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 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.

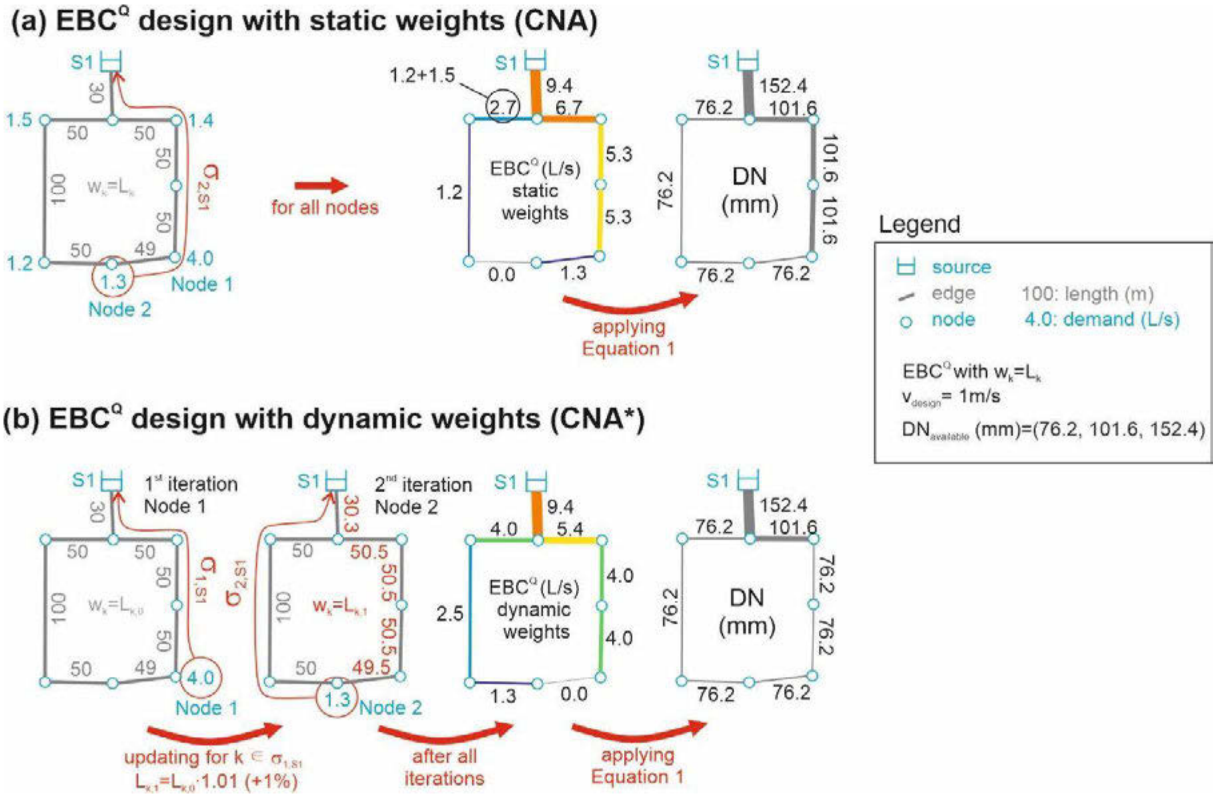

Figure 1. Multi-objective design procedure based on CNA [6](licensed under CC-BY 4.0)

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.

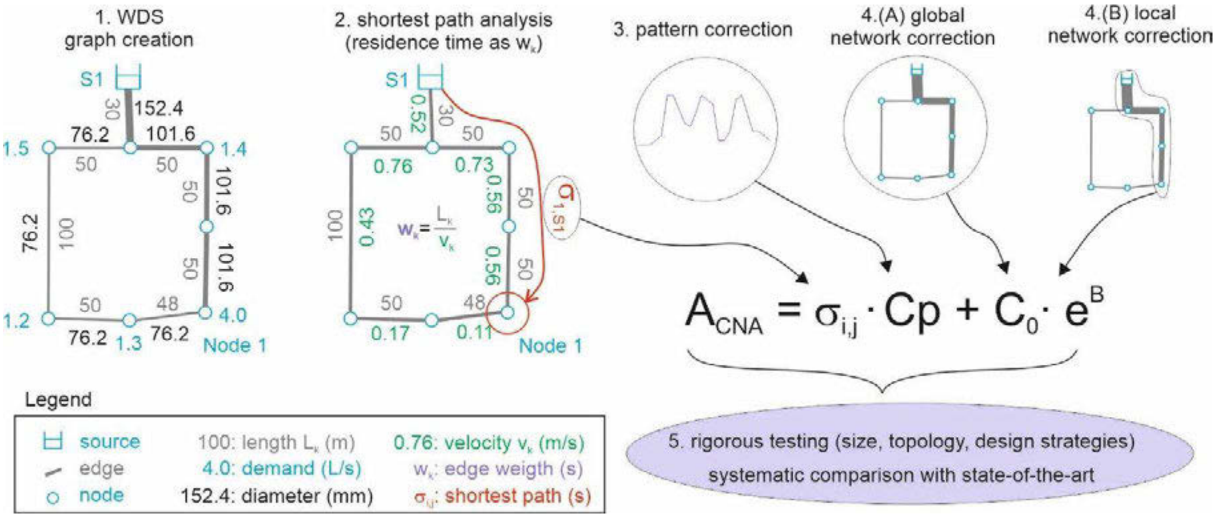

Figure 2. Water quality assessment with CNA [7] (licensed under CC-BY 4.0)

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 ($0 \le Q_{alt} \le 1$) 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.

