# Sensor Placement Optimization in Dynamic Water Distribution Network by Considering the Importance of Nodes and Uncertain Contamination Entrance

**Siroos harif**
University of Sistan and Baluchestan

**Mohsen Dehghani Darmian** ( mohsen.dehghanidarmian@gmail.com )
Water Research Institute https://orcid.org/0000-0002-8955-4485

**Gholamreza Azizyan**
University of Sistan and Baluchestan

**Mohammad Givehchi**
University of Sistan and Baluchestan

**Research Article**

**Keywords:** Contaminant detection, Water distribution systems, Water quality, Sensor placement strategy, Simulation-Optimization model

**Posted Date:** June 23rd, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1640174/v1

# Sensor Placement Optimization in Dynamic Water Distribution Network by Considering the Importance of Nodes and Uncertain Contamination Entrance

**Siroos harif[a], Gholamreza Azizyan[b], Mohammad Givehchi[c], Mohsen Dehghani Darmian[d*]**

**a- Civil Engineering Department, University of Sistan and Baluchestan, Zahedan, Iran, Email: siroos.harif@pgs.usb.ac.ir**

**b- Civil Engineering Department, University of Sistan and Baluchestan, Zahedan, Iran, Email: g.azizyan@eng.usb.ac.ir**

**c- Civil Engineering Department, University of Sistan and Baluchestan, Zahedan, Iran, Email: m_givehchi@yahoo.com**

**d- Department of Hydraulic Engineering and Hydro-Environment, Water Research Institute, Tehran, Iran, Email: Mohsen.dehghanidarmian@gmail.com**

## Abstract

Urban infrastructure is heavily reliant on water distribution systems. Contaminants entering the water distribution systems (WDS) are one of the most dangerous events that may occur either deliberately or by accident. Water can become polluted with chemicals or biological agents as a result of water flow. This can cause sickness or even death among people who drink the water. Using an Early Warning Detection System (EWDS) is one of the most effective ways to minimize negative consequences on public health. EWDS are sensors that can reduce the damage due to detecting the contamination. The main challenge is to arrange the sensors in the network in the most efficient way. In this study, the Non-Dominated Sorted Genetic Algorithm-II (NSGA-II), a multi-objective optimization approach, is developed to determine

* Corresponding author. Email: Mohsen.dehghanidarmian@gmail.com

the optimal placement of quality sensors in water distribution networks by balancing four conflicting objectives. 1. Sensor detection likelihood, 2. Sensor expected detection time, 3. Sensor detection redundancy, and 4. The affected population before detection. A contamination matrix with 1000 contaminants event was generated, which represented the total possible combinations of pollutants, then the optimal Pareto fronts are obtained for each two conflicting objectives. The importance coefficients are proposed and applied to minimize the pollution detection time and the number of infected populations based on the contamination risks and the node demands, respectively. Moreover, the sensitivity analysis of the results obtained from different objective functions related to the number of sensors installed in the network was conducted.

**Keywords:** Contaminant detection, Water distribution systems, Water quality, Sensor placement strategy, Simulation-Optimization model

## 1. Introduction

A majority of the Earth's surface is covered by water, but less than 3% of this is freshwater. Approximately 0.01 % of freshwater is available for human consumption; the remainder is bound in glaciers and ice (Ahmed et al., 2015). Several people argue that safe water is an essential human right as well as necessary for health (WHO., 2010., Keramati et al., 2019, Massoudinejad et al., 2018, Yousefi et al., 2018). However, many people around the world lack access to this fundamental need (Nabeela et al., 2014). The United Nations reports that more than 780 million people in developing countries do not have access to safe and adequate drinking water (UNISEF and WHO., 2012), while 2.3 billion people are suffering from water-related diseases (UNESCO., 2003). There are an estimated 10 deaths a day caused by diarrhea in Europe, according to the WHO (WHO., 2015). 80% of population of Pakistan has not access to safe drinking water and forced to use unsafe drinking water (Daud et al., 2017). Water is

essential for all of the things we need in everyday life, including drinking, hygiene, and food preparation, and is used in almost every building, including homes, schools, hospitals, commercial and service buildings, restaurants, and plants. Drinking water distribution networks (WDN) are an essential part of a civil infrastructure that distributes clean water (safe water) from the main supply to billions of people (customers) and ensures the quality of life (Wang & Zhou., 2017). A large number of people use water from the WDN every day and rely on the safety and the quality of water in their lives or work (de Winter et al., 2019). There are several components to this system, such as a network of pipes, intersections of pipes as nodes, as well as reservoirs, storage tanks, pumps, and valves (Adedoja et al., 2018b). There are many types of the water distribution networks, and they can be very large consist of hundreds of kilometers of pipes and many delivery points (de Winter et al., 2019). WDN problems can have an enormous impact on society if they arise. Consequently, water quality of the WDN is directly related to customer health and well-being, making it a critical infrastructure for society (He et al., 2018).

