District metered area design through multi-criteria and multi-objective optimization

Running title: **DMA design through MC & MO optimization**

Bruno M. Brentan[1], Silvia Carpitella[2], Joaquín Izquierdo[3], Edevar Luvizotto Jr[4], Gustavo Meirelles[1]

[1]Engineering School, Federal University of Minas Gerais – UFMG, Belo Horizonte, Brazil

[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic

[3]FluIng-IMM - Universitat Politècnica de València, Valencia, Spain

[4]Laboratory of Computational Hydraulics - University of Campinas, Campinas, São Paulo, Brazil

Abstract

The design of district metered areas (DMA) in potable water supply systems is of paramount importance for water utilities to properly manage their systems. Concomitant to their main objective, namely deliver quality water to consumers, the benefits include leakage reduction and prompt reaction in cases of natural or malicious contamination events. Given the structure of a water distribution network (WDN), graph theory is the basis for DMA design, and clustering algorithms can be applied to perform the partitioning. However, such sectorization entails a number of network modifications (installing cut-off valves and metering and control devices) involving costs and operation changes, which have to be carefully studied and optimized. Given the complexity of WDNs, optimization is usually performed using metaheuristic algorithms. In turn, optimization may be single or multiple-objective. In this last case, a large number of solutions, frequently integrating the Pareto front, may be produced. The decision maker has eventually to choose one among them, what may be tough task. Multi-criteria decision methods may be applied to support this last step of the decision-making process. In this paper DMA design is addressed by: i) proposing a modified k-means algorithm for partitioning; ii) using a multi-objective particle swarm optimization to suitably place partitioning devices; iii) using fuzzy analytic hierarchy process (FAHP) to weight the four objective functions considered; and iv) using technique for order of preference by similarity to ideal solution (TOPSIS) to rank the Pareto solutions to support the decision. This joint approach is applied in a case of a well-known WDN of the literature, and the results are discussed.

Keywords: Graph theory, k-means algorithm, metaheuristic, multi-objective optimization, fuzzy AHP, TOPSIS, decision-making, district metered areas, water distribution systems

1. Introduction and literature review

The complex structures of water distribution networks (WDNs) are responsible for delivering drinkable water to citizens. Topography of cities, dynamics of consumption, and highly looped networks may cause lack of service uniformity, and make systems management a complex task. Large efforts have been devoted to improve WDN management, mainly regarding leakage. In [1], an extensive literature review is presented, which collects the most important contributions to leak management in water distribution systems. Among the alternatives for better controlling pipe pressure, the creation of district metered areas (DMAs) stands out. Different pressurized zones can be better operated by splitting the network into smaller parts, the so-called DMA [2] When efficiently operated, DMAs are capable to control the hydraulic pressure in water pipes, thus reducing leakage, what is one of the most important objectives for water utilities. Furthermore, a well partitioned network can react promptly in anomalous scenarios as contamination intrusion (either natural or malicious), for which correct network management pursued by isolating the non-affected area is crucial for health security.

DMA creation can be basically divided into the two following phases: identification of DMA zones, and installation of metering and control devices, as well as cut-off valves.

Identification of DMA zones could be approached by using experts' knowledge. However, advances in applied mathematics and computational models in water systems have made it possible the use of the graph theory and Data Mining for this task [2-6].

Since water network models are built in terms of nodes and links simulating, respectively, the users and pipes composing the water systems, these models can be seen as graphs, with the consequent possibility of applying graph theory [7]. Graph connectivity is explored as an important feature for DMA creation by means of algorithms such as the deep-first search (DFS) [8] and the breadth-first search (BFS) [9] in water networks [10]. Comparing both tools, the authors of [11] propose the division of a Mexican water network into independent but interconnected networks. The DFS and BFS algorithms are also explored in [12] for water network clustering. An important point to be highlighted about this work consists in the proposal of DMA creation to study contamination spread. The BFS is applied in [13] to define the nodes of a graph that constitute each DMA, and the shortest path distance from each source is computed to determine the set of boundary pipes for each DMA.

By linking graph theory and data analysis, community algorithms have also been proposed to identify DMA regions [2,5,13-14]. Specifically, [2] combines the concept of accumulated shortest path value and social community detection algorithms to identify DMAs. A community structure detection algorithm is also applied in [5]. The study uses a community structure to create physical boundaries for DMAs. Community detection is performed to maximize the matrix modularity. In the realm of data mining and clustering algorithms, gaussian mixture models (GMM) are applied in [14] to identify DMAs in water networks.

The purely data mining algorithms are capable to cluster high-dimensional datasets; however, the straightforward application to WDNs is not easy. This difficulty originates from the fact that features characterizing each node should be exhaustively defined to avoid clustering physically unconnected nodes. Coordinates of nodes are commonly used as input together with physical and hydraulic features. However, using coordinates is not enough to guarantee the connection of all the nodes in a cluster. This evidence can indeed lead to the design of unfeasible layouts, with consequent losses of time and money.

By coupling graph theory and data mining algorithms, this paper proposes a modified k-means algorithm for DMA region identification. Such a modified algorithm considers the connectivity of nodes during the clustering stage by guaranteeing the interconnection of all the nodes within a cluster. The main advantage of this proposal with respect to the simple k-means algorithm regards the correction of clusters by considering information about real links of the water network.

Despite there are many water utilities still using operators' expertise to identify DMAs zones, data mining tools represent an effective support for decision makers to accomplish this task.

Let's move now to the second phase of DMA design. After identifying the regions and building a DMA structure, control devices must be installed in an optimal way. In this direction, a trade-off has to be reached between the cost of control devices and the operational parameters of the water network. Indeed, if, on the one hand, installing many devices improves network control, on the other hand, this has an impact on project costs and operation. From the optimization point of view, placing control devices and cut-off valves can be studied as an optimization problem subjected to a set of operational constrains. By using social community detection for DMA region identification, optimization is put to work at two levels in [3]. The authors use Particle Swarm Optimization (PSO) to select the position of control devices and cut-off valves at the first level and, successively, they obtain the operational point of control devices at the second level. The concept of minimal background leakage is applied in [15] to locate the best locations for cut-off valves.

Since DMA creation modifies the hydraulic behaviour of water networks, many objectives could be explored during the process of optimal placement of valves and DMAs entrances. The traditional approach that minimises the surplus pressure in the system, then reducing background leakage, has been widely explored in the literature [16,17]. Even prior to the introduction of the concept of DMA took place, the objective of reducing leakage by means of pressure reducing valves (PRV) can be considered as a pioneer approach for optimization of water systems. With new challenges and requirements for water companies, governmental institutions guidelines, and global goals, the main objectives of DMA creation have also become the reduction of the energy spent on operation, and the improvement of water quality and resilience of systems.

Regarding optimization, single and multi-objective approaches have been proposed in the literature by combining various objectives for optimal DMA's entrance and cut-off valve installation [3,6]. While single objective approaches attribute different weights to the objectives, multi-objective solutions are capable to find a trade-off among all the objectives. Both cases present their own challenge. In the single objective approach, weights change throughout the search space, by attributing different importance to the objectives. Nevertheless, a single solution is obtained from the optimization process and it can be directly implemented in the water system. In contrast, multi-objective solutions are not unique, and water utilities' managers have to decide about which solution will be eventually implemented. In this case, an analysis supported by the use of multi-criteria decision-making (MCDM) methods can be useful to help decision makers in this task, by providing an answer in terms of which solutions, among those belonging to the set of the non-dominated solutions of the Pareto front, represent optimal trade-offs according to the evaluation of differently weighted criteria.

Regarding MCDM approaches commonly used in the literature, the FAHP (Fuzzy Analytic Hierarchy Process), first proposed in [17] as an evolution of the traditional AHP [18], represents a valuable way to manage situations affected by uncertainty by taking advantage of the fuzzy set theory [19,20]. Regarding the field of application of the mentioned method, a wide review (190 related papers) has been published by Kubler et al. [21], which recognise the FAHP as an effective tool for criteria weight calculation.

