Citation for published version:
Brentan, BM, Campbell, E, Goulart, T, Manzi, D, Meirelles, G, Herrera, M, Izquierdo, J & Luvizotto Jr., E 2018, 'Social network community detection and hybrid optimization for dividing water supply into district metered areas', Journal of Water Resources Planning and Management, vol. 144, no. 5, 04018020. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000924

DOI:
10.1061/(ASCE)WR.1943-5452.0000924

Publication date:
2018

Document Version
Peer reviewed version

Social network community detection and hybrid optimization for the Battle of the Water Networks DMA (BWNDMA)

PhD Bruno Brentan, PhD Enrique Campbell, MSc Thaisa Goulart [3], MSc Daniel Manzi[4], MSc Gustavo Meirelles[5], PhD Manuel Herrera[6], PhD Joaquín Izquierdo[7] and PhD Edevar Luvizotto Jr[8]

[1]Computational Hydraulic Laboratory, University of Campinas, Campinas, (Brazil)

[2]Berliner Wasserbetriebe, Berlin (Germany)

[3]Computational Hydraulic Laboratory, University of Campinas, Campinas, (Brazil)

[4]Computational Hydraulic Laboratory, University of Campinas, Campinas, (Brazil)

[5]Computational Hydraulic Laboratory, University of Campinas, Campinas, (Brazil)

[6]EDEn, University of Bath, Bath (UK)

[7]FluIng-IMM, Universitat Politècnica de València, Valencia, (Spain)

[8]Computational Hydraulic Laboratory, University of Campinas, Campinas, (Brazil)

[1]brunocivil08@gmail.com, [2]enrique.campbell@bwb.de,
[3]a.m.herrera.fernandez@bath.ac.uk, [4]jizquier@upv.es

Abstract: Water supply utilities need to properly manage their systems to guarantee a quality supply. One way to manage large systems is through division into district metered areas (DMAs). Graph clustering with an unknown number of subdivisions, as in social network theory, has proven highly efficient in this sectorization problem. Several physical and hydraulic features may easily be used as criteria to suitably divide the network. In this work, we use social network community detection algorithms to define several DMA scenarios. Configurations mainly depend on nodal demand and elevation, but adaptations may be needed to guarantee full supply in future scenarios related to system growth – and rehabilitation actions may also be required. The problem associated with pipes and valves is first solved with three optimization methods. The best solutions then enter a new optimization process, where tank dimensions and valve set points are defined. This complex optimization-segregation approach enables an improvement in the hydraulic efficiency of the E-town network at an affordable cost, and this approach also determines the measures needed to meet the dry season requirements.

Key-words: water distribution networks, DMA definition, rehabilitation, PSO, GA, soccer league competition

INTRODUCTION

Water supply systems (WSSs) have a fundamental function in urban design: guaranteeing citizens access to safe drinking water (Di Nardo et al. 2013). Management of WSSs becomes more complex when future demand challenges are taken into account. Segregating water distribution networks into district metered areas (DMAs) enables better management by improving efficiency and safety through strategic rule implementations.

To establish a suitable DMA configuration, several aspects must be considered simultaneously, including: topology; elevation; size; loop configuration; costs; and resilience – and this makes any manual/empirical approach unfeasible (Diao et al. 2012). Several automatic approaches, such as the graph decomposition theory applied to DMA divisions (Swamee and Sharma, 2008), or multiple source decomposition with pre-located influence zones (Tzatchkov et al. 2008) have been implemented in the last decade and demonstrated better performances than empirical approaches. Following in the wake of neural networks and other machine learning tool applications, Herrera et al. (2010) propose the use of semi-supervised learning methods with information on different supply constraints and gathering the reality of hydraulic zones in a single matrix. Accordingly, spectral clustering is applied to obtain the DMA configuration. Diao et al. (2012) present an automatic DMA boundary selection method based on a social network community detection algorithm. In a similar line, Di Nardo et al. (2013) present a social network analysis based on complex system decomposition.

However, division into DMAs is not the end of the story. The status of the boundary valves and their location at DMA entrances must be optimized to eventually achieve a reliable configuration. The main constraint is supplying the required quality and quantity, as well as the satisfaction of service pressure (nominal pressure) and the control of tank levels. Meta-heuristic optimization methods have been broadly used to find the optimal solution for several water supply network problems (Montalvo et al., 2014, Marchi et al., 2014). The problem of optimal valve placement and optimal operation has been highlighted (Nicolini et al., 2009, Brentan et al., 2017).

A fully automatic water network partitioning algorithm requires the use of a method to identify the DMAs and an optimization procedure to identify the boundaries and entrances of each DMA. Di Nardo et al. (2014) present a hybrid algorithm that links graph partition methods with genetic algorithms (GAs) to create an automatic method for DMA design. De Paola et al. (2014) define DMAs using the shortest path concept from the graph theory domain in a combination with the NSGA-II algorithm to carry out an optimization procedure in which topological and operational aspects are considered. Brentan et al. (2017) present a set of hydraulic criteria implemented in a hybrid algorithm compound by a social network community detection algorithm and particle swarm optimization (PSO).

For the task using the Battle of Water Network District Metered Areas (BWNDMA) considered in this work, we present an alternative to achieve a solution for the DMA configuration problem of the E-Town network that takes into account future demands and pipe rehabilitation.

Our approach has two main phases. In the first phase, the distribution network is decoupled from the trunk network through a process based on the shortest path concept derived from graph theory. DMAs are then defined over the distribution network by means of a community detection algorithm and a herein proposed community recursive merging process. Once the DMAs are defined, a set of entrances and boundary valves, and the set points of the pressure reducing valves (PRV) located at the entrance(s) of each DMA, are defined using a multilevel optimization approach. In this phase, the rehabilitation of the network, including pipe duplication and/or replacement and installation of new tanks, is also considered. Such a process is based on three algorithms: GAs, PSO, and soccer league competition (SLC).

Our approach relies on several advantages, the first being the use of a community detection algorithm that can efficiently deal with extremely large networks (even of millions of nodes). A second advantage is the optimization process itself, which is conducted using a multi-level approach, splitting the problem into smaller and better manageable subproblems.

PROBLEM AND APPROACH DESCRIPTION

BWNDMA description

The BWNDMA was a water distribution analysis competition proposed by the Water Distribution System Analysis Congress 2016. The main objective of the challenge was to improve the management of the system of the E-town network for a future scenario, creating DMAs in a water network with more than 11,000 nodes and about 14,000 pipes. E-town is supplied by three water sources (Cuzca, Bochica and Bachue) which can cope with the supply during the rainy season (from March to May and from September to November). However, for the dry season (December to February and June to August) the capacity of the sources is reduced and the use of aquifer sources is necessary.

The problem should be solved for the rainy season and the operational changes (valve closures) must be defined to reach a feasible scenario for the dry season. For the rainy solution, the number of DMAs, pressure uniformity (*PU*), demand similarity (*DS*), implementation costs, and water quality are used as parameters to evaluate the quality of the solutions. To find feasible solutions, some structural improvements are allowed, namely: pipe replacements or duplication; PRV installations; and the construction of new tanks.

Algorithm's overview

The proposed methodology can be divided into two phases: the DMA design phase and the optimization phase. Firstly, the hydraulic and topological data should be prepared. The water network model is built and then the trunk network is identified. The trunk network transports water from the sources to distribution networks. This identification is important to facilitate pipe replacement and duplication. Once the trunk network is identified, the DMA design core searches for the best configuration of node clusters when considering hydraulic and topological criteria. In this step, the *Walktrap* community detection algorithm (Pons and Latapy, 2006) is applied to identify the nodes with similar features, which will eventually integrate the communities. During this process, several communities are fused to generate new communities until a set of criteria is met. Data preparation and DMA configuration correspond to the step of defining DMAs in the distribution network. Together with the DMAs, the boundary pipes are also identified. With this information, it is possible to start the optimization steps.

The optimization core is responsible for achieving a feasible scenario, taking into account the constraints of the problem. The methodology used for the optimization core divides the problem into three levels to reduce the dimensionality of the problem. The first level (Optimization level 1) uses GAs, PSO, and SCL to find the best implementation of PRVs and new pipes, minimizing the cost and the *PU* parameter. To guarantee the water supply for the future scenario with the increase of consumption, the optimal installation of new tanks is tackled in the second level (Optimization level 2). Finally, the operational points of PRVs and flow control valves (FCV) are adjusted to improve the operation of the system, thus minimizing the *PU*. For each level of the optimization process, a solution vector must be defined (Optimization level 3) Figure 1 presents a flowchart of the full algorithm, which considers both the sectorization and the multi-level optimization processes.

SECTORIZATION PROCEDURE

In the computer domain, social networks are graphs intended to represent relations among social actors through a set of dyadic ties. Topologically speaking, a social network and a water supply network are equivalent (both are formed by nodes connected by links), the latter can be represented as a social network, and all the algorithms/concepts derived from the social network theory field of study can be implemented over water supply networks. In this work, we use the “shortest path” concept and a “community detection” algorithm.

Depending on the network topology, the sectorization is conducted in one of two ways: if the network has many sources (and these sources are located within the meshing space) the DMAs can be established around these sources; in contrast, when the number of sources is limited and/or are located outside the meshing area, the network must be supplied by a main conduction system (or trunk network) and, therefore, the DMAs must be established around the latter. Once the trunk network is defined, it is uncoupled from the distribution network, and communities of nodes are detected in this distribution network. Such communities are recursively merged in a fusion process herein proposed until a partition is found, in which each DMA satisfies the predefined set of constraints.

Trunk Network Detection Algorithm

As described above, in a WSS, the trunk network corresponds to a group of continuously connected pipes (a stem) that transports water from the sources to the pipes in the distribution network. The appropriate distinction of the latter from the distribution network is a key aspect in a sectorization design, as the closure of at least one pipe of the trunk network could dramatically affect the resilience of the entire system. There are several general criteria to distinguish the trunk network from the distribution network, such as: diameters, connections, and locations. In general, the connection to the trunk network is restricted to medium and large diameter pipes. However, in some cases, especially in the case of WSSs with multiple sources, the span of the trunk network is not so clear. Therefore, graph theory concepts can be used to distinguish the level of importance of each pipe in the supply of the network.

The core idea is to generate a ranking of pipes. Such ranking is based on the role of each pipe in the supply of the entire network. To assess the role of each pipe, a hydraulic simulation is conducted with EPANET (Rossman, 2000) for the most critical scenario (the instant of highest demand). The direction of the flow in each pipe is then retrieved and stored in a square matrix (see Equation 1). This matrix enables calculating the number of nodes/pipes that can be reached from each node, this being the accumulated shortest path value (ASPV).

$$\begin{array}{cc} & \begin{array}{cccc} Node_1 & Node_2 & \cdots & Node_n \end{array} \\ \begin{array}{c} Node_1 \\ Node_2 \\ \vdots \\ Node_n \end{array} & \left(\begin{array}{cccc} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{array}\right) \end{array} \tag{1}$$

The entries of this matrix are calculated as follows, where IN stands for initial node and EN stands for end node:

- If $IN = EN \rightarrow 0$
- If IN cannot reach $EN \rightarrow \infty$
- In any other case → nodes in the shortest path

The flow in each pipe is then multiplied by its corresponding ASPV (the result is represented by ASPV*). The trunk network is expected to be formed by pipes with high and less frequent AVSP* values, whereas the pipes in the distribution network are expected to have low and highly frequent AVSP* values (Campbell et al., 2016).

Community detection algorithm

Community detection algorithms based on the social network theory are aimed at revealing structural network modules based on a modularity index proposed by Newman (2006). One of these algorithms corresponds to the *Walktrap* algorithm, which is based on the mathematical concept known as a random walk. In comparative terms, with respect to other community detection algorithms, the tests performed by Kesiban Orman et al. (2011) classify this algorithm as the second best algorithm for graph community detection, just after the Infomap algorithm (Rosvall & Bergstrom, 2008). In contrast, results obtained by Savić et al. (Newman & Girvan, 2004) when comparing the *Walktrap* algorithm with other algorithms, namely propagation (Raghavan, 2007) and greedy modularity optimization (Clauset et al., 2004), show that the former has significant advantages in terms of quality and evolutionary stability.

This algorithm corresponds to a stochastic process where the position of a given particle at a certain instant relies solely on its position at a previous instant and on a random variable (which determines its subsequent direction and step length). If random walks of a given length are performed over a graph, the resulting Markov matrix reflects the probability of going from one node to another in a given number of steps. These probabilities gather enough information about the topology of the network, and reveal a multiple arrangement of communities with different values of modularity. All these arrangements are located in a hierarchy of partitions, from which any particular partition may be selected. It is noticeable that the partition of maximum modularity can generate extremely small communities, whose implementation could be economically unfeasible. This is why a recursive merging process is proposed, to ensure that all the DMAs comply with a series of pre-established constraints. In this merging process, the total demand of each DMA is computed (other characteristics can be used, for example, pipe length), and a matrix containing the results of the sum of the demands for every pair of directly connected communities is then computed. The pair of communities with the lowest result is merged into one new DMA, and the new collection of DMAs is re-enumerated. The process continues until no new mergers are possible (a new merger would exceed the maximum demand per DMA). The matrix in equation (2) illustrates the merging process.

$$\begin{array}{c c} & \begin{array}{cccc} DMA_x & DMA_y & \cdots & DMA_n \end{array} \\ \begin{array}{c} DMA_x \\ DMA_y \\ \vdots \\ DMA_n \end{array} & \left(\begin{array}{cccc} 0 & DMA_x + DMA_y & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{array} \right) \end{array} \tag{2}$$

The entries of this matrix are calculated as follows:

- If DMA_x and DMA_y are connected then $d = DMA_x + DMA_y$
- If $DMA_x = DMA_y$, then $d = 0$.

NETWORK OPTIMIZATION PROCEDURE

Optimization routine

Due to the complexity of the problem, the optimization procedure is divided into three levels. In the first, the method searches for the best location for PRVs and their pressure setting, considering the previous boundary pipes defined in the sectorization procedure. Boundary pipes correspond to the pipes with start and end nodes not belonging to the same DMA. The method also evaluates the replacement of pipes with a diameter greater than 152 mm, an imposed condition in BWNDMA. Considering the possibility of installing parallel pipes, the trunk network is omitted in this first optimization process.

The initial topology configuration of the E-town network does not fulfill the future demand scenario, which makes the 168-hour simulation process extremely difficult from the computational viewpoint. To reduce the processing time, this optimization step was performed with only maximum and minimum demands, both with 60% water volume in all tanks. With this procedure, it is expected that minimum and maximum pressure constraints during the entire 168-hour period almost match desirable pressures. In this first step, pressure uniformity is also considered in the objective function (Eq. 3) to define the best PRV characteristics.

$$F_1 = \left(\sum_{i}^{NP} L_i . C_{D_i} + \sum_{j}^{NV} K_{D_j} + \sum_{i}^{NP} L_i . C_{D_i}^{Par}\right) . Pen_{min} \, . \, Pen_{neg} \, . \, Pen_{max} \, . \, Pen_{VRP} . PU \qquad (3)$$

F_1- objective function of the initial network design;

NP - number of new pipes installed;

C_{D_i}- new pipe replacement cost associated to its diameter, D_i;

$C_{D_i}^{Par}$ - new parallel pipe installation cost associated to its diameter, D_i;

L_i - length of new pipe;

NV - number of PRVs installed;

K_{D_j}- PRV cost associated to its diameter, D_j;

Pen_{min}- penalty for pressure below 15 m in demand nodes;

Pen_{neg}- penalty for negative pressure in nodes without demand;

Pen_{max} - penalty for pressure above 60 m;

Pen_{VRP} - penalty for exceeding the maximum number of PRVs in a DMA (two in this study);

PU - network pressure uniformity.

Moosavian and Roodsari (2014) and Mora-Meliá et al. (2015) present comparisons among bio-inspired algorithms, where they observe that their effectiveness depends on the problem characteristics. Therefore, PSO, GA, and SLC algorithms are used to obtain their best solutions, merging the results of each to achieve a better solution.

At this level, the problem was codified using mixed binary and discrete variables. Each solution vector contains information about the status of the boundary pipes and a discrete number that correspond to available pipe diameters. For the binary positions, if the value of a position is 1, it means that the corresponding boundary pipe is open and requires the installation of a control device. A similar procedure is used for the installation of new pipes or for the replacement of pipes. The position of the solution vector corresponding to an index pipe contains a discrete value that corresponds to an available diameter for this pipe in the set of candidate diameters. With this mixed vector of decision variables it is possible to calculate the terms in the objective function F_1 that corresponds to the implementation costs.

Usually, bio-inspired algorithms are unable to treat constrained problems and require the use of constraint-handling, such as the common approach of penalty functions (Mezura-Montes and Coelho, 2011; Coelho, 2002) in transforming these problems into unconstrained problems. Following the general proposal by Parsapoulos and Vrahatis (2002), the penalty function can be written as:

$$Pen = \sum_{i=1}^{N_c} \beta_i . \left| x_i^s - x_i^{lim} \right|^k . \tag{4}$$

Here β_i and k are the two penalty factors to adjust the variable x_i^s to meet the constraint limit x_i^{lim} in a problem with N_c constraints.

Pressure uniformity is used to evaluate the quality of the network partition process, since it measures the difference between the nodal pressure and both the minimal required value and the average hourly value, as in (5). Based on an equation by Alhimiary et al. (2007):

$$PU = \sum_{j=1}^{M} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\left(P_{i,j} - P_{min}\right)}{P_{min}} + \frac{\sqrt{\sum_{i=1}^{N} \frac{\left(P_{i,j} - P_{av_j}\right)^2}{N}}}{P_{av_j}} \right], \tag{5}$$

where $P_{i,j}$ is the nodal pressure at time step j at node i for a network with N demand nodes simulated during M time steps. P_{min} is the minimal required pressure for demand nodes (15 m) and P_{av_j} is the average network pressure at time step j.

At the end of this first step, the results of the three methods are evaluated and the best method is chosen to feed the following steps. Since the simulation is carried out with PRVs installed in all boundary pipes, an accurate analysis is necessary to define, from among the open valves, which will effectively operate.

As a first approach, the PRVs with higher flows in each DMA are selected, and a simulation is made to evaluate the results. This simulation aims to identify DMAs with high pressure zones, and where pipes must be closed to achieve the pressure constraints.

While the steady-state simulation helps define PRV locations and the pipes to be replaced or candidates for parallel installation, this hydraulic approach cannot show the influence of level oscillations in the tanks, thus hampering a full rehabilitation evaluation.

Therefore, the second optimization level is made on a 24-hour basis, considering the addition of adjacent tanks, pipe closures, and a new setting for PRVs and FCVs. Since the demand pattern repeats through the week (168 hours), it is necessary that the initial tank level remains the same at the end of the simulation. Eq. (6) presents the objective function for this second optimization level.

$$F_2 = \left(\sum_{tq=1}^{Ntq} C_{V_{tq}} \right) . Pen_{min} \, . \, Pen_{neg} \, . \, Pen_{max} \, . Pen_{lv} \tag{6}$$

where:

F_2- objective function of the final rainy season configuration;

Ntq - number of new tanks installed;

$C_{V_{tq}}$- cost of new tank *tq* associated with its volume;

Pen_{lv}- penalty for the difference between initial and final tank levels;

The new tanks should be considered to enable the full supply of the water network in a future demand horizon (2022). The need to expand the storage capacity (thus guaranteeing better quality of water service) is associated with increasing demand and with the daytime oscillation level. The adjusted result obtained in the first stage is submitted to the second optimization level, which will define the final modifications of the network topology.

In this step, a mixed approach of discrete and continuous variables is used. The discrete variables of the solution vector correspond to positions in a list reflecting the available volume for the new tanks, while the continuous variables correspond to the set points of PRVs and FCVs. To solve this optimization problem, GA, PSO and SLC are once again applied to achieve the optimal solution.

After this optimization, the final configuration for the rainy season is achieved. However, the topology of the network changes from the rainy to the dry season. Water sources reduce their capacity and two wells are activated to guarantee the water supply.

The final level of the optimization process achieves the new settings for PRVs and FCVs, and defines opening/closure of pipes for the dry season. The solution vector is programmed with mixed binary and continuous variables. The binary variables correspond to the status of pipes and valves, and the continuous variables correspond to the new set points for the valves. The objective function (7) minimizes the change of pipe statuses and valve settings, allowing easier maneuvers and thus fulfilling the operational constraints.

$$F_3 = \left(\sum_{lk}^{Nlk} Op_{lk}\right) . Pen_{min} \, . \, Pen_{neg} \, . \, Pen_{max} \, . Pen_{lv} \tag{7}$$

Here

F_3 - objective function of the dry season optimization;

Op_{lk} – the operational change at the link lk;

Optimization algorithms

The use of bio-inspired algorithms to solve optimization problems in water distribution problems has gained space in the scientific literature. For the sake of robustness, this work applies two classical bio-inspired algorithms from among the many developed: namely, PSO (Eberhart and Kennedy, 1995), based on the behavior of a flock of birds or a school of fish searching for food; and GA (Goldberg and Holland, 1988), based on the evolutionary competition of species. Also applied is a recent proposal by Moosavian and Roodsari (2013), a SLC optimization methodology based on the competitive environment among teams and players in soccer leagues.

GAs and PSO are widely applied in water distribution problems: and SLC is already highly promising in this field despite its relative newness. The results produced by the three algorithms were equivalent in all the cases, what enhances the robustness and reliability of the obtained results.

The implementation of PSO for the first level is made with 300 particles. This value is selected considering the number of variables (number of valves and pipes to be replaced). The second level is developed with 150 particles, since the number of variables is lower than the first level. Finally, only 50 particles are used to optimize the tank diameters. The maximum number of iterations is 2000 for the three levels. Regarding GAs, the population size for the first optimization level is 480 individuals, the second level uses 240 individuals, and the last level uses 80 individuals. For all levels, the algorithm ran 1000 generations with an elitism rate of 10%. The implementation of SLC for the BWNDMA optimization problem follows the same proportions shown for PSO and GA. The first level is conducted with 60 teams, the second level with 30 teams, and the last level with 10 teams. For all levels, each team has eight main and eight reserve players.

RESULTS AND DISCUSSION

Sectorization and trunk network identification

The uniform distribution of demands among the DMAs and the uniformity of pressure in the network requirements, elevation, and demand at each node of the DMA are considered in the sectorization process. As a first approach, the trunk network is defined and uncoupled, thus disconnecting the pipes which are linked with this network and generating isolated DMAs. This procedure presents hydraulic and management problems, such as water supply disruption and micro-DMA creation, given the spatial distribution of the nodes.

However, the importance of the trunk network is linked to the capacity of water transport from the water source to consumers. This structure has a specific treatment in the first step of the optimization process, where the possibility of installing parallel pipes is considered. Figure 2a shows the defined trunk network in red, and Figure 2b presents the nodal elevation for the E-town. The analysis of these figures enables us to identify a certain trend about the level of the trunk network. Given the low efficiency of the sectorization without the trunk network, the entire network is used, reaching the goal of 15 DMAs as defined by the social community method. Figure 3a presents the DMA configuration results in colors, while Figure 3b shows a block diagram for DMA interconnection.

With the communities defined through the social community method, the entrance and exit of each DMA must be defined. Since boundary pipes are considered to have a PRV installed, the settings of each are optimized to reach the constraints with minimum costs. In addition, an initial dimensioning of the network pipes is also made.

With this initial solution, the entrances of each DMA are established manually, considering the flow through each PRV as a decision parameter, and using a 24-hour period to evaluate the configuration. This procedure shows the necessity to improve the capacity of the Bochica and Cuza pipelines. Therefore, the trunk network obtained in the sectorization study is duplicated so that all demand nodes can be satisfied with minimal pressure.

This configuration of the entrance and exit of each DMA is achieved only after an analysis of the hydraulic performance of the system for a 24-hour period. Note that DMA #3 involves the trunk network. Therefore, this is the main DMA of the system and from where all distribution occurs. In addition, DMA #3 is the only DMA capable of supplying DMAs #0, #1, #2 and #14. The other DMAs configure a distribution loop, thus reinforcing water supply reliability.

Table 1 shows the DMA characteristics (demand, average elevation, and entrances) obtained once the segregation and the first optimization process are finished. The importance of the DMA#3 as a supplier for eight DMAs is noteworthy. This is produced by the architecture of DMA#3, which corresponds to the trunk network with a small distribution network (see Figure 3a).

Network optimization for rehabilitation and operation in the rainy season

With the final configuration of DMAs, the tank levels are evaluated for the week horizon. It is observed that some tanks are emptying while others are overflowing. This happens because the capacity of the three water sources are not fully used, thus overloading the tanks. To solve this problem, a third pipeline is created from the Cuza reservoir, and its flow capacity is increased almost to the maximum allowed (1600 l/s in the rainy season).

The most important tank in this DMA configuration is Tank #1, placed in DMA #3. The importance of this tank is related to the most elevated end nodes of this sector, which, in turn, require more capacity from the tank. The periodic oscillation of this tank can improve the water quality of several DMAs and contribute to the pressure control in the network. Figure 4 shows the oscillation of the Tank #1 level during the week.

Despite most tanks oscillating during the week and returning to the initial level, with this configuration, Tanks #2 and #1 drain. Tank #2 recovers its level in low consumption periods, but not enough to restore its initial level. The most critical situation is in Tank #11. This tank has the highest elevation and so the network pressure must be high to feed it suitably. Since only 164 nodes (1.5% of the original network) are supplied from this tank, its level control was neglected.

Figure 5 shows the level oscillation for these two tanks (Tank #11 – red line and Tank #2 – green line) during the week.

The configuration obtained in this second optimization level meets minimum pressure constraints for the 168-hour period. However, the maximum pressure limit violation remains a problem. Therefore, a final evaluation for PRV settings and pipe closures was made, thus achieving a plausible solution for the rainy season. Table 2 shows the evolution in pressure constraints and pressure uniformity of the network and considering only maximum and minimum consumption periods.

This table points to the efficiency of the optimization-segregation approach when compared with the initial configuration of the E-town network. It is important to note the increase in negative pressure nodes between the original network and the sectorized network. This happens mainly because of high demand values in a network with many closed pipes, thus guiding the flow to non-optimized pipes.

The installation of parallel pipes (duplicating the trunk and third Cuza pipeline) considerably decreases the negative pressure in the network, since this action reduces hydraulic head loss in the trunk network. However, with this action mainly occurring at low elevation nodes, the available hydraulic head increases the pressure, thus inducing high pressure values for nodes above 60 m, despite most of these nodes being placed in the trunk network.

Finally, an aspect regarding pressure uniformity (*PU*) in the network can be highlighted. The sectorized network without optimization is unable to deliver all the demands, since some nodes become disconnected. This prevents *PU* calculation at this stage. Similar values are found if the *PU* value is compared between the original network and the final network with defined PRVs and pipe closures. However, the final network has a better hydraulic performance.

More than the hydraulic performance of the final network, a quality analysis of the E-town network is required. This analysis is made considering the water age at nodal demands. The age map in Figure 6 presents the state at hour 168 (the end of the simulation and the most critical moment for water age). It is possible to observe that most nodes are under 30 hours, enabling us to affirm the high performance of the DMA partition and tank use. Only 6% of the water network is older than 60 hours (the maximum allowed). Furthermore, the comparison between the original water network and the optimized network points to a substantial improvement in the water quality.

The total costs of rehabilitation for future demand are presented in Table 3. It is very important to highlight that the main costs are associated with the pipes (implantation and replacement) while only two tanks should be built. Also the low cost of valve implantation can be observed, corresponding to the maximal efficiency of the DMA entrance definition (which is able to satisfy most constraints at a low cost).

Network optimization for rehabilitation and operation in the dry season

The topological changes during the dry season require new statuses for pipes and set points of valves. Pressure distribution and water age are affected by this new network configuration. While for the rainy season, the DMA configuration can supply all nodes with pressure above that required for the dry season (even after the optimization process) a feasible solution that meets pressure constraints was not achieved.

The reduction in the water source availability induces the network to lower pressures and the use of two pump stations is required. Even with these two new water sources, DMA #8 is disconnected from the network for 24 hours. This occurs because the high elevation of this area prevents guaranteed demand fulfillment by feeding the tanks. Figure 7 shows the level variation of Tank #1 (the most important supply tank). It can be seen how its draining during 168 hours harms the supply process during the week. In addition, high pressure in the lower area of the trunk network harms water pumping, since the pump station near this pipeline cannot work at full power, thus reducing the water availability.

GENERAL DISCUSSION

The relation between DMA segregation and the optimal rehabilitation of large water distribution networks is shown in this work. The high performance of the optimization process is closely linked to DMA definition. This is because the increase in demand of the existing infrastructure is unable to supply all nodes, and the disconnection of nodes or links significantly impairs the hydraulic simulation.

A previous optimization process, using the maximal and minimal demand to determine the entrance and exit of each DMA and the new diameters for pipe replacement, is an interesting approach to reduce the computational effort. Furthermore, the previous identification of the trunk network enables the initialization of the first solution of the optimization methods with larger diameters, thus facilitating hydraulic simulations.

Despite the steady state defining new diameters or PRV placement, this approach is not useful to evaluate the tank level behavior and, consequently, is not useful to determine the need for new tanks. The use of extended period simulation for 24 hours, considering the initial and final tank levels as a problem constraint, is useful and guarantees high optimization performance. Moreover, the extended period simulation enables the PRV and FCV set point definition and pipe statuses to find the topology with the lowest constraint violation.

Finally, changes of season require a new evaluation of the network without diameter changes. This makes the optimization process difficult because the change of pipe or valve statuses slows the convergence when compared with the pipe replacement problem.

CONCLUSIONS

The Battle of Water Networks District Metered Area (BWNDMA) presents a large DMA creation problem jointly with the rehabilitation of the network to fulfill future demand. The importance of DMA creation coupled with optimal pipe replacement or new pipe installation, as well as PRV placement and new tank dimensioning, is evidenced by the reduction in the constraint violations presented in this work.

The community detection algorithm can congregate nodes by distance, elevation, and demand criteria. The high performance of this method when applied to DMA creation problems is the strong point of this work, since network partition using this technique generates important and not obvious divisions of the network.

While the optimization process presented in this work was unable to satisfy all the constraints (especially maximum pressure) the complex approach of the optimization-segregation process enables DMA creation with good indicators for pressure and demand uniformity.

Acknowledgments

The authors wish to express a special gratitude and remembrance towards Professor Rafael Perez García, who recently passed away. The use of English has been supervised by John Rawlins.

REFERENCES

Alhimiary, H. A., and Alsuhaily, R. H. (2006). Minimizing leakage rates in water distribution networks through optimal valves settings. In Proceedings of World Environmental and Water Resources Congress 2007 (pp. 1-13).

Brentan, B. M., Campbell, E., Meirelles, G. L., Luvizotto, E., & Izquierdo, J. (2017). Social Network Community Detection for DMA Creation: Criteria Analysis through Multilevel Optimization. Mathematical Problems in Engineering, 2017.

Brownlee, J. (2011). "Clever algorithms: nature-inspired programming recipes". Jason Brownlee.

Campbell, E., Izquierdo, J., Montalvo, I., and Pérez-García, R. (2016). “A novel water supply network sectorization methodology based on a complete economic analysis, including uncertainties”. WATER, 8(5), 179.

Civicioglu, P., & Besdok, E. (2013). A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms. Artificial intelligence review, 1-32.

Coello, C. A. C. (2002). Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. Computer methods in applied mechanics and engineering, 191(11), 1245-1287.

Daoud, S., Chehade, H., Yalaoui, F., & Amodeo, L. (2014). Efficient metaheuristics for pick and place robotic systems optimization. Journal of Intelligent Manufacturing, 25(1), 27-41.

Eberhart, R. C., and Kennedy, J. (1995). "A new optimizer using particle swarm theory". In Proceedings of the sixth international symposium on micro machine and human science (Vol. 1, pp. 39-43).

Di Nardo, A., Di Natale, M., Santonastaso, G. F., Tzatchkov, V. G., and Alcocer-Yamanaka, V. H. (2013). "Water network sectorization based on graph theory and energy performance indices". Journal of Water Resources Planning and Management, 140(5), 620-629.

Di Nardo, M. Di Natale, G.F. Santonastaso, S. Venticinque, An automated tool for smart water network partitioning, Water Resour. Manag. 27(13) (2013) 4493–4508.

Diao, K., Zhou, Y., and Rauch, W. (2012). "Automated creation of district metered area boundaries in water distribution systems". Journal of Water Resources Planning and Management, 139(2), 184-190.

Goldberg, D. E., & Holland, J. H. (1988). "Genetic algorithms and machine learning". Machine learning, 3(2), 95-99.

Herrera, M., Canu, S., Karatzoglou, A., Pérez-García, R., and Izquierdo, J. (2010). "An approach to water supply clusters by semi-supervised learning". In: Proceedings of International Environmental Modelling and Software Society (IEMSS).

A. Marchi et al., "Battle of the Networks II, "Journal of Water Resources Planning and Management, 140, n. 7, pp 1-14, 2014.

Mezura-Montes, E., & Coello, C. A. C. (2011). Constraint-handling in nature-inspired numerical optimization: past, present and future. Swarm and Evolutionary Computation, 1(4), 173-194.

I. Montalvo, J. Izquierdo, R. Pérez-García and M. Herrera, "Water Distribution System Computer- Aided Design by Agent Swarm Optimization," Computer-Aided Civil and Infrastructure Engineering, vol. 29, n. 6, pp 433-448, 2014.

Moosavian, N., and Roodsari, B. K. (2013). "Soccer league competition algorithm, a new method for solving systems of nonlinear equations". International Journal of Intelligence Science, 4(01)

Moosavian, N., and Roodsari, B. K. (2014). "Soccer league competition algorithm: A novel meta-heuristic algorithm for optimal design of water distribution networks". Swarm and Evolutionary Computation, 17, 14-24.

Mora-Melia, D., Iglesias-Rey, P. L., Martinez-Solano, F. J., and Ballesteros-Pérez, P. (2015). "Efficiency of evolutionary algorithms in water network pipe sizing". Water Resources Management, 29(13), 4817-4831.

Newman, M. E. (2006). "Finding community structure in networks using the eigenvectors of matrices". Physical review E, 74(3), 036104.

M. Nicolini and L. Zovatto, “Optimal location and control of pressure reducing valves in water networks.” Journal of Water Resources Planning and Management, 135(3), 178-187, 2009.

Parsopoulos, K. E., and Vrahatis, M. N. (2002). "Particle swarm optimization method for constrained optimization problems". Intelligent Technologies–Theory and Application: New Trends in Intelligent Technologies, 76, 214-220.

Pons, P., and Latapy, M. (2006). "Computing communities in large networks using random walks". J. Graph Algorithms Appl., 10(2), 191-218.

Rahmani, F., Ardeshir, A., Behzadian, K., and Jalilsani, F. (2014). "Optimal Rehabilitation Strategy", In: Water Distribution Systems Considering Reduction In Greenhouse Gas Emissions.

Roshani, E., and Filion, Y. R. (2014). "Water distribution system rehabilitation under climate change mitigation scenarios in Canada". Journal of Water Resources Planning and Management, 141(4), 04014066.

Rossman, L. A. (2000). "EPANET 2.0 User's Manual". Drinking Water Research Division, Risk Reduction Engineering Laboratory, U.S. Environmental Protection Agency

Swamee, P. K., and Sharma, A. K. (2008). "Design of water supply pipe networks". John Wiley and Sons.

Tzatchkov, V. G., Alcocer-Yamanaka, V. H., and Ortíz, V. B. (2006, August). "Graph theory based algorithms for water distribution network sectorization projects". In: Proc. of the 8th Annual Water Distribution Systems Analysis Symposium, WDSA, Cincinnati, Ohio, USA.

Table 1 - DMA final configuration: hydraulics and topology of each DMA

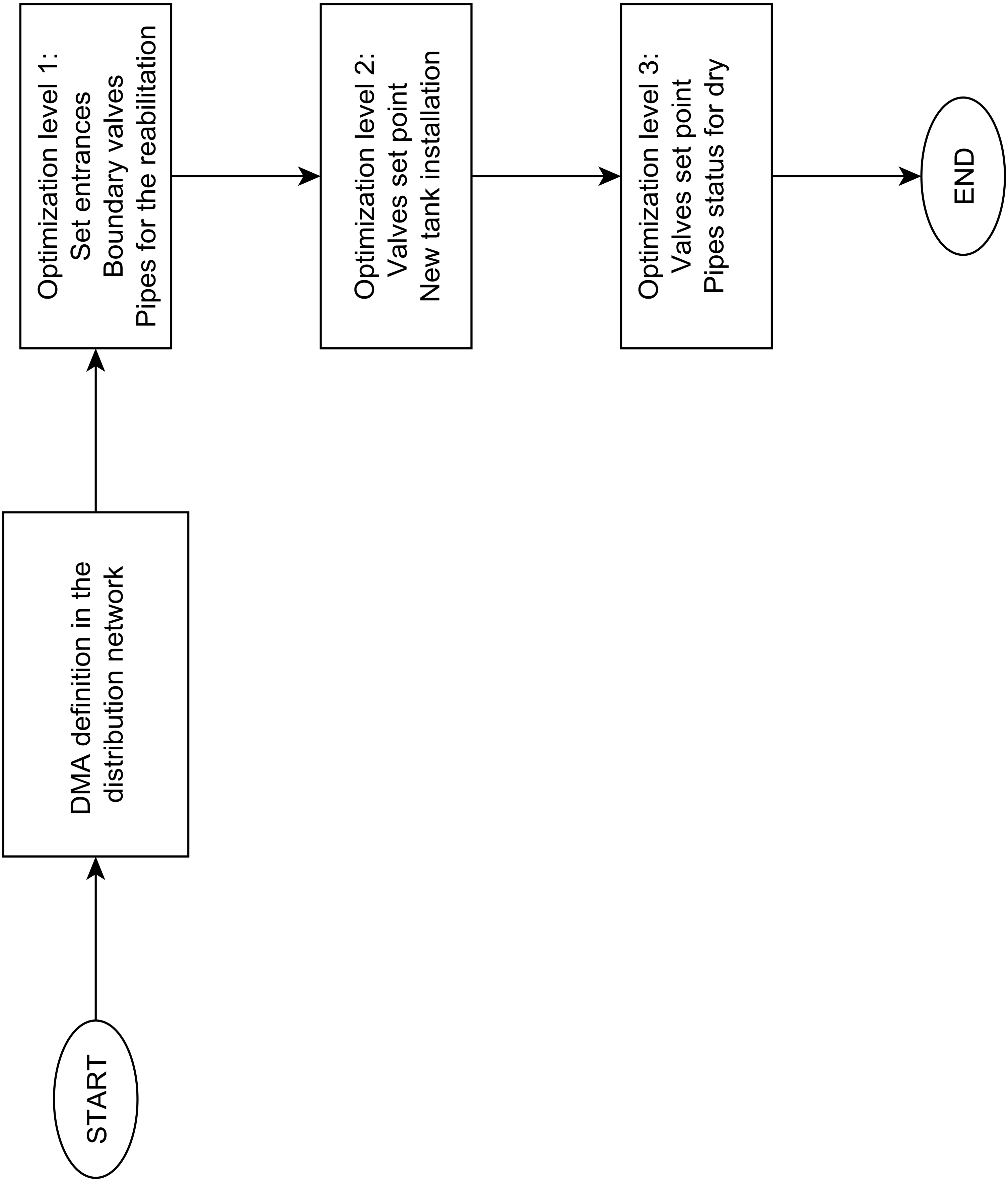
DMA	Number of nodes	Demand [l/s]	Average Elevation [m]	DMA entrance	DMA exit	Number of PRVs
0	344	103.86	0.91	2	-	1
1	1614	123.84	9.90	3	-	2
2	1052	148.44	2.30	3	0	2
3	2780	133.65	14.39	14	1, 2, 9, 10, 11, 12, 13, 14	1
4	1364	165.53	2.82	10, 6	7	2
5	638	132.69	0.92	8	-	1
6	1514	98.30	5.83	9	4	2
7	508	139.95	1.14	4	-	1
8	2872	175.37	13.29	10	5	2
9	1622	122.13	7.33	3, 13	6, 13	2
10	2230	137.62	5.10	3, 11	4, 8, 12	2
11	1740	116.78	14.54	3, 12	10	2
12	1065	83.90	10.12	3, 10	11	2
13	1712	102.30	11.30	3, 9	9	2
14	1115	132.93	20.24	3	3	1

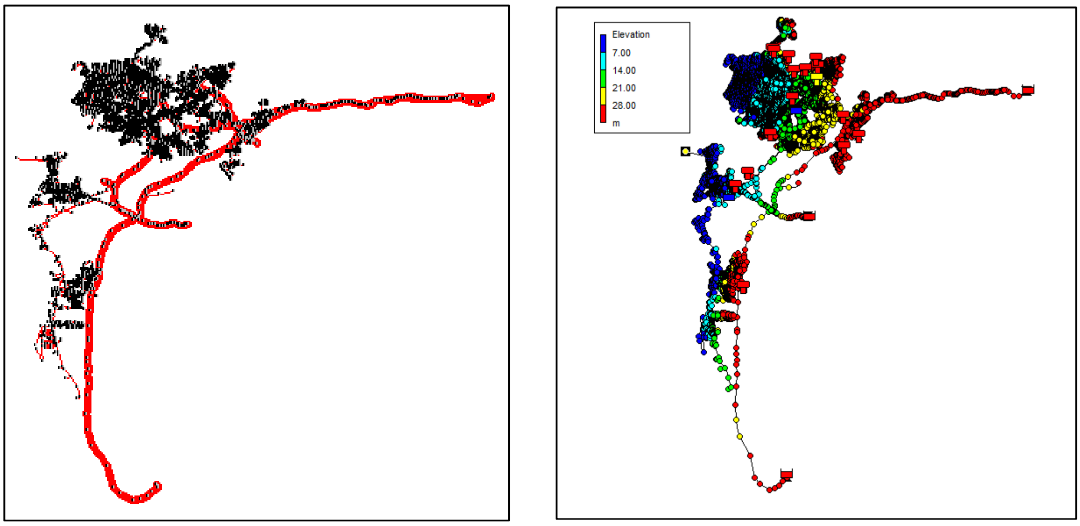
Table 2 - Hydraulic constraints evolution during the segregation-optimization process

	Original Network	**Sectorized Network**	**Duplicated Trunk Network**	**Third Cuza Pipeline**	**Final PRV settings and Pipe Closure**
Demand nodes with pressure under 15 m	445	8815	95	146	143
Nodes with negative pressure	79	2420	6	0	0
Nodes with pressure above 60 m	3304	1003	6072	6123	4763
Minimum pressure	-22.93	-	-8.92	0	0
Maximum pressure	91.72	117.35	92.32	101.23	103.48
Pressure uniformity	3.6	-	4.17	4.23	3.98

Table 3- Costs obtained for network rehabilitation

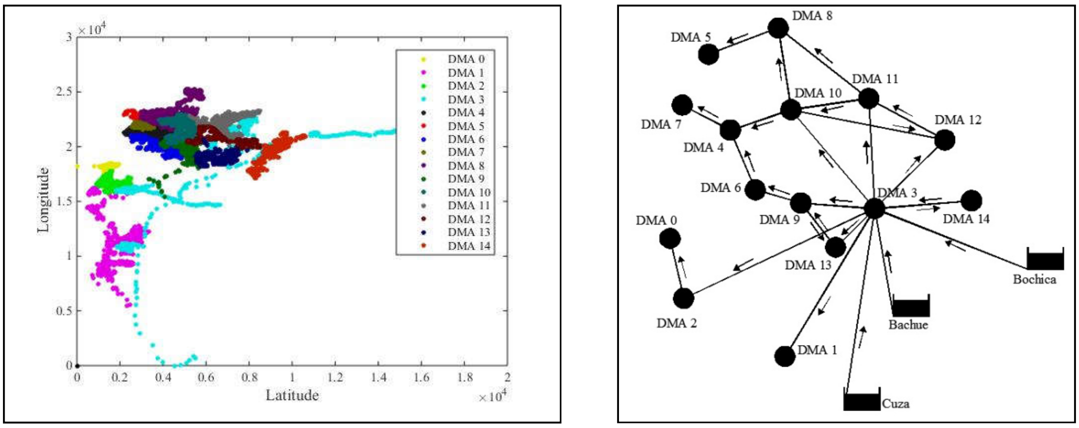
Implementation Costs	
Pipes	12,896,666
Tanks	814,742
Valves	159,614
Total	13,871,022


START
DMA definition in the distribution network
Optimization level 1:
Set entrances
Boundary valves
Pipes for the reabilitation
Optimization level 2:
Valves set point
New tank installation
Optimization level 3:
Valves set point
Pipes status for dry
END



(a) **(b)**

Figure 2 - E-town map a) WDN highlighting the trunk network. b) Nodal elevation



(a) **(b)**

Figure 3 - DMA distribution: a) Geographical distribution; b) Block diagram

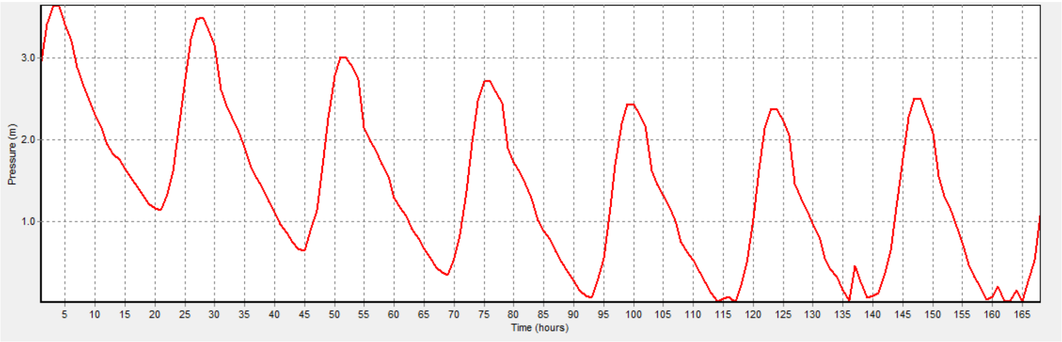


Figure 4 - Level fluctuation during a week in Tank #1

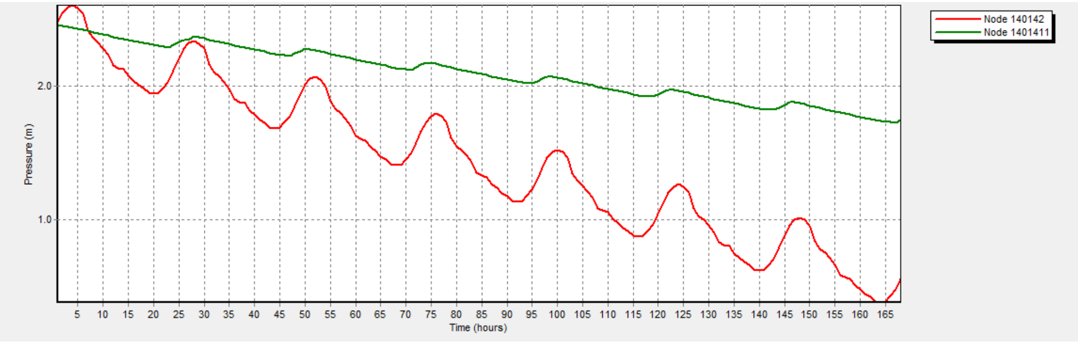


Figure 5 - Level decreasing during a week in Tanks #11 and #2

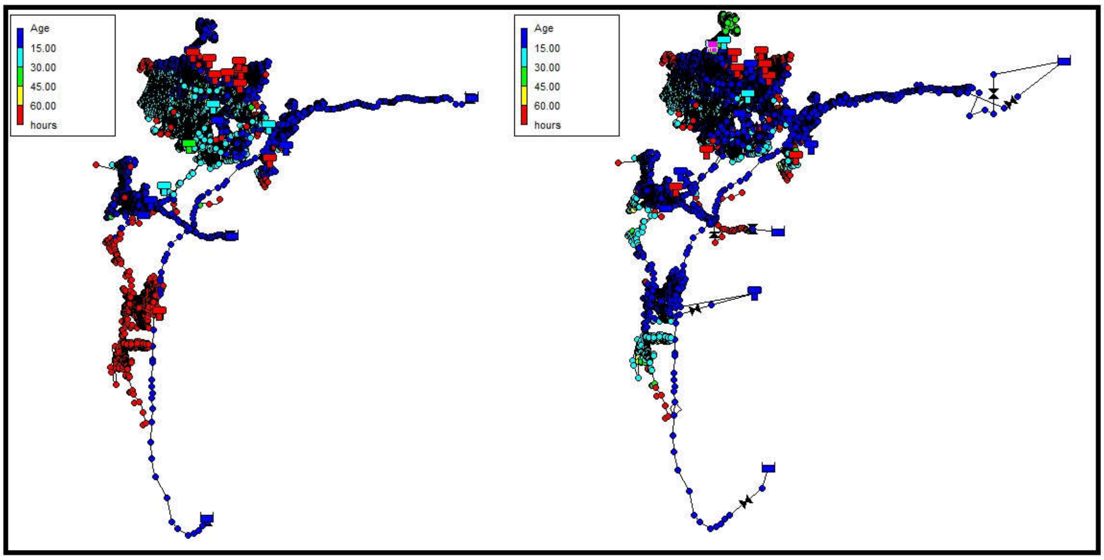


Figure 6 - Age map for the original water distribution network (left) and the final optimized network (right)

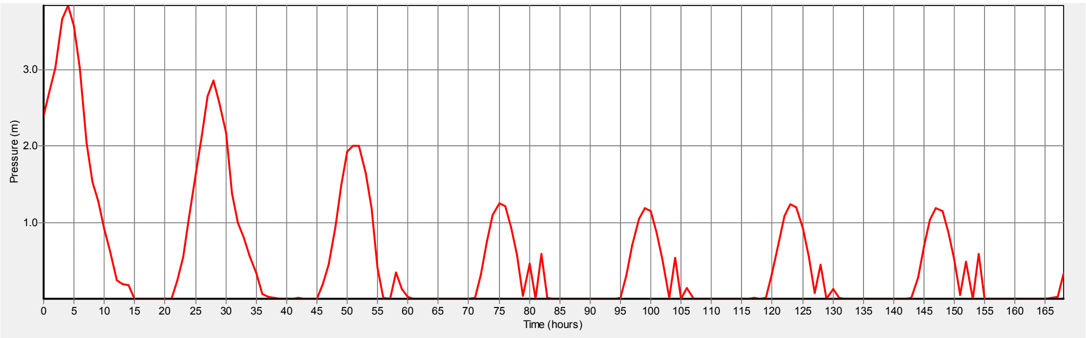


Figure 7 - Tank level for Tank #1 during one week

TABLE CAPTION LIST

FIGURE CAPTION LIST

Figure 1 - Flowchart with complete optimization procedure.

Figure 2 - E-town map a) WDN highlighting the trunk network. b) Nodal elevation

Figure 3 - DMA distribution: a) Geographical distribution; b) Block diagram

Figure 4 - Level fluctuation during a week in Tank #1

Figure 5 - Level decreasing during a week in Tanks #11 and #2

Figure 6 - Age map for the original water distribution network (left) and the final optimized network (right)

Figure 7 - Tank level for Tank #1 during one week