A city's WDNs can be an easy target for accidental or intentional contamination due to its extensive geographic coverage, multiple points of access, backflow, infrastructure aging, and designed sabotage (Yang& Boccelli., 2016, Costa et al., 2013, Perelman et al., 2013, Hart& Murray., 2010, Huang & McBean., 2009). One of the biggest threats to a population is chemical contamination within a water network, which may cause widespread diseases and death (de Winter et al., 2019). The pollutants can enter the WDN at any point and move through the entire system with water flow over time, affecting large parts of the population. Public health and the environment may be severely impacted by pollutions in WDNs, in addition to significant economic losses and negative social impacts. In 1993, more than 400,000 consumers in Milwaukee (USA) were affected by an intrusion that left them hospitalized (Corso et al., 2003). During an event in Walkerton, Ontario in 2000, more than 2300 people

were affected (Hrudey et al., 2003). In 2007, the water supply in Ruyang, China poisoned 71 people as well (ChinaNews). Similar incidents have occurred in Colorado (Falco & Williams., 2009), West Virginia (Rosen et al., 2014), Ohio (Grayman et al., 2016), Hong Kong (ChinaNews., 2016), and the United States (Bush., 2002).

Studies have confirmed that attacks on water distribution networks are real, as they have occurred previously and might happen again (Adedoja et al., 2018b). Thus, it is necessary to identify and warn about contamination events within the WDS and protect water quality against accidental and intentional contamination events. The first step toward achieving this goal is improving the physical security of the system (e.g. fencing, guarding, additional alarms, surveillance instruments, locks, etc.) (Preis & Ostfeld., 2008). The second step is using monitoring sensors to monitor water quality, a contamination warning system, (CWS) is also known as an early warning system (EWS) (Grayman et al., 2016). It is possible to gain the maximum level of safety by monitoring all nodes in the system systematically (Adedoja et al., 2018b, Yazdi., 2018). Sensors cannot be installed at every point of the network due to the high cost, budget constraints, and maintenance issues (Adedoja et al., 2018a,b). We need to optimally use only a few sensors at the right locations in the network to achieve effective monitoring. Several algorithms and optimization models have been proposed and utilized to determine the optimal locations for a set of sensors. Sensor optimization investigation in the network started in 1990. Mixed-integer linear programming (MILP) was developed and employed to locate the quality sensors in the network optimally (Lee & Deininger., 1992, Kumar et al., 1997).

Furthermore, in recent years, the optimal location of sensors in the networks has been extensively investigated. The issue based on the number of objective functions is classified as single objective optimization problems (Kessler et al., 1998, Woo et al., 2001, Al-Zahrani & Moeid., 2001, Ostfeld & Salomons., 2004, Ostfeld & Salomons., 2005, Berry et al., 2006, Berry

et al., 2009, Propato., 2006, Propato & Piller., 2008, Shastri & Diwekar., 2006, Comboul & Ghanem., 2013, Schwartz et al., 2014, Cheifetz et al., 2015, Ohar et al., 2015) and multiple objective optimization problems (de Winter et al., 2019, Preis & Ostfeld., 2008, McKenna et al., 2006, Ostfeld & Salomons., 2008, Dorini et al., 2008, Eliades & Polycarpou., 2008, Gueli., 2008, Huang et al., 2008, Wu & Walski., 2006, Ostfeld et al., 2008, Bazargan-Lari., 2014, Afshar & Miri Khombi., 2015, Rathi & Gupta., 2016, Hu et al., 2017, Tinelli et al., 2017, Nazempour et al., 2018, Naserizade et al., 2018, He et al., 2018, Khorshidi et al., 2019a,b).

According to previous researches, sensors have been placed based on a certain number of pollution entrance scenarios, and a sensor's location has been determined by these scenarios. Furthermore, nodes have been given equal weight in water distribution networks as well. In this investigation, the genetic algorithm is used to generate a contamination matrix that represents all possible combinations of pollutants entrance. Then, a multi-objective optimization approach is applied to determine the optimal location of the sensors in the WDN in order to maximize sensor detection likelihood and detection redundancy, as well as minimize sensor expected detection time, and percentage of affected nodes. By using the importance coefficient, the detection time is minimized based on the amount of damage caused by the pollution entering the WDN. As a result, pollutants that infect more network nodes and cause more damage to the WDN should have a shorter detection time. Furthermore, the importance coefficient based on the node demands is used to calculate the number of infected nodes (the fourth objective function). This coefficient indicates that nodes with higher demands are more important than those with lower demands and that sensors should be placed near these nodes to detect contamination before entering.

