efficiency of a network is the relation between revenue water and the total supplied volume during a period of time. In the calibrated model, volumetric efficiency is 63.3% with 33m of minimum pressure in the critical point. If the minimum pressure required is lowered to 20m, volumetric efficiency improves by 5.6%.

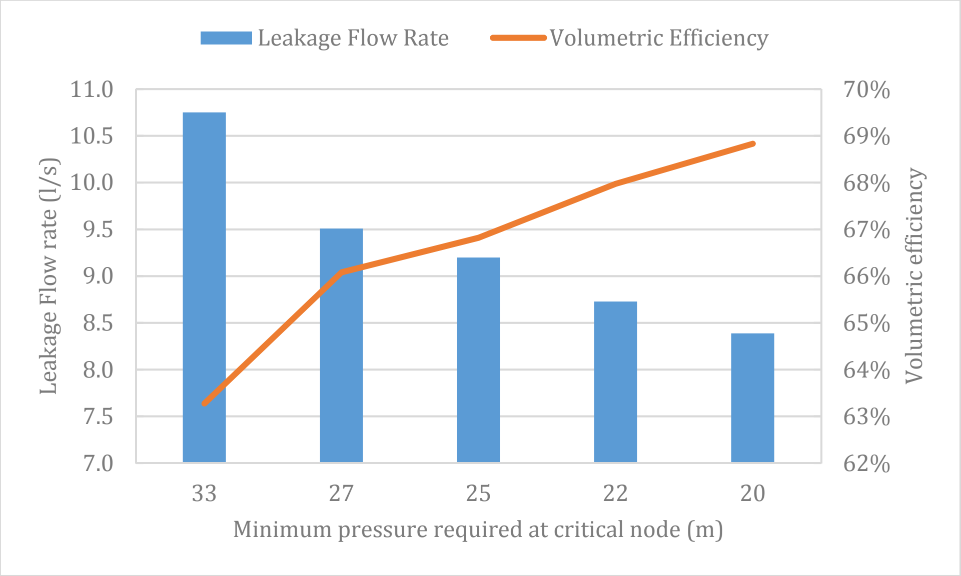


Figure 4. Relation between leakage flow rate and volumetric efficency in Benimaclet network.

As regards energy recovery, the PRE in the Vicente Zaragoza inlet is calculated as in equation (1) without considering the turbine efficiency yet. The average pressure drop is 10.1m taking in count upstream pressure is 40.61 m and average downstream pressure is 30.5m. As mentioned above, it is expected that MNF distribution among the inlets is the same as the average flow distribution during all day. Therefore, MNF for Vicente Zaragoza inlet is 6.5 l/s and, as result, the PRE is 643W, enough to operate the actuator of the PRV.

3.4 Economic assessment

With the proposed operation there is a reduction in the average annual input volume of 39,201 m^3. Considering fee for typical consumption set by local government in 2020 and currently in force (0.5584 €/m^3), it would stop losing an amount of 21,890€ per year.

Furthermore, water price in the water treatment plant inlet is 0.1712 €/m^3 to which must be added the Hydrographic Confederation tax (regulatory agency for the basin supplying the city) of 0.0206 €/m^3. Thus, leakage reduction will result in saving of 7,518 €/m^3 on the cost of bringing the water to the water treatment plant.

In this work only the unit price of the turbine has been taken into account in the economic assessment. From HIDRIC catalogue information, where is displayed the price list of its turbines, it has been related maximum recoverable power of each turbine model and its unit price. The equation (6) is the result of applying a regression analysis by the method of least squares.

$$C = 1148 + 0.06 \cdot PRE^{1.54} \tag{6}$$

Where C is the turbine unit price in euros and PRE is potential recovered energy inlet in watts. The PRE in the Vicente Zaragoza inlet is 643 W, thus the unit price of the turbine could be estimated as 1,147 €.

However, installation of a turbine implies an initial investment that includes the building of the extension or the new chamber, all necessary connections for the incorporate the turbine into the network, PRV controller, the turbine itself and all staff necessary to generate energy.

4 CONCLUSIONS

Energy efficiency is a crucial factor considering currently energy situation. As water industry is one of the main consumers of energy, relation water-energy has started to be studied in the last few years. Related to this matter is leakage in WDN which its reduction directly involves a reduction in consumed energy. The leakage volume is diminished using PRVs, which need energy to be controlled. This energy come from the network itself by using small turbines. In this way the energy that initially will be dissipated can be converted into electricity.