As shown by many authors [22-24], the method has also been integrated with other MCDM techniques aimed at ranking the alternatives under evaluation. In this regard, the TOPSIS (technique for order of preference by similarity to ideal solution) effectively works across various application areas [25]. Such a technique was developed by Hwang and Yoon [26] as a simple way to solve decision-making problems by ranking various decision alternatives [27]. With respect to other MCDM methods, TOPSIS allows to get the final ranking even if the set of solutions is large, as it is often the case of solutions belonging to the Pareto front. Moreover, this method is able to obtain the ranking of alternatives without pairwise comparing them, what may be a really high time-consuming task when we have to deal with a huge number of solutions. This was already shown in a previous conference contribution [28] we aim to develop in this paper, which is a substantial extension of that contribution. Among other new aspects, the TOPSIS was applied to rank optimal solutions, but the objectives were considered as having the same mutual importance. On the contrary, in the present extended research we integrate the FAHP to attribute different weights to the objectives on the basis of judgments provided by an expert in the field of water distribution management. The weights were used to combine various objectives in a single objective problem and now we also rank the Pareto solutions using TOPSIS. Additionally, regarding the optimisation problem, the modified k-means algorithm is applied to identify DMA regions in a real-size benchmark water network. The boundary pipes are identified as possible positions for control and cut-off valves. In this context, the objective of the MCDM application to the multi-objective problem consists in selecting solutions representing optimal trade-offs (among the set of optimal solutions belonging to the Pareto front) under the perspective of the considered evaluation criteria weighted by means of the FAHP. We note that the same weights were used to build the objective function for the single objective optimization approach presented in [28], which we also consider now for comparison with the results regarding the multi-objective optimisation performed in the present research.

The flowchart of Figure 1 synthetises the entire procedure implemented in the present paper.

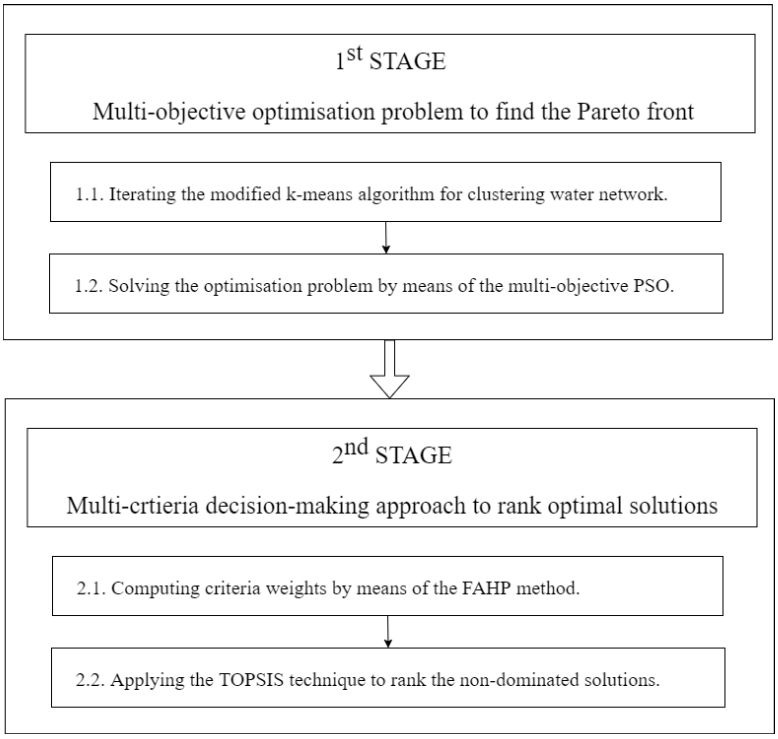

Figure 1. Flowchart summarising the proposed procedure

2. Mathematical methods: clustering based on a modified k-means algorithm and optimization

In this section we present the mathematical methods used in our approach for DMA design in a WDN, leaving the integrate multi-criteria approach for the next section. First of all, a modification of the traditional k-means algorithm is presented for partitioning the nodes of the network into clusters on the basis of the physical connections between nodes. After that, this section formalises the mathematical formulation of objective functions with relation to the problem of optimal valve placement for DMA design. Lastly, the multi-objective evolution of the PSO approach is described as a way to solve the optimisation problem. The output of the whole stage will be a set of non-dominated solutions, that is the Pareto front, to be successively treated by means of the integrated multi-criteria approach.

2.1 Clustering process using a modified k-means algorithm

As stated in the introduction, DMA design starts by clustering the nodes of the network. Among the various methods suitable for this step, a simple and effective one is the k-means algorithm. The classical k-means is based on grouping samples by a similarity measure, being the Euclidean distance between samples and centroids the most used metric. To exemplify the process, let's take a set with m data points $\chi = [\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_m]$ where each point $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n}]$. 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 with all the other centres. When all groups are identified, new position of each centre is calculated as the mean value of all the points belonging to the corresponding 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.

In this work, a modification of k-means is proposed to consider the physical links between the nodes. Before replacing the centre, a verification is performed to identify the links between one node and the others. Since k-means was not originally developed for graph clustering, the algorithm does not use connectivity information for clustering. In this sense, it may occur that some nodes end up clustered in a certain group with no physical connection to this group. When this kind of wrong classification happens, a clustering post-processing is performed, and unconnected nodes are assigned to the neighbour group they have real physical connection with. Concluding this stage, the centres are recalculated, based on the average value of the cluster's point position. This modification is important to guarantee the identification of physically feasible DMA's.

One important task on using clustering algorithms, such as k-means, is defining the number of clusters. In [29] various mathematical and engineering criteria for DMA design are evaluated. Mathematical criteria evaluate the quality of the clustering process in terms of external and internal measures. Among those mathematical criteria, the Davies-Bouldin (DB) criterion [30] 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 DB criterion is written as:

$$DB = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}\{D_{i,j}\}, \tag{1}$$

where $D_{i,j}$ is the ratio of distances within the same cluster i and the distances between clusters i and j, written as

$$D_{i,j} = \frac{\bar{d}_i+\bar{d}_j}{d_{i,j}}, \tag{2}$$

where $\bar{d}_i$ is the mean distance between the points belonging to cluster i and the centre of this cluster; $\bar{d}_j$ is the mean distance between the points belonging to cluster j and the centre of this cluster; and $d_{i,j}$ is the distance between the centres i and j. The lower DB the better the clustering, since internally the points are near to the respective centres while the centres are far from each other.

2.1. Mathematical formulation of optimal valve placement for DMA design

After identifying each DMA region, optimal places for the corresponding entrances, where control devices and flow meters have to be installed, should be identified. Furthermore, cut-off valves should also be installed on other pipes linking the DMAs. These devices are crucial for effective isolation and control of DMAs. Cut-off valves and control devices, however, modify the hydraulics of the system, usually reducing pressure, and consequently leakage, but also impairing the system resilience. Furthermore, cutting pipes make the water take alternative paths to reach the users. As a result, as pointed by [31], DMA creation can increase the water age, with a negative impact on the water quality. Bearing in mind these hydraulic modifications of the water system, not only should the cost be minimized, but also the benefits should be maximized and the drawbacks minimized.

The problem can be thus formulated by means of four objective functions, corresponding to minimize the cost (F_1), minimize pressure management (F_2) as proposed in [32], maximize resilience of the system (F_3) as proposed in [33], and minimize water quality (F_4), as proposed in [34]:

$$F_1 = \sum_{j=1}^{Nb} c(d_j), \tag{3}$$

$$F_2 = \sum_{t=1}^{T}\frac{1}{N_n}\sum_{i=1}^{N_n}\left(\frac{P_{i,t}-P_{min}}{P_{min}} + \frac{\sqrt{\frac{\left(P_{i,t}-P_{m_t}\right)^2}{N_n}}}{P_{m_t}}\right), \tag{4}$$

$$F_3 = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{Nn} q_{i,t}\left(h_{i,t}-h^*_{i,t}\right)}{\sum_{k=1}^{Nr} Q_{k,t}H_{k,t}+\sum_{j=1}^{Np} {}^{P_{j,t}}/_{\gamma}-\sum_{i=1}^{Nn} q_{i,t}h^*_{i,t}}, \tag{5}$$

$$F_4 = \frac{\sum_{i=1}^{Nn}\sum_{t=1}^{T} k_{i,t}q_{i,t}\left(WA_{i,t}-WA_{max}\right)}{\sum_{i=1}^{Nn}\sum_{t=1}^{T} q_{i,t}}. \tag{6}$$

The variables are as follows:

- F_1 calculates the accumulated cost, $c(d_j)$, of control devices and flow meters installed, which depends on the diameter d_j of the boundary pipes, these being Nb in total.

- F_2 is called pressure uniformity. This parameter measures the distance of the operational pressure $P_{i,t}$ of the network from the minimal operational pressure P_{min} and from the average pressure of the network P_{m_t}. It is composed by the sum for all simulation time steps, totaling T time steps, and for each node i from a total of N_n nodes.
- F_3 is called resilience index and can be seen as a relation between the required and the available power in the system; $q_{i,t}$ and $h_{i,t}$ are, respectively, the flow and hydraulic head at node i at time step t; $h^*{}_{i,t}$ is the required hydraulic head to deliver the demand at node i time step t; $Q_{k,t}$ and $H_{k,t}$ are the outlet flow and hydraulic head of the reservoir k at time step t in a network with Nr reservoirs; $P_{j,t}$ is the power of pump j at time step t in a network with Np pumps.
- Finally, F_4 is a quality parameter related to the water age; $k_{i,t}$ is a Boolean variable, equal to 1 when the water age $WA_{i,t}$ at node i at time step t is greater than a water age limit WA_{max}.