## 2. Material and methods

In this research for simulating the effect of a contamination event distributed through the water network, the EPANET 2.0 software was used. EPANET is a computer program

performing an extended period simulation of hydraulic and water quality behavior within a drinking water distribution system. In the quality analysis, the concentration is calculated at each node and time step (Rossman., 2000). The quality analysis in EPANET are advection, diffusion, and reactions that expressed as follows:

$$\frac{\partial c_i}{\partial t} = -u_i \frac{\partial c_i}{\partial x} \pm R(c_i) \tag{1}$$

In this equation, $c_i$ is the concentration of the contaminant in the ith pipe at time t and location x (mg/L). $u_i$ is the velocity of water in the i$^{th}$ pipe (m/s) and $R(c_i)$ is the term for the reaction rate which includes the wall and bulk reactions.

The EPANET includes Programmers Toolkit that is a dynamic-link library that allows users to customize the EPANET computing engine according to their requirements (Rossman., 2000). The output of this can be used as the input to MATLAB. Then this software is used for optimizing the location of sensors in distribution networks. The methods of Evolutionary computation such as genetic algorithms are very useful and powerful tools in optimization problems and solving searches with respect to their unbiased nature allowing them to perform very well in situations with little domain knowledge (Gong et al., 2014, Gong et al., 2015). The NSGA-II, an improved version of NSGA (Srinivas & Deb., 1994), is one of the most efficient algorithms for many-objective evolutionary algorithms that has been widely applied in the optimal design of water distribution networks and has exhibited high performance.

The NSGA-II has three specific characteristics, fast crowded distance estimation, fast non-dominated sorting, and simple crowded comparison (Yusoff et al., 2011). The NSGA-II algorithm uses a nondominated sorting approach for fitness assignments based on the concept of Pareto domination and optimality. for many-objective minimization problems the concept of Pareto domination and optimality can be expressed as follows (Bekele & Nicklow., 2007):

Minimize: $$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)) \tag{2}$$

Subject to: $$g(x) = (g_1(x), g_2(x), \ldots, g_n(x)) \leq 0 \tag{3}$$

In this equation f(x) is the vector-valued function, x is the decision vector, and g(x) is a vector of constraints.

If we have two decision vectors like a and b, it is said that a dominates b: if 1) the solution of a is no worse than the solution of b in all objectives and 2) the solution of a is strictly better than the solution of b in at least one objective (Bekele & Nicklow., 2007, Kannan et al., 2008).

$$\text{If} \quad \forall i \in \{1,2,\ldots,n\}: f_i(a) \leq f_i(b) \tag{4}$$

$$\text{And} \quad \exists i \in \{1,2,\ldots,n\}: f_i(a) < f_i(b) \tag{5}$$

All individuals in front number 1 are not dominated by any other individuals. Individuals dominated only by the individuals in front number 1 are assigned front number 2, and so on (Kannan et al., 2008). in many-objective optimization there are two goals: finding a set of solutions from all the sets of solutions to yield optimal values with respect to all the objective functions and to guarantee that the set of solutions is as varied as possible (Preis & Ostfeld., 2008). A brief description of NSGA-II implemented in this study is described in Fig. 1 and more details can be found in (Kalyanmoy., 2001, Deb et al., 2002).