The setup of installation determines the operation of turbines. In this work the proposed installation has a turbine with a PRV in series and a main PRV in parallel. It is expected that the turbine works with a flow rate equal to MNF if the WDN has only one inlet or the MNF in that entrance if there are more than one. In this way, turbine will operate not only during daytime but also during night-time.

The study of different turbine models lets define minimum flow and drop pressure for turbine performance. This, joined to other criteria, allows the selections of network to be built and studied. The hydraulic model for this works represents non-revenue water as pressure-dependent demands by using emitter coefficients. The dynamic pressure management considered is based on the minimum pressure required at critical node of the network. By means of the setpoint curve is possible to know the PRV setting for specific input flow and, consequently, is possible proposed a PVR control based on time periods.

The network under study in this work is Benimaclet that has a volumetric efficiency of 63.3% and consists of two inlets with a fixed pressure output. The dynamic pressure management proposed is to keep one of the inlets with a fixed pressure output and other inlet with a pressure modulation based on time. This modulation involves a fixed a lower pressure output between 2 a.m. and 6 a.m. than daytime.

In conclusion, implementing dynamic pressure management in the Benimaclet DMA could lead to a reduction in the leakage rate. Consequently, it results in savings in water cost at the inlet of drinking water treatment plant and in lost revenue due to leakages. In fact, only with the savings in the cost paid at the entrance to the drinking water treatment plant, the price of the turbine could be amortised over a period of one year. These outcomes are reached by setting the minimum required pressure at 27 m. Undoubtedly, these savings can be greater if minimum required pressure lowered to 25 m as local regulations mandates.

5 REFERENCES

[1] M. Iglesias-Castelló, P. L. Iglesias-Rey, and F. J. Martínez-Solano, "Potentially recoverable energy assessment in water distribution networks," in WDSA/CCWI Joint Conference Proceedings, 2018.

[2] M. N. Sharif, H. Haider, A. Farahat, K. Hewage, and R. Sadiq, "Water–energy nexus for water distribution systems: a literature review," Environ. Rev., vol. 27, no. 4, pp. 519–544, Dec. 2019.

[3] G. Zaragoza, M. Buchholz, P. Jochum, and J. Pérez-Parra, "Watergy project: Towards a rational use of water in greenhouse agriculture and sustainable architecture," Desalination, vol. 211, no. 1–3, pp. 296–303, Jun. 2007.

[4] A. F. Colombo and B. W. Karney, "Impacts of Leaks on Energy Consumption in Pumped Systems with Storage," J. Water Resour. Plan. Manag., vol. 131, no. April, pp. 146–155, 2005.

[5] G. Germanopoulos and P. Jowitt, "LEAKAGE REDUCTION BY EXCESS PRESSURE MINIMIZATION IN A WATER SUPPLY NETWORK.," Proc. Inst. Civ. Eng., vol. 87, no. 2, pp. 195–214, Jun. 1989.

[6] B. P. W. Jowitt and C. Xu, "Optimal Valve Control in Water˜Distribu tion Networks," J. Water Resour. Plann. Manag., vol. 116, no. 4, pp. 455–472, 1990.

[7] A. Carravetta, G. Del Giudice, O. Fecarotta, and H. M. Ramos, "Energy Production in Water Distribution Networks: A PAT Design Strategy," Water Resour. Manag., vol. 26, no. 13, pp. 3947–3959, Oct. 2012.

[8] H. Nautiyal, Varun, and A. Kumar, “Reverse running pumps analytical, experimental and computational study: A review,” Renew. Sustain. Energy Rev., vol. 14, no. 7, pp. 2059–2067, 2010.

[9] F. J. Martínez-Solano, P. L. Iglesias-Rey, and S. X. M. Arce, “Simultaneous Calibration of Leakages, Demands and Losses from Measurements. Application to the Guayaquil Network (Ecuador),” in Procedia Engineering, 2017, vol. 186.

[10] C. F. León Celi, P. L. Iglesias-Rey, and F. J. Martínez-Solano, “Energy optimization of supplied flows from multiple pumping stations in water distributions networks,” in WDSA 2016: 18th Water Distribution Systems Analysis Conference, 2016.