Since optimal valve placement is an engineering problem, operational constraints should be considered during the process. Minimal operational pressure for demand nodes and non-negative pressure for connection nodes are considered as constraints. Bearing in mind the stated optimization problem, and the fact that a heuristic algorithm is used for solving it, the constraints are handled by penalty functions, as presented in equation (7).

$$C_1 = \delta \left| P_{i,t} - P_{min,i} \right| \tag{7}$$

Here $P_{min,i}$ is the minimal operational pressure if the node i is a demand node, or zero otherwise and δ is the penalty coefficient, responsible for amplifying the fitness value of a solution that violates the minimal operational pressure. Non-negative pressure at junctions is treated similarly. In this work, δ is set to 10^6, so as to guarantee convergence.

For the single objective approach, the penalty function is summed to the combined objective functions. For the multi-objective approach, penalties, obviously, only affect objective functions F_2 and F_3.

2.2. Particle Swarm Optimization

Optimization problems in water distribution analysis have been explored with heuristic algorithms. Among the classical heuristic algorithms, PSO [35], a swarm-based metaheuristic, has an important place on solving complex optimization problems. The algorithm is based on a flock of birds traveling in the search for food. The principle of PSO is the improvement of solutions, guiding their search for optimal solutions. Each particle of the swarm is evaluated according to its position on the search space. The position changes through iteration using its search velocity, which is updated following a linear combination of three parameters,

$$\mathbf{v}_i^{t+1} = w\mathbf{v}_i^t + c_1 r_1\left(\mathbf{x}_i^t - \mathbf{p}_i^t\right) + c_2 r_2(\mathbf{x}_i^t - \mathbf{g}^t), \tag{8}$$

where $\mathbf{v}_i^{t+1}$ is the search velocity of particle i at iteration $t+1$; $w\mathbf{v}_i^t$ is called inertia term, and is responsible to avoid roaming by partially maintaining the particle search direction of the last iteration given by its previous velocity $\mathbf{v}_i^t$ weighted by the inertia coefficient, w; $c_1 r_1\left(\mathbf{x}_i^t - \mathbf{p}_i^t\right)$ is the cognitive term and is calculated as the difference between the last particle position and the best position already occupied by the particle, weighted by a cognitive coefficient c_1 and multiplied by a random number r_1; finally, the third term, $c_2 r_2(\mathbf{x}_i^t - \mathbf{g}^t)$, is called social term, and is calculated by the difference between the last position of the particle and the best position ever occupied by a swarm particle (the swarm leader), weighted by a social coefficient c_2 and multiplied by a random number r_2. The main goal of the random numbers is to avoid local optimal points, leading the particles to explore the search space.

Different from single objective optimization algorithms, multi-objective approaches do not find just one optimal solution, but a set of compromise solutions, the so-called Pareto front. In [36] an extension from single-objective PSO to multi-objective PSO (MOPSO) problems is proposed.

The main structure of MOPSO follows the proposal of [35], by using the original equations of PSO. The position and velocity of the particles are randomly initialized. A particle's position vector represents a possible solution. Each particle is evaluated under the objective functions. After evaluating all particles, the non-dominated solutions are stored. To identify 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.

With the primary Pareto front, a new velocity and position for each particle is calculated and the objective functions are re-evaluated. This process is repeated until a convergence criterion is reached, such as a 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.

3. Integrated multi-criteria approach

This section presents the MCDM methodologies used in this paper to support the optimisation problem. The main objective consists in treating the solutions belonging to the Pareto front resulting from the former stage. Specifically, we aim to obtain a final ranking of solutions to the ones representing the best trade-offs under the considered evaluation criteria. As analysis criteria, we create a direct correspondence with the objectives of the optimisation problem and apply a MCDM technique to derive their mutual degree of importance, while simultaneously manage uncertainty of input evaluations. Once criteria weights have been derived, their values will be used with the rest of the input data for the application of another MCDM method able to rank large sets of solutions. Specifically, the FAHP will be applied to calculate criteria weights, and the TOPSIS will be implemented to select the optimal solution representing the best trade-off under the perspective of the previously weighted criteria. These techniques are described next.

3.1. The FAHP to calculate criteria weights

As previously said, the fuzzy set theory is a helpful support tool in solving those situations involving human judgments, thus affected by uncertainty of evaluations. Its main advantage consists in the possibility of representing linguistic variables through fuzzy numbers rather than crisp values, with associated a degree of membership $\mu(x)$ varying between 0 and 1.

As recalled in a previous work [37], there are various types of fuzzy numbers, TFN (Triangular Fuzzy Numbers) $\tilde{n}$ and TrFN (Trapezoidal Fuzzy Numbers) $\tilde{m}$ being the most common [38,21]:

$$\tilde{n} = (a,\ b,\ c), \tag{9}$$

$$\tilde{m} = (d,\ e,\ f,\ g), \tag{10}$$

where $a \leq b \leq c$ and $d \leq e \leq f \leq g$. Addition, multiplication and inversion (considered in this study) are examples of common algebraic operations than can be performed with TFNs $\tilde{n}_1$ and $\tilde{n}_2$:

$$\tilde{n}_1 \oplus \tilde{n}_2 = (a_1 + a_2,\ b_1 + b_2,\ c_1 + c_2), \tag{11}$$

$$\tilde{n}_1 \odot \tilde{n}_2 = (a_1 \times a_2,\ b_1 \times b_2,\ c_1 \times c_2), \tag{12}$$

$$\tilde{n}_1^{\ -1} = (\tfrac{1}{c_1},\ \tfrac{1}{b_1},\ \tfrac{1}{a_1}). \tag{13}$$

Taking advantage of the use of fuzzy numbers, the FAHP method can be implemented for the purpose of our research by performing three steps in sequence, as suggested in [39]: 1) building the hierarchy structure representing the decision-making problem under analysis; 2) collecting fuzzy pairwise comparisons from experts with respect to evaluation criteria; 3) calculating the vector of criteria weights that represent their mutual importance.

Concerning the stage of fuzzy pairwise comparisons collection, fuzzy input evaluations (translating linguistic judgments attributed by the expert or decision-making team) have to be collected in a FPCM (Fuzzy Pairwise Comparison Matrix), $\tilde{X}$, which is a squared, reciprocal matrix:

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{x}_{n1} & \cdots & \tilde{x}_{nn} \end{bmatrix}, \quad (14)$$

the generic element $\tilde{x}_{ij}$ expressing the degree of preference of criterion i with respect to criterion j with a certain level of uncertainty. Reciprocity implies that for each element $\tilde{x}_{ij} = (x_1,\ x_2,\ x_3)$ of the matrix $\tilde{x}_{ji} = (\frac{1}{x_3},\ \frac{1}{x_2},\ \frac{1}{x_1})$ holds (for convenience, we have omitted here the subindices).

Linguistic variables used to express pairwise comparisons about the relative importance between a pair of criteria refer to the fuzzy version of the Saaty scale (Figure 2). The considered variables and the associated TFNs are: equal (EQ), (1,1,2); moderate (M), (2,3,4); strong (S), (4,5,6); very strong (VS), (6,7,8); and extreme (EX) importance, (8,9,9). The TFNs (1,2,3), (3,4,5), (5,6,7) and (7,8,9) correspond to intermediate values. The diagonal elements, $\tilde{x}_{11}, \tilde{x}_{22}, \dots, \tilde{x}_{nn}$, of matrix $\tilde{X}$, express the comparison between an element and itself. As a consequence, they have all associated an evaluation of "equal", what corresponds to the TFN (1,1,2) according to the presented scale.

