

# Multi-objective optimization of water distribution networks via NSGA-II and Pseudo-Weights

Samar Ben Ammar  
Département du Génie Rural, Eaux et Forêts,  
National Institute of Agronomy of Tunisia  
Tunis, Tunisia  
ba.samar94@yahoo.com

Mario G.C.A. Cimino  
Department of Information Engineering  
University of Pisa  
Pisa, Italy  
mario.cimino@unipi.it

Pierfrancesco Foglia  
Department of Information Engineering  
University of Pisa  
Pisa, Italy  
pierfrancesco.foglia@unipi.it

Federico A. Galatolo  
Department of Information Engineering  
University of Pisa  
Pisa, Italy  
federico.galatolo@ing.unipi.it

Issam Nouri  
Département du Génie Rural, Eaux et Forêts,  
National Institute of Agronomy of Tunisia  
Tunis, Tunisia  
inouri@yahoo.fr

**Abstract**—Managing water distribution networks via pumps scheduling programs is a multi-objective optimization problem with dynamic and various site-specific challenges. Metaheuristics-based approaches, with respect to mathematical solvers, offer data-driven strategies for manageable and adaptive control. Some evolutionary approaches are suitable for multi-criteria decision making and decentralized coordination on programmable logic controllers. This paper focuses on the development of a testbed and an early assessment of an approach based on NSGA-II and Pseudo-Weights. The experimental studies are based on a physically developed case study, and on a scalable case study with realistic water demand and source patterns. The testbed has been publicly released.

**Keywords**—water distribution network, genetic algorithm, multi-objective optimization, multi-criteria decision making

## I. INTRODUCTION AND BACKGROUND

The optimal management of Water Distribution Networks (WDNs) often faces various site-specific challenges: network topology and complexity, number of electrical pumps, heterogeneous water sources, availability of service reservoirs, local electricity tariffs for peak/off-peak periods, pumps maintenance cost, local hourly water demand, seasonal weather conditions, medium-term socio-economic context such as growth of population and urbanization, and so on [1].

The task of determining the operating time of each pump, in order to meet the predicted water demand, and according to other management criteria, is known as Pump Scheduling (PS) [2]. Two common criteria are to reduce the pumping cost and maintenance, and to guarantee a cyclic sources exploitation. Usually, the scheduling horizon is one day, after which the same or a revised schedule is applied. To manage the PS task complexity, a WDN is partitioned into subnetworks [2]. The PS program provides a real-time control of sources, pumps, and tanks of the local distribution system, according to a decision-making process supported by information technology [3][4].

In the literature, different approaches have been proposed for optimal PS, such as linear/nonlinear programming, dynamic programming, metaheuristic algorithms. It is well-known that

metaheuristics are computationally slower than mathematical solvers, but are relatively easier to implement, and are capable of handling both real and discrete types of decision variables in multimodal search spaces [5]. For this reason, various approaches have been experimented to the PS, such as Simulated Annealing, Hill Climb, Ant Colony Optimization, Shuffled Frog Leaping, and Genetic Algorithms (GAs). Among them, GAs have been extensively considered [5][7][8][9][10].

A significant design aspect is the scheme of representation of PS, affecting the size and complexity of the decision space: it can be an *implicit scheme* (e.g., expressed in terms of tank level-controlled triggers) or an *explicit scheme* (e.g., expressed in terms of time-controlled triggers).

Although some researches model the PS as a multi-objective optimization problem [5][7], most of them find one particular Pareto-optimal solution at a time. Thus, for multiple solutions, the GA has to be applied multiple times. In order to support a more flexible and adaptive decision making, this research focuses on the ability of some GAs to find multiple Pareto-optimal solutions in one single simulation run [5]. In essence, since GAs work with a population of solutions, it can be extended to maintain diverse sets of solutions. In Nondominated Sorting GA (NSGA)[5], there are representative solutions (called “non-dominated”) that are superior or equal to the others (called “dominated”), with respect to all objectives. In NSGA, the selection operator works differently than in conventional GA: the fitness assignment, based on dummy values, allows multiple optimal points to co-exist in the population.

The choice of one solution over the others requires a decision-making process. One typical decision-making is to scalarize the vector of objectives into one objective, via a weight vector. However, the obtained solution largely depends on the weights, and this sensitivity makes the determination of such parameters expensive. A method to reduce the parametric sensitivity of decision-making is to move from the space of solutions to another space, simpler and more manageable from the human decisor. In this paper, the pseudo-weight method is used, in which a candidate solution is represented by a calculated weight for each objective. In the normalized pseudo-weight

space, the best solution can be identified by assigning a target pseudo-weight vector expressing the relative importance among the objectives.

The paper is structured as follows: Section II formally describes the methodology. Section III illustrates the case studies. Section IV draws the conclusions. The main contributions of this paper are the following: (i) to propose a multi-objective optimization for water distribution networks based on NSGA-II and Pseudo-Weights; (ii) to show how different decision-making strategies can be easily compared; (iii) to develop and publicly release a testbed; (iv) to outline future developments of this research.

## II. METHODOLOGY

Fig. 1 illustrates a reference scenario of water distribution. Let us consider  $N$  tanks, with related pumps  $p_i$ ,  $i = 1, \dots, N$ . Each pump with a binary (on/off) command to supply the main tank with a constant flow  $f_i$ . A pump behavior is specified via an hourly scheduling program in a daily horizon, imparted via a relay,  $p_i(t) \in \{0,1\}$ ,  $t = 0, \dots, 23$ . Each tank is supplied by an external source  $s_i(t)$ . The total daily flow of the external sources is assumed to be sufficient to cover the total daily water demand at the main tank.

Let  $v_i(t)$  be the current volume of the  $i$ -th tank, measured via a sensor. Let  $v_0(t)$  be the main tank volume. The main tank is equipped with  $M$  independent output valves having constant flow, to supply the water demand  $d_j(t)$ ,  $j = 1, \dots, M$ . On the figure top, a Programmable Logic Controller (PLC) solves the PS problem, by taking the expected daily water demand  $\{d_j(t)\}$  and the tank volumes  $\{v_i(t)\}$  as inputs, and providing the PS solution as an output.

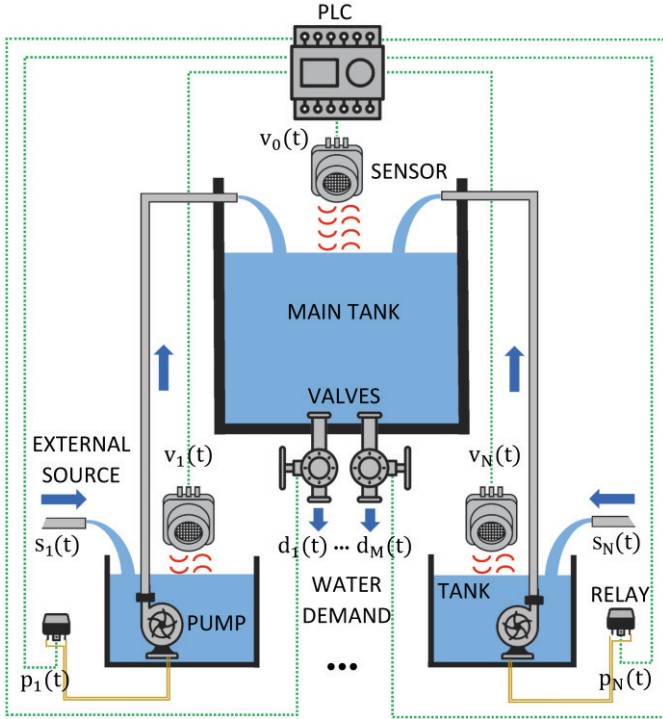


Fig. 1. Illustration of a reference scenario of water distribution

With the *explicit* scheme, the solution is specified as a sequence of 24 binary commands per pump,  $p_i(t)$ . With the *implicit* scheme, the solution is a pair of thresholds volumes ( $v_i^{on}, v_i^{off}$ ): initially the pump is turned off; when the level is under  $v_i^{off}$  / upper  $v_i^{on}$ , then the pump is switched off / on, respectively.

The optimal PS is represented as a multi-objective optimization problem with three objective functions to minimize: (1) daily pumping energy cost, (2) daily number of pump switches, (3) daily volume surplus/deficit of the main tank. Table I summarizes the motivation/description of each objective, whereas Equations (1-3) define them formally.

TABLE I. DEFINITION OF EACH OBJECTIVE TO OPTIMIZE

Objective	Motivation/description
Total daily pumping cost ( $F_{PC}$ )	The pumping energy cost is directly proportional to the pump water flow (l/hour).
Total daily number of pump switches ( $F_{SW}$ )	The number of pump switches is a measure of mechanical wear.
Daily volume variation of the main tank ( $F_{\Delta V}$ )	It is the absolute difference between the initial and final main tank volume. It measures the acyclicity of the process.

$$F_{PC} = \sum_{i=1}^N \sum_{t=0}^{23} f_i \cdot p_i(t) \quad (1)$$

$$F_{SW} = \sum_{i=1}^N \sum_{t=0}^{23} |p_i(t+1) - p_i(t)| \quad (2)$$

$$F_{\Delta V} = \sum_{t=0}^{23} \left( \sum_{i=1}^N f_i \cdot p_i(t) - \sum_{j=1}^M d_j(t) \right) \quad (3)$$

Three types of constraints are also imposed on the problem: (4) daily water availability, (5) lower tank volume, (6) upper tank volume. Table II summarizes the motivation/description of each constraint, whereas Equations (4-6) define them formally.

TABLE II. DESCRIPTION OF EACH CONSTRAINT

Constraint	Motivation/description
Daily water availability	The total daily water supplied by external sources is greater than or equal to the total daily water demand.
Lower tank volume	The tank volume cannot be lower than a predefined threshold, under which the pump is switched off.
Upper tank volume	The tank volume cannot be higher than a predefined threshold, over which the water surplus is lost (water overflow).

To minimize the objective functions (1-3) under the constraints (4-6), the constrained NSGA-II [11] is used. Specifically, NSGA-II is an evolutionary algorithm for multi-objective problems, based on the notion of *dominancy* between solutions. Solution  $x$  dominates  $y$  if none of objectives in  $x$  is worse than in  $y$  and at least one objective in  $x$  is better than in  $y$  [11]. At each generation, the population of PS solutions is divided into the feasible and infeasible sub-populations,

according to the total constraints' violation, and then the feasible population is ranked based on the notion of dominance.

$$\sum_{i=1}^N \sum_{t=0}^{23} s_i(t) \geq \sum_{j=1}^M \sum_{t=0}^{23} d_j(t) \quad (4)$$

$$p_i(t) = 0, \forall i \neq 0 \wedge v_i(t) \leq \underline{v}_i \quad (5)$$

$$v_i(t+1) = \begin{cases} \min\{\bar{v}_i; v_i(t) - f_i \cdot p_i(t) + s_i(t)\}, & i \neq 0 \\ \min\{\bar{v}_i; v_i(t) - \sum_{j=0}^M d_j(t) + \sum_{i=0}^N f_i \cdot p_i(t)\}, & i = 0 \end{cases} \quad (6)$$

The Pareto fronts of  $l$  levels are formed, the infeasible chromosomes are stored in front  $l+1$ . The crowding distance between chromosomes is also considered to preserve diversity. The parent selection is based on a combined criterion: total constraint violation, non-domination rank and crowding distance. After obtaining a set of non-dominated solutions, a single solution has to be chosen for implementation.

The decision-making approach used is based on pseudo-weights [12]. For each  $o$ -th objective function, a pseudo-weight of a solution is calculated as follows:

$$w_o = \frac{(\bar{F}_o - F_o)/(\bar{F}_o - \underline{F}_o)}{\sum_{h=1}^3 (\bar{F}_h - F_h)/(\bar{F}_h - \underline{F}_h)} \quad (7)$$

where  $\bar{F}_o$  and  $\underline{F}_o$  are the maximum and minimum values of the  $o$ -th objective function, respectively. In (7) the sum of all elements  $w_o$  of the vector is forced to one. Thus, the pseudo-weight vector represents a relative importance of each objective function for the solution. This means that solutions that are closer to the minimum objective value have a higher weight value for that objective. To accomplish different strategies, the decisor establishes different target pseudo-weight vectors. In the reference scenario, the following four different strategies and related targets have been considered: (a) balanced: (1/3, 1/3, 1/3); cost-saving: (.9, .05, .05), switch-saving (.05, .9, .05), volumes-cyclicity (.05, .05, .9). For each strategy, the best solution is found in the space of pseudo-weights, as the solution with the closest target pseudo-weight vector.

### III. CASE STUDIES

In order to assess the effectiveness of the proposed approach, a simulation test bed has been developed for the modeling and execution of pilot case studies. The testbed is based on a recent multi-objective optimization framework supporting a variety of approaches [13]. A hydraulic simulation model has been developed in order to support the features illustrated in Fig. 1. The overall testbed has been publicly released as a GitHub repository [14]. To extend the features of the hydraulic simulator, as a future task, the testbed could be integrated with the WNTR framework, which contains the *EpanetSimulator* and the *WNTRSimulator* [15].

To understand the potential of the proposed approach, this Section shows an experiment on an interpretable pilot case study of the reference scenario, as well as another experiment on a scalable case study.

#### A. Interpretable case study

This first case study is based on a comparative analysis of implicit and explicit schemes across four different strategies. The case study is intentionally small to allow an easy interpretation of the scheduling results. It uses  $N=2$  pumps and  $M=2$  valves for demand. It has also been physically developed in a hydraulic lab as a prototype, using a Raspberry PI© programmable controller, for the purpose of integration and life-cycle testing.

Specifically, the following hyperparameters have been set in the optimization framework: population size: 100, number of generations: 1000, crossover rate 0.9, selection: binary tournament, and mutation rate = 1 / chromosome length (48 or 4 genes, in case of explicit or implicit scheme, respectively). The following parameters have been set in the hydraulic simulator, according to the physical prototype (volumes are in  $l$ , flows in  $l/h$ ):  $\bar{v}_0 = 30$ ,  $\bar{v}_1 = \bar{v}_2 = 14$ ,  $\underline{v}_0 = 10$ ,  $\underline{v}_1 = \underline{v}_2 = 4$ ,  $v_0(0) = 30$ ,  $v_1(0) = v_2(0) = 4$ ,  $f_1 = f_2 = 3$ . Fig. 2 shows the water demand. For the sake of simplicity/interpretability,  $d_1=s_1$  and  $d_2=s_2$ .

Fig. 3 shows the optimal PS and simulation results with the explicit scheme and strategies (a) balanced, (b) cost-saving, (c) switch-saving, (d) volumes-cyclicity. Fig. 4 shows the results generated with the implicit scheme. As expected, Fig. 3a and particularly Fig. 3d show a better volumes-cyclicity with respect to Figures 3b-c, which is consistent with the corresponding strategies. The volumes-cyclicity is better achieved with the implicit scheme: in Figures 4a-d,  $F_{AV}$  is less than or equal with respect to the corresponding Figures 3a-d. In contrast, the cost-saving is better with the explicit scheme: in Figures 3a-d,  $F_{PC}$  is less than or equal with respect to the corresponding Figures 4a-d. The switch-saving is generally better with the implicit scheme, except if cost-saving is the preferred strategy: note in Fig. 4b that pump  $p_1$  is characterized by frequent switches. Such switches are caused by the very close switching thresholds,  $v_1^{off} \approx v_1^{on}$ , which results from the cost-saving strategy. In a balanced strategy, the implicit scheme is more effective.

Overall, with respect to the explicit scheme, the implicit scheme always achieves minimum (optimum) values of  $F_{AV}$ , as well as of  $F_{SW}$  except for the cost-saving strategy. On the other hand, the  $F_{PC}$  value achieved via the implicit scheme is equal or higher to the corresponding value achieved via the explicit one.

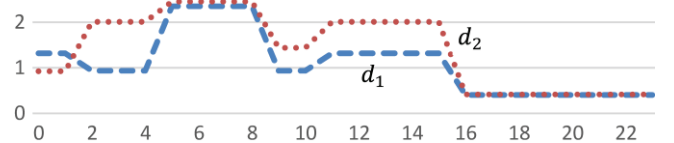


Fig. 2. Water demand, equal to the external sources  $s_1, s_2$ .

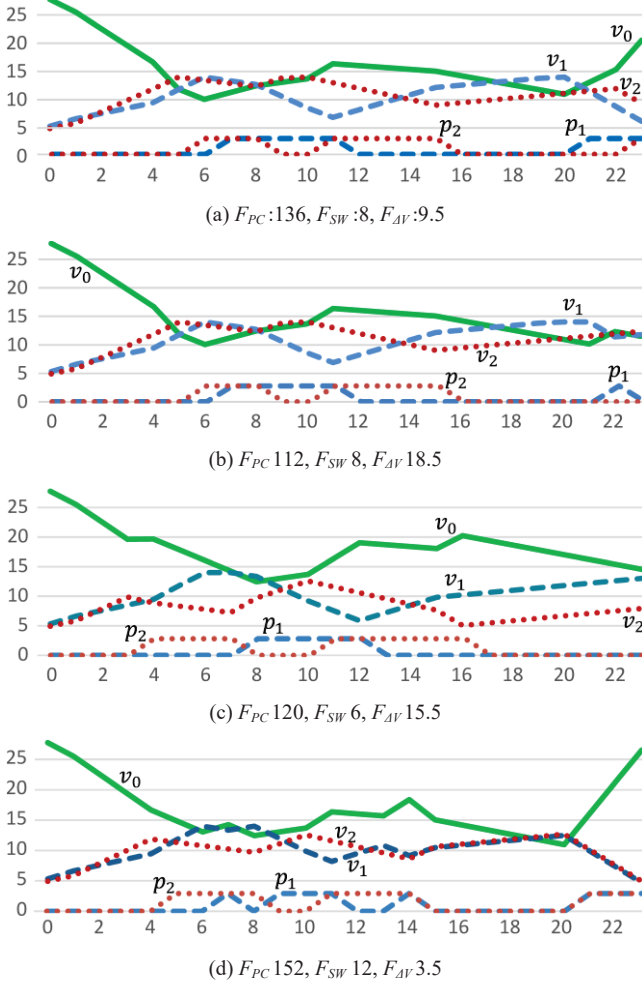


Fig. 3. Optimal PS and simulation results, with explicit scheme and strategies (a) balanced, (b) cost-saving, (c) switch-saving, (d) volumes-cyclicity.

Hence, the implicit scheme tends to find better solutions from the point of view of mechanic wear and process cyclicity, at a cost that in some cases is higher. In this pilot case study, the implicit pump scheduling scheme can be appreciated for its effectiveness with the balanced, switch-saving, and volumes-cyclicity strategies, whereas the explicit scheduling can be better exploited to manage the cost-saving strategies.

### B. Scalable case study

In order to evaluate the effectiveness of the proposed approach when varying the problem size, in this case study the considered problem complexity differs by one order of magnitude with respect to the previous one. Specifically,  $N=20$  pumps/sources and  $M=20$  valves for demand have been used. Half of the water production (i.e., 10 sources) is based on natural sources, whereas the other half is based on water desalination technology. The natural sources are characterized by a daily constant supply. The average flow is  $s_i=0.8$ , with 10% variation among sources. The artificial water production is based on a combined water desalination and electricity generation system, a sustainable technology in which a humidification-dehumidification (HDH) process is integrated with photovoltaic-thermal modules [16].

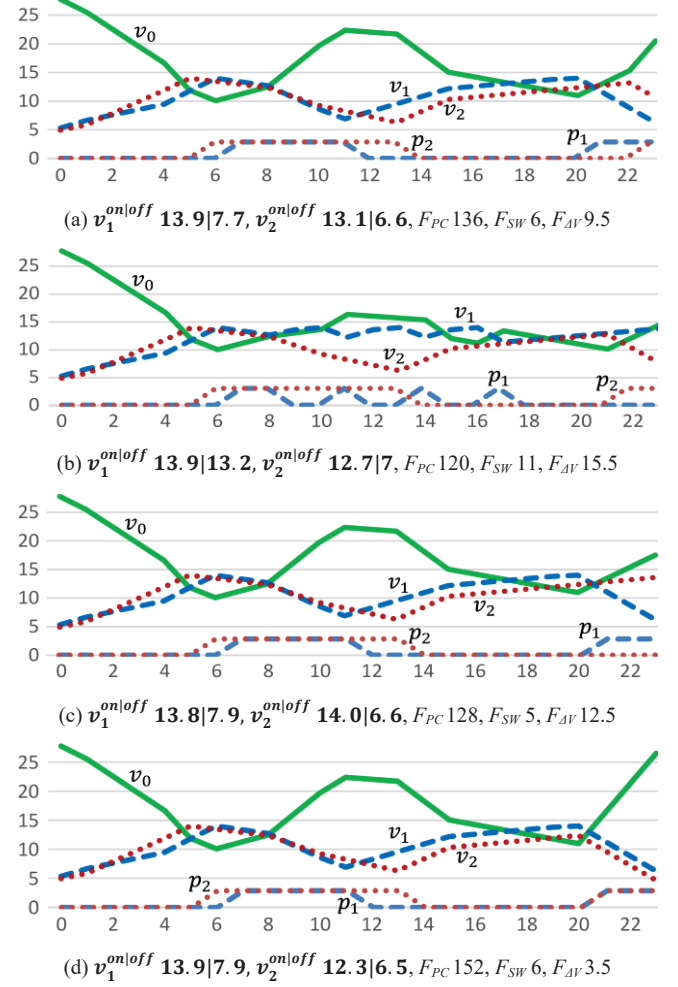


Fig. 4. Optimal PS and simulation results, with implicit scheme and strategies (a) balanced, (b) cost-saving, (c) switch-saving, (d) volumes-cyclicity.

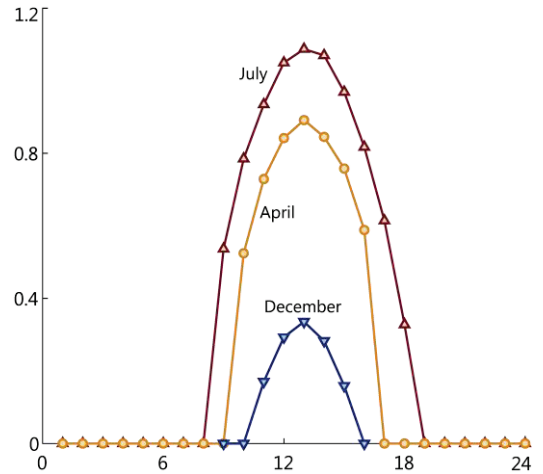


Fig. 5. Hourly profiles of water flow for HDH-based sources  $s_i$  (adapted from [16])

Fig. 5 shows the hourly profiles of water flow for some representative months of the year. In the scenario, the production in July has been considered, with 10% variation among sources.





*Resour Manage* 31, 1283–1304. <https://doi.org/10.1007/s11269-017-1577>.

- [6] Abkenar, S.M.S., Stanley, S.D., Miller, C.J., Chase, D.V. and McElmurry, S.P., 2015. Evaluation of genetic algorithms using discrete and continuous methods for pump optimization of water distribution systems. *Sustainable Computing: Informatics and Systems*, 8, pp.18-23.
- [7] De Wrachien, D., Mambretti, S. and Orsi, E., 2017. Optimization of pump operations in a complex water supply network: New Genetic Algorithm frameworks. *International Journal of Sustainable Development and Planning*, 12(1), pp.79-88.
- [8] Behandisha, M. and Wub, Z.Y., 2014. Concurrent pump scheduling and storage level optimization using meta-models and evolutionary algorithms. *Procedia Engineering*, 70, pp.103-112.
- [9] Nouri, I., 2014. Multi-objective tool to optimize the water resources management using genetic algorithm and the Pareto optimality concept. *Water resources management*, 28(10), pp.2885-2901.
- [10] Abkenar, S.M.S., Chase, D.V., Stanley, S.D. and McElmurry, S.P., 2013, June. Optimizing pumping system for sustainable water distribution network by using Genetic Algorithm. In *2013 International Green Computing Conference Proceedings* (pp. 1-6). IEEE.
- [11] Deb, K., Pratap, A., Agarwal S., Meyarivan, T. "A fast and elitist multiobjective genetic algorithm: NSGA-II," in *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, April 2002, doi: 10.1109/4235.996017.
- [12] Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons, Inc.: New York, NY, USA, 2001; ISBN 047187339X
- [13] Blank J., and Deb K., "Pymoo: Multi-Objective Optimization in Python," in *IEEE Access*, vol. 8, pp. 89497-89509, 2020, doi: 10.1109/ACCESS.2020.2990567
- [14] Galatolo, F. A. 2021. Optimal Water Distribution Network, GitHub repository: <https://github.com/galatolofederico/optimal-wdn>
- [15] Klise, K.A., Murray, R. and Haxton, T., 2018, July. An Overview of the Water Network Tool for Resilience (WNTR). In *WDSA/CCWI Joint Conference Proceedings* (Vol. 1).
- [16] Gabrielli, P., Gazzani, M., Novati, N., Sutter, L., Simonetti, R., Molinaroli, L., Manzolini, G. and Mazzotti, M., 2019. "Combined water desalination and electricity generation through a humidification-dehumidification process integrated with photovoltaic-thermal modules: Design, performance analysis and techno-economic assessment". *Energy Conversion and Management*: X, 1, p.100004.
- [17] Rasekh, A. and Brumbelow, K., 2014. "Drinking water distribution systems contamination management to reduce public health impacts and system service interruptions". *Environmental Modelling & Software*, 51, pp.12-25.