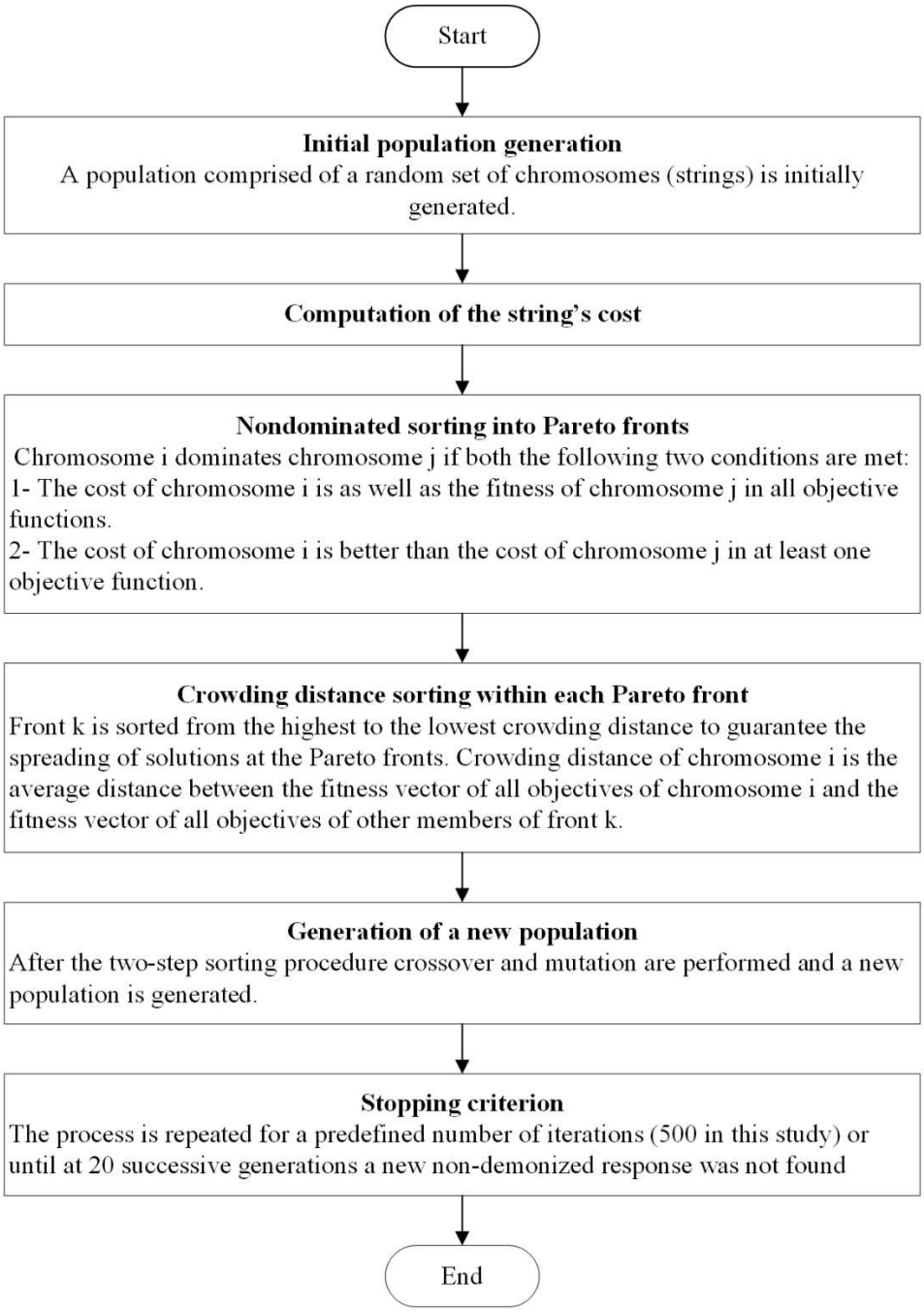



**Fig. 1.** NSGA-II methodology flowchart

## 2.1. Objective Functions

The goal of this study is to optimize the locations of quality sensors in the network as part of the battle of the water sensors networks (BWSN) using the NSGA-II. The assumptions of the research are given as below:

1- Contamination events can occur at any time of the day.

2- Each contamination event can occur at one point in the network.

3- pollutant does not react with other materials in water and is stable.

4- All nodes can be a possible location for the quality sensor placement

5- Response time of the system is assumed to be 30 minutes, in other words when any of the sensors detect contamination, after 30 minutes the network is closed and during this time the monitoring station will collect data.

6- Monitoring stations record concentrations every minute.

7- The sensors can detect contamination greater than 0.001 mg/l without any errors.

8- Pollutant injection time is between 30 and 180 minutes.

9- Injection mass rate ranges from 10 to 100 gr/min.

Four specific objectives are considered for the optimal locations of water quality sensors in the network:

1- Sensor detection likelihood ($f_1$), 2- Sensor expected detection time ($f_2$), 3- Sensor detection redundancy ($f_3$), and 4- Percentage of affected consumer nodes ($f_4$).

**2.1.1. Sensor detection likelihood ($f_1$)**

In a specific network of sensors, the probability of detection is described as bellow:

$$f_1 = \frac{1}{TS}\sum_{i=1}^{TS} d_i \tag{6}$$

$$d_i = \begin{cases} 1 & C(i,t) \geq C_{min} \\ 0 & C(i,t) < C_{min} \end{cases}$$

In this equation, TS denotes the total number of contamination scenarios considered.

**2.1.2. Sensor expected detection time ($f_2$)**

For each pollution incident, the elapsed time from the start of the contamination event, to the first identified presence of contaminant by a sensor is defined as the time of detection by a

sensor. $t_i$ is the time of the first detection by $i^{th}$ sensor. The time of detection ($t_d$) is the minimum detection time among all sensors present in this design.

$$t_d = min\, t_i \tag{7}$$

The sensor expected detection time is computed by:

$$f_2 = E(t_d) = \frac{1}{\sum_{i=1}^{TS} d_i} \sum_{i=1}^{TS} IF_i \times td_i \times I[t_d(i,t)] \tag{8}$$

$$I[t_d(i,t)] = \begin{cases} 1 & t_d(i,t) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$IF = \frac{1}{N} \sum_{i=1}^{N} NP_i$$

$$d_i, NP_i = \begin{cases} 1 & C(i,t) > C_{min} \\ 0 & \text{otherwise} \end{cases}$$

where E(td) represents the mathematical expectation of the minimum detection time $t_d$. IF is the impact factor, so contamination with the greatest impact should be identified more quickly.

**2.1.3. Sensor detection redundancy ($f_3$)**

The development of sensors that detect the contamination in water distribution networks in real-time is still ongoing research. Although, there will be an uncertainty in sensor detections. Then, to avoid false-positive sensor detections and to increase the reliability of sensor detections it is necessary to increase redundancy among sensors. a triply redundant measure that at least three sensors were required to detect the presence of contaminant concentration at a maximum time of 30 minutes between the first and third detections is considered. For each event, the redundancy of a sensor network design is equal to 1 if all of the following conditions are true, and if at least one of them is false the redundancy is equal to 0.