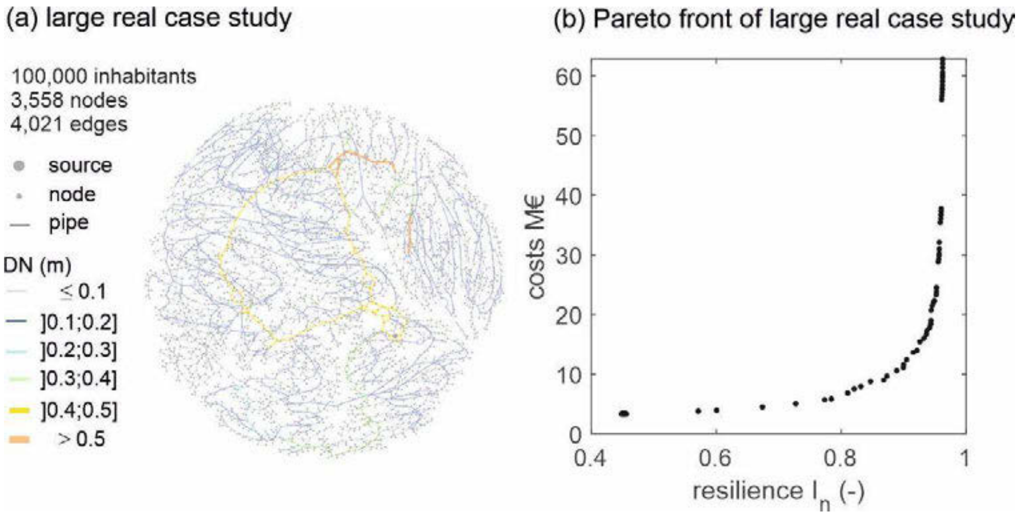

Figure 3. Anonymized layout of large real WDS (a) and Pareto front of optimal design solutions from meta-heuristic optimization

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 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.

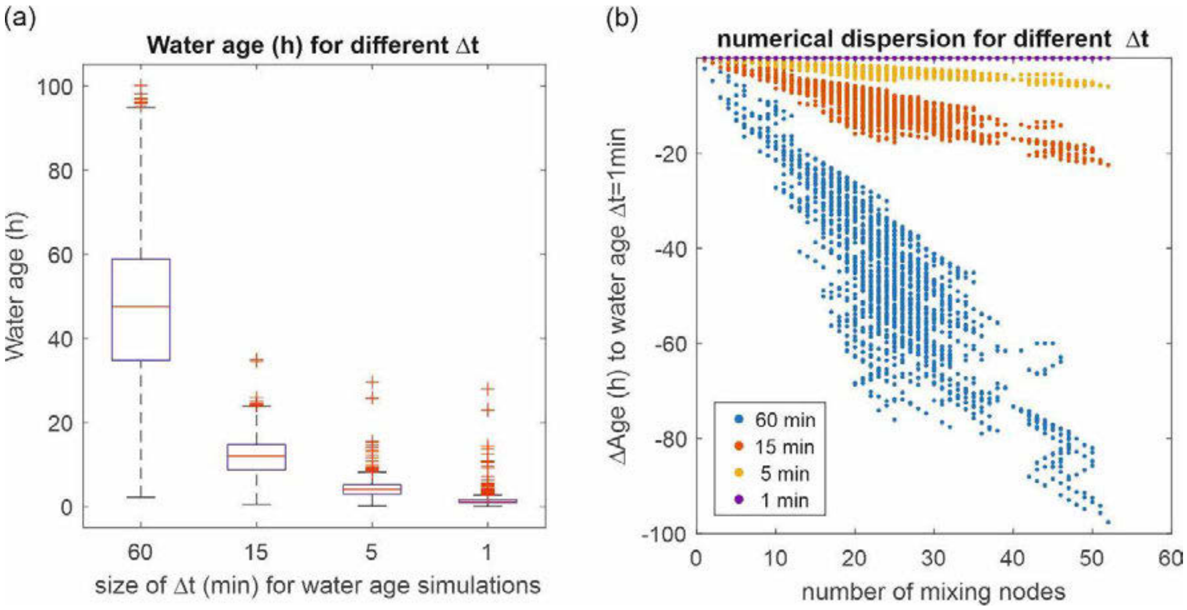

Figure 4. Impact of different Δt on the water age (a); impact of number of mixing nodes on the water age difference ΔAge to the results with Δt =1min (b)

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.

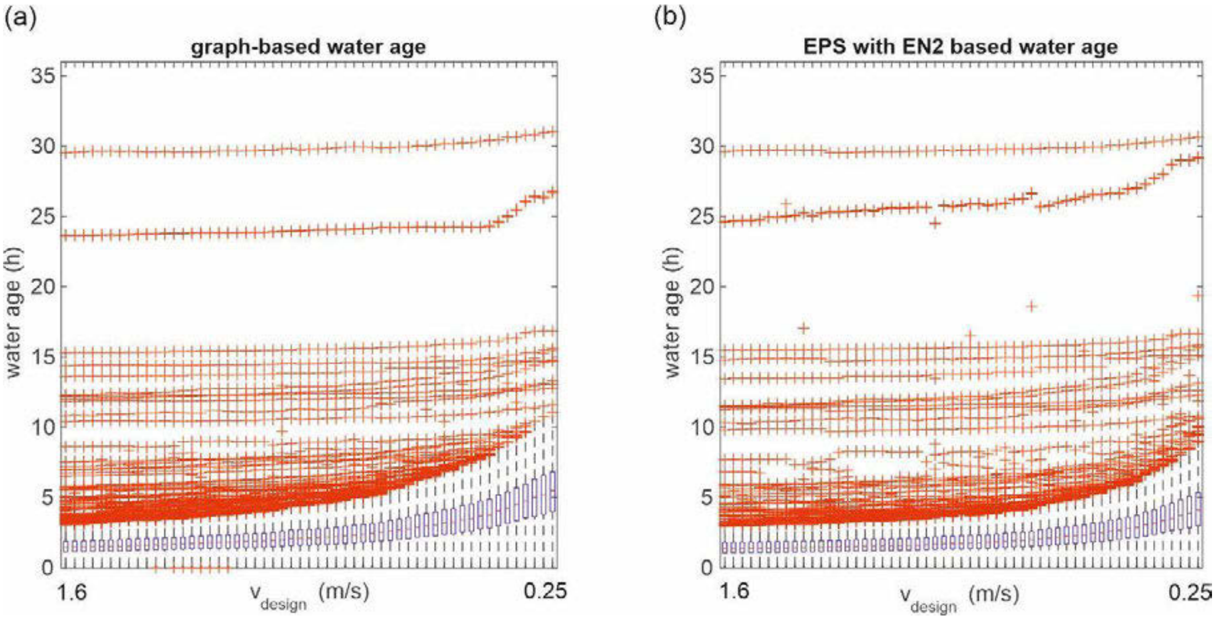

Figure 5. Results for graph-based water age (a) in comparison with results from EPS with Epanet2 (EN2) (b) for 55 graph-based design solutions

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.

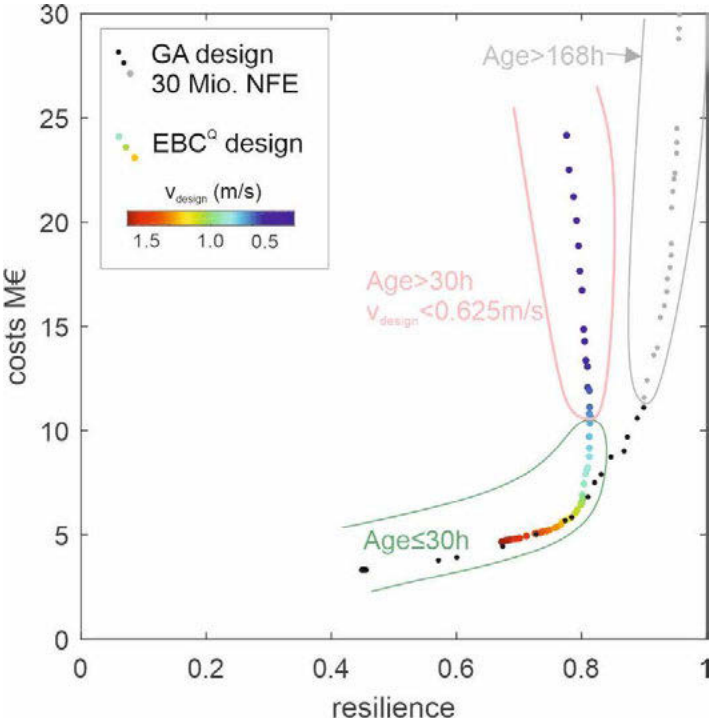

Figure 6. Results from graph-based design (coloured dots) in comparison with results from GALAXY Genetic Algorithm (GA) with number of function evaluations NFE = 30 Mio.

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.