DMA segmentation and multi-objective optimization for trading-off water age, excess pressure and pump operational cost in water distribution systems

By Mohamad Zeidan¹, Pu Li², and Avi Ostfeld, Fellow ASCE³

Abstract
This study presents a heuristic multi-objective approach for segmenting and operating water distribution systems (WDSs). The methodology employs a two-pronged strategy: the first is a heuristic method for dividing the network into clusters (i.e., district metering areas) based on connectivity analysis. The second is the application of the evolutionary multi-objective optimization method NSGA-II for trading off the operational cost, excess pressure (serving as a proxy to leakage reduction), and water age (acting as a surrogate to water quality) in the WDS. Three example applications of increasing complexities with various clusters partitioning are explored, showing a clear tradeoff between the objectives. This study introduces an unprecedented heuristic approach for jointly solving the multi-objective problem, under given system partitioning. However, by enforcing a-priori clustering formation (rather than including it in the optimization), optimality, completeness and precision are compromised in favor of computational speed and effort. Thus, additional sensitivities need to be conducted outside of the optimization as for the clusters impact. Challenges of extending this study are in embedding the clusters formations in the optimization considering other objectives such as residual capacity, developments of other optimization frameworks outside of the generic link of simulation-optimization, and uncertainty inclusion (e.g. in demands). All data and codes are attached as supplementary to this work for allowing full replications and comparisons.

¹ MSc, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel; PH: +972-4-8292630; E-mail: ar5medes@gmail.com
² Professor, Process Optimization Group, Institute of Automation and Systems Engineering, Technische Universität Ilmenau; Ilmenau 98693, Germany; PH: +49-367-769-1423; E-mail: pu.li@tu-ilmenau.de
³ Professor (corresponding author), Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel; PH: +972-4-8292782; E-mail: ostfeld@technion.ac.il

Keywords: water distribution systems, leakage, water age, design, optimization, multi-objective, genetic algorithm.
Introduction

Water leakage remains a major cause of water loss in water distribution systems, particularly in developing countries. In addition to self-evident economic losses, water leakage poses environmental, sustainability and health risks (Puust et al., 2010). On the other hand, water consumption in many developed countries is decreasing due to population decline, leading to increased water age (e.g., Japan, Netherlands, and Central Eastern Europe). According to the United Nations’ World Population Prospects, it is estimated that between 2019 and 2050, 55 countries or areas are expected to see their populations decrease by at least one percent. The largest relative reductions in population size over that period, with losses of around 20 percent or more, are expected in Bulgaria, Latvia, Lithuania, Ukraine, and the Wallis and Futuna Islands (United Nations 2019; Nations U. 2015; Ubarevičienė et al., 2016; Haartsen and Venhorst 2010).

Water age is widely used as a surrogate for water quality degradation in WDSs. Wang et al. (2014) showed that water age directly affects water chemistry, causes depletion of disinfectant residual, increases total organic carbon, and decreases dissolved oxygen. Water age diminishes water quality as retention time increases due to more interactions between water within the pipe bulk flow and its wall, including attached biofilms, thus raising the likelihood of contaminant formulations.

In addition, as noted by Wu et al. (2005), water with higher age may have little or no residual disinfectant due to substance decay and reactions with network materials. Consequently, the disinfection efficiency decreases, and biological activity that may yield unpleasant tastes and odors increases. Moreover, an increase in water age can be
associated with the occurrence of opportunistic pathogens in a WDS (Wang et al., 2014; Murray et al., 2009; Klise et al., 2015).

Management problems of WDSs involve competing objectives such as minimizing design and operational costs, maximizing reliability, minimizing risks and minimizing deviations from water quantity, pressure, and water quality objectives. As a consequence, such problems are inherently of a multi-objective nature. For such problem, there is no single optimal solution, but a set of compromising solutions, which form a Pareto optimal solution set. The consideration of multiple objectives for optimizing water distribution systems is preferable to using a single one, since a broader range of solutions is explored, thus enabling more possible realistic decisions.

In attempt to effectively overcome leakage problems, utilities consider partitioning distribution networks into district metered areas (DMAs) (Savic and Ferrari, 2014). In addition, the subdivision of a looped WDS into numerous confined DMAs improves water distribution networks management by simplifying the water balance calculation and reducing water security risks, since contaminants’ movement in the system is limited.

In this study, a heuristic multi-objective approach is introduced for the automatic creation of DMAs, and pumps-scheduling optimization in WDSs. The two-pronged approach suggested in this research, automatically define clusters boundaries based on the networks topology, using the community structure algorithm (CSA). Furthermore, it disperses the connections and mutual interaction between said clusters by closing part of the connecting pipes, in accordance with the networks connectivity, thus generating
DMAs boundaries and feedlines. This stage is accompanied by scheduling the pumps’ operating patterns, to competitively reduce operational cost, leakage, and water age using the NSGA-II (Deb et al., 2002) multi-objective genetic algorithm linked with EPANET. The proposed methodology is demonstrated through three case studies of increasing complexity.

**Literature review**

This literature review sets the background for this study by briefly describing water distribution systems optimization, leakage in water distribution systems, water age, and district metered areas.

**Water distribution systems optimization**

The most common design optimization approach involves minimizing capital costs associated with water distribution system infrastructure (Zecchin et al., 2007) often defined through pipe diameters cost, augmented by a multi-objective optimization, such as system’s reliability (Kapelan et al., 2005; Prasad and Park, 2004), or pressure deficit (Farmani et al., 2005; Perelman et al., 2008). However, as stated by Walski (2001), cost does not necessarily reflect the most important criteria whereby WDSs should be evaluated. Moreover, numerous alternative designs can be considered equal, given the uncertainty of cost and benefit estimates. In operational optimization of WDSs, the operational cost minimization prevails, attained by tank level control (Van Zyl et al., 2004) or pump scheduling (Ulanicki et al., 2007; Carrijo et al., 2004; Kurek and Ostfeld, 2013; Price and Ostfeld, 2016).
The utilization of multi-objective optimization is enabling the operator the ability to select favorable solutions from multiple selections (Biscos et al., 2003). This kind of inclusion may lead to suboptimal solutions regarding the parameters of interest, such as water age and leakage. Fu et al. (2012) recognized that water age is uncorrelated to levels of capital and operating costs, emphasizing that solutions with a wide range of water age can be found at any cost level. Their findings also showed that capital and operational costs have distinct relationships with water age and leakage, and thus should be studied as separate objective functions.

Genetic algorithms (Holland 1975; Goldberg 1989) enables the inclusion of pumps, control valves, consumer demands, storage and other devices and considerations for the overall design/operation of water distribution systems. The non-dominated sorting genetic algorithm-II (NSGA-II) (Deb et al., 2002) was selected herein as the optimization algorithm, as it is considered one of the more successful multi-objective evolutionary algorithms (MOEAs) for WDSs management problems (Wang et al., 2015).

**Leakage in water distribution systems**

The most used definition of leakage is the amount of water that outflows from a pipe network by means other than through a controlled action (Ofwat 2018). Among the factors that bolster leakages are bad pipe connections, internal or external pipe corrosion, mechanical damage, ground movement, high water pressure, damage due to excavation, pipe age, winter temperature, defects in pipes, ground conditions and poor quality of workmanship (Puust et al., 2010).
Commonly, leakage is related to pressure as described by the orifice equation:

$$q = AC_d \sqrt{2gh}$$  \hspace{1cm} (1)

where $q$ is the flow rate, $A$ is the orifice area, $C_d$ is a flow rate coefficient, $g$ is the gravity acceleration constant, and $h$ is the pressure head over the orifice.

When applied to pipe leakage, equation (1) gets the form of flowrate per unit length of pipe:

$$q = \beta h^\alpha$$  \hspace{1cm} (2)

where $\beta$ is the leakage coefficient, and $\alpha$ quantifies the leakage rate. Several field studies (e.g., Van Zyl and Clayton, 2007) reported that $\alpha$ can vary between 0.5 and 2.79 with a median of 1.15. Hence, leakage in water distribution systems is much more sensitive to pressure than might be intuitively assumed.

Significant research has been focused on the optimal placement and management of network pressure control devices (i.e., pumps and control valves) to reduce leakage (Vairavamoorthy and Lumbers, 1998; Araujo et al., 2006; Liberatore and Sechi, 2009; Ali 2015, Nicolini and Zovatto, 2009; Creaco and Pezzinga, 2015a, b). Most methods mainly suggest introducing pressure reducing valves (PRVs) to manage leakage in the system (Nicolini et al., 2011). For a higher solution flexibility, Saldarriaga et al., (2016) incorporated pump optimization in addition to PRVs placement. However, in certain cases, the tendency to pump at low night tariffs may contradict the objective of leakage minimization (Price and Ostfeld, 2014).

Other studies show that incorporating pipe replacements can be an effective way to reduce leakage in water distribution networks (Creaco and Walski, 2017). Additionally,
incorporating micro hydropower turbines within water supply networks has been shown as a viable option for pressure reduction which in turn reduces water losses from leakages (Carravetta et al., 2014; Gallagher et al., 2015; Fecarotta et al., 2015; Sitzenfrei and Rauch, 2015; Corcoran et al., 2016; Bonthuys et al., 2020).

In this study one of the objective functions is the minimization of network pressure, towards minimizing network leakage.

**Water age**

Water age is the average time required for water to travel from a source to a particular location within a WDS (Letterman 1999), which can differ in a looped network.

Following Murray et al. (2009), the average system water age $f_{wa}$ can be quantified as:

$$f_{wa} = \frac{1}{nn \times T} \sum_{i=1}^{nn} \sum_{t=1}^{T} A_{i,t}$$ (3)

where $nn$ are the total number of network nodes, $T$ is the total hydraulic simulation time, and $A_{i,t}$ is the water age at junction $i$ at time step $t$.

Klimek et al. (2015) suggested that water age is only meaningful when it is above a given threshold that indicates potential water quality problems. Machell and Boxall (2014) investigated the correlation between simulated water age and water quality characteristics, suggesting water age as a useful surrogate for water quality. Quintiliani et al. (2019) and Schwetschenau et al. (2019) incorporated water age as a water quality surrogate in WDS optimization.
Water age management is an important subject in the operation of WDSs, especially for regions with decreasing water demands, which causes water age to increase. In this work, the management of water age and pressure are considered simultaneously for WDSs which are partitioned into district metered areas.

**District metered areas**

A district metered area (DMA) is a subsystem of the water distribution system with a defined boundary (Morrison et al., 2007). Defining district boundaries is the basis for DMA design. Yet, the intricacy of WDSs complicates this task, especially within large-scale, highly looped systems. As a result, in practice, DMA design largely remains an empirical, manual process, although attempts were made in the literature for DMA automated construction, as briefly reviewed below.

Swamee and Sharma (1990) developed a method for decomposing multisource WDSs with predefined locations and influence zones of all water sources into single-source subsystems. Other studies developed topological approaches for defining DMA boundaries based on graph theory and machine learning (Clauset et al., 2004; Tzatchkov et al., 2007; Deuerlein 2008; Izquierdo et al., 2009; Nardo and Natale, 2010; Herrera et al., 2010; Perelman and Ostfeld, 2011; Diao et al., 2013; Scibetta et al., 2014; Giustolisi and Ridolfi, 2014; Ferrari et al., 2014; Giudicianni et al., 2020). Empirical and heuristic attempts for guiding DMAs design considering criteria such as size, pressure, leakage level, water quality, cost, reliability, and others were also suggested (Morrison et al., 2007; Grayman et al., 2009; Savic and Ferrari, 2014; Vasilic et al., 2020).
Important to mention herein is the 2016 battle of the water networks district metered area (BWNDMA) (Saldarriaga et al., 2019) competition on DMA formations for an existing network. The competition involved design requirement constraints related to cost, pressure uniformity, water quality, and other network and demand related restrictions. Various approaches were proposed by the groups ranging from engineering judgment to topological analysis and optimization methods (Gilbert et al., 2017; Salomons et al., 2017; Martínez-Solano et al., 2018; Rahman and Wu, 2018; Rahmani et al., 2018; Pesantez et al., 2020).

One of the most exciting outcomes of the BWNDMA is the comparison between the performances of the group methods ranging from advanced DMA partitioning computerized optimization algorithms to pure engineering judgement. Surprisingly enough, engineering judgement prevailed. The ability to even find a feasible DMA partitioning was extremely difficult, and only engineering experience and knowledge succeeded to overcome this barrier, as stated in Saldarriaga et al. (2019): "at the end, a person or persons had to decide whether the optimization algorithm’s solution was good enough to solve the problem because some parts of it cannot be modeled or simulated properly." In this study DMAs are selected using an engineering judgement based rational, as will be further elaborated below.

The remaining of this study includes a description of the model formulation, the solution methodology, and analysis of three example applications.
Model formulation

The model formulation includes a description of the stages utilized for DMAs creation: WDS mapping, community structure identification, dendrogram cutting, and boundary valve identification, followed by an outline of the selected multi-objective optimization method for trading-off operational cost, pressure, and water age.

Mapping the water distribution system

The water distribution system is mapped onto an undirected graph \( G = (V, E) \), in which the vertices \( V \) represent the consumers, sources, and tanks, and the edges \( E \) represent the connecting pipes, pumps, and valves (Perelman and Ostfeld, 2011).

An adjacency matrix (a square matrix used to represent a finite graph, where the elements of the matrix indicate whether pairs of vertices are adjacent) is used to represent the network’s finite graph. The undirected graph is sufficient for this type of topological analysis; thus, the graph is undirected and the adjacency matrix is symmetric.

Community structure identification

The algorithm proposed for WDS segregation by Clauset et al. (2004) is used in this research to subdivide the graph \( G \) into several clusters. This algorithm was selected since its run-time efficiency excelled that of competing algorithms (Diao et al., 2013). The algorithm proved itself to be fast and efficient, detecting communities and cutting dendrogram as desired.
Community structure is defined as the gathering of vertices into groups such that a higher density of edges exists within groups, rather than between them. The algorithm uses the modularity $Q$ to quantify the quality of the graph division into communities (Novák et al., 2010). In this study, the modularity is considered only as an initial state estimation value to direct the division process, rather than a conclusive value, and does not dictate the number of the clusters.

Detaching the communities’ selection from the optimization using a heuristic method is proposed for excluding the clusters stage partitioning from the genetic algorithm scheme. This substantially simplifies the entire optimization framework and leaves only the pump operation schedule as decision variables.

Modularity is a property of the network and its proposed division into communities. Considering two merging vertices/communities $v$ and $w$, modularity is defined as:

$$ Q = \frac{1}{2m} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v c_w) $$

(4)

where: $A_{vw} =$ element of the network’s adjacency matrix; thus:

$$ A_{vw} = \begin{cases} 1 & \text{If vertices } v \text{ and } w \text{ are connected} \\ 0 & \text{Otherwise} \end{cases} $$

(5)

Suppose the vertices are divided into communities such that vertex $v$ belongs to the community $c_v$; the parameters $c_v$ and $c_w$ represent two different communities. $k_v =$ degree of vertex $v$, defined as the number of edges connected to that vertex.

$$ k_v = \sum_w A_{vw} $$

(6)

$m =$ the total number of edges in the network; where:

$$ m = \frac{\sum_{vw} A_{vw}}{2} $$

(7)

$\delta(c_v c_w) =$ a function defined as:

$$ \delta(c_v c_w) = \begin{cases} 1 & \text{If } c_v = c_w \\ 0 & \text{Otherwise} \end{cases} $$
\[
\delta(c_v c_w) = \begin{cases} 
1 & \text{Otherwise} \\
0 & \text{otherwise}
\end{cases}
\] 

(8)

The algorithm proposed by Clauset et al. (2004) utilizes a greedy optimization scheme, starting with each vertex as a single community and repeatedly merging two communities whose union results in the largest increment to the modularity \( Q \). This step is repeated until there is only one community left, which contains all the graph junctions.

Consider a network with \( n \) vertices. After \((n - 1)\) such joins there will be only one community left, and the algorithm terminates. However, considering that the graph of modularity \( Q \) has a single peak over the course of the algorithm (Diao et al., 2013), the method could stop as soon as the largest modularity increment from merging two communities \( i \) and \( j \) is zero or negative, since there is no potential increment to the modularity value. In other words, joining more communities would not lead to a better division.

The CSA merging process is illustrated through a simple 7-junction network displayed in Fig. 1, where the network was divided into 2 and 4 clusters. The modularity value throughout the algorithm is presented as a graph in Fig. 2 (b), where points A and B indicates the modularity value for clustering the network into 4 and 2 clusters, respectively. The dendrogram is illustrated in Fig. 2 (a), where the points A and B mentioned earlier are presented as two cuts at the dendrogram branches, referred to as cross sections. Each cross section defines where the CSA ceased, and the joint nodes beneath it.
**Dendrogram cutting**

A dendrogram is a tree diagram, frequently used to illustrate the arrangement of clusters produced by a hierarchal partitioning. In this study, the dendrogram describes the categorized community structure of a water distribution network. The dendrogram can assist decision-makers in determining the number of required clusters, and/or the minimum/maximum number of vertices allowed in each cluster. Different clustering formations can be generated for management purposes associated with various objectives such as cost, reliability, and redundancy. The dendrogram can be cut at any branch, creating different clusters layout with different numbers of internal vertices, as shown in Fig. 1.

In this study the number of clusters is selected a priori through topological considerations, yet with conducting sensitivities on these selections, as described through the example applications.

**Boundary valve identification**

Once clusters’ boundaries are defined, their mutual interconnections are minimized for forming district metering areas (DMAs). Such DMAs can result an improved control on contaminants spread and pressures.

The following steps are taken in this process:

In step 1, identifying the bridge-pipes (i.e., pipes which connect two different clusters on each end). This is accomplished through exploring the nodes connecting links index (NCLI) (a matrix containing each edge and its pair of nodes) and by examining the
dendrogram graph [Fig. 2 (a)]. This simple procedure dramatically decreases the optimization problem dimension, as will be further elaborated below.

In step 2 the bridge-pipe are sorted in ascending order according to their diameters; hence, the bridge – pipe with the smallest diameter will be the first in the list and the largest last.

In step 3, the algorithm goes through the now sorted bridge-pipe list, and examines each pipe starting from the smallest (first in list). Each iteration, the initial status of the selected pipe is modified from 1-open to 0-closed (i.e., eliminating the pipe from the network). Thereafter, a breadth-first search (BFS) is applied to the residual network for inspecting its connectivity. In case eliminating the pipe leads to multiple undirected graphs (i.e., not all communities are connected to each other), its status is altered from 0-closed back to 1-open, and the algorithm continues to next pipe in the list.

The algorithm ceases, once all pipes in the list are checked, resulting in a new network layout. The algorithm flowchart scheme is given in Fig. 3. In other words, the algorithm initiates by closing the smallest pipes in the system as long as all the junctions remain connected to each other in one graph. The rationale behind this approach is to feed the DMAs using the biggest pipes in the system, for reducing velocities and head-losses in the system. The outcome of this step is the networks’ spanning tree with largest diameters.

**Multi-objective optimization**

The goal in a multi-objective optimization scheme is to find from all the problem
feasible solutions, the set which yields optimal values with respect to all the considered 358
objective functions. This set of solutions is termed the Pareto optimal solution set or 359
the non-dominated solution set. Each solution in the Pareto optimal set is optimal in the 360
sense that it is not possible to improve one objective without making at least one of the 361
others worse.

The multi-objective optimization problem solved in this study is the minimization of 364
the system’s operational cost, pressure, and water age, defined as:

\[ C = Ec \sum_{j=1}^{k} \sum_{i=1}^{T} \phi_{j,i} p_{j,i} \]  

(9) 366

where: \( C \) is the operational cost in $, \( Ec \) the electricity cost in $/kWh, \( \phi \) tariff coefficient 367
unitless, \( P \) is the pump's power in kWatt, \( k \) is the number of pumps, and \( T \) is the number 368
of time steps during the simulation.

\[ \bar{H}_p = \frac{\sum_{j=1}^{n} \sum_{i=1}^{T} H_{p_{j,i}}}{n} \]  

(10) 370

where: \( \bar{H}_p \) is the networks average pressure head, \( H_{p_{j,i}} \) is the head pressure for node \( j \) at 371
time step \( i \) in meters, and \( n \) represents the nodes number.

\[ \bar{WA} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{T} WA_{j,i}}{n} \]  

(11) 373

where: \( \bar{WA} \) is the networks average water age, and \( WA_{j,i} \) is the water age for node \( j \) at 374
time step \( i \) in hours.

The decision variables in this study are the pump speed coefficients for each of the 377
operational time steps, including zero for shutdown. The change in pump speed will be 378
reflected in the pump power as well. It is assumed that there are adjustable frequency 379
drives (AFDs) attached to each of the pumps, enabling control over the pump’s speed and power.

For the first two case studies, explored below, all pumps work according to the same pattern; therefore, 24 decision variables (for each case study) are representing a 24 hours pattern for all the pumping stations. In the third case study, each pump station (consisting of three pumps each) operates according to its own pattern, hence, with 48 decision variables in total.

While retaining a minimum pressure of 30 meters at all demand nodes, and a proper tank level cycle, the optimization was accomplished through utilizing the multi-objective genetic algorithm (NSGA-II) provided in Matlab®. It is essential to mention that there is no guarantee that the solution found will be the most accurate or optimal for the given problem, rather a "good enough" solution in most cases.

The GA parameters such as population size and maximum number of generations have a significant impact on the optimization process, its outcomes, and running durations. For instance, as further described below, a population size of 50, with 30 generations was sufficient to solve the small M1 network at a running time duration of about 25 minutes, where for the Wolf network a population of 100, with 80 generations was required, resulting a running time of about 360 minutes.

**Example applications**

In this section, the methodology is demonstrated on three WDSs example applications of increasing complexities. Moreover, since a-priori topological based clustering
formation was imposed, additional sensitivities need to be conducted outside of the optimization as for the clusters impact on the system’s performance.

It is important to note that the networks in all case studies are not built with a spine main topology, but rather in a more interconnected way, so as to challenge the developed model with real complicated, large-scale, highly looped systems. EPANET is used to simulate the system’s behavior. All system’s input files are provided as supplementary data.

The investigated networks are:

(1) M1 - a simple network.

(2) MMOD - a modified version of the MOD network.

(3) Wolf – a modified version of the Colorado Spring network.

**Example 1 – M1 – a simple network**

The network of the first example is a small illustrative system, for testing and demonstrating the proposed model.

The system layout is given in Fig. 4 (a). It consists of one constant head source, a pumping station, a storage tank, and 38 consumers. A demand pattern P1 was defined as fractions of the max demand, as well as electricity tariff variations (with a base electricity tariff of 0.16 $/kw-hr) for four-time durations during a day (Tables 1 and 2).
Following the community structure scheme, the network’s highest modularity value (i.e., the best topological partition) is achieved when it is divided into six clusters, as presented in Fig. 5 (a).

As the method initiates [i.e., the right side of the horizontal axis in Fig. 5 (a)], 39 communities are defined, as there are 39 vertices, thus every vertex is a community. The modularity value is set to zero. When the algorithm progresses, additional vertices are merged, decreasing the number of communities, and increasing their modularity, until six communities are left. The modularity value slightly decreases at five communities and drops dramatically at four communities. As the algorithm terminates [i.e., the left side of the horizontal axis in Fig. 5 (a)], a one community remains with a modularity value of zero. The network will thus be segregated into six clusters, as shown in Fig. 6 (a).

After the community’s identifications, it is required to isolate the communities from one another through minimizing their interconnections. In this example, ten pipes were marked as bridge pipes (i.e., pipes which connect different communities) and the initial status of five of them were altered from one (open) to zero (closed) as detailed in Table 3. The closed pipes (marked with red X) and the clusters boundaries are shown in Fig. 6 (a).

The multi-objective genetic algorithm is applied to the network layout, resulting in a three-dimensional cloud of points, creating the 3D Pareto front as given in Fig. 7. Each point at the Pareto front represents a non-dominated solution of the three objective functions of: (1) operational cost, (2) mean pressure, and (3) mean water age. The 24-
hour solution vector can be traced back for each such point, thus retrieving its corresponding operational pattern. This operational pattern can then be implemented into the system for modeling the WDS hydraulics. The pump’s power and tank’s water level retrieved from one of the Pareto front solutions are presented in Fig. 8. Note that the pumps power increases between 19-24 hours when the energy tariff is low, while decreasing at 6-12 and 12-18 hours when the energy tariff is high.

The interconnections between water tanks and pumps allows head and pressure flexibility in system operation. Since the tank acts as a buffer by receiving the excessive pumped water, it allows pumping water at higher flowrates than the current demand. Therefore, different solutions with different flowrates can be attained in the system while meeting consumer demands.

Fig. 9 describes the mean pressure in the system as a function of operational cost. Observing Fig. 9, it looks counterintuitive to get less pressure when energy cost increases. The reason for that is that increasing the pump’s power results in higher velocities and thus greater head losses in the network. Since M1 is a small network with few pipelines, where five of them were closed in the community isolation stage, just few available flow paths remained to provide the required flows. As a result, velocities increase causing head losses to rise as well.

Following the same rational, the more money invested in energy, the higher the water velocity in the pipes, resulting a reduction in water age (Fig. 10). Although closing some pipes could increase the flow paths between the source and the consumers, in this case, the velocity increment acts as a counterbalance. It is important to explore the
change of the average water age in the system to ensure that there is no significant increase, which would negatively affect the system water quality. In this example, the average water age decreased as shown in Fig. 10, yet the drop is negligible from an engineering perspective.

**Example 2 – MMOD**

The second example (MMOD) is a modification of the MOD network downloaded from the University of Exeter benchmark networks cite. The MMOD network consists of 270 junctions and its layout is presented in Fig. 4 (b). A storage tank of 30 meters in diameter with a maximum level of 20 meters was added to its center. In addition, a pump was attached at each of its four constant head sources. A demand pattern (P1) was defined as fractions of the max demand, as well as electricity tariff variations (Tables 1 and 2).

The CSA was implemented on the MMOD system, generating the modularity chart shown in Fig. 5 (b), where the modularity is maximized at 19 communities. As the entire network consists of 203 junctions such a division will create communities of about ten nodes each. Moreover, multiple pipelines are eliminated; hence, the optimization process would probably fail to supply the min-required pressure and consequently fail to reach a feasible solution.

To further explore the solutions sensitivities to the modularity figure selection, the number of communities (i.e., clusters) was manually set to three, four and five. The cluster-numbers were selected manually, not considering the maximum modularity value, as of the above reasoning. In other words, the CSA was utilized only as a
partitioning method for the chosen clusters-number, regardless of the highest modularity value.

Results from the community structure for five clusters are shown in Fig. 6 (b). Once the communities are identified, there is a need to isolate them from each other and sparse their feedlines. The resulted bridge-pipes between the five clusters are detailed in Table 3 and marked with red in Fig. 6 (b).

Fig. 11 presents the tradeoff between minimum pressure and cost for three, four, and five clusters. Unlike the M1 example, the current network has sufficient redundancy to provide the desired flows, thus the head losses in the network are insignificant compared to the added pressure.

Fig. 12 presents the tradeoff between mean water age and operational cost. It can be seen from Fig. 12 that the more communities are generated, the higher the water age for the same operational cost. This happens since additional communities increase the number of closed pipes in the isolation process. This results in fewer edges available to flow through, leading to longer paths from the sources to the consumers, hence a higher water age. The Pareto front for three clusters scenario is allocated at the low operation cost range, unlike the case of four and five clusters. Since it is a Pareto front, only the non-dominated solutions are preserved. Nevertheless, it should be noted here (as in M1) that the differences in water age are small from an engineering perspective.

Supplementary Figure S1 provides 3D Pareto front surfaces for three, four, and five clusters partitioning.
Example 3 – Wolf

The Wolf system layout is presented in Fig. 4 (c), while Table 2 shows the consumers demand pattern (P2). The system consists of four constant head sources with six pumping stations and 1782 junctions. No storage tanks exist in this example.

Sensitivities are explored for 5, 10, 20 and 46 communities, where 46 is the highest partitioning modularity value, as shown in Fig. 5 (c). An example of the system segregation into ten clusters is presented in Fig. 6 (c); Table 3 details the resulted bridge pipes for ten clusters, and supplementary Figures S2-S3 describe the analysis outcome for 5, 10, 20, and 46 communities.

Note in Fig. S2 that the number of communities have almost no influence on the mean pressure change versus the operational cost up to an operational cost of about $1,000,000. Beginning from this figure differentiations start, with a maximum gap of about 10 m at about $2,000,000, attributed to 5 communities with about 48 m average pressure to 46 communities with about 58 m average pressure. This is due to the very high looped layout of the Wolf system, where as the number of communities' increases, its impact on pressure drop decreases. In Fig. S3 water age is described as a function of operational cost, displaying a higher reduction in water age as the number of communities decrease. The reduction of water age for fewer number of communities is attributed to the increase in flow paths flexibilities as the number of communities decrease (i.e. less pipes are closed as the number of communities is reduced). The mean water age as a function of mean pressure is presented in Fig. S4. Hereby, initially there is a steep decrease in water age as the mean pressure increases until reaching a plateau.
Supplementary Figure S5 provides 3D Pareto front surfaces for 5, 10, 20, and 46 clusters partitioning.

Conclusions

This study presented a methodology for multi-objective optimization for minimizing operational cost, mean pressure, and mean water age. The optimization is conducted using the multi-objective genetic algorithm NSGA-II (Deb et al., 2002), under an a-priori network segmentation into clusters through a heuristic pipelines closure procedure post community partitioning using a graph theory topology method (Clauset et al., 2004). Having a-priori clusters formation, substantially simplifies the overall multi-objective optimization procedure, yet sensitivities need to be conducted for clusters partitioning (i.e., number and layout).

Three example applications of increasing complexities were explored. Tradeoffs were found between the three objectives of cost, pressure, and water age demonstrating the optimization capabilities, yet the degree of tradeoffs was very much case dependent. Sensitivities on the impact of clusters partitioning were also investigated.

The new contribution of this study is in defining and solving in a one framework a multi-objective optimization problem incorporating operational cost, access pressure, and water age objectives. In addition, while in previous research the initial hydraulics of the system dictated the DMAs formations, the method introduced herein defines the DMAs from a topological engineering based rational. The suggested DMA division follows the simple logic of supplying flow to DMAs through the system largest pipes (i.e., through a spanning tree holding the system’s major pipes).
Future directions which can extend this work are in: (1) embedding the clusters formation directly into the optimization problem including minimum velocities/head loss constraints, as well as reliability considerations (i.e., there is an obvious clear tradeoff between clusters layout and residual capacity of a system once a failure occurs), (2) development and application of analytical simplified optimization methods other than the general framework of linking simulation to optimization, (3) test the proposed methodology using other evolutionary computational methods, (4) incorporate pressure reducing valves (PRVs) locations and settings, and (5) integrate uncertainty into the analysis (e.g., inclusion of demands uncertainties using robust optimization).

**Supplemental data**

Figures S1 to S5 are provided as supplementary files.

**Data availability statement**

All data, models, and code generated or used during the study appear in the submitted article.

**Acknowledgments**

Financial support from the Deutsche Forschungsgemeinschaft (DFG) to this work is gratefully acknowledged, as well as from the Israel Science Foundation (grant No. 555/18).
References


Environmental Resources Congress: Critical Transitions in Water and Environmental Resources Management, pp. 4475.


List of Tables

Table 1: Energy tariff pattern coefficients

Table 2: Demand pattern 1 used at the M1 and MMOD networks and demand pattern 2 used at the Wolf network

Table 3: The bridge pipes with their properties (0 open, 1 closed) for networks M1, MMOD, and Wolf
List of Figures

Fig. 1: Dendrogram cutting at different branches, cross section A (left), B (right)  

Fig. 2: The dendrogram cutting illustration including: (a) cross sections A and B, and 
(b) the modularity values throughout the algorithm  

Fig. 3: The proposed community isolation method  

Fig. 4: The water distribution systems layouts for the three case studies: (a) M1, (b) 
MMOD, and (c) Wolf  

Fig. 5: Modularity of the community structure algorithms for the: (a) M1, (b) MMOD, 
and (c) Wolf networks  

Fig. 6: The (a) M1, (b) MMOD, and (c) Wolf networks segregated into 6, 5 and 10 
clusters, respectively  

Fig. 7: Pareto front surface for network M1 after segmentation into six clusters  

Fig. 8: Pump power pattern and tank level during a 24-hour simulation for the M1 
system  

Fig. 9: Pareto surface projection on the cost-pressure plane for M1 network  

Fig. 10: Pareto surface projection on the cost-water age plane for M1 network  

Fig. 11: Pareto surface projection on the cost-pressure plane for MMOD network  

Fig. 12: Pareto surface projection on the cost-water age plane for MMOD network
### Table 1: Energy tariff pattern coefficients

<table>
<thead>
<tr>
<th>Time of day [hour]</th>
<th>Energy tariff pattern coefficient [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00-06:00</td>
<td>0.25</td>
</tr>
<tr>
<td>06:00-12:00</td>
<td>0.80</td>
</tr>
<tr>
<td>12:00-18:00</td>
<td>1.50</td>
</tr>
<tr>
<td>18:00-24:00</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Table 2: Demand pattern 1 used at the M1 and MMOD networks and demand pattern 2 used at the Wolf network

<table>
<thead>
<tr>
<th>Time [hour]</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1 demand multiplier [-]</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time [hour]</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1 demand multiplier [-]</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time [hour]</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 2 demand multiplier [-]</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time [hour]</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 2 demand multiplier [-]</td>
<td>0.6</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Table 3: The bridge pipes with their properties (0 open, 1 closed) for networks M1, MMOD, and Wolf

<table>
<thead>
<tr>
<th>ID</th>
<th>M1 Diameter [mm]</th>
<th>Status (0/1)</th>
<th>ID</th>
<th>MMOD Diameter [mm]</th>
<th>Status (0/1)</th>
<th>ID</th>
<th>Wolf Diameter [mm]</th>
<th>Status (0/1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>203</td>
<td>0</td>
<td>4</td>
<td>100</td>
<td>0</td>
<td>2029</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>254</td>
<td>0</td>
<td>16</td>
<td>100</td>
<td>0</td>
<td>2185</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>254</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>0</td>
<td>2257</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>65</td>
<td>254</td>
<td>1</td>
<td>27</td>
<td>100</td>
<td>0</td>
<td>7051</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>305</td>
<td>0</td>
<td>33</td>
<td>100</td>
<td>0</td>
<td>7095</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>305</td>
<td>1</td>
<td>82</td>
<td>100</td>
<td>0</td>
<td>7132</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>305</td>
<td>0</td>
<td>151</td>
<td>100</td>
<td>0</td>
<td>7141</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>305</td>
<td>1</td>
<td>204</td>
<td>100</td>
<td>0</td>
<td>7157</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>356</td>
<td>1</td>
<td>210</td>
<td>100</td>
<td>0</td>
<td>7165</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>356</td>
<td>1</td>
<td>213</td>
<td>100</td>
<td>0</td>
<td>7453</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>231</td>
<td>0</td>
<td>7846</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>259</td>
<td>0</td>
<td>7947</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>261</td>
<td>0</td>
<td>21216</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>269</td>
<td>0</td>
<td>21244</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>121</td>
<td>0</td>
<td>21275</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>185</td>
<td>0</td>
<td>21407</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>240</td>
<td>0</td>
<td>22245</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>272</td>
<td>0</td>
<td>22278</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>154</td>
<td>0</td>
<td>22352</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>281</td>
<td>1</td>
<td>2049</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>65</td>
<td>1</td>
<td>2126</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>162</td>
<td>0</td>
<td>2213</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>164</td>
<td>0</td>
<td>2266</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>180</td>
<td>1</td>
<td>21415</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>291</td>
<td>1</td>
<td>22325</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>71535</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2027</td>
<td>16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2084</td>
<td>16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>7016</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>7098</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>71564</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>71737</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NA = not available
Identify Bridge Pipes List

Sort Pipes in ascending order

\[ n = \text{length(list)}; \]
\[ i = 1; \]

\[ i \leq n \]

No

Open pipe(i)
Status = 1;

is there only one network?

Yes

\[ i++ \]

Main Pipe List

Yes

Close pipe(i)
Status = 0;
Figure 5

(a) 

(b) 

(c)
Closed pipes

(a)

Bridge-pipes

(b)

(c)
Figure 7
Click here to access/download
Supplemental Materials File
S2.pdf
Click here to access/download
Supplemental Materials File
S4.pdf
Click here to access/download
Supplemental Materials File
S5.pdf
Click here to access/download
**Supplemental Materials File**
Codes and Networks.zip