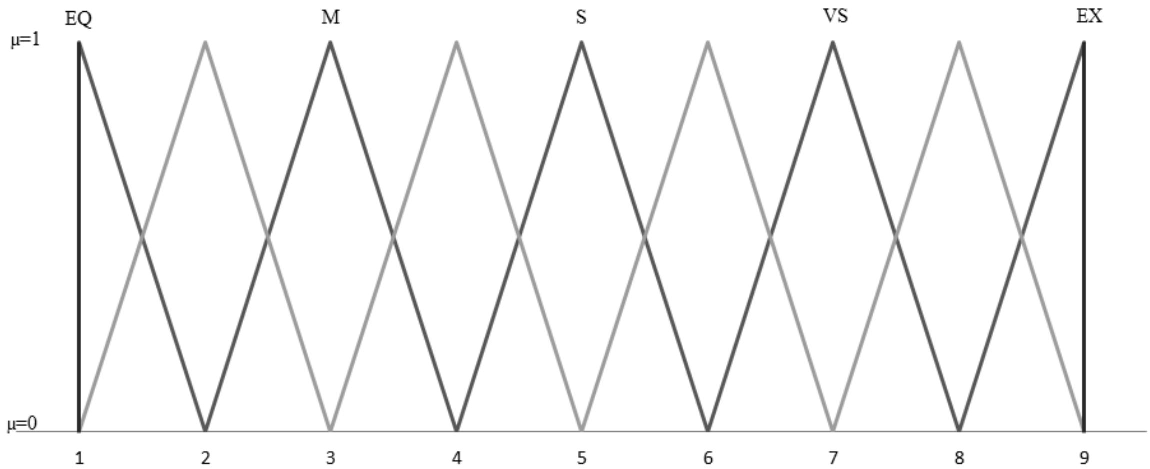

Figure 2. Fuzzy version of the Saaty scale [38]

Once filled in the FPCM $\tilde{X}$, the literature offers several approaches to calculate the vector of weights. Chang [40] proposes to derive crisp weights from the input matrix, by means of the extent analysis method. The value of fuzzy synthetic extent with relation to the i^{th} element of matrix $\tilde{X}$ can be calculated as follows:

$$S_i = \sum_{j=1}^{m} \tilde{x}_{ij} \odot \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{x}_{ij}\right]^{-1}, \quad (15)$$

in our case $n = m$, since the FPCM $\tilde{X}$ is a square matrix.

With relation to two fuzzy pairwise comparisons (e.g. two TFNs $\tilde{n}_1$ and $\tilde{n}_2$), we aim to establish the degree of possibility that $\tilde{n}_1 \geq \tilde{n}_2$, defined in [25]:

$$V(\tilde{n}_1 \geq \tilde{n}_2) = \mu(x^*) = \begin{cases} 1 & \text{if } b_1 \geq b_2 \\ 0 & \text{if } a_2 \geq c_1 \\ \frac{a_2 - c_1}{(b_1 - c_1) - (b_2 - a_2)} & \text{otherwise} \end{cases}, \quad (16)$$

where x^* is the ordinate of the highest intersection point P between the two membership functions $\mu_{\tilde{n}_1}$ and $\mu_{\tilde{n}_2}$ of the two considered TFNs (Figure 3). In order to compare the two TFNs $\tilde{n}_1$ and $\tilde{n}_2$, both values $V(\tilde{n}_1 \geq \tilde{n}_2)$ and $V(\tilde{n}_2 \geq \tilde{n}_1)$ have to be calculated.

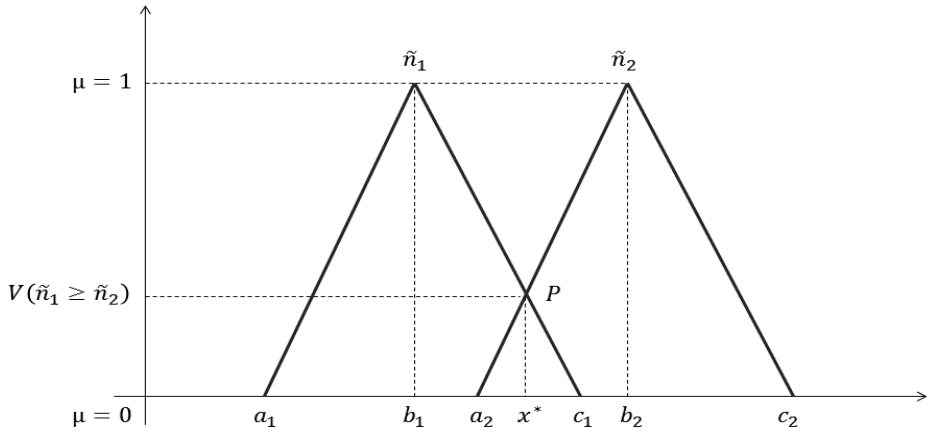

Figure 3. Representation of the degree of possibility that $\tilde{n}_1 \geq \tilde{n}_2$ [40,41]

We can also determine the possibility degree that a fuzzy number $\tilde{n}$ is greater than k fuzzy numbers $\tilde{n}_i (i = 1 \dots k)$ as:

$$V(\tilde{n} \geq \tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_k) = V[(\tilde{n} \geq \tilde{n}_1)\ and\ (\tilde{n} \geq \tilde{n}_2)\ and \dots and (\tilde{n} \geq \tilde{n}_k)] = min\, V(\tilde{n} \geq \tilde{n}_i), i = 1 \dots k. \quad (17)$$

In this way, we can link each criterion X_i given in the FPCM $\tilde{X}$ to the related value of fuzzy synthetic extent by defining:

$$x^{*\prime}(X_i) = min\, V(S_i \geq S_k), \quad (18)$$

for $k = 1 \dots n, k \neq i$. The vector of crisp and not normalized weights is lastly given by:

$$W' = \left(x^{*\prime}(X_1), x^{*\prime}(X_2), \dots, x^{*\prime}(X_n)\right)^T, \quad (19)$$

having the former weights to be normalised with respect to their total to obtain the final vector of normalised crisp weights:

$$W = \left(x^*(X_1), x^*(X_2), \dots, x^*(X_n)\right)^T. \quad (20)$$

The last operation consists in checking the consistency ratio (CR) of the FPCM $\tilde{X}$. To such an aim, each fuzzy value $\tilde{x}_{ij}$ of the matrix needs to be defuzzified and transformed into a crisp value x_{ij} by means of the graded mean integration approach [40]:

$$G(\tilde{x}_{ij}) = x_{ij} = \frac{x_1 + 4x_2 + x_3}{6}. \quad (21)$$

After having defuzzified each value of the matrix, consistency can be easily verified with the proper threshold [42]. We underline that checking consistency represents a fundamental issue in this kind of application. Indeed, if judgments were not consistent, this would have a negative impact on the whole quality of final decision. In such a case, experts should be asked to formulate new judgments until the condition of consistency is met [43].

3.2. The TOPSIS to rank optimal solutions

The TOPSIS method is capable to rank even large sets of alternatives, such as the set of optimal solutions of the decision-making problem under analysis, for example, those belonging to a populated Pareto front. 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 the following five main steps.

- Building the weighted normalized decision matrix, for which the generic element u_{ij} is calculated as:

$$u_{ij} = w_j \cdot z_{ij}, \forall i, \forall j; \tag{22}$$

  where w_j is the weight of criterion j and z_{ij} is the score of the generic solution i under criterion j, normalized by means of the equation:

$$z_{ij} = \frac{g_{ij}}{\sqrt{\sum_{i=1}^{n} g_{ij}^2}}, \forall i, \forall j. \tag{23}$$

- Identifying the positive ideal solution A^+ and the negative ideal solution A^-, calculated through the following equations:

$$A^+ = (u_1^*, \dots, u_k^*) = \{(u_{ij} | \, j \in I'), (u_{ij} | \, j \in I'')\}; \tag{24}$$

$$A^- = (u_1^-, \dots, u_k^-) = \{(u_{ij} | \, j \in I'), (u_{ij} | \, j \in I'')\}; \tag{25}$$

  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:

$$S_i^+ = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_j^*)^2}, i = 1, \dots n; \tag{26}$$

$$S_i^- = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_j^-)^2}, i = 1, \dots n. \tag{27}$$

- 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:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^*}, 0 \le C_i^* \le 1, \forall i. \tag{28}$$

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

4. Case study

The DMA design methodology proposed is applied to the literature benchmark water network called EXNET [44]. It is a large-size water network since it supplies around 400,000 consumers and the network is composed of 1,891 nodes and 2,465 pipes. The network is fed by two reservoirs and five injection nodes (well pumps). Each node is used as a data point for the clustering analysis, and is endowed with its topological features, namely geographical position, elevation and base demand.

4.1 DMA identification

To define the number of clusters, the modified k-means algorithm is applied by varying k from 2 to 15. The best number of clusters, which minimises the intra-criterion and maximizes the inter-criterion, is nine, as shown in Figure 4. Figure 5 shows the clustered network with EPANET, using latitude and longitude coordinates. This configuration results in 136 boundary pipes that are candidates to be selected as DMA entrances.

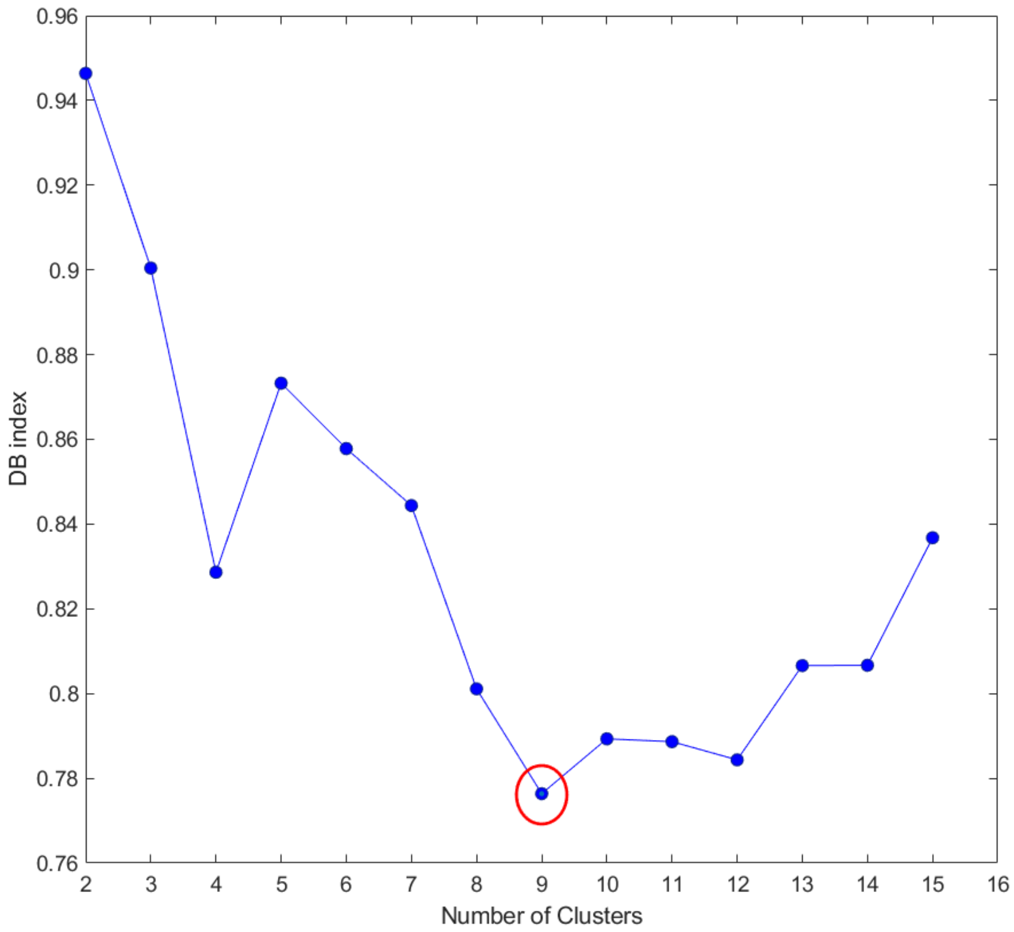

Figure 4: Davies-Bouldin index for various numbers of clusters applying modified k-means algorithm.

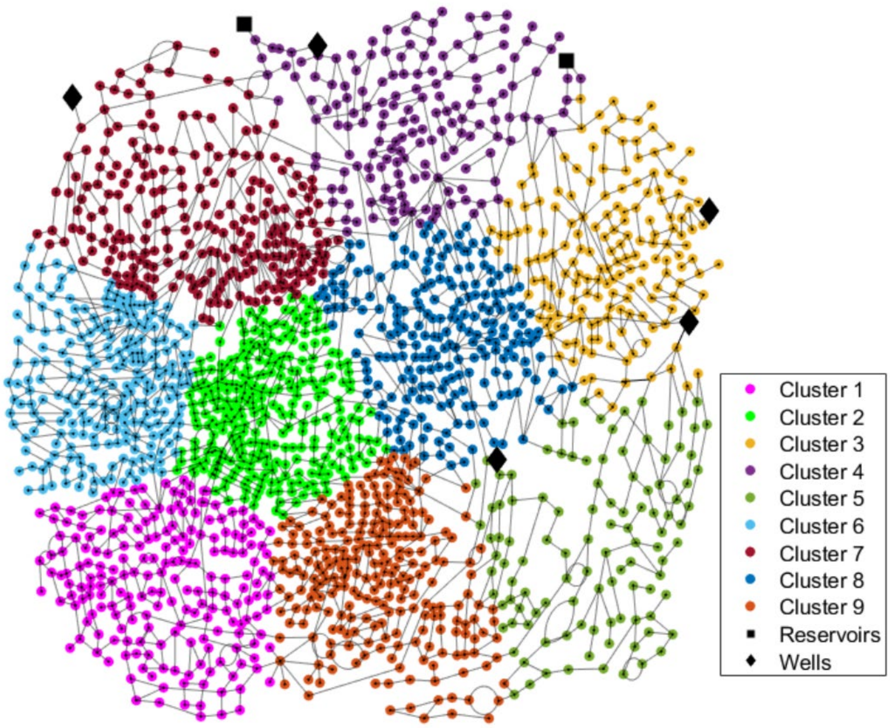

Figure 5. DMA regions identified by the modified k-means algorithm

4.2 Fuzzy AHP to identify weights for the objective functions

As previously stated, the FAHP technique is applied to determine weights of the objective functions, treated as evaluation criteria for single objective optimization. The same weights will be used later within TOPSIS. The four evaluation criteria used for the analysis (F_1, cost, F_2, lack of pressure, F_3, resilience, and F_4, water age/quality) have been pairwise compared by using the linguistic scale of Figure 2. The responsible of the safety management system of a water distribution utility was involved in such a task, given his relevant background for providing effective pairwise comparisons between pairs of criteria characterising the topic under evaluation. Tables 1 and 2 respectively report the collected linguistic evaluations and the FPCM of input for the FAHP application (the last column reporting the final vector of normalised criteria weights).

Table 1. Linguistic evaluations provided by the expert

	F1	**F2**	**F3**	**F4**
F1	-	EQ	M	M
F2	-	-	M/S	EQ/M
F3	-	-	-	EQ/M
F4	-	-	-	-

Table 2. FPCM and vector of criteria weights

$\widetilde{X}$	**F1**	**F2**	**F3**	**F4**	**weights**
F1	(1, 1, 2)	(1, 1, 2)	(2, 3, 4)	(2, 3, 4)	36.12%
F2	$(\frac{1}{2}, 1, 1)$	(1, 1, 2)	(3, 4, 5)	(1, 2, 3)	36.12%
F3	$(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$	$(\frac{1}{5}, \frac{1}{4}, \frac{1}{3})$	(1, 1, 2)	(1, 2, 3)	17.42%
F4	$(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$	$(\frac{1}{3}, \frac{1}{2}, 1)$	$(\frac{1}{3}, \frac{1}{2}, 1)$	(1, 1, 2)	10.34%

Table 3 summarises: the values of fuzzy synthetic extent for each criterion, calculated by means of formula (11); the related degrees of possibility, obtained through formula (12); and the components of the non-normalised vector of weights, achieved by formula (14).

Table 3. Synthesis of FAHP main results

Values of fuzzy synthetic extent
$S_1 = (6.00, 8.00, 12.00) \odot \left(\frac{1}{33.33}, \frac{1}{21.92}, \frac{1}{15.87}\right) = (0.18, 0.36, 0.76)$
$S_2 = (5.50, 8.00, 11.00) \odot \left(\frac{1}{33.33}, \frac{1}{21.92}, \frac{1}{15.87}\right) = (0.16, 0.36, 0.69)$
$S_3 = (2.45, 3.58, 5.83) \odot \left(\frac{1}{33.33}, \frac{1}{21.92}, \frac{1}{15.87}\right) = (0.73, 0.16, 0.37)$
$S_4 = (1.92, 2.33, 4.50) \odot \left(\frac{1}{33.33}, \frac{1}{21.92}, \frac{1}{15.87}\right) = (0.06, 0.11, 0.28)$

Degrees of possibility to compare values of fuzzy synthetic extent							
$V(S_1 \geq S_2)$	1	$V(S_2 \geq S_1)$	1	$V(S_3 \geq S_1)$	0.4822	$V(S_4 \geq S_1)$	0.2861
$V(S_1 \geq S_3)$	1	$V(S_2 \geq S_3)$	1	$V(S_3 \geq S_2)$	0.5014	$V(S_4 \geq S_2)$	0.3145
$V(S_1 \geq S_4)$	1	$V(S_2 \geq S_4)$	1	$V(S_3 \geq S_4)$	1	$V(S_4 \geq S_3)$	0.7865
Components of the non-normalised vector of weights							
$x^{*\prime}(\mathrm{F}_1) = V(S_1 \geq S_2, S_3, S_4) = \min(1; 1; 1) = 1$							
$x^{*\prime}(\mathrm{F}_2) = V(S_2 \geq S_1, S_3, S_4) = \min(1; 1; 1) = 1$							
$x^{*\prime}(\mathrm{F}_3) = V(S_3 \geq S_1, S_2, S_4) = \min(0.4822; 0.5014; 1) = 0.4822$							
$x^{*\prime}(\mathrm{F}_4) = V(S_4 \geq S_1, S_2, S_3) = \min(0.2861; 0.3145; 0.7865) = 0.2861$							

It is lastly possible to get the normalised vector of weights $W = (0.3612, 0.3612, 0.1742, 0.1034)^T$, already presented in the last column of Table 2, to verify the consistency of the FPCM. After having defuzzified the values of the matrix through the graded mean integration approach, we can affirm that the level of consistency of judgments is acceptable being the CR index equal to 0.0642, that is, within the allowed threshold of 0.08 established in [27].

4.3 Single-objective solution

PSO is run with 400 particles, using inertia weight varying from 1.2 to 0.8 and cognitive and social coefficients equal to 1.95, following the literature suggestion [35], resulting in a scenario with 55 entrances. In the present work, the number of entrances is not limited, allowing more than one entrance by DMA. Table 4 summarises the results for each objective function for the best particle.

Table 4. Objective function values for best solution of single optimization

Objective Function	Optimal solution	Original Network
F_1 (cost – US$)	120,542	-
F_2 (pressure uniformity – [ ])	89.28	92.21
F_3 (resilience [ ])	0.88	0.95
F_4 (water age index [h])	4.00	3.69

The optimized layout improves pressure uniformity by reducing the operational pressure of the system. Consequently, network resilience is reduced whereas water age grows. This occurs because, when cutting pipes, water has to follow a longer path to reach some consumers. Figures 6a and 6b present a pressure and water age surface plot for the maximum demand time.

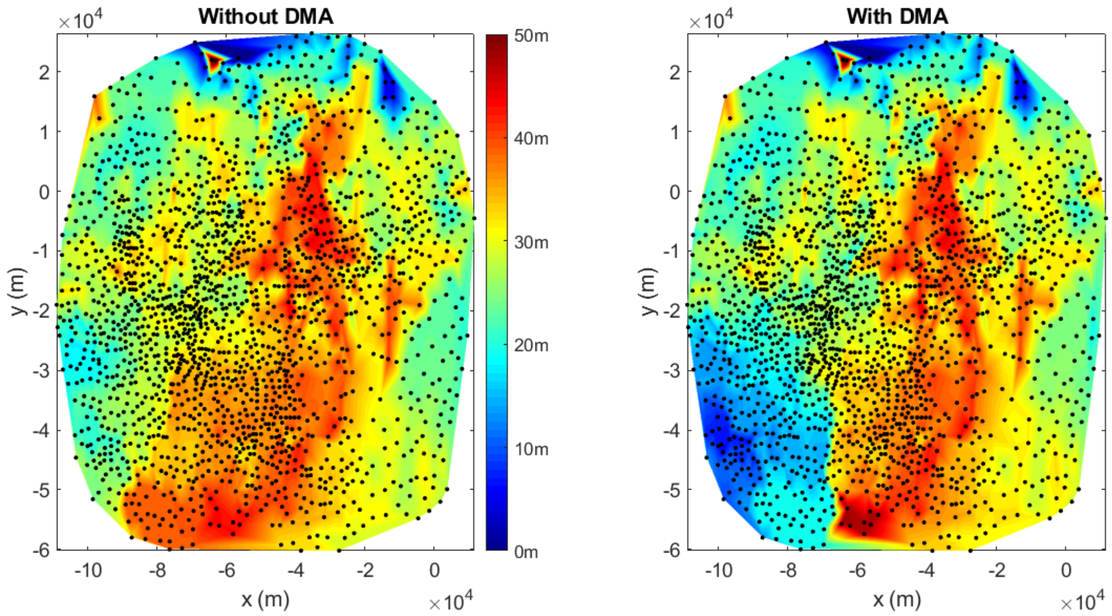

Figure 6a. Pressure surface comparing original network and optimized solution (single objective)

In terms of pressure, the south-west region is the most affected by the DMA enforcement, what significantly reduces the pressure. Broadly speaking, it is possible to appreciate pressure improvement in the west region.

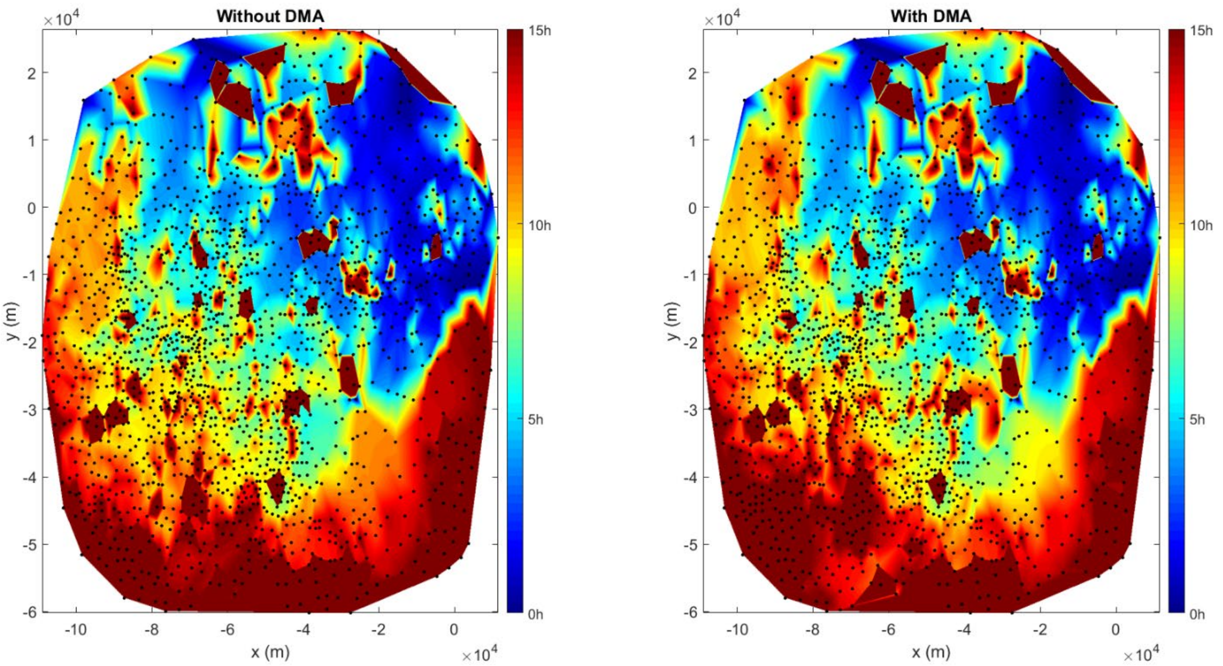

Figure 6b. Water age surface comparing original network and optimized solution (single objective)

In terms of water quality, the southern region reduces its water age. This can be observed by the light red on the optimized surface. At the same time, the water age is harmed in the western region. This relation between improvement of pressure management and water age harming is expected because, when closing a pipe to control pressure, water reaches some nodes following a longer path.

4.4 Multi-objective approach and TOPSIS ranking of the Pareto solutions

The application of MOPSO is applied considering the four abovementioned objective functions and uses 400 particles to explore the search space. The algorithm parameters take the same values as for the single-objective PSO. After several iterations, the algorithm results in a Pareto front with 7 non-dominated solutions. This is a small Pareto front, that usually has hundreds or thousands of solutions. The complexity of the hydraulic network considered, together with the many non-feasible solutions discarded, reduced significantly the number of possible solutions.

With this recognition, by applying the TOPSIS method, we want to provide a structured framework able to deal with general situations, including those cases in which the number of alternatives to be ranked is large. In the particular case study herein analysed, the TOPSIS method has been applied to rank the seven solutions belonging to the Pareto front, being the application suitable and extendable for cases characterised by higher numerousness. Obviously, cost, lack of pressure and water age (respectively F_1, F_2 and F_4) will be minimised, whereas resiliency (F_3) will be maximised. Differently from the application presented in [27], the evaluation criteria have not herein assigned the same importance, but input weights will be those obtained previously through the FAHP. Results of the TOPSIS application are reported in Table 5.

Table 5. TOPSIS results

Ranking position	**# Pareto solution**	**F$_1$** $w_1 = 36.12\%$	**F$_2$** $w_2 = 36.12\%$	**F$_3$** $w_3 = 17.42\%$	**F$_4$** $w_4 = 10.34\%$	**Closeness coefficient value**
1	6	6.30E+04	8.31E+01	3.58E+00	0.87E+00	0.9233
2	7	6.32E+04	8.31E+01	3.58E+00	0.86E+00	0.9152
3	5	6.55E+04	8.29E+01	3.56E+00	0.85E+00	0.6858
4	3	7.00E+04	8.34E+01	3.61E+00	0.79E+00	0.0998
5	1	7.02E+04	8.29E+01	3.60E+00	0.79E+00	0.0873
6	2	7.05E+04	8.28E+01	3.60E+00	0.78E+00	0.0627
7	4	7.08E+04	8.28E+01	3.60E+00	0.78E+00	0.0534

4.5 Discussion of results

The solutions in the first positions of the final ranking obtained by applying TOPSIS are characterised by higher values of the closeness coefficient, as can be appreciated by observing the last column of Table 5. Higher values of the closeness coefficient show that those alternatives have large distance to the negative ideal solution and small distance to the positive ideal solution; these ideal solutions have been previously identified within the set of input data by means of formulas (24) and (25). Similar solutions appear in Table 5, such as 1 and 2, or 6 and 7. This happens by the closeness of those solutions in the Pareto front. This ranking approach shows the interest of MCDMs to select trade-off scenarios under the considered criteria. The first solution shows the best cost while the second highest pressure uniformity and lowest resilience. That means, the best hydraulic and operation conditions will appear in the most expensive scenario. This is because more installed valves allow to reduce even more the operational pressure. The relation between 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. Comparing with the single objective solution, it is possible to observe that multi-objective solutions dominate the single objective solution, since better values for three of the four objectives are reached. The selected solution (Solution 6) has a lower cost, and better pressure uniformity and water quality index than the single objective one. As expected, this solution also exhibits a lower resilience index.

To evaluate the hydraulic and quality effects of the selected solution from the Pareto's front, Figures 7a and 7b present pressure and water age maps, and compare the chosen optimal solution with the network with no DMA structure. Pressure is significantly reduced in the south-west region of the network, as observed also for the single-objective solution. However, for the multi-objective solution, also the eastern side of the network has the pressure reduced, something is not occurring for the single-objective solution. In the case of water age, one crucial indicator of water quality in the network, the southern side of the network has increased this parameter in most of the pipes. In the northern and central areas of the network it is possible to observe a reduction on water age, thus increasing the quality of water. This happens due to the changes on the topology, which increases the flow of a set of pipes.

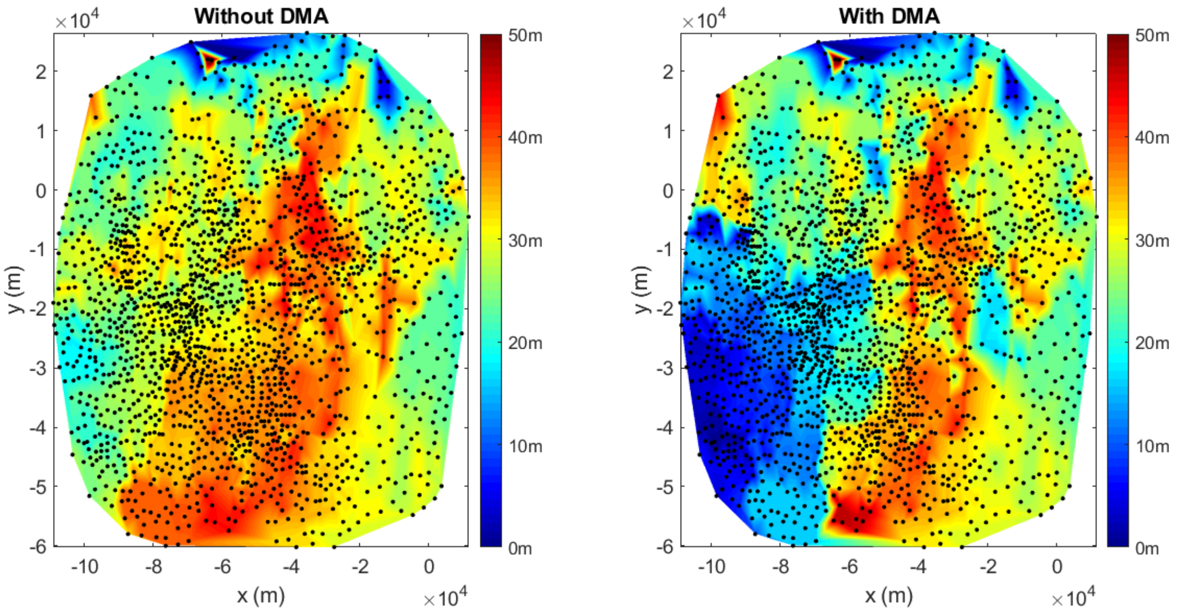

Figure 7a. Pressure surface comparing original network and optimized solution (multi objective)

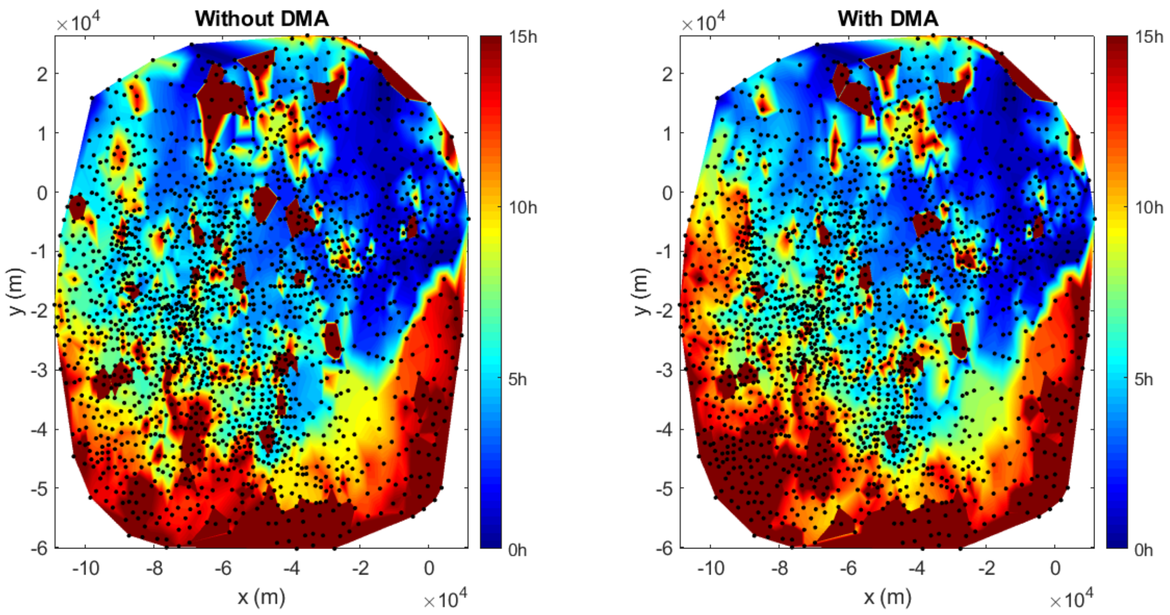

Figure 7b. Water age surface comparing original network and optimized solution (multi-objective)

Comparing single and multi-objective optimization, the average water age on the demand nodes is the same, 8.61h. This value is slight lower than the one for the original network, 8.76h. Usually, closing pipes for DMA creation increases water age, thus harming the water quality. However, in this case study, the consideration of water quality during the optimization process allows to reduce the water age, improving the water quality. In terms of pressure, the average pressure in the single and multi-objective approaches are respectively 36.85m and 34.63m, while in the original network this value is 38.98m. This comparison shows that the multi-objective approach is able to define better management for pressure, reducing the water age.

5. Conclusions

The present work proposes a fully automated algorithm for DMA design based on clustering analysis, multi-objective optimization and multi-criteria analysis, which is compared with a weighted single-objective approach. The clustering analysis is undertaken through a modified k-means algorithm evaluated under the Davies-Bouldin criterion, resulting in nine DMAs. The weighted single optimization found a feasible solution that could be implemented. However, the multi-objective optimization for entrance location is conducted with MOPSO and finds 7 non-dominated solutions in a trade-off between various objectives. In addition, an integrated MCDM approach, making use of the FAHP and TOPSIS, is applied to firstly weight objectives and to secondly rank the non-dominated solutions. The aim consists in identifying that optimal solution representing the best trade-off in fulfilling the objectives to be matched. Operational and hydraulic criteria are used to evaluate the solutions. The selected solution from TOPSIS has better hydraulic and water quality parameters with lower cost, when compared with the single-objective solution. Even though for the use case addressed, the optimization process of multi-objective results in a reduced number of Pareto solutions, the multi-level (multi-objective and multi-criteria analysis) algorithm is able to handle hundreds of solutions without high computational effort. The multi-level algorithm finds a feasible and high-performance solution, guaranteeing low cost and good efficiency of the system.

References

1. Rahman, N.A., Muhammad, N.S., Mohtar, W.H.M.W. Evolution of research on water leakage control strategies: where are we now? *Urban Water J.* 2018; 15(8): 812-826

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

3. Brentan B, Campbell E, Goulart T, Manzi D, Meirelles G, Herrera M, ... Luvizotto Jr E. Social network community detection and hybrid optimization for dividing water supply into district metered areas. *J Water Res Plan Manag*. 2018; 144(5):04018020.

4. Di Nardo A, Di Natale M. A heuristic design support methodology based on graph theory for district metering of water supply networks. *Eng Opt*. 2011; 43(2):193-211.

5. Diao K, Zhou Y, Rauch W. Automated Creation of District Metered Area Boundaries in Water Distribution Systems. *J Water Res Plan Manag*. 2013; 139:184–190.

6. Gomes, R., Marques, A.S., Sousa, J. Decision support system to divide a large network into suitable District Metered Areas. *Water Sci Technol.* 2012; 65(9): 1667–1675.

7. Giudicianni C, Di Nardo A, Di Natale M, Greco R, Santonastaso G, Scala A. Topological Taxonomy of Water Distribution Networks. *Water* 2018; 10:444.

8. Tarjan R. Depth-first search and linear graph algorithms. *SIAM J Comput*. 1972; 1:146–160.

9. Pohl IS. *Bi-Directional and Heuristic Search in Path Problems*. Ph.D. Thesis, Stanford Linear Accelerator Center; Stanford University, Stanford, CA, USA, 1969.

10. Tzatchkov VG, Alcocer-Yamanaka VH, Bourguett Ortíz V. Graph Theory Based Algorithms for Water Distribution Network Sectorization Projects. In *Water Distribution Systems Analysis Symposium; American Society of Civil Engineers*: Cincinnati, OH, USA, 2008; 1–15.

11. Lifshitz R, Ostfeld A. Clustering for Analysis of Water Distribution Systems. *J Water Res Plan Manag*. 2018; 144:04018016.

12. 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.

13. Ciaponi C, Murari E, Todeschini S. Modularity-Based Procedure for Partitioning Water Distribution Systems into Independent Districts. *Water Resour Manag*. 2016; 30:2021–2036.

14. Chatzivasili S, Papadimitriou K, Kanakoudis V, Patelis M. Optimizing the formation of DMAs in a water distribution network applying Geometric Partitioning (GP) and Gaussian Mixture Models (GMMs). In *Multidisciplinary Digital Publishing Institute Proceedings*, 2018; 2(11):601.

15. Laucelli DB, Simone A, Berardi L, Giustolisi O. Optimal Design of District Metering Areas for the Reduction of Leakages. *J Water Res Plan Manag*. 2017; 143:04017017.

16. Araujo LS, Ramos H, Coelho ST. Pressure control for leakage minimisation in water distribution systems management. *Water Res Manag*. 2006; 20(1):133-149.

17. Creaco E, Franchini M, Alvisi S. Optimal placement of isolation valves in water distribution systems based on valve cost and weighted average demand shortfall. *Water Res Manag*. 2010; 24(15):4317-4338.

18. Van Laarhoven P, Pedrycz W. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets Syst*. 1983; 11(1):199–227.

19. Saaty TL. *The Analytic Hierarchy Process: Planning, Priority Setting and Resource Allocation*, McGraw-Hill, New York, 1980.

20. Zadeh L. Fuzzy sets. *Inf Control*. 1965; 8:338–353.

21. Klir GJ, Yuan B. *Fuzzy sets and fuzzy logic – Theory and applications*. Prentice Hall PTR, Upper Saddle River, New Jersey, 1995.

22. Kubler S, Robert J, Derigent W, Voisin A, Le Traon Y. A state-of the-art survey & testbed of fuzzy AHP (FAHP) applications. *Expert Syst Appl*. 2016; 65:398–422.

23. Büyüközkan G, Çifçi G. A combined fuzzy AHP and fuzzy TOPSIS based strategic analysis of electronic service quality in healthcare industry. *Expert Syst Appl*. 2012; 39(3):2341–2354.

24. Kaya T, Kahraman C. An integrated fuzzy AHP–ELECTRE methodology for environmental impact assessment. *Expert Syst Appl*. 2011; 38(7):8553–8562.

25. Ka B. Application of Fuzzy AHP and ELECTRE to China Dry Port Location Selection. *Asian J Shipp Logist*. 2011; 27(2):331–353.

26. Behzadian M, Otaghsara SK, Yazdani M, Ignatius J. A state-of the-art survey of TOPSIS applications. *Expert Syst Appl*. 2012; 39(17):13051-13069.

27. Hwang C-L, Yoon K. *Multiple-Attribute Decision Making. Methods and Applications. A state-of-the-Art Survey*. Springer-Verlag, Berlin, Heidelberg, New York, 1981.

28. Brentan B, Carpitella S, Izquierdo J, Luvizotto Jr E, Meirelles G. A multi-objective and multi-criteria approach for district metered area design: water operation and quality analysis. In *Proceedings of the 21th International Conference on Mathematical Modeling in Engineering & Human Behaviour*, Valencia, Spain, 2019; 110-117.

29. Novarini B, Brentan BM, Meirelles G, Luvizotto Jr E. Optimal pressure management in water distribution networks through district metered area creation based on machine learning. *RBRH*. 2019; 24.

30. Davies DL, Bouldin DW. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, *PAMI*-1(2), 1979; 224-227.

31. Brentan BM, Campbell E, Meirelles GL, Luvizotto Jr E, Izquierdo J. Social network community detection for DMA creation: criteria analysis through multilevel optimization. *Math Probl Engrg*. 2017: 9053238.

32. Alhimiary H, Alsuhaily R. Minimizing leakage rates in water distribution networks through optimal valves settings. In *Proc World Environ Water Res Congress*, 1–13. Reston, VA: ASCE, 2007.

33. Todini E. Looped water distribution networks design using a resilience index based heuristic approach. *Urban water* 2000; 2(2):115-122.

34. Saldarriaga J, Bohorquez J, Celeita D, Vega L, Paez D, Savic D, ..., Kapelan Z. Battle of the water networks district metered areas. *J Water Res Plan Manag*. 2019; 145(4):04019002.

35. Eberhart R, Kennedy J. Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, 1995; 4:1942-1948.

36. Coello CC, Lechuga MS. MOPSO: A proposal for multiple objective particle swarm optimization. In *Proceedings of Congress on Evolutionary Computation*. CEC'02 (Cat. No. 02TH8600), 2002; 2:1051-1056. IEEE.

37. Zimmermann HJ. *Fuzzy set theory and its applications*. Kluwer, Boston, MA, 1985.

38. Carpitella S, Ocaña-Levario SJ, Benítez J, Certa A, Izquierdo J. A hybrid multicriteria approach to GPR image mining applied to water supply system maintenance. *J Appl Geoph*. 2018; 159:754–764.

39. Durán O, Aguiló J. Selección de máquinas de control numérico usando Fuzzy AHP. *Espacios* 2006; 27(1). https://www.revistaespacios.com/a06v27n01/06270141.html.

40. Chang D-Y. Application of the extent analysis method on fuzzy AHP. *Eur J Oper Research*. 1996; 95:649–655.

41. Kutlu AC, Ekmekçioğlu M. Fuzzy failure modes and effects analysis by using fuzzy TOPSIS-based fuzzy AHP. *Expert Syst Appl*. 2012; 39(1):61–67.

42. Saaty TL. A scaling method for priorities in hierarchical structures. *J Math Psych*. 1977; 15(3):234–281.

43. Benítez J, Carpitella S, Certa A, Ilaya-Ayza AE, Izquierdo J. Consistent clustering of entries in large pairwise comparison matrices. *J Comput Appl Math*. 2018; 343:98-112.

44. Farmani R, Savic DA. Walters GA. "EXNET" Benchmark Problem for Multi-Objective Optimization of Large Water Systems. *Modelling and Control for Participatory Planning and Managing Water Systems*, IFAC workshop, Venice, Italy, 2004.