1. $| t_1 - t_2 | \leq 30\text{min}$
2. $| t_1 - t_3 | \leq 30\text{min}$
3. $| t_2 - t_3 | \leq 30\text{min}$

Where $t_1$, $t_2$, and $t_3$ are the detection times by sensor1, sensor 2, and sensor 3 respectively.

The redundancy ($f_3$) for a sensor network design is:

$$f_3 = \frac{1}{\sum_{i=1}^{TS} d_i} \sum_{r=1}^{TS} R_r \tag{9}$$

$$R_r = \begin{cases} 1 & | t_1 - t_3 | \leq 30\text{min} \\ 0 & \text{otherwise} \end{cases}$$

### 2.1.4. Percentage of affected consumer nodes ($f_4$)

in each water distribution network, several nodes have no demand, in other words, water is not taken from these nodes. These nodes are considered inactive nodes and nodes that have demand are considered active nodes (internal nodes). Therefore, in this section, the percentage of nodes that have demand and have been infected before system response has been calculated. This objective function that should be minimized can be formulated as follows:

$$f_4 = \frac{\sum_{i=1}^{N} NP_i \times D_i \times IF_i}{N \times TS} \times 100 \quad (10)$$

$$NP_i = \begin{cases} 1 & C(i,t) > C_{min} \\ 0 & \text{otherwise} \end{cases} \qquad D_i = \begin{cases} 1 & D(i,t) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$IF = \frac{\sum_{t=0}^{t=t_d+30} D(i,t)}{\sum_{i=1}^{N} \sum_{t=0}^{t=t_d+30} D(i,t)}$$

Where $NP_i = 1$ if pollution concentration of node i is higher than allowable ones, otherwise $NP_i = 0$ and N denotes the total number of nodes in the water distribution network. $D_i$ is the demand of nodes and IF is the impact factor. According to this factor, the nodes that have more demands should be identified before being infected.

**2.2. Case study**

In this study, in order to design water quality sensors by using the optimization method and the model proposed above example network 3, shown in Fig. 2, from EPANET was used. This system consists of 91 nodes (consumer and internal nodes), 120 pipes linking the nodes, three elevated storage tanks, two pumping stations, and two constant head sources: a lake and a river. Time steps for water quality and hydraulic simulations were 1 min and the demand flow patterns were 24 h.

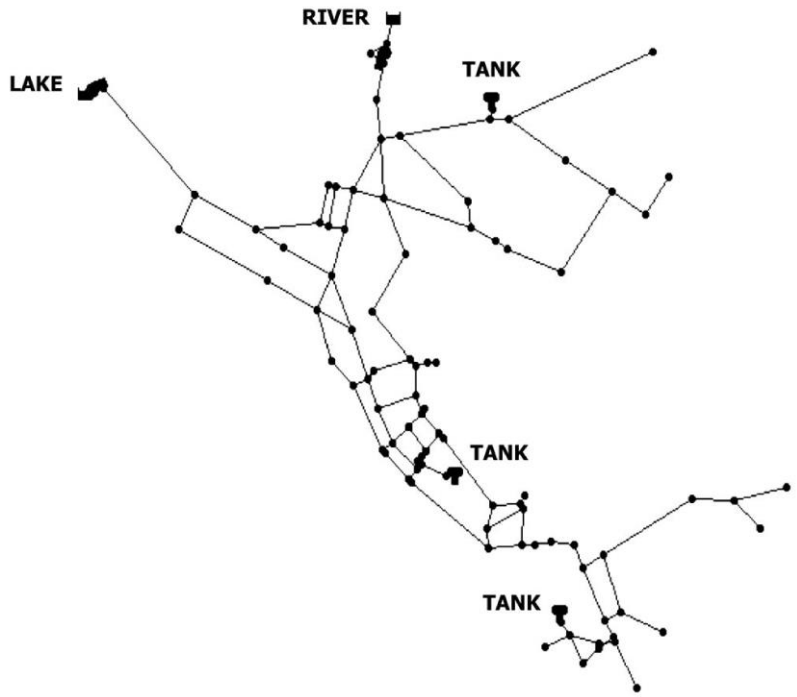



**Fig. 2.** Layout of EPANET example 3

### 2.3. Construction of contamination matrix

To evaluate the fitness function of the sensor network detection likelihood (f1), the sensor network detection redundancy (f3), and to evaluate the cost function of sensor expected detection time (f2), and percentage of affected consumer nodes (f4) a contamination matrix [35] should be produced. To construct a contamination matrix, assumptions about the number, starting time, mass rate, duration of injection, and location must be considered. In this study the following assumptions are considered:

1- number of injections: 1 node

2- starting time of injection: the time is randomly selected from the simulation time of a water distribution system.

3- mass rate of injection: randomly selected between 10 to 100 gr/min.

4- duration of injection: randomly selected between 30 to 180 min.

5- location of injection: randomly selected from all nodes excluding dead ends

As mentioned, contamination events can occur at any node at any time with any mass rate and duration times. Therefore, as the size of the system increases, the number of possible contamination events increases. To deal with this problem, Ami Preisa et al [26], developed a heuristic procedure which used a small sample of contaminants that represented the total potential contaminants. Instead of using the entire contamination matrix, they used a smaller number of contamination samples to illustrate the probable contamination events and found similar results to those achieved using the full matrix. They used the following formulas to create a set of contaminants that included a reduced contamination matrix:

Minimize: $\sum_{i=1}^{5} |AS_i - AN_i| + |\sigma S_i - \sigma N_i|$ (11)

Subject to: $q_j > 0 \quad j = 1,2,\ldots,N$ (12)

Where $AS_i$ and $S_i$ are average and standard deviation values of the geographical x coordinate of a sampled contamination events set. $AN_i$ and $N_i$ are average and standard deviation values of the geographical x coordinate of the water distribution system nodes. $q_j$ is outgoing flow from node j, and N is the total number of system nodes. In these equations, i=2 refers to the geographical y coordinate; i=3 represents the injection mass rate; i=4 expresses the injection starting time; and i=5 is the injection duration.

Equation 11 states that when searching for the sample, the nodes located at the endpoints of the network should not be used because if the infection enters the network in these nodes, the infection will not spread in the network.

By solving the optimization problem presented in Equations 10 and 11, we come to a sample of pollution, which represents the total pollution of the pollution matrix. This optimization problem is solved using a genetic algorithm.

The contamination matrix in this study has 62208 contamination events, which is obtained as follows:

injecting at 81 nodes (The number of dead-end nodes is 10, which is subtracted from the total number of network nodes), every 30 min four mass injection rates of 20, 40, 60, and 80 gr/min, at four injection durations of 40, 80, 120, and 160 minutes (i.e. 81*48*4*4=62208 events).

To create a reduced matrix by equations 10 and 11, 1000 contamination events were selected and this reduced matrix was used as contamination events that are injected into the network and optimization is performed based on these contamination events. This study's general outline is presented in Figure 3.

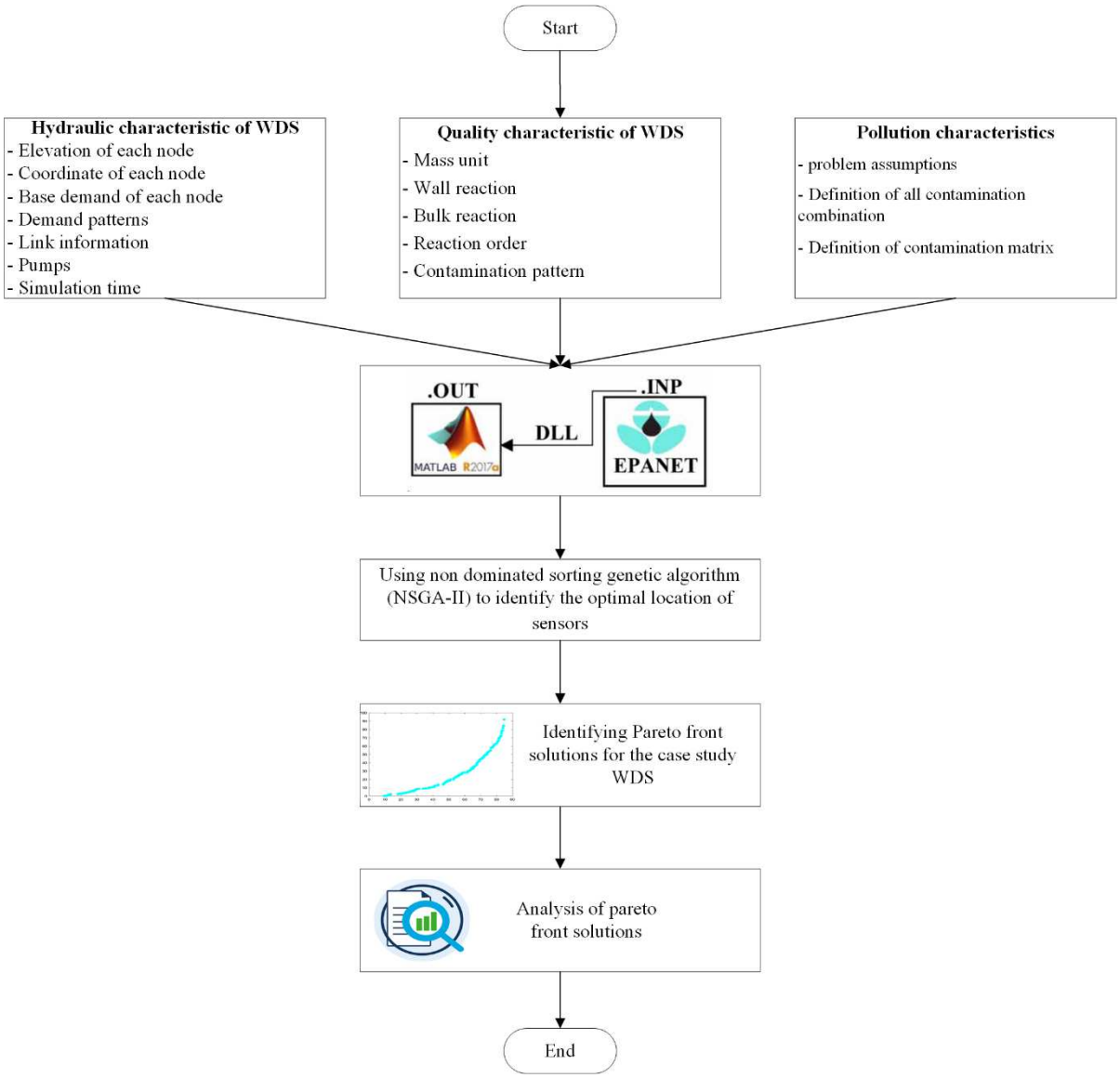



**Fig. 3.** simulation–optimization flowchart of the research

## 3. Results and Discussion

In EPANET example 3, four sensors have been placed and sensitivity analysis has been done. Genetic algorithm parameters in this study are as follows: crossover and mutation probability of 0.75 and 0.1, respectively were used. The number of populations was 1000 and the maximum number of iterations (generations) was set to 300. There were two stop conditions for the NSGA-II algorithm: either if the number of iterations reaches its maximum of 300, or if at 20 successive generations, a new non-demonized response was not found.

### 3.1. Base run results

Figs. 4 to 9 show base run results for placing four sensors in EPANET Example 3. in fig. 4.a. optimal Pareto front for maximizing sensor detection likelihood ($f_1$) versus minimizing sensor expected detection time ($f_2$) is summarized. The interaction between the two objective functions is such that in order to maximize sensor detection likelihood in the network, sensors should be installed in the downstream nodes of the network, Installing sensors downstream of the network increases detection time. On the other hand, in the second objective function, sensors should be installed in nodes close to the entry point of infection into the network to minimize the time of detection, so the layout of sensors according to the two objective functions is in conflict with each other.

Figure 4. b shows the selected locations of the sensors in the network for the three solutions from the Pareto front. In order to place the sensors in the network, according to the different points provided by the Pareto front, three different solutions (locations) were selected. The first solution: the best answer for the objective function number one and the worst answer for the objective function number 2, the second solution: the best answer for the objective function number two and the worst answer for the objective function number one, and finally the third solution where both objective functions have an intermediate state. The sensor locations according to solution 1.2.1 indicate the best detection likelihood equal to 77.4% but the worst

solution for the detection time (120.6 min). Solution 1.2.2 shows the location of the sensors according to the minimum contamination detection time (3.28 minutes) but the worst detection likelihood is equal to 8.4%. In solution 1.2.3, the method presented by Young (1993) is used to select the optimal point of Pareto front for locating the sensors in the distribution network. According to the Young's method, the optimal point has a detection likelihood equal to 51.7% and a contamination detection time of 60.07 minutes.

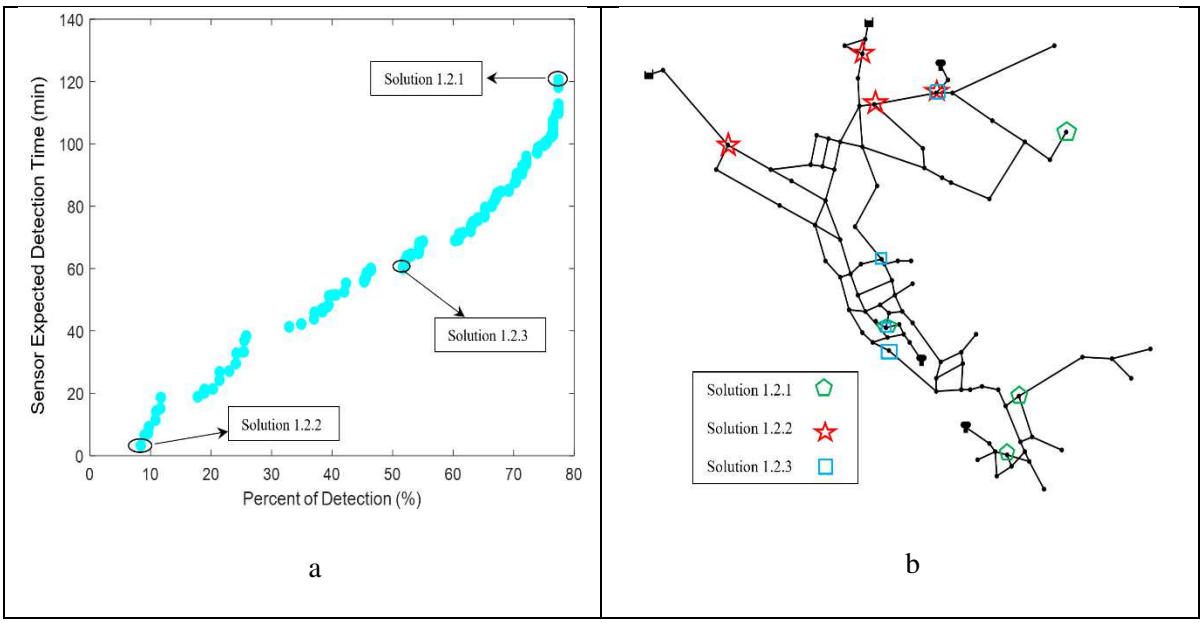



**Fig. 4.** Optimal Pareto front for (a) detection likelihood versus sensor expected detection time, and (b) selected locations of the sensors according to Pareto front

As expected, the optimal location of the sensors according to solution 1.2.1 is the downstream points of the network and according to solution 1.2.2, the points are close to the injection areas.

Figure. 5.a. an optimal Pareto front for maximizing sensor detection likelihood ($f_1$) versus maximizing sensor detection redundancy ($f_3$) are summarized. The interaction between the two objective functions is such that in order to achieve the maximum detection likelihood, the sensors must be spread in the downstream nodes of the network to cover more nodes, while in

the objective function 3, in order to achieve maximum redundancy, the sensors must have a short distance from each other. Therefore, the location of the sensors in the two objective functions is in conflict with each other.

Figure 5. b. shows the selected locations of the sensors in the network for the three solutions from the Pareto front. The location of the sensors according to solution 1.3.1 indicates the best percentage of detection likelihood equal to 77.4% but the worst value of redundancy (0%). Solution 1.3.2 shows the location of the sensors according to the best redundancy value of 97.87%, but the worst detection likelihood (33%). In solution 1.3.3, the method presented by Young (1993) is used to select the optimal point of Pareto front for locating the sensors in the distribution network. According to the Young's method, the optimal point has a detection likelihood equal to 65.3% and the redundancy value of 52.68%.

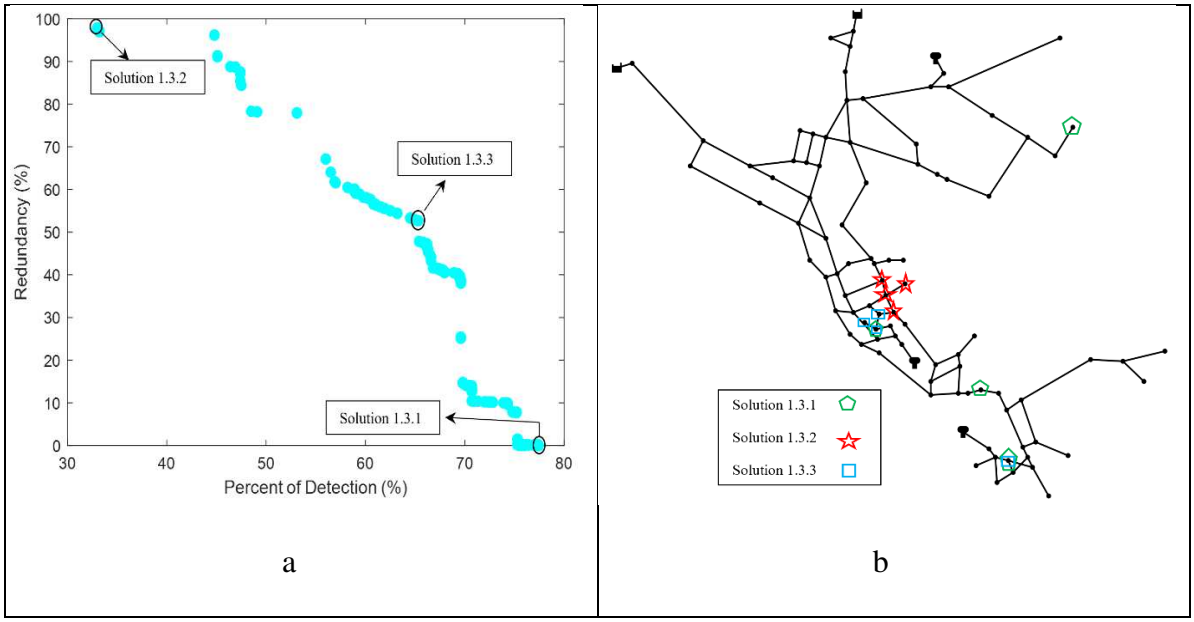



**Fig. 5.** Optimal Pareto front for (a) detection likelihood versus sensor detection redundancy, and (b) selected locations of the sensors according to Pareto front

Figure. 6. a. optimal Pareto front for maximizