Efficient resizing and topological optimization of real-world water distribution networks in a multi-criteria decision-making framework

Minaei, Amin; Zecchin, Aaron C.; Hajibabaei, Mohsen; Mitrovic, Djordje; van Laarhoven, Karel; Vertommen, Ina; Alexander, Brad; Blokker, Mirjam; Savic, Dragan; Creaco, Enrico

DOI
10.1111/mice.70036

Publication date
2025

Document Version
Final published version

Published in
Computer-Aided Civil and Infrastructure Engineering

Received: 31 March 2025 | Accepted: 30 July 2025

DOI: 10.1111/mice.70036

RESEARCH ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

Efficient resizing and topological optimization of real-world water distribution networks in a multi-criteria decision-making framework

Amin Minaei[1,2] | **Aaron C. Zecchin**[3] | **Mohsen Hajibabaei**[2] | **Djordje Mitrovic**[4] | **Karel van Laarhoven**[4] | **Ina Vertommen**[4] | **Brad Alexander**[5] | **Mirjam Blokker**[4,6] | **Dragan Savic**[4,7] | **Enrico Creaco**[8] | **Robert Sitzenfrei**[2]

[1]Austrian Academy of Sciences, Vienna, Austria

[2]Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Innsbruck, Austria

[3]School of Architecture and Civil Engineering, the University of Adelaide, South Australia, Australia

[4]KWR Water Research Institute, Nieuwegein, The Netherlands

[5]Optimatics, Adelaide, South Australia, Australia

[6]Faculty of Civil Engineering and Geosciences, Delft University of Technology, CM Delft, The Netherlands

[7]Centre for Water Systems, University of Exeter, Exeter, UK

[8]Dipartimento di Ingegneria Civile e Architettura, Università degli Studi di Pavia, Pavia, Italy

Correspondence
Robert Sitzenfrei, Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria.
Email: robert.sitzenfrei@uibk.ac.at

Abstract

This study addresses complex multi-objective optimization challenges in large-scale, real-world water distribution networks (WDNs). The primary objectives are to improve a water quality index (water age) and network resilience by optimizing pipe diameters and network topology as decision variables. The proposed approaches leverage the non-dominated sorting genetic algorithm II (NSGA-II) producing Pareto optimal alternatives for water utility decision-makers. To enhance computational convergence runtime and solution quality of optimization, novel techniques are employed. These include advanced NSGA-II constraint handling, search space reduction, graph theory-based formulation of decision variables, constraints, and objective functions, as well as multi-stage and hydraulic-free optimization strategies. Furthermore, soft constraints are relaxed and integrated into Pareto decision-making spaces to provide a comprehensive, multi-criteria decision-making framework. The approaches are applied to a real case study, and the results demonstrate optimization performance improvements, with efficiency increasing by approximately 20% (in terms of convergence speed). Additionally, water age is reduced by 52% while achieving favorable results in the hydraulic and topological criteria. These findings highlight the effectiveness of the proposed methodologies in addressing WDN optimization challenges.

1 | INTRODUCTION

The use of computational search to optimize water infrastructure is well-established. Optimization in water distribution network (WDN) design was initially tackled with a single-objective linear programming algorithm (technique), which selects pipe diameters (decision variables) for small synthetic networks, minimizing design costs (objective function) while keeping nodes' piezometric heads higher than desirable values (a system constraint)

(Alperovits & Shamir, 1977). Since then, developments on WDN optimization algorithms' main elements, which are objective function, decision variables, constraints, and techniques, that is, linear or nonlinear mathematical modeling, have advanced to higher degrees of complexity (Mala-Jetmarova et al., 2018).

Regarding advancements in objective functions, optimization techniques have evolved from single-objective to multi-objective algorithms for solving WDN problems with conflicting goals, typically balancing total network cost with hydraulic performance (resilience or reliability) to be minimized and maximized (Minaei et al., 2019; Todini, 2000; Zheng et al., 2016).

The most used decision variables in optimization problems have been pipe diameters, the sizing of which is the engineers' main objective (Cisty et al., 2016; Dandy et al., 1996; Muhammed et al., 2017). In this context, researchers recently introduced a pipe removal action to decision variables in a novel approach aimed at optimizing real-world WDN topology. Notably, Vertommen et al. (2022) partly changed a large-scale WDN topology from looped to tree shape through the pipe removal strategy for improving water age at demand nodes (DNs). However, they did not develop a method through which water age is calculated within the optimization process, and instead they introduced network volume (pipes diameter times pipes length as a substitute objective with water age). Minaei et al. (2023) proposed pipe removal interventions for decoupling an old WDN from its neighboring systems, that is, sewer and road networks, to optimize the WDN layout minimizing the risk of cascading failures between neighboring systems. However, they did not propose efficiency improvement methods for topology change optimization, which is a very challenging optimization problem.

Constraints in WDN optimization problems have been classified into three categories (Mala-Jetmarova et al., 2018):

- **Hydraulic constraints** dictated by physical laws governing fluid flow in pipe networks, such as the conservation of mass.
- **System constraints** imposed by operational and design requirements, including minimum/maximum pressure at DNs.
- **Constraints on decision variables**, such as limitations on pipe diameters.

Approaches for handling these constraints include integrating hydraulic solvers (e.g., EPANET) for automated hydraulic constraint enforcement, using a standard penalty function (Kang & Lansey, 2012), applying a penalty function with a self-adaptive penalty multiplier (Wu & Simpson, 2002), or employing (modified) constraint tournament selection (Minaei et al., 2020).

While hydraulic constraints are essential in optimization models, Mala-Jetmarova et al. (2018) found that 5% of models in the literature impose no constraints, mainly in multi-objective optimization, where pressure requirements are treated as objectives rather than constraints. Nearly half (48%) of the models include only one constraint, typically the minimum pressure requirement. About 25% incorporate two constraints, while 13% and 9% of models feature exactly three and four or more (up to 10) system constraints, respectively.

Various evolutionary, nature-inspired, and metaheuristic optimization algorithms—such as genetic, harmony search, and spiral dynamic algorithms—have been developed to tackle complex optimization problems beyond WDNs (Adeli & Cheng, 1994; Siddique & Adeli, 2014, 2015). The stochastic nature of WDN optimization problems makes them well-suited to these types of algorithms.

Among the optimization algorithms´ techniques, the genetic algorithm (GA), an evolutionary method following Darwin´s evolution law (Goldberg, 2013), is a well-established technique for solving WDNs problems in the literature. When it comes to multi-objective optimization problems, the non-dominated sorting genetic algorithm II (NSGA-II; Deb et al., 2002) has gained popularity and has been used by many researchers (Jafari et al., 2021; Wang et al., 2023; Zhao et al., 2019). These algorithms solve optimization problems stochastically, where solutions and decision variables represent chromosomes and genes in GA.

Diverse case studies from small to large sizes have been used in the literature concerning WDN optimization problems to show the application of optimization models. For instance, Alperovits and Shamir (1977) demonstrated the application of the optimization algorithm to a two-loop network supplied by gravity (incl. seven junctions). More recently, Santonastaso et al. (2021) demonstrated the applicability of such methods for finding efficient quality detection points in a network with 999 junctions. Optimizations for pressure and leakage management in two real WDNs, one with 10,100 and another with 992 junctions were presented by Ulusoy et al. (2022) and Shahhosseini et al. (2023), respectively. Mottahedin et al. (2023) solved a leakage rehabilitation optimization problem for a network with 2869 junctions, and Moeini and Abokifa (2024) applied Bayesian optimization to chlorine management in WDNs demonstrating its effectiveness for a network with 113 junctions.

Looking at all applications, two important challenges emerge: The first concern is the degree to which optimization results are applicable in real industrial projects. In this regard, Walski (2014) noted the lack of practical considerations in many WDN optimization studies, which often leads to non-applicable master plans for WDN design. The second challenge is the inefficiency of the stochastic optimization models in identifying near-global optimal solutions. As the size of the networks grew, inefficiency became more evident. This is because the search space for large networks is very large, and exploring such spaces often incurs incredibly high computational costs. Hence, some studies paid attention to the optimizations' computational efficiencies and ways to improve optimization convergence (Bi et al., 2015; Zecchin et al., 2005; Zheng et al., 2017).

Various strategies have been proposed to enhance the computational efficiency of complex optimization problems. These strategies generally fall into two categories: inside-engine methods, which focus on improving the core mechanics of the optimization algorithms, and outside-engine methods, which target the initial and boundary conditions of the problem itself. Inside-engine approaches involve modifying the algorithmic framework or key parameters to enhance performance. For instance, Hao et al. (2022) improved the global search capability of NSGA-II by integrating Lévy distribution into the non-dominated sorting and solution diversification processes. Similarly, efforts to optimize tournament selection, mutation, and crossover parameters have been made to further enhance NSGA-II's performance (Antkiewicz & Myszkowski, 2024; Carles-Bou & Galán, 2023; Yi et al., 2020).

On the other hand, outside-engine methods focus on engineering the problem setup itself—by refining case study formulations, manipulating decision variables, interchanging and tuning the number of objectives and system constraints, or developing surrogate models. Riyahi et al. (2023) categorized these methods into four distinct groups:

- **Graph topology-informed methods**: These methods utilize graph theory algorithms, such as the shortest path algorithm, or topological indices to develop surrogate models for optimization problems (Pudasaini & Shahandashti, 2020; Sitzenfrei et al., 2020). Instead of directly generating optimal solutions, optimization algorithms are primarily used to validate the results of the surrogate models.
- **Expert choice methods**: In this approach, expert judgment is leveraged to generate the initial population for optimization algorithms based on the specific nature of the problem (Minaei et al., 2020).
- **Multi-stage optimization methods**: This method involves a multi-stage optimization process where suboptimal solutions are initially generated using different parameter scenarios. These suboptimal (warm start) solutions are then refined in subsequent optimization stages (Cisty et al., 2016; Hajibabaei et al., 2024; McClymont et al., 2013).
- **Gene coding scheme**: This approach enhances convergence speed by employing a novel gene coding scheme inspired by topological and hydraulic metrics (Diao et al., 2022).

In all the efficiency-improving studies, the primary assumption is the fixed topology of the network over a design period. This is an important assumption as topology changes within an optimization process increase immensely the number of infeasible solutions generated in every optimization generation, leading to a slow optimization convergence (Simpson et al., 1994). Additionally, the studies on improving the efficiency of optimization for minimizing the water age in real-world WDNs are few due to the necessity for extended period simulation of such problems. This study aims to address this critical research gap by developing efficiency-enhancing methods that consider changes in large-scale network topology while improving both water age and resilience. In this study, the efficiency improving approaches focus only on the outside-engine methods and not improving or developing a new engine core of NSGA-II. Moreover, we propose a comprehensive multi-criteria post-optimization scheme increasing the applicability of designed master plans out of optimizations. Ultimately, our research will provide a clear roadmap for future studies on efficient optimization algorithms for various types of WDN optimization problems.

2 | METHODOLOGY

The methodology includes the following five subsections: (1) hierarchical, topological, and graph characteristics of WDNs; (2) reference study's optimization (RSO) approach; (3) strategies for enhancing the computational efficiency of RSO; (4) post hoc multi-criteria decision making; and (5) summary and computational assessment for the design approaches, described in detail in the following subsections. The first subsection enables readers to understand the technical topological, graphical, and hydraulic terms used in the subsequent subsections. The second subsection explains the original study work on which improvements are carried out by the current study. Subsection three explains the ways through which the computational efficiency and flexibility of the original study's design approach can be improved. The fourth subsection explains how every solution on a Pareto front can be assessed by different criteria, and the fifth subsection summarizes differences between efficiency-improving approaches and

TABLE 1 Properties of primary, secondary and tertiary water distribution network (WDN) parts in real-world WDN (Vertommen et al., 2022).

	Explanation	Regulation	Design shape
Primary	Pipelines with $Dj \geq 300$ mm, primarily responsible for bulk water transport without direct customer connections (CCs)	Based on Dutch drinking water law, guaranteeing reliable water distribution and system resilience	Looped configuration
Secondary	Pipelines with 160–300 mm diameter, responsible for distributing water. These pipes may include CCs as needed	Robust system design ensuring uninterrupted water delivery during hydraulic disturbances, maintaining water quality, and, whenever possible, protecting public health and reliability.	The system uses loops for redundancy, aiming to reduce pipe diameters and promote unidirectional flow for improved efficiency and control
Tertiary	Pipelines with 40–160 mm diameters, typically used for CCs in the WDN	Meeting fire flow and water quality standards	Branched, avoiding sediment with flushing

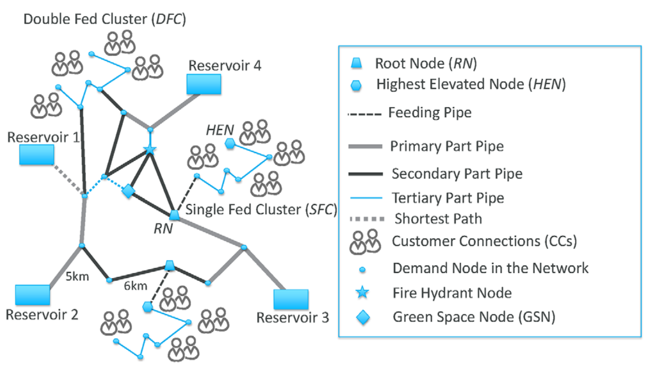

FIGURE 1 A conceptualized water distribution network (WDN) with different components.

computational aspect for convergency of the optimization models.

2.1 | Hierarchical, topological, and graph characteristics of WDNs

Different sections of a WDN serve distinct functions and have varying levels of importance in ensuring water supply to customers. Vertommen et al. (2022) classified the complex topology of large-scale WDNs into three categories based on Dutch definitions:

1. Primary network
2. Secondary network
3. Tertiary network

A detailed explanation of these classifications is provided in Table 1. Figure 1 shows a conceptualized multi-reservoir WDN in which primary, secondary, and tertiary parts are identified. Among the three parts, the tertiary has great importance due to the inclusion of a high number of downstream customer connections (CCs). Moreover, these parts are usually at the furthest points from reservoirs, hence having the largest values of the water age (Note all symbols are defined in the nomenclature in this paper).

The tertiary part includes clusters that can be either single-fed clusters (SFCs), which are fed by only one pipe connected to the secondary part of the network or multiple-fed clusters where they are fed by more than one pipe connected to the secondary part (like the double fed cluster in Figure 1). However, SFCs are important as the CCs in SFCs are at a higher water supply risk, meaning that the failure of feeding pipes isolates customers from reservoirs due to the lack of redundant paths.

Junctions in a layout could have different functionalities. Some of these relate to DNs that receive water over a 24-h period. Moreover, they could be fire hydrant or green space nodes (GSNs) that are operational only at specific times over an extended period. Different types of junctions in a WDN model are depicted in Figure 1. Among the junctions, root nodes (RNs) play the role of a feeding source for the associated SFCs. Hence, an RN needs to have sufficient piezometric head (hydraulic pressure plus elevation) to feed the associated SFC. The highest elevated node (HEN) in an SFC has a higher probability of being fully or partially unsupplied, compared to other DNs in the associated SFC. Hence, a higher piezometric head of the RN compared to the HEN guarantees a supply to all nodes of the associated SFC.

A WDN can be demonstrated as a graph G, consisting of a set of vertices (e.g., DNs) and edges (e.g., pipes). The shortest path index is a well-established graph metric used to approximate the path of a water particle traveling from a source, such as a reservoir, to a DN. The shortest path, $\sigma_{z,k}$, in G is the path between demand nodes z and k that has the minimal sum of positive edge weights, also referred to as the minimal path length; for example, in Figure 1, among all available paths from reservoirs to the GSN, the shortest path is associated only with Reservoir 1 as indicated by the dots. The shortest path length can be defined in various ways depending on the weight assigned to each pipe, such as the shortest Euclidean distance, least friction losses, or shortest residence time (Sitzenfrei et al., 2020).

In line with mimicking the flow path, another relevant graph metric is the edge betweenness centrality (EBC) for each pipe j, which measures how frequently an edge j appears in the shortest path from a source s to every demand node $z \in$ DNs (Brandes, 2008). Specifically, for WDN analysis, Sitzenfrei et al. (2020) introduced a modified version, demand-weighted EBC (EBCQ(j)), which adjusts the shortest path by incorporating the demand Q_z of DN z to EBCQ(j) along $\sigma_{s,z}$,

$$EBCQ\,(j) = \sum_{z\epsilon DNs} \sigma_{s,z}\,(j) \cdot Q_z, \left[0, \sum_{z\epsilon DNs} Q_z\right] \tag{1}$$

The interval in front of Σ in Equation (1) shows the range of values that can be taken on by $EBCQ(j)$. This changes from 0, meaning that edge j is never located in the shortest path, to $\sum_{z\epsilon DNs} Q_z$ meaning that edge j is in the shortest path of all DNs connected to the sources.

2.2 | RSO approach

Vertommen et al. (2022) proposed an approach for solving a real-world WDN problem concerning high water age levels at the consumption nodes, prompting the need for a network redesign. This redesign involved optimizing both the pipe diameters and the overall topology to effectively tackle the water quality issue. Hence, Vertommen et al. (2022) formulated an optimization problem using GA, to solve the problem with one objective, where n_p, SP and PP are the number of pipes, secondary and primary parts, respectively, and $P_1 - P_4$ stand for the penalty functions (explained by Equations 3–6), respectively; $D_j \in cd$ meaning diameters belong to the set of commercial pipe diameters (cd) ranging from the smallest available diameters to the largest, and L_j refers to the length of pipe $j; j = 1$, …, n_p; As can be seen, only the pipes belonging to secondary and primary sections participate in the calculation of the objective function.

$$\text{Minimize} \sum_{j=1}^{n_p} D_j \cdot L_j + P_1 + P_2 + P_3 + P_4, \forall_j \in SP \cup PP \tag{2}$$

The pipe removal action completes through assigning the fictitious 0.0001 mm diameter to pipes within the optimization process. Hence, decision variables involve pipe resizing with the set of commercial diameters plus topology changes through pipe removal action.

In their study, the hydraulic constraints were met through using hydraulic solver (EPANET 2.2). There are four system constraints in the problem referring to both hydraulic and topological aspects of the network handled by penalty functions and evaluated by Equations (3)—(6).

In Equation (3), c_1 is a penalty factor that could take different ranges of values, from small values to infinity, depending on the importance of the associated constraint defined by decision-makers. p_{enf} stands for enforced pressure at RNs, ΔZ_i represents the elevation difference between RN and the highest elevated node for the associated SFC i, where $i = 1, \ldots, n_{SFCs}$, and n_{SFCs} is the number of SFCs. p_i refers to the water pressure at RN i, and $i = 1, \ldots, n_{RNs}$ where n_{RNs} is the number of RNs. SFCs and RNs stand for the SFCs and RNs sets, respectively, and these two sets have the equal number of members.

$$P_1 = c_1 \max\left\{0, \max_{i \in RNs\ \&\ SFCs} \left\{p_{enf} + \Delta Z_i - p_i\right\}\right\} \tag{3}$$

In Equation (4), c_2 is a penalty factor, p_z is the pressure at every DN z where z belongs to the DNs set, DNs, $z = 1, \ldots, n_{DNs}$, and n_{DNs} stands for the number of DNs. Note that every type of nodes, introduced in Figure 1, has a set and the sum of all the nodes´ sets is the set of all junctions in the network.

$$P_2 = c_2\ \max\left\{0, \max_{z \in DNs} \{-p_z\}\right\} \tag{4}$$

In Equation (5), c_3 is a penalty factor, and p_{des} is a desirable pressure, defined by national drinking water regulations.

$$P_3 = c_3 \sum_{i \in RNs\ \&\ SFCs} \max_{i \in RNs\ \&\ SFCs} \{0, p_{des} + \Delta Z_i - p_i\} \tag{5}$$

In Equation (6), c_4 is a penalty factor and the constraint implies that the number of CCs in every SFC ($N_{CCs}^{SFC_i}$) should be less than the allowable number of CCs ($N_{CCs}^{allowable}$).

$$P_4 = c_4 \sum_{i=1}^{n_{SFCs}} \left(\max_{i \in SFCs}\left\{0, N_{CCs}^{SFC_i} - N_{CCs}^{allowable}\right\}\right)^2 \tag{6}$$

Every case study has a subset of pipes that, for practical reasons, cannot be changed (fixed pipes). These pipes are not represented as decision variables in RSO. Hence, the decision variables exclude fixed pipes from the design process within the optimization.

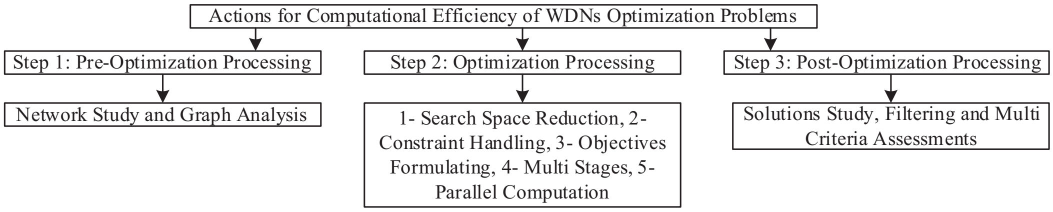

FIGURE 2 Different steps and actions for improvements in the computational efficiency and practicality of real-world WDN optimization problems.

2.3 | Strategies for enhancing the computational efficiency of RSO

The RSO refers to a very complex and constrained optimization problem, and therefore its application to real-world WDNs is computationally burdensome. Hence, the current study proposes some actions to enhance the computational efficiency and flexibility of RSO. These actions are applied in three steps: (1) pre-optimization processing, (2) optimization processing, and (3) post-optimization processing (see Figure 2).

In Step 1, a comprehensive study of the problem's nature, case-study regulations, and practical considerations is carried out to find decision variable constraints, for example, fixed pipes in RSO. Furthermore, engineering judgment and graph analysis were used to select "out of design" pipe candidates. For example, if a problem includes a snapshot hydraulic simulation and a pipe is connected to a dead-end node without demand, the pipe does not need to be redesigned within the optimization process.

In Step 2, there are five actions that could improve the optimization´s computational efficiency. Having understood the decision variable constraints in the previous step, a search space reduction can be applied by imposing limitations on decision variables. While constraints on decision variables could bring benefits to computational efficiency in optimization, hydraulic and system constraints could be important impediments to solution quality. Hence, constraints should be handled with frugality, and efforts should be made to minimize the number of systems and hydraulic constraints. Objectives should be formulated with the least need for hydraulic simulation (e.g., graph-based objectives) as hydraulic simulation arguably impacts optimization run duration. Parallel computing is another technique to cope with computational costs.

In Step 3, the focus is on solutions obtained in Step 2, checking feasibility, followed by sorting and prioritization through multi-criteria assessments.

The next sections describe four computational efficiency approaches. Therein, the optimization run can go through either one or multiple times "Stages" in "Step" 2 (Figure 2). For example, pre-optimization processing completes (Step 1) to enter optimization processing (Step 2). In Step 2, optimization is run to obtain Stage 1 results and use those results to initiate subsequent optimization run (Stage 2).

2.3.1 | Approach 1 (AP1)

This approach refers to one-stage optimization leading to flexible decision making as the problem with two objectives is solved by the NSGA-II engine, minimizing a graph-based water age (GWA) index and maximizing a hydraulic resilience index of the network as shown in Equation (7) where *GWA* and *GRF* stand for the water travel time, calculated only by graphical properties of the network (Sitzenfrei, 2021), and generalized resilience failure (Creaco et al., 2016a) indices for the whole network, respectively.

$$\text{Minimize } (GWA, -GRF) \tag{7}$$

The mathematical model indices are calculated by Equations (8) and (9) where for every DN z, $n_{p,z}$ refers to the number of pipes belonging to the associated shortest path ($\sigma_{s,z}$); for every pipe j, the $\frac{L_j}{V_j}$ refers to the pipe j length divided by the pipe j velocity (calculated by a hydraulic solver, i.e., EPANET), which is water particle residence time in pipe j and is considered as weights for shortest path calculation.

$$GWA = \frac{\sum_{z=1}^{n_{DNs}} \sum_{j=1}^{n_{p,z}} \left(\frac{L_j}{V_j}\right)}{n_{DNs}} \tag{8}$$

$$GRF = I_r + I_f \tag{9}$$

The GWA index represents the average water travel time over the whole network, which gives a good approximation for the real value of water age in the network (Sitzenfrei, 2021), named GWA in this study. GRF is equal to the sum of two dimensionless resilience and failure indices. The resilience index I_r is non-negative and expresses the ratio of the power excess supplied to users to the power excess exiting from WDN sources. The failure index I_f is non-positive and shows the ratio of the power deficit supplied to users to the desired power.

The two indices are calculated as follows where Q_0 (m^3/s) and H_0 (m) are the vectors of outflow and head from source, in turn; q_{user} (m^3/s) is the vectors of real outflow from nodes to users, calculating the demand after reducing the leakage with the pressure-driven approach. H (m) is nodal heads, referring to the operational real heads of the network. d (m^3/s) and H_{des} (m) are the nodal demand vectors and desired nodal heads for full demand satisfaction (the theoretical design factors), respectively; finally, Qp (m^3/s) and Hp (m) are the pump flows vectors and heads, respectively.

$$I_r = \frac{\max\ \left(q_{user}^T H - d^T H_{des}, 0\right)}{Q_0^T H_0 + Q_p^T H_p - d^T H_{des}} \tag{10}$$

$$I_f = \frac{\min\ (q_{user}^T H - d^T H_{des}, 0)}{d^T H_{des}} \tag{11}$$

It should be noted that the numerator and denominator in Equations (10) and (11) have units of $\frac{m^4}{s}$ and are proportional to hydraulic power (i.e., flow × head), but they omit the constant physical parameters such as fluid density and gravitational acceleration. In summary, at pressure deficit conditions, $I_r = 0$ and $-1 < \mathrm{GRF} < 0$ for the network; at pressure surplus conditions, $I_f = 0$ and $0 < \mathrm{GRF} < 1$ for the network.

The outflow is calculated based on three conditions: (1) when the pressure head at a node is lower than the minimum required head, the node is considered out of services, that is, no demand is satisfied; (2) when the pressure head of a node is higher than the minimum head but lower than the service head, the node is considered partially satisfied; and (3) when the pressure head of a node is higher than the service head, the node not only meets the full design demand but also contributes to increased network resilience during emergencies (Creaco et al., 2016a). In this context, a minimum pressure ($p_{\min}$) is used to determine the minimum head as described in Wagner et al. (1988).

The decision variables are the same as those in RSO. However, AP1 imposes additional constraints on the decision variables by leveraging a graph theory-based approach. This method excludes pipes that lie on the shortest paths from being considered for removal during the optimization process. This aligns with a search space reduction strategy aimed at improving the computational efficiency of the optimization. The shortest path is determined through network decomposition, where each pipe's weight is determined as the ratio length/diameter, based on the assumption that a water particle prefers paths with shorter lengths and large diameters (Hajibabaei et al., 2023). Additional constraints on the decision variable include not considering pipes in clusters connected to non-DNs and fixed pipes for design decisions, further reducing the search space.

Regarding the four system constraints of the problem, Equations (3)–(6), AP1 categorizes them into the two groups of (1) soft and (2) hard constraints. In this study, a network design solution cannot be considered operable if it violates hard constraints. If, however, it violates soft constraints, it can still function, though such solutions are not desirable from a design perspective. Constraints one and three are relevant with two different level of significance as $p_{enf} < p_{des}$, meaning that meeting constraint three will certainly meet constraint one; hence, AP1 considers constraint one and three as hard and soft constraints. Other hard and soft constraints are constraints two and four as meeting the former one makes a network without negative pressures at nodes, and meeting the latter one just decreases the number of CCs under water supply risk. The soft constraints are considered as post-optimization criteria for evaluating Pareto fronts' solutions, and therefore only the two system hard constraints are modeled in the optimization. Hence, AP1 has only two system constraints, and the two hydraulic constraints (conservation of mass and energy) are met automatically in the hydraulic solver.

In contrast to the way constraints are handled in RSO, AP1 handles the constraints through a tournament selection method in NSGA-II, which self-adaptively finds feasible solutions within the optimization process. In this regard, Minaei et al. (2020) showed the superiority of this method for the optimization efficiency, compared to the penalty factors method. The improved NSGA-II compares a couple of individuals X and Y that are members of the population and δ_X and δ_Y refer to the X and Y violations; on this basis, Y is dominated by X only if:

- X meets feasibility constraints ($\delta_X = 0$), and Y does not meet ($\delta_Y \neq 0$);
- X and Y both do not meet the feasibility constraints, but $\delta_X < \delta_Y$;
- X and Y both meet the feasibility constraints, but the X solution dominates the solution of Y;
- otherwise, Y dominates X.

Please refer to Minaei et al. (2020) for learning more about the method.

2.3.2 | Approach 2 (AP2)

This approach refers to a one-stage optimization problem formulation with two graph-based objectives and no hydraulic and system constraints, generating a wide variety of layouts for WDNs. To the best of our knowledge, WDN optimization modeling without hydraulic and system constraints is rare. However, it has the advantage of searching beyond limited feasible spaces but the potential shortcoming of generating a high number of infeasible solutions. The two objectives are the topological resilience index (Equation 12) and the network cost estimated by Equation (13). The topological resilience index invented by Creaco et al. (2016b) is calculated with Equation (12) where, C_u refers to a topological resilience index in this study, and n_{pl} is the number of pipes in loops (pipes in branched structures are not included in n_{pl}); C_l is the loop diameter uniformity index calculated by the ratio of the mean to the maximum diameter of the generic lth of the n_l loops (i.e., a loop with five pipes, average diameter 110 mm and maximum diameter 250 mm has $C_l = 0.44$).

$$C_u = \frac{n_{pl}}{n_p} \times \frac{\sum_{l=1}^{n_l} C_l}{n_l} \tag{12}$$

The cost function is calculated with Equation (13) where uc_j is the unit cost for pipe j depending on the diameter of pipe. In this study, cost refers to the monetary value of pipes across the entire network, including both old and resized pipes (this does not represent the total replacement cost).

$$Cost = \sum_{j=1}^{n_p} uc_j \cdot L_j \tag{13}$$

In AP2 optimization, the key points are the correlation of objectives with the ones introduced in AP1 and the degree of conflict between them. These issues are assessed, and their impact will be shown in the results section. Decision variables and associated constraints are similar to those in AP1.

2.3.3 | Approach 3 (AP3)

So far, two approaches have been explained. The approaches are different in terms of techniques for improving the computational efficiency in RSO. For example, by trying AP1 on the RSO and analyzing the AP1 results, one could understand how relaxing soft system constraints, graph-informed objective function and decision variable constraints, and NSGA-II with self-adaptive constraint handling could improve the optimization speed of convergence and the solutions quality. By trying AP2 and comparing the output results, one could understand how much hydraulic-free optimization could improve computational efficiency.

In line with trying different efficiency-improving techniques, AP3 refers to two optimization stages where the first stage is the same as the optimization in AP2. The solutions from stage one are used as "warm solutions" for initiating the stage two optimization, which is exactly AP1 optimization. By trying AP3, one could understand the impacts of initiating the optimization with warm solutions on the computational efficiency of designs generations.

2.3.4 | Approach 4 (AP4)

In this approach, using a graph theory-based method, optimal diameters are designed for the different network topologies obtained from AP2 optimization (referring to Step 2 in Figure 2 including one stage optimization). The graph theory method for pipe diameter design refers to the state-of-the-art approach by Sitzenfrei et al. (2020) and Hajibabaei et al. (2023) where pipes' diameter selection is completed through the following two tasks:

Task 1: Graph-based network decomposition: The process begins with network decomposition using a graph-theoretic framework. In WDNs with multiple sources, DNs are associated with specific reservoirs or sources. This association is based on nodal head values, ensuring that each node is supplied by the most appropriate source. Consequently, the first task involves estimating the nodal heads and decomposing the networks generated by AP2 into subnetworks. This graph-based estimation of nodal heads considers the elevation of each source node and the energy dissipation along the shortest path from each source to every DN. The detailed methodology for this estimation process is elaborated in Hajibabaei et al. (2023).

Task 2: Graph-based pipe diameter selection: Once the network is decomposed into subnetworks, the graph-based design approach proposed by Sitzenfrei et al. (2020) is applied to each subnetwork to determine the optimal diameter for each pipe from a topological point of view, neglecting topographical arrangements. This method utilizes the EBCQ metric (Equation 1), which provides a robust estimation of the flow in each pipe (edge). By estimating the flow in each pipe within a subnetwork, the continuity equation is applied to compute the next larger available diameter for each pipe j, in Equation (14), where $D_{available}$ refers to the available discrete diameter of pipe j (mm), and V_{design} refers to design velocity (m/s) defined by reasonable ranges, which can be found in national drinking water regulations of the case study. If the calculated diameter exceeds the maximum available size in the predefined set, the maximum available diameter is assigned.

$$D_j = \left\lceil \sqrt{\frac{4}{\pi} \cdot \frac{EBCQ\,(j)}{V_{design}}} \right\rceil \in D_{available} \tag{14}$$

Following the design of individual subnetworks, the optimal solutions for all network topologies are compiled into a post-processing pool. These solutions undergo feasibility analysis and non-dominated sorting based on the objective functions of AP1, resulting in the formation of the relevant Pareto front (see Step 3 in Figure 2).

As seen, EBCQ refers to graph theory design, which needs deep graph analysis and study of the network. Hence, this has the potential to give optimal solutions to existing networks before applying any optimization. This relates to Step 1 in Figure 2. However, as EBCQ design is applied to every solution generated from the optimization in AP2, this falls in the post-processing for AP4 (Step 3 in Figure 2). Readers can refer to pseudocodes provided in the online repository of the University of Innsbruck (https://doi.org/10.48323/qezag-gkb63) for a better understanding of AP4.

2.4 | Post hoc multi-criteria decision making

The optimization results in the form of Pareto fronts undergo post-processing to be assessed by different criteria explained in the following:

1. Criterion 1 (hard constraints): The sum of values in Equations (3) and (4) that refer to the strict testing of solution feasibility. These are the sum of two system constraints introduced as hard constraints in AP1.
2. Criterion 2 (pressure deficit of RN from desirable pressure): This has been introduced in Equation (5) where the value of this criterion is obtained without considering the penalty factor, referring to one of the soft constraints, thus showing the influence of the constraint relaxation on the optimization performance.
3. Criterion 3 (number of SFCs with more than maximum allowable number of CCs): This refers to another soft constraint of this study (Equation 6).
4. Criterion 4 (hydraulic resilience): This refers to the GRF index introduced in Equation (9).
5. Criterion 5 (topology resilience): This is the criterion introduced in Equation (12) expressing the performance of every design solution in terms of uniformity of diameters of pipes in loops and the loops number. This criterion only focuses on the topology and graph features of every solution.
6. Criterion 6 (GWA): This is the one introduced in Equation (8), which will show the superiority of every solution in terms of GWA.
7. Criterion 7 (cost): For every solution, the design cost is calculated by the cost function (Equation 13). This will contribute to decision-makers selecting a design based on an available budget.
8. Criterion 8 (CCs population in SFCs): Besides the number of SFCs, the size of these clusters has great importance in decision making for minimizing the number of CCs affected by the failures of the feeding pipes.
9. Criterion 9 (number of CCs to removed pipes): In the RSO algorithm, Vertommen et al. (2022) allowed the pipes with CCs to be candidates for removal. The reason for this is the existence of a redundant path for supplying such customers. They have also added such customers size to the number of CCs in the closest SFC to avoid this undesired design outcome. For the sake of comparability with previous methodologies, this study allows the optimization algorithm to remove pipes with CCs; however, the number of CCs associated with removed pipes is considered a criterion for evaluating solutions.
10. Criterion 10 (Total number of customers with high risk of supply): In this study, the customers without redundant paths for water supply are considered at a high risk of receiving substandard supply service where N_{CCs}^{HRS} is the total number of CCs with a high risk of substandard supply service because of being connected via only one supply path, and N_{CCs}^{RPs} stand for the number of CCs associated with removed pipes.

$$N_{CCs}^{HRS} = N_{CCs}^{RPs} + N_{CCs}^{SFCs} \tag{15}$$

As seen, the criteria can be categorized into the three groups, (1) feasibility constraints (CR1–CR3), (2) performance metrics (CR4–CR7), and (3) customer risk indicators (CR8–CR10). Moreover, criteria and optimizations' objectives are interchangeable. For example, topology resilience is an optimization objective in AP2, while this is a criterion for the outputs of AP1. This way Pareto fronts provide further insights for decision-makers needing to select the best design.

Even though the post hoc evaluation could significantly contribute to the applicability of the design, there could remain a certain level of practical infeasibility when it comes to the construction phase, as real projects are usually logistically complex, cost-intensive, and often constrained by existing urban infrastructure, service continuity requirements, and regulatory approvals. Hence, the mentioned criteria should be modified and adapted to every particular case, making the master plan as practically feasible and possible.

2.5 | Summary and computational assessment for the design approaches

The methodology proposed four approaches for improvements of RSO in two aspects of solutions quality and

TABLE 2 The summary of design approaches used in this work.

Step 1		Step 2															Step 3		
		Optimization																	
		Stage 1							Stage 2										
Pre-processing		Obj*		Con*			DV*		Obj		Con			DV			Post-processing		
	Graph analysis	1	2	HC	SC	DVC	PR*	TD*	1	2	HC	SC	DVC	PR	TT		EBCQ-D*	FSF&S*	CA*
AP1	√	GWA*	HR*	2	2	3	√	√	–	–	–	–	–	–	–		–	–	√
AP2	√	Cost	TR*	–	–	3	√	√	–	–	–	–	–	–	–		–	–	√
AP3	√	Cost	TR	–	–	3	√	√	GWA	HR	2	2	3	√	√		–	–	√
AP4	√	Cost	TR	–	–	3	√	√	–	–	–	–	–	–	–		√	√	√

Abbreviations: CA*, criteria assessment; EBCQ-D*, demand edge betweenness centrality design; FSF&S*, feasible solution finding and sorting; GWA, graph-based water age; HR*, hydraulic resilience; PR*, pipe resizing; TD*, topology design; TR*, topology resilience.

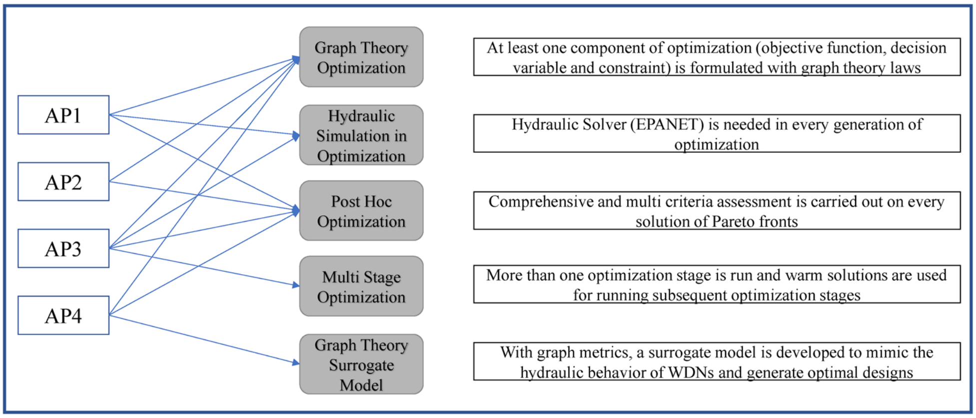

FIGURE 3 Table 2 demonstration and fundamental theoretical connections and distinctions between the approaches.

convergence computational speed. To accomplish this, different approaches have been employed including (i) the two-objective optimization using GWA and hydraulic resilience of a network, (ii) search space reductions by imposing constraints on decision variables, (iii) graph theory design, (iv) multi-stage optimization together with hydraulic- and constraint-free optimizations. Table 2 and Figure 3 show the differences between approaches. Numbers in every row in Table 2 define the number of optimization components including objectives (Obj), constraints (Con) which are 1- hydraulic constraints (HC), 2- system constraints (SC) and 3- decision variables constraints (DVC).

As shown in Table 2 and Figure 3, all approaches use graph analysis in the pre-optimization step. The only approach with two optimization stages refers to AP3. All approaches solve the optimization problem with two objectives. Only the AP2 and AP4 include hydraulic-simulation-free optimizations (either for objective or constraint calculations). The decision variable types in all approaches are the same, which are pipe resizing and a network's topology design. Only AP1 and AP3 have system constraints. Two hydraulic constraints are used for all approaches that perform hydraulic simulations.

Regarding boundary conditions and limitations, while AP1 and AP3 are the most complete approaches in terms of constraint modeling in optimization and the generation of a high number of feasible solutions, they always need to use a hydraulic solver, which can negatively impact the speed of convergence. On the other hand, AP2 and AP4 leverage hydraulic-free optimization, which can accelerate convergence but are weaker in terms of generating feasible solutions. Hence, the results from the approaches will show whether graph theory in optimization contributes to a fast search for hydraulically feasible solutions. This will provide guidance to practitioners in selecting appropriate methods for practical applications.

Overcoming the computational burden of the proposed design processes is important and challenging as the scale and complexity of the problem are very large and high. To improve this issue, a parallel computation technique was used to run all the optimizations and simulations facilitated by the state-of-the-art approach to parallel

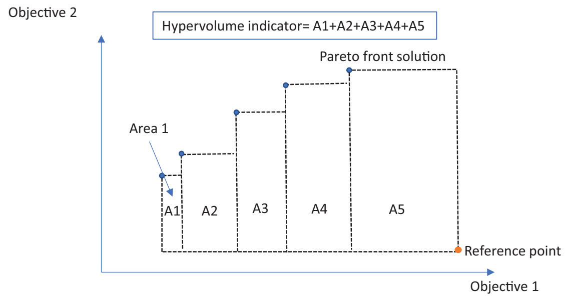

FIGURE 4 Illustrative hypervolume indicator for a Pareto front sample with five solutions: A1–A5 refer to Areas 1 until 5.

computing (Eliades et al., 2016). The criteria for assessing the computational performance of the processes are the number of objective function evaluations (N.F.E) and the time (hours) taken to meet the convergence criterion. N.F.E is the number of individuals in NSGA-II times the number of generations.

A hypervolume indicator is used (Wang et al., 2023) as the convergence criterion. The integral of areas below Pareto fronts regarding reference point (Figure 4) is calculated from generation to generation within an optimization process. Typically, this is done by adding a small margin (e.g., 10%) to the worst objective values during the optimization run to ensure the hypervolume is properly bounded. The change in area after a certain number of generations should be less than a threshold value (θ) so that the convergence criterion is met and the optimization terminates.

3 | CASE STUDY

An anonymous and very large-scale WDN in City X was employed to show the application of the proposed optimal design approaches. The utility's goal is to transition toward an alternative network design that improves water quality, for instance, by reducing water residence times. The network layout is shown in Figure 5A. The network model has 5191 links, including isolation valves and pipes. It has 4798 junctions, including demand, fire hydrant and GSNs, and six reservoirs. Among the six reservoirs, only Reservoirs 3, 4, 5, and 6 refer to real ones feeding the network, and the rest are pumping stations modeled as reservoirs. The peak water demand observed on the day with the highest demand in the past 10 years at 9:00 a.m. was increased by 10% in the model to account for potential future demand growth. This adjusted demand is used by the water utility to evaluate its service requirements under this future scenario.

The maximum allowable CCs in an SFC are 50 where enforced and desirable pressure values at RNs are 22.5 and 28 m, respectively. The minimum and service pressures for calculating the nodes' real outflows are 5 and 25.49 m, respectively, while leakages are neglected in this study. The current average GWA is 9.02 h (Equation 8), and the hydraulic resilience is about 92%, showing that the network is currently performing well in terms of water supply. As regards the hard constraints, the network meets the hard constraints, but the soft constraints are not met. There are five SFCs with more than 50 CCs, and the deficit violation of the desirable pressure at the RN reaches the values of 2.26 m (Criterion 2) for all RNs. Even though the network has many SFCs, most of them are not considered because of their small sizes. Among the SFCs, SFC4 has the maximum number of CCs where the total number of CCs under high supply risk is 397 people for the existing network.

TABLE 3 Commercial pipe diameters sizes and materials, polyvinyl chloride (PVC), ductile cast iron and fiberglass (FNG) (Vertommen et al., 2022).

ID	Outer D_j (mm)	Roughness (mm)	Unit cost (€/m)
PVC-100	110	0.05	110
PVC-150	160	0.05	160
FNG-200	222	0.10	222
FNG-250	274	0.10	274
FNG-300	326	0.10	326
FNG-400	429	0.10	429

The existing network's monetary value is around €26.86 million (obtained by Equation 13). There are six different pipe types with different materials and diameters to design the network as shown in Table 3.

4 | RESULTS AND DISCUSSION

The NSGA-II Pareto fronts converged using a population of 1000 individuals, the mutation rate of 0.07, and θ = 0.01 (the threshold for hypervolume index change, which is an expert choice). The choice of population size is an important aspect affecting the optimization process performance. The larger it is, the better the coverage of trials in the search space, but when it goes beyond a threshold value, it could negatively affect computational efficiency. Hence, a population size of 1000, as well as other NSGA-II parameters like crossover and tournament selection parameter rates, were tuned through initial trial-and-error runs (more information about these parameters can be found in the pseudocodes in the online repository of the University of Innsbruck: https://doi.org/10.48323/qezag-gkb63). All the Pareto fronts from

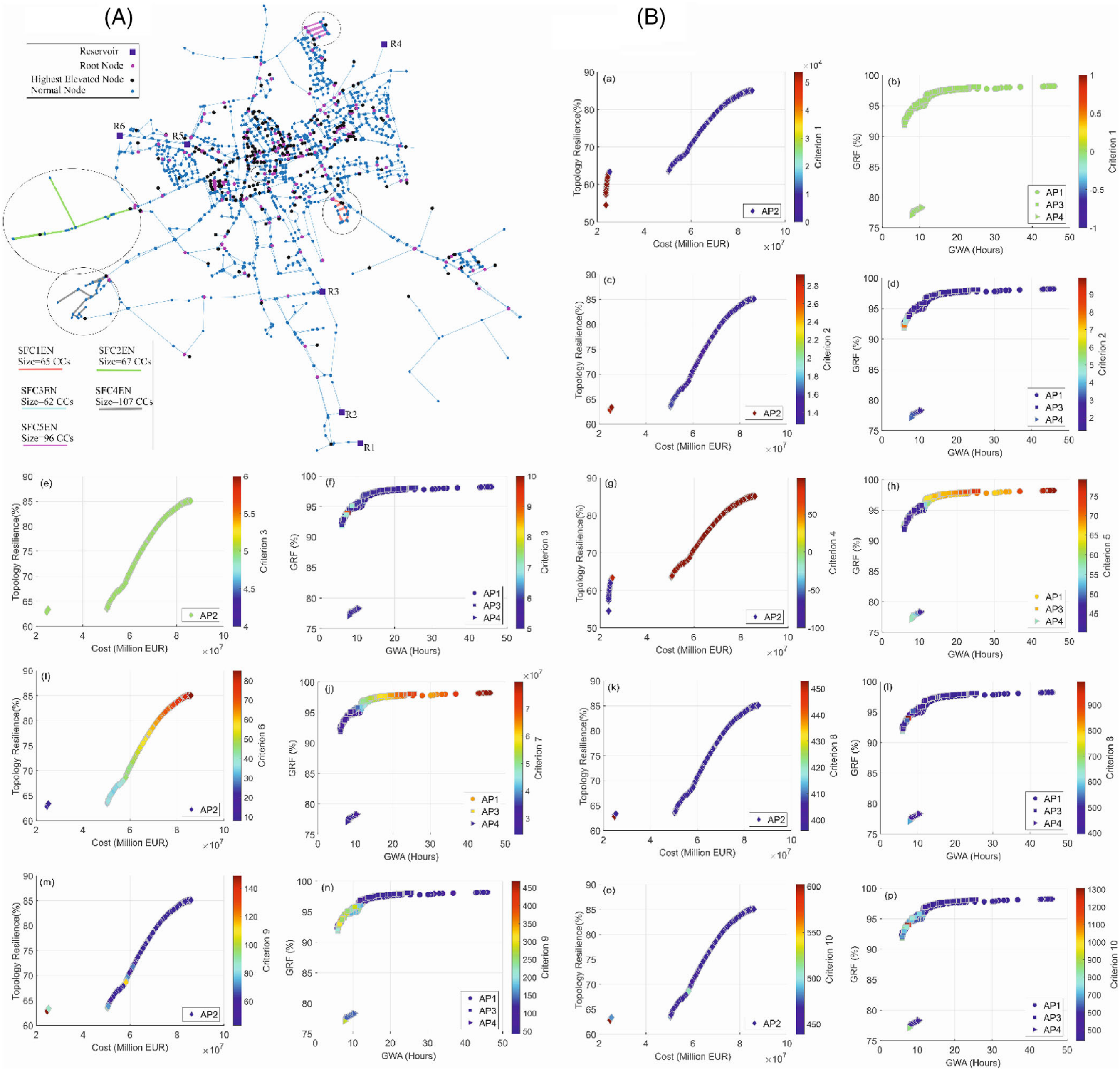

FIGURE 5 (A) The case study layout, including the single-fed clusters and junctions' properties. (B) Multi-criteria result for assessments of Pareto front solutions from different approaches.

different approaches converged after a certain N.F.E., and design solutions were generated after a certain processing time.

All design optimization algorithms were coded in MATLAB R2021b. The optimization runs for the four approaches were run on the same computer, which has an Intel(R) Core (TM) i9-14900K 3.2 GHz processor and 128 GB of RAM. To ensure unbiased results, identical memory and CPU resources were allocated for every optimization run.

4.1 | Approaches evaluations with the quality of solutions

Figure 5B presents the Pareto fronts for all approaches where every Pareto front solution is evaluated by different criteria. Each criterion is represented as a third dimension for every solution, displayed on the right side of each graph, with values indicated by the markers' face colors. The Pareto front for AP2 is shown separately in all graphs because its optimal design process focuses on balancing the objectives of cost and topology resilience. In contrast, the optimal design processes in AP1, AP3, and AP4 aim to balance two different objectives: GWA and GRF.

As illustrated, all Pareto fronts demonstrate well-formed trade-offs, validating the effective selection of conflicting objectives across all the proposed multi-objective optimization models.

Comparing Pareto fronts in Figure 5B(a) shows that AP2 includes some infeasible solutions (the red diamonds), whereas all solutions in AP1, AP3, and AP4 are feasible. Among the 1000 optimal solutions from AP2 Stage 1, only 82 were infeasible (91.8% frequency of feasible solution generation) demonstrating a high frequency feasible solution. The infeasible solutions in AP2 are primarily associated with cost-effective designs that feature smaller diameters and fewer loops, compared to more expensive designs. Hence, they cannot provide a diverse set of cost-effective solutions. The significant jump in AP2 solutions from a design costing approximately €25 million to one costing €50 million is due to the existence of 12 large diameters in the secondary part of the network for the expensive designs, which are removed in the cheap designs. Practically, this provides valuable insight for decision-makers when there are budget limitations and limited feasible options for a low-cost network rehabilitation plan. For example, by investing an additional amount (around €2.5 million) they may not improve the topology resilience of the design, but they can decrease the number of CCs at risk of supply failure. Therefore, the final decision depends on the policy and priorities of the decision-makers.

An interesting aspect of AP2 solutions is the number of feasible designs despite the absence of hydraulic and system constraints. This highlights the advantages of graph-informed decision-variable constraints and the strong correlation of AP2 objectives with those in other approaches. Such an approach offers promising potential for real-world WDN optimization, providing a faster method for generating near-optimal solutions (the approaches processing run time is explained and compared in the next section). The infeasible solutions in AP2 are displayed only in Figure 5B(a) and 5B(g) to illustrate the infeasibility effects on hydraulic resilience. In all other graphs for AP2, the infeasible solutions have been filtered out.

Examining the Pareto fronts in Figure 5B(b), it is evident that the solutions from AP1 and AP3 dominate those from AP4. This suggests that while EBCQ-D is a computationally efficient method for pipe sizing, it requires further development to effectively address combined topology and diameter design problems. Additionally, the approach falls short, compared to others in generating diverse solutions and feasible outcomes. Among the 301,000 generated solutions in AP4, only 364 are feasible (0.0003%), demonstrating a very low frequency of feasible solutions generation in AP4.

The differences between AP1 and AP3 solutions are minimal, with AP3 demonstrating a slight advantage at the knee points. This observation indicates that two-stage optimization does not significantly enhance the performance efficiency of the current study's optimization problem in terms of solution quality.

Regarding Criterion 2 (Figure 5B(c) and 5B(d)), none of the solutions achieve a value of 0, which would indicate meeting the desirable pressure for RNs. In general, the design solutions perform reasonably well in this aspect, except for some designs in AP1 and AP3 that exhibit short water ages, with the criterion values ranging between 7 and 9 m. This may be attributed to specific configurations following pipe removal, which lowers the RNs' pressure while maintaining short water ages.

For Criterion 3 (number of SFCs with more than 50 CCs), AP2 solutions show minimal violations, with values of 5 or higher corresponding to infeasible solutions that are not displayed in Figure 5B(e). In contrast, AP1, AP3, and AP4 generate more diverse solutions for this criterion, with no significant violations observed. As expected, the hydraulic resilience is low for infeasible solutions of AP2 (Figure 5B(g)); however, solutions with high topology resilience have high hydraulic resilience. Similarly, topology resilience (Criterion 5) shows higher values for those solutions with higher values for GRF, which means there is a good correlation between these two types of resilience. The correlation between other optimization objectives is well shown in Figure 5B(i) and 5B(j), where the longer water age results in higher cost. Figure 5B(k) and 5B(l) evaluates the Pareto front solutions in terms of CCs sizes in SFCs with more than 50 CCs. As expected, the AP2 solutions are not very diverse; however, AP1 includes a design with a maximum of 700 CCs in SFCs. Compared to the CCs associated with removed pipes (Criterion 9), AP2 solutions show higher diversity for Criterion 8. In this regard, solutions from AP1, AP3, and AP4 have higher CCs associated with removed pipes. Consequently, the total CCs under risk are shown in Figure 5B(o) and 5B(p) where the highest number goes to 1300 for AP3.

4.2 | Approaches evaluations with the speed of convergence

Table 4 shows optimization and post-processing computational times for different approaches. The pre-processing duration in this study is not considered as it requires expert input that could be different from study to study. For every 1000 individuals, the

TABLE 4 Computational costs of design approaches.

	N. F. E			Time (h)					
	Optimization			Optimization		Post processing			
Approach	Stage 1	Stage 2	Sum.	Stage 1	Stage 2	EBCQ-D*	FSF&S*	CA*	Sum
AP1	202,000	–	202,000	56.11	–	–	–	0.31	≈56.42
AP2	459,000	–	459,000	19.41	–	–	0	0.31	≈20
AP3	459,000	750,000	1,209,000	19.41	208.33	–	–	0.31	≈228.05
AP4	459,000	–	459,000	19.41	–	0.27	80	0.11	≈100

Abbreviations: AP1, Approach 1; CA, criteria assessment; EBCQ-D, discharge edge betweenness centrality design; FSF&S*, feasible solution finding and sorting.

objective function evaluation time was measured multiple times, and the most frequently occurring value was selected to calculate the total computational runtime required to achieve convergence of the Pareto fronts.

As shown in Table 4, AP2 exhibits the fastest computational time, requiring approximately 20 h, while AP3 demonstrates the slowest computational time, taking around 228.05 h. This highlights the advantage of hydraulic-free optimization approach and the relaxation of soft constraints, which significantly accelerates the solution generation process. AP3 has been identified in the literature as a fast approach for diameter design problems. However, its limitations in this context may be due to the low quality of some individuals among the warm-start solutions. Since Stage 1 lacks a filtration process for hard and soft constraints, it can produce solutions that are feasible in terms of layout or diameters but not both. These differences can cause the optimization in Stage 2 to start far from the global optimal solution, thereby potentially misleading the NSGA-II algorithm. Readers should bear in mind that the convergence criterion is strictly enforced in this study, which could be another reason for the observed slow convergence.

Despite being a hydraulic-free optimization approach, AP4 does not achieve the fastest convergence due to a computationally intensive post-processing step. In Stage 1, 1000 designs with varying configurations are generated, and the EBCQ-D approach generated 301 optimal designs for each configuration, resulting in 301,000 total designs. During the EBCQ-D procedure, V_{design} is increased from 0.5 to 3.5 m/s with the intervals of 0.01 m/s, resulting in 301 potential solutions for each obtained network's topology out of AP2.

While solo diameter design for 1000 configurations takes just 16.33 min using EBCQ-D, identifying feasible solutions and sorting the Pareto front requires a further 80 h. However, AP4 achieves the fastest criteria assessment time, at around 6.83 min (0.11 h), due to the smaller number of solutions in its Pareto fronts, compared to other approaches.

4.3 | Comparisons between selected solutions from approaches and GWA validation

In line with the goal of the original study (Vertommen et al., 2022), solutions from different approaches with the shortest GWA are selected and compared in Table 5. As regards SFCs with more than 50 CCs (CR3), AP1 performs the best and has the same value as the existing network.

As for hydraulic resilience (CR4), AP1 and AP3 solutions make better solutions while leaning the network. They perform better than AP2 and AP4 because of the optimization objectives in AP1 and AP3, where GRF is one of the main objectives.

Topology resilience (CR5) is improved by AP1 and AP2, and the AP2 solution is the best. This is because the criterion is one of the direct objectives of optimization in AP2.

The best rank for GWA (CR6) goes for AP1, while the least cost designs (CR7) are achieved by AP2 and AP4. The best rank for the size of SFCs with more than 50 CCs (CR8) and the number of customers associated with removed pipes (CR9) is achieved by AP1, while the solutions from AP3 and AP4 bring about the highest number of CCs under water supply risk (CR10).

Looking at all the results of different approaches, it could be understood that the performance of those that have worked well on problems with fixed topology cannot be guaranteed for topology change optimization. There is no definitive answer to the question of which approach is the best globally for all optimization efficiency problems. However, graph theory optimization implementation together with goal-oriented graph-informed decision variable constraints, relaxing soft constraints and multicriteria Pareto assessments, could significantly cast light on the road toward increasing the efficiency of WDNs optimization problems.

One important argument in the current study could be the validity of how GWA was calculated and used as one objective in the optimization models. For this, 100 solutions were selected randomly from AP1 Pareto front,

TABLE 5 Comparisons of selected design solutions from different approaches and the existing network.

Name	CR*2 (m)	CR3	CR4 (%)	CR5 (%)	CR6 (h)	CR7 (M EUR)	CR8	CR9	CR10
EX	2.26	5	92	53	9.02	26.86	397	0	397
AP1	4.21	5	92	60	5.87	25.68	397	81	477
AP2	2.85	5	87	63	8.15	24.66	453	149	602
AP3	4.72	7	91	47	5.96	26.11	637	209	864
AP4	3.33	6	77	57	7.85	24.10	573	290	863

Abbreviations: AP1, Approach 1; CR*, criteria, EX*, existing.

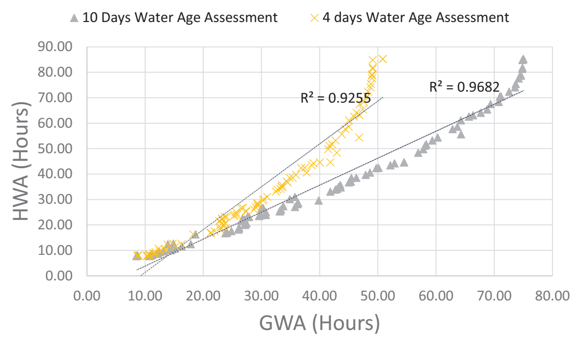

FIGURE 6 Hydraulic and graph-based water ages for 100 random solutions of Approach 1 (AP1) and coefficient of determinations values for two simulation times.

and the hydraulic-based water age (HWA) was calculated by EPANET 2.2 and compared with GWA on the graphs shown in Figure 6. For this, water ages for two simulation periods of 10 and 4 days were analyzed assuming a fixed demand pattern. As seen, for both simulation periods, the coefficient of determinations shows high values with 0.96 and 0.92 for 10- and 4-days simulations, respectively. It is worth mentioning that the two-timing experiments on the validity of GWA provide strong evidence of its low sensitivity to the extended period intervals and topology variations as the solutions in Figure 6 contain different topologies. We also ran a sensitivity analysis to see how the 95% confidence interval for $R^2 = 0.96$ changes with different sample sizes, 10, 50, 100, 200, 300, 400, and 500. The results showed that once the sample size exceeds 100, the interval stabilizes, ranging between 0.93 and 0.97. This suggests that increasing the sample size beyond 100 has little impact on the reliability of the estimate.

HWA for RSO solution was calculated roughly the same as the one calculated for the current study´s best solution in terms of water age (AP1 solution with the lowest GWA). While the RSO solution was obtained after about 250,000 N.F.E, the best solution in this study was achieved only after 202,000 N.F.E. Vertommen et al. (2022) reported that hydraulic experts from the water utility calculated the HWA for RSO solution and existing designs and a decrease from 7.5 to 3.6 h was achieved. This means that the current study's optimization result improved the water age by around 52% with around 20% higher efficiency than RSO and better in terms of hydraulic resilience with 92% against 74%, respectively. Accordingly, the improvement in the efficiency relates only to the speed of convergence during optimization. As shown in the literature, constraint handling through tournament selection in NSGA-II has a greater impact on solution quality than on computational runtime (Minaei et al., 2020). Hence, when comparing the key differences between AP1 and RSO, the efficiency gains mainly come from graph-informed constraints on decision variables, simpler objective function evaluations, fewer hard constraints, and more effective optimization post-processing using soft constraints.

5 | SUMMARY AND CONCLUSION

Pipe diameter and topology optimization of WDNs is a complex and recently developed concept in the literature of real-world WDN optimization problems. Moreover, large-scale WDN optimization problems aiming to maximize water quality and hydraulic resilience are highly constrained and computationally expensive decision-making problems. This study, for the first time, proposed novel approaches for the efficiently solving such optimization problems by using a broadly adopted evolutionary multi-objective optimization algorithm called NSGA-II. The approaches leverage various techniques that have already been introduced in the literature for pipe resizing problems, including (1) graph theory design where optimal pipe diameters are generated by the use of a surrogate model following only graphical laws, (2) multi-staged optimization where outputs of prior stage become warm solutions for initiating the subsequent stage, (3) tournament selection method for handling constraints in NSGA-II rather than penalty factors method, and (4) decreasing constraints of the problem through relaxing soft constraint and defining constraints-inclusive objective functions. Accordingly, this study developed a high-performance single-stage evolutionary optimization algorithm where search space reduction is achieved by imposing a graph-informed constraint on decision variables, excluding shortest path pipes from removal decisions. Another strong feature of the model is its GWA objective function, which serves as a robust surrogate for HWA. The model is equipped with a comprehensive hydraulic resilience indicator, generating numerous alternatives along a Pareto front, ranging from solutions with pressure deficit to pressure surplus. Soft constraints are relaxed and serve as criteria for assessing alternatives, while hard constraints are handled self-adaptively through the tournament selection method. Another important feature of this study is the multi-criteria decision-making model, which allows many exchangeable objectives and criteria to be applied in evaluating alternatives from the Pareto front. Finally, the superior approach (AP1 solution with the lowest GWA) managed to provide a design solution that improved the water age by around 52% and 20% higher efficiency than the reference study's approach. Moreover, with the superior solution, other topological and hydraulics aspects of the network including hydraulic and topology resilience, the number of SFCs with more than 50 CCs, desirable piezometric heads at RNs of the SFCs, cost and the total number of CCs at high risk of supply could be kept at a good level of desirability for decision-makers.

This study introduced a road map to assist future WDN optimization studies where the three methodology steps (pre-processing, optimization processing, and post-processing) could be used to achieve high-quality, computationally efficient optimal and practical solutions. Additionally, a parallel computing technique is used for large-scale and time-consuming optimization problems. However, there are important simplifications, limitations and assumptions in the study as follows:

- The real-world WDN was simplified such that many components and details were not simulated, including leakages, pumps, valves, and so forth.
- Uncertainties in variables and parameters can significantly impact the results. These include uncertainties in the strict hypervolume indicator, different network characteristics (velocity, demand, and flow conditions), and NSGA-II parameters such as crossover, mutation and tournament selection rate operators.
- The performance of only one evolutionary optimization algorithm, NSGA-II, was assessed and others such as water drop algorithms, particle swarm, simulated annealing, and so forth, were not specifically considered. The outside-engine approach presented in this work will be applied on other optimizers in future endeavors.
- The degree of acceptability of relaxing soft constraints in practical engineering applications was not analyzed in this study.
- Water age and the uniformity of improvements in DNs were not assessed.
- There are different parallel computing techniques and heuristic simplifications, such as reducing the number of NSGA-II generations or using adaptive population sizes, which were not investigated.
- The operating cost was not considered for evaluating the cost function.
- The sensitivity of GWA to different demand patterns and network topologies has not been analyzed.

These issues should receive appropriate attention, and each one opens a new window for future research work. Also, new research prospects were identified with showing the weak points of previous approaches, which worked well for pipe diameter design problems. For example, graph theory pipe diameter design could produce 301,000 solutions in only 16.33 min, while the scale of achieving optimal solutions in this study is in days. However, this method has not yet worked well when it comes to a combined diameter-topology design of real-world WDNs. This is a great future research direction to advance graph theory design, which can derive many optimal diameters, and topology designs in the scale of minutes without using optimization.

NOMENCLATURE

δ_X	Constraint violation of individual X
δ_Y	Constraint violation of individual Y
N_{CCs}^{HRS}	Total number of CCs with a high risk
N_{CCs}^{RPs}	Number of CCs associated with removed pipes
$N_{CCs}^{SFC_i}$	Number of CCs in single-fed cluster i
$N_{CCs}^{allowable}$	Number of allowable CCs
AP	Approach
c	Penalty factor for penalty function
CCs	Customer connections
cd	Commercial diameters
C_l	The ratio of the mean to the maximum diameter of the generic lth of the n_l loops
Con	Constraints
C_u	Loop diameter uniformity index
d	Nodal demands vector
$D_{available}$	Available discrete diameter
D_j	Pipe j diameter
DNs	Demand nodes
DVC	Decision variable constraints
EBC	Edge betweenness centrality
$EBCQ(j)$	Demand-weighted EBC for node j
G	Graph
GRF	Generalized resilience failure index
GSN	Green space node
GWA	Graph-based water age
H	Vector of nodal heads
H_0	Vector of heads from sources

HC	Hydraulic constraints
H_{des}	Vector of desired nodal heads
HEN	Highest elevated node
H_p	Pump head vector
I_f	Failure index
I_r	Resilience index
L_j	Pipe j length
$N.F.E$	Number of objective function evaluations
n_{DNs}	Number of demand nodes
n_l	Number of loops
n_p	Number of pipes
$n_{p,z}$	Number of pipes belonging to the shortest path between source node and demand node z
n_{pl}	Number of pipes in loop
n_{SFCs}	Number of single-fed clusters
$NSGA\text{-}II$	Non-dominated sorting genetic algorithm II
Obj	Objectives
P	Penalty function
p_{des}	Desirable pressure at RNs
p_{enf}	Enforced pressure at RNs
p_i	Pressure at root node i
$p_{\min}$	Minimum pressure
PP	Primary part
SP	Secondary part
Q_0	Vectors of outflow from sources
Q_p	Pump flows vector
q_{user}	Vector of outflow from nodes to users
Q_z	Demand at demand node z
RNs	Root nodes
RSO	Reference study's optimization
s	Source node
SC	System constraints
$SFCs$	Single-fed clusters
SPI	Shortest path index
uc_j	Unit cost for pipe j
V_{design}	Design velocity
v_j	Pipe j velocity
$WDNs$	Water distribution networks
ΔZ_i	Elevation difference RN_i and HEN_i for SFC_i
θ	Threshold value for optimization convergency
$\sigma_{s,z}$	Shortest path between source nodes and node z
$\sigma_{z,k}$	Shortest path between nodes z and k

ACKNOWLEDGMENTS

The authors from the University of Innsbruck were funded by the Austrian Science Fund (FWF), and by the Vice Rectorate for Research at the University of Innsbruck, the Austrian Academy of Sciences (ÖAW), the Federal Ministry of Education, Science and Research (BMBWF) in the Republic of Austria (OeAD-GmbH—Austria's Agency for Education and Internationalization; Mobility Programmes and Cooperation) and the University of Innsbruck for providing financial support through the Early-Stage Funding, DOC Fellowship (AP845023), Marietta Blau scholarship and doctoral scholarship which have facilitated the doctoral education of Mr. Amin Minaei and his research stays at KWR Water Research Institute in the Netherlands and the University of Adelaide in Australia. Dragan Savic has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [951424]).

REFERENCES

Adeli, H., & Cheng, N. T. (1994). Augmented Lagrangian genetic algorithm for structural optimization. *Journal of Aerospace Engineering*, *7*(1), 104–118.

Alperovits, E., & Shamir, U. (1977). Design of optimal water distribution systems. *Water Resources Research*, *13*(6), 885–900.

Antkiewicz, M., & Myszkowski, P. B. (2024). Balancing Pareto Front exploration of Non-dominated Tournament Genetic Algorithm (B-NTGA) in solving multi-objective NP-hard problems with constraints. *Information Sciences*, *667*, 120400.

Bi, W., Dandy, G. C., & Maier, H. R. (2015). Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. *Environmental Modelling & Software*, *69*, 370–381.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, *30*(2), 136–145.

Carles-Bou, J. L., & Galán, S. F. (2023). Self-adaptive polynomial mutation in NSGA-II. *Soft Computing*, *27*(23), 17711–17727.

Cisty, M., Bajtek, Z., & Celar, L. (2016). A two-stage evolutionary optimization approach for an irrigation system design. *Journal of Hydroinformatics*, *19*(1), 115–122.

Creaco, E., Franchini, M., & Todini, E. (2016a). Generalized resilience and failure indices for use with pressure-driven modeling and leakage. *Journal of Water Resources Planning and Management*, *142*(8), 04016019.

Creaco, E., Franchini, M., & Todini, E. (2016b). The combined use of resilience and loop diameter uniformity as a good indirect measure of network reliability. *Urban Water Journal*, *13*(2), 167–181.

Dandy, G. C., Simpson, A. R., & Murphy, L. J. (1996). An Improved genetic algorithm for pipe network optimization. *Water Resources Research*, *32*(2), 449–458.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

Diao, K., Berardi, L., Laucelli, D. B., Ulanicki, B., & Giustolisi, O. (2022). Topological and hydraulic metrics-based search space reduction for optimal re-sizing of water distribution networks. *Journal of Hydroinformatics*, *24*(3), 610–621.

Eliades, D. G., Kyriakou, M., Vrachimis, S., & Polycarpou, M. M. (2016). EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB. *Proceedings of the 14th International Conference on Computing and Control for the Water Industry (CCWI)*, The Hague, The Netherlands.

Goldberg, D. E. (2013). *Genetic algorithms*. Pearson Education India.

Hajibabaei, M., Hesarkazzazi, S., Minaei, A., Savić, D., & Sitzenfrei, R. (2023). Pareto-optimal design of water distribution networks: An improved graph theory-based approach. *Journal of Hydroinformatics*, *25*(5), 1909–1926.

Hajibabaei, M., Minaei, A., Shahandashti, M., & Sitzenfrei, R. (2024). Graph-based warm solutions for optimal resilience enhancement of water distribution networks. *Engineering Proceedings*, *69*(1), 102.

Hao, J., Yang, X., Wang, C., Tu, R., & Zhang, T. (2022). An improved NSGA-II algorithm based on adaptive weighting and searching strategy. *Applied Sciences*, *12*(22), 11573.

Jafari, H., Nazif, S., & Rajaee, T. (2021). A multi-objective optimization method based on NSGA-III for water quality sensor placement with the aim of reducing potential contamination of important nodes. *Water Supply*, *22*(1), 928–944.

Kang, D., & Lansey, K. (2012). Revisiting optimal water-distribution system design: Issues and a heuristic hierarchical approach. *Journal of Water Resources Planning and Management*, *138*(3), 208–217.

Mala-Jetmarova, H., Sultanova, N., & Savic, D. (2018). Lost in optimisation of water distribution systems? A literature review of system design. *Water*, *10*(3), 307.

McClymont, K., Keedwell, E. C., Savić, D., & Randall-Smith, M. (2013). Automated construction of evolutionary algorithm operators for the bi-objective water distribution network design problem using a genetic programming based hyper-heuristic approach. *Journal of Hydroinformatics*, *16*(2), 302–318.

Minaei, A., Haghighi, A., & Ghafouri, H. R. (2019). Computer-aided decision-making model for multiphase upgrading of aged water distribution mains. *Journal of Water Resources Planning and Management*, *145*(5), 04019008.

Minaei, A., Hajibabaei, M., Savic, D., Creaco, E., & Sitzenfrei, R. (2023). Optimal rehabilitation planning for aged water distribution mains considering cascading failures of interdependent infrastructure systems. *Journal of Hydroinformatics*, *25*(5), 2084–2105.

Minaei, A., Sabzkouhi, A. M., Haghighi, A., & Creaco, E. (2020). Developments in multi-objective dynamic optimization algorithm for design of water distribution mains. *Water Resources Management*, *34*(9), 2699–2716.

Moeini, M., & Abokifa, A. (2024). Chlorine dosage management in drinking water systems: Comparing Bayesian optimization to evolutionary algorithms. *Journal of Hydroinformatics*, *26*(11), 2720–2738.

Mottahedin, A., Minaei, A., Sitzenfrei, R., & Creaco, E. (2023). Multistep approach for the rehabilitation of a deteriorated water distribution network operating intermittently. *Journal of Water Resources Planning and Management*, *149*(12), 04023067.

Muhammed, K., Farmani, R., Behzadian, K., Diao, K., & Butler, D. (2017). Optimal rehabilitation of water distribution systems using a cluster-based technique. *Journal of Water Resources Planning and Management*, *143*(7), 04017022.

Pudasaini, B., & Shahandashti, M. (2020). Topological surrogates for computationally efficient seismic robustness optimization of water pipe networks. *Computer-Aided Civil and Infrastructure Engineering*, *35*(10), 1101–1114.

Riyahi, M. M., Bakhshipour, A. E., & Haghighi, A. (2023). Probabilistic warm solutions-based multi-objective optimization algorithm, application in optimal design of water distribution networks. *Sustainable Cities and Society*, *91*, 104424.

Santonastaso, G. F., Di Nardo, A., Creaco, E., Musmarra, D., & Greco, R. (2021). Comparison of topological, empirical and optimization-based approaches for locating quality detection points in water distribution networks. *Environmental Science and Pollution Research*, *28*(26), 33844–33853.

Shahhosseini, A., Najarchi, M., Mahdi Najafizadeh, M., & Hezaveh, M. M. (2023). Performance optimization of water distribution network using meta-heuristic algorithms from the perspective of leakage control and resiliency factor (case study: Tehran water distribution network, Iran). *Results in Engineering*, *20*, 101603.

Siddique, N., & Adeli, H. (2014). Spiral dynamics algorithm. *International Journal on Artificial Intelligence Tools*, *23*(6), 1430001.

Siddique, N., & Adeli, H. (2015). Harmony search algorithm and its variants. *International Journal of Pattern Recognition and Artificial Intelligence*, *29*(8), 1539001.

Simpson, A. R., Dandy, G. C., & Murphy, L. J. (1994). Genetic algorithms compared to other techniques for pipe optimization. *Journal of Water Resources Planning and Management*, *120*(4), 423–443.

Sitzenfrei, R. (2021). Using complex network analysis for water quality assessment in large water distribution systems. *Water Research*, *201*, 117359.

Sitzenfrei, R., Wang, Q., Kapelan, Z., & Savić, D. (2020). Using complex network analysis for optimization of water distribution networks. *Water Resources Research*, *56*(8), e2020WR027929.

Todini, E. (2000). Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water*, *2*(2), 115–122.

Ulusoy, A.-J., Mahmoud, H. A., Pecci, F., Keedwell, E. C., & Stoianov, I. (2022). Bi-objective design-for-control for improving the pressure management and resilience of water distribution networks. *Water Research*, *222*, 118914.

Vertommen, I., Mitrović, D., van Laarhoven, K., Piens, P., & Torbeyns, M. (2022). Optimization of water network topology and pipe sizing to aid water utilities in deciding on a design philosophy: A real case study in Belgium. *Water*, *14*(23), 3973.

Wagner, J. M., Shamir, U., & Marks, D. H. (1988). Water distribution reliability: Simulation methods. *Journal of Water Resources Planning and Management*, *114*(3), 276–294.

Walski, T. (2014). How does water distribution design really work? *World Environmental and Water Resources Congress 2014*, Portland, OR.

Wang, P., Zecchin, A. C., & Maier, H. R. (2023). Improved selection strategy for multi-objective evolutionary algorithms with application to water distribution optimization problems. *Computer-Aided Civil and Infrastructure Engineering*, *38*(10), 1290–1306.

Wu, Z. Y., & Simpson, A. R. (2002). A self-adaptive boundary search genetic algorithm and its application to water distribution systems. *Journal of Hydraulic Research*, *40*(2), 191–203.

Yi, J.-H., Xing, L.-N., Wang, G.-G., Dong, J., Vasilakos, A. V., Alavi, A. H., & Wang, L. (2020). Behavior of crossover operators in NSGA-III for large-scale optimization problems. *Information Sciences*, *509*, 470–487.

Zecchin, A. C., Simpson, A. R., Maier, H. R., & Nixon, J. B. (2005). Parametric study for an ant algorithm applied to water distribution system optimization. *IEEE Transactions on Evolutionary Computation*, *9*(2), 175–191.

Zhao, Z., Liu, B., Zhang, C., & Liu, H. (2019). An improved adaptive NSGA-II with multi-population algorithm. *Applied Intelligence*, *49*(2), 569–580.

Zheng, F., Zecchin, A. C., Maier, H. R., & Simpson, A. R. (2016). Comparison of the searching behavior of NSGA-II, SAMODE, and Borg MOEAs applied to water distribution system design problems. *Journal of Water Resources Planning and Management*, *142*(7), 04016017.

Zheng, F., Zecchin, A. C., Newman, J. P., Maier, H. R., & Dandy, G. C. (2017). An adaptive convergence-trajectory controlled ant colony optimization algorithm with application to water distribution system design problems. *IEEE Transactions on Evolutionary Computation*, *21*(5), 773–791.

How to cite this article: Minaei, A., Zecchin, A. C., Hajibabaei, M., Mitrovic, D., Laarhoven, K., Vertommen, I., Alexander, B., Blokker, M., Savic, D., Creaco, E., & Sitzenfrei, R. (2025). Efficient resizing and topological optimization of real-world water distribution networks in a multi-criteria decision-making framework. *Computer-Aided Civil and Infrastructure Engineering*, 1–19. https://doi.org/10.1111/mice.70036