# MODELLING FOR ENGINEERING & HUMAN BEHAVIOUR 2019

im²

Instituto Universitario de Matemática Multidisciplinar
Polytechnic City of Innovation

Edited by
R. Company, J.C. Cortés,
L. Jódar and E. López-Navarro

July 10th - 12th 2019

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
50 anys

CIUDAD POLITÉCNICA DE LA INNOVACIÓN

# Modelling for Engineering & Human Behaviour 2019

València, 10 – 12 July 2019

This book includes the extended abstracts of papers presented at XXIst Edition of the Mathematical Modelling Conference Series at the Institute for Multidisciplinary Mathematics "Mathematical Modelling in Engineering & Human Behaviour".

I.S.B.N.: 978-84-09-16428-8
Version: 15/11/19
Report any problems with this document to ellona1@upvnet.upv.es.

**Edited by:** R. Company, J. C. Cortés, L. Jódar and E. López-Navarro.
Credits: The cover has been designed using images from kjpargeter/freepik.

Instituto Universitario de Matemática Multidisciplinar

This book has been supported by the European Union through the Operational Program of the [European Regional Development Fund (ERDF) / European Social Fund (ESF)] of the Valencian Community 2014-2020. [Record: GJIDI/2018/A/010].

GENERALITAT VALENCIANA
Conselleria d'Hisenda i Model Econòmic

Fons Europeu de Desenvolupament Regional
Una manera de fer Europa
UNIÓ EUROPEA

# Contents

22

# A multi-objective and multi-criteria approach for district metered area design: water operation and quality analysis

Bruno Brentan[♭], Silvia Carpitella[♮], Joaquín Izquierdo[♮1], Edevar Luvizotto Jr[♯] and Gustavo Meirelles[♭]

(♭) Engineering School,
Federal University of Minas Gerais,
(♮) FluIng-IMM,
Universitat Politècnica de València,
(♮) Laboratory of Computational Hydraulics,
University of Campinas.

## 1 Introduction

Water distribution network (WDN) operation may be improved by district metered area (DMA) design [1]. A first step to create DMAs uses graph theory and non-supervised learning, where physical features of the WDN, such as node coordinates, elevation and demand, are used for clustering purposes [2]. A second step is related to the necessary isolation of the clustered elements. For isolation purposes, it is important to determine the DMA entrances and, consequently, the needed cut-off valves. Closure of pipes and definition of DMA entrances can be set as an optimization problem with the costs associated to the valves, which are linked to pipe diameters, as a primary objective. However, placement and operation of pressure reducing valves (PRVs) change the hydraulic conditions, and the optimization process should respect operation limits, such as minimum and maximum pressure and minimum and maximum tank levels. The optimization process can be written as:

$$\sum_{i=1}^{N_v} c(D_i) \quad \text{s.t. } P_{\min} \le P_{t,j} \le P_{\max} \text{ and } T_{\min,k} \le T_{t,k} \le T_{\max,k}, \tag{1}$$

where $c(D_i)$ is the cost of a valve placed in a pipe with diameter $D_i$; $N_v$ is the number of valves; $P_{\min}$ and $P_{\max}$ are the limit pressures allowed; $P_{t,j}$ is the operational pressure at time step $t$ in pipe $j$; $T_{\min,k}$ and $T_{\max,k}$ are the minimum and maximum tank levels allowed for tank $k$; and $T_{t,k}$ is the operational level of tank $k$ at time step $t$.

Constrained problems are frequently handled by using penalty functions. However, as discussed in [3], penalty approaches modify the search space, impairing the search process by the creation

[1]e-mail: jizquier@upv.es

of new local minima. To solve this problem, a bio-inspired algorithm widely applied in water distribution problems [4], adapted for a multi-objective approach, is applied. In this context, constraints become objectives to be reached, which turns the problem unconstrained.

Moreover, as observed in [5], such crucial water distribution parameters as resilience, pressure uniformity and water quality strongly depend on DMA configurations. These parameters are known to depend on pressures and water tank levels and, together with cost, will be the other objectives of our optimization.

A multi-objective approach gives a set of solutions, the so-called Pareto front. To select, within that front, which non-dominated solution will be implemented may be hard task. To help this process, this work presents: a) multi-level optimization process for entrance location and set point definition of PRVs, and b) a post-processing based on a multi-criteria method, which ranks the non-dominated solutions based on the relative importance of the said four main objectives: implementation cost, resilience, pressure uniformity and water quality.

Among the wide range of MCDM (multi-criteria decision-making) methods used in the literature, the *Technique for Order of Preference by Similarity to Ideal Solution* (TOPSIS) effectively works across various application areas [6]. Such a technique was developed by Hwang and Yoon [7] as a simple way to solve decision-making problems by means of the ranking of various decision alternatives [8, 9]. In this context, the objective of the TOPSIS application to the multi-objective problem consists in selecting the solution representing the best trade-off (among the set of optimal solutions belonging to the Pareto front) under the perspective of the considered evaluation criteria.

# 2 Clustering process based on a k-means algorithm

The first step for DMA design is to cluster the nodes of the network. Among the various methods suitable for this step, a simple and effective one is the k-means algorithm. The method uses the Euclidean distance between samples and centroids, and clusters are defined according to the smallest distances. For a simple explanation, let's take a set with $m$ data points $\chi = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m]$ where each point $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}]$. Taking a pre-defined number of clusters $k$, the method starts distributing randomly the $k$ centres in the data space. The Euclidean distance $d_{i,j}$ between each center $j$ and each data point $\boldsymbol{x}_i$ is computed. The data points are classified as belonging to cluster $j$ if the distance $d_{i,j}$ is minimum when compared for all other centres. After the classification step, the centres are replaced to the mean value of all points belonging to a cluster. The process is repeated (distance calculation, point classification, and centre replacement) until the distance between the centres at iteration $t-1$ and $t$ is smaller than a tolerance value.

# 3 The non-dominated sorting genetic algorithm (NSGA-II)

Different from single objective optimization algorithms, multi-objective approaches do not find just one optimal solution, but a set of compromise solutions, so-called Pareto front. Among several algorithms proposed for multi-objective optimization, population based algorithms, such as NSGA-II [10], are widely applied for engineering problems, highlighting their applications in the water distribution domain [11, 12].

NSGA-II evaluates all the $N$ possible solutions composing a population. The solutions are evaluated for all the objectives finding non-dominated solutions. Considering two solutions $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$, it is said that $\boldsymbol{x}_a$ dominates $\boldsymbol{x}_b$ if and only if both conditions a) and b) below are satisfied.

a) $\boldsymbol{x}_a$ is no worse than $\boldsymbol{x}_b$ for all objectives, and

b) $\boldsymbol{x}_a$ is strictly better than $\boldsymbol{x}_b$ at least for one objective.

For each single solution $\boldsymbol{x}_p$ it is possible to know the number $n_p$ of solutions dominating $\boldsymbol{x}_p$ and the set of solutions $S_p$ dominated by $\boldsymbol{x}_p$. By definition, non-dominated solutions have $n_p = 0$, and integrate the so-called primary Pareto front. Now, for all solutions $q$ in the sets $S_p$ included in that primary Pareto front, the value of $n_q$ is reduced by one and those solutions $q$ with new $n_q = 0$ for all $p$ are collected into a new set, so-called secondary Pareto front. This procedure is repeated until finding new high-level fronts.

The algorithm starts with a random population $P_0$ of size $N$. Individuals are evaluated for all the objectives and the non-dominated front is found as described next. Genetic operations (binary tournament selection, recombination, mutation) are used to create a new set of solutions. The new population, now sizing $2N$, is sorted according to non-domination, and the Pareto fronts for all the levels are found; if the number of solutions belonging to the primary Pareto front is smaller than $N$, all solutions are preserved and the new population is completed with the higher-level Pareto fronts, according to the ranking. Otherwise, the $N$ first solutions of the primary Pareto front are selected. One important feature of population-based algorithms is the maintenance of the solution spread. This is fundamental for good convergence to a Pareto-optimal set [10]. To this purpose, NSGA-II uses a crowded-comparison approach, based on crowding distances (see [10]).

The process re-starts with the new population by re-evaluating each solution under all objectives and re-ranking solutions based on the non-domination criterion. The algorithm stops when reaching some termination criteria, such as maximum number of iterations, or no improvements in the Pareto front. The method results in a set of non-dominated solutions with an optimal compromise relation for all the objectives. However, for practical problems, the evaluation of the Pareto front by experts could be hard task. To help, a post-processing step, based on MCDM is proposed.

# 4 The TOPSIS to rank solutions

As mentioned, TOPSIS is a MCDM method aimed at ranking various alternatives, such as the solutions of the decision-making problem under analysis. The method calculates distances from each solution to a positive ideal solution and to a negative ideal solution. The solution representing the best trade-off under the considered criteria is the one characterised by the shortest distance to the positive ideal solution, and the farthest to the negative one.

First of all, the TOPSIS technique needs the preliminary collection of the following input data to be applied: a decision matrix (collecting the evaluations $g_{ij}$ of each alternative $i$ under each criterion $j$), the weights of criteria (representing their mutual importance), and their preference directions (to establish if criteria have to be minimised or maximised).

The implementation of the procedure is led by following five main steps:

- Building the weighted normalized decision matrix, for which the generic element $u_{ij}$ is calculated as:
$$u_{ij} = w_j \cdot z_{ij}, \ \forall i, \forall j; \tag{2}$$
where $w_j$ is the weight of criterion $j$ and $z_{ij}$ is the score of the generic solution $i$ under the criterion $j$, normalized by means of the equation:
$$z_{ij} = \frac{g_{ij}}{\sqrt{\sum_{i=1}^{n} g_{ij}^2}}, \ \forall i, \forall j. \tag{3}$$

- Identifying the positive ideal solution $A^*$ and the negative ideal solution $A^-$, calculated through the following equations:
$$A^* = (u_1^*, \ldots, u_k^*) = \{(u_{ij} | j \in I'), (u_{ij} | j \in I'')\}; \tag{4}$$
$$A^- = (u_1^-, \ldots, u_k^-) = \{(u_{ij} | j \in I'), (u_{ij} | j \in I'')\}; \tag{5}$$
$I'$ and $I''$ being the sets of criteria to be, respectively, maximized and minimized.

- Computing the distance from each alternative $i$ to the positive ideal solution $A^*$ and to the negative ideal solution $A^-$ as follows:
$$S_i^* = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_j^*)^2}, \ i = 1, \ldots, n; \tag{6}$$
$$S_i^- = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_j^-)^2}, \ i = 1, \ldots, n. \tag{7}$$

- Calculating, for each alternative $i$, the closeness coefficient $C_i^*$ which represents how the solution $i$ performs with respect to the ideal positive and negative solutions:
$$C_i^* = \frac{S_i^-}{S_i^- + S_i^*}, \ 0 \leq C_i^* \leq 1, \ \forall i. \tag{8}$$

- Obtaining the final ranking of alternatives on the basis of the closeness coefficients calculated above. In particular, with relation to two generic solutions $i$ and $z$, solution $i$ must be preferred to solution $z$ when $C_i^* \geq C_z^*$.

# 5 Case study

The methodology proposed is applied to the literature water network called Exnet [13]. The system supplies 400,000 consumers approximately, requiring a delivery minimum pressure of 20m. The network is composed by 1,891 nodes, and 2,465 pipes. Two reservoirs and five injection nodes (wells) feed the network. For clustering analysis, each node is used as a data point, endowed with its topological features, namely geographical position, elevation and base demand. The number of clusters is defined using the Davies-Bouldin (BD) criterion [14], which evaluates the final clustered data, considering the distances among data points in a cluster and the corresponding centre (intra-criterion), and the distances among centres (inter-criterion). The best cluster number minimizes the intra-criterion and maximizes the inter-criterion. Varying from two to 15 clusters, the best DB criterion is found to be nine clusters. Fig. 1 shows the clustered network.

Once clustered, the network should pass by the optimization step in order to define the entrances and, consequently, those pipes where PRVs will be installed. The application of NSGA-II at this step results in a Pareto front with 115 non-dominated solutions, as shown in Fig. 2.

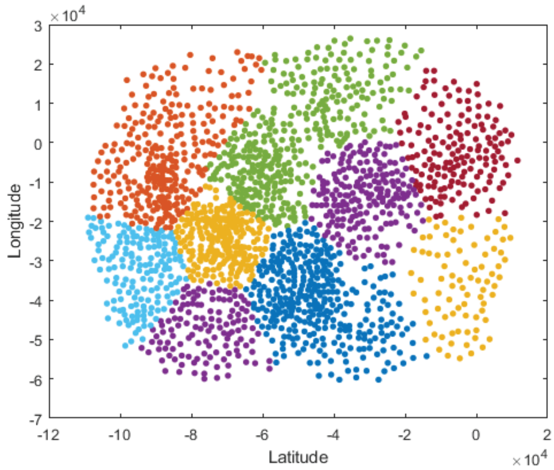


Figure 1: Clustered Exnet with optimal BD criterion, resulting in 9 clusters.

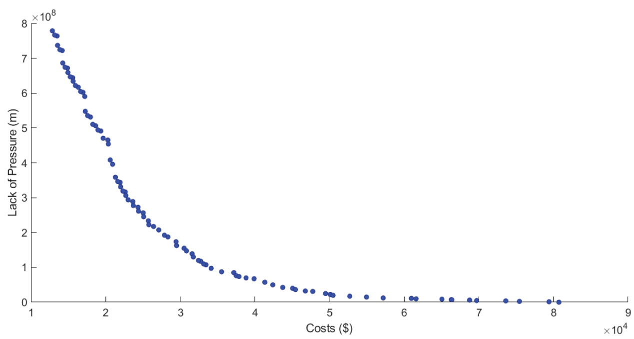


Figure 2: Pareto's front showing bi-objective problem of PRV placement in clustered networks.

The TOPSIS method described in the previous section has been applied to rank the 115 solutions belonging to the Pareto front. Obviously, resiliency is maximized whereas the other four criteria (pressure uniformity, dissipated energy, lack of pressure, and cost) are minimised. Moreover, at this stage of analysis, all the criteria have been considered as having the same importance. It means a weight equal to 20% has been assigned to each criterion. Results of TOPSIS application are reported in Table 1. Just the first five positions are given for sake of brevity.

| **Ranking position** | **Pareto Solution** | **Resilience** | **Pressure Uniformity** | **Energy dissipated** | **Lack of Pressure** | **Costs** | **Closeness Coefficient value** |
|---|---|---|---|---|---|---|---|
| 1 | 5 | 2,12E-01 | 8,30E+01 | 6,16E+03 | 7,54E+04 | 2,20E+06 | 0,758828 |
| 2 | 82 | 2,12E-01 | 8,30E+01 | 6,16E+03 | 7,54E+04 | 2,20E+06 | 0,758828 |
| 3 | 43 | 2,21E-01 | 8,38E+01 | 5,98E+03 | 7,94E+04 | 1,20E+06 | 0,754085 |
| 4 | 90 | 2,21E-01 | 8,38E+01 | 5,98E+03 | 7,94E+04 | 1,20E+06 | 0,754085 |
| 5 | 37 | 2,22E-01 | 8,38E+01 | 5,96E+03 | 8,07E+04 | 0,00E+00 | 0,750996 |

Table 1: TOPSIS results.

The solutions in the first positions present higher values of closeness coefficient, since they have large distance to the negative ideal solution and small distance to the positive ideal solution. Similar solutions appear in Table 1, such as 5 and 82, or 43 and 90. This happens by the closeness of those solutions in the Pareto front, with identical rounded values. Solution 37 exhibits an implementation cost equal to zero. This is a solution for the multi-objective problem from the mathematical point of view. From the engineering point of view, this means that all the boundary pipes remain open, thus resulting in a non-segregated network. Despite solution 37 is the last of the five top solutions, it still can help decision makers to find the benefits of DMA creation on that network.

To provide readers with an effective comparison of the results in terms of the values of the considered parameters, we also provide in Table 2 the last five positions of the ranking, those with the lowest closeness coefficient. It is possible to note as these last positions present higher associated costs (an objective to be minimised) and lower values of operation parameters (objectives to be maximise, instead).

| **Ranking position** | **Pareto Solution** | **Resilience** | **Pressure Uniformity** | **Energy dissipated** | **Lack of Pressure** | **Costs** | **Closeness Coefficient value** |
|---|---|---|---|---|---|---|---|
| 111 | 11 | 0,00E+00 | 1,92E+06 | 1,71E+04 | 1,42E+04 | 7,22E+08 | 0,249465 |
| 112 | 10 | 0,00E+00 | 1,92E+06 | 1,82E+04 | 1,29E+04 | 7,79E+08 | 0,249008 |
| 113 | 113 | 0,00E+00 | 1,92E+06 | 1,82E+04 | 1,29E+04 | 7,79E+08 | 0,249008 |
| 114 | 93 | 0,00E+00 | 1,92E+06 | 1,80E+04 | 1,32E+04 | 7,67E+08 | 0,248811 |
| 115 | 13 | 0,00E+00 | 1,92E+06 | 1,80E+04 | 1,35E+04 | 7,64E+08 | 0,248037 |

Table 2: TOPSIS results: last five positions of the ranking.

This ranking approach shows the interest of MCDMs to select trade-off scenarios under the considered criteria. The first solution shows the best pressure uniformity and lack of pressure, but the highest cost and lowest resilience. That means, the best hydraulic and operation conditions will appear in the most expensive scenario. The relation of resilience and pressure uniformity can also be highlighted. Scenarios with lower pressure uniformity present lower resilience, since resilience is calculated based on overpressure, and pressure uniformity tries to minimize overpressure.

## 6 Conclusions and future developments

The present work proposes a fully automated algorithm for DMA design based on clustering analysis, multi-objective optimization and multi-criteria analysis. The clustering analysis is done by a k-means algorithm evaluated under the Davies-Bouldin criterion, resulting in nine DMAs. Multi-level optimization for entrance location and set point definition of pressure reducing valves achieve network clustering. NSGA-II finds 115 non-dominated solutions in a trade-off between various objectives. In addition, a MCDM is applied to rank the non-dominated solutions, to identify the one representing the best trade-off in fulfilling the objectives to be matched. Operational and hydraulic criteria are used to evaluate the solutions.

Regarding MCDM, the TOPSIS method has been applied to obtain the final ranking of non-dominated solutions. In particular, this application has been carried out under the evaluation of four criteria: implementation cost, resilience, pressure uniformity and water quality. In the presented case study, we assumed these criteria as having the same weight, in other terms, the same degree of mutual importance.

Results point to solution number 5 as the best trade-off among all the 115 non-dominated solutions, since it is the first in the ranking. While this solution embodies the best operational criteria, is the most expensive and least resilient.

Future developments of the present work may regard a further integration between the multi-objective and the multi-criteria perspectives, though with a different purpose: for example, instead of getting just a rank of the non-dominated solutions, the application of a MCDM method to classify them into proper clusters will be worth it.

# References

[1] Campbell, E., Izquierdo, J., Montalvo, I., Ilaya-Ayza, A., Pérez-García, R. and Tavera, M., A flexible methodology to sectorize water supply networks based on social network theory concepts and multi-objective optimization. *Journal of Hydroinformatics*, 18(1): 62-76, 2016.

[2] Brentan, B., Campbell, E., Goulart, T., Manzi, D., Meirelles, G., Herrera, M., ... and Luvizotto Jr, E., Social network community detection and hybrid optimization for dividing water supply into district metered areas. *Journal of Water Resources Planning and Management*, 144(5), 04018020, 2018.

[3] Lima, G. M., Luvizotto Jr, E., Brentan, B. M. and Ramos, H. M., Leakage control and energy recovery using variable speed pumps as turbines. *Journal of Water Resources Planning and Management*, 144(1), 04017077, 2017.

[4] Montalvo, I., Izquierdo, J., Schwarze, S. and Pérez-García, R., Multi-objective particle swarm optimization applied to water distribution systems design: an approach with human interaction. *Mathematical and Computer Modelling*, 52(7-8): 1219-1227, 2010.

[5] Brentan, B. M., Campbell, E., Meirelles, G. L., Luvizotto, E. and Izquierdo, J., Social network community detection for DMA creation: criteria analysis through multilevel optimization. *Mathematical Problems in Engineering*, 2017.

[6] Behzadian, M., Otaghsara, S.K. Yazdani, M. and Ignatius, J., A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39(17): 13051-13069, 2012.

[7] Huang, C.-L. and Yung, K., *Multiple-Attribute Decision Making. Methods and Applications. A state-of-the-Art Survey.* Springer-Verlag, Berlin, Heidelberg, New York, 1981.

[8] Carpitella, S., Certa, A., Enea, M., Galante, G., Izquierdo, J., La Fata C.M. and Vella, F., Combined HACCP and TOPSIS-based approach to prioritize risks in the salmon manufacturing process: an applicative case. Proceedings of the 22th Summer School "Francesco Turco", Palermo, Italy, September 13-15, 150-156, 2017.

[9] Carpitella, S., Certa, A., Izquierdo, J. and La Fata, C.M., k-out-of-n systems: an exact formula for stationary availability and multi-objective configuration design based on mathematical programming and TOPSIS. *Journal of Computational and Applied Mathematics*, 330: 1007-1015, 2018.

[10] Deb, K., Agrawal, S., Pratap, A. and Meyarivan, T., A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In International conference on parallel problem solving from nature (pp. 849-858). Springer, Berlin, Heidelberg, 2000.

[11] Atiquzzaman, M., Liong, S. Y. and Yu, X., Alternative decision making in water distribution network with NSGA-II. *Journal of water resources planning and management*, 132(2): 122-126, 2006.

[12] Preis, A. and Ostfeld, A., Multiobjective contaminant sensor network design for water distribution systems. *Journal of Water Resources Planning and Management*, 134(4): 366-377, 2008.

[13] Farmani, R., Savic, D. A. and Walters, G. A., Exnet benchmark problem for multi-objective optimization of large water systems. *Modelling and control for participatory planning and managing water systems*, 2004.

[14] Davies, D. L. and Bouldin, D. W., A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2): 224-227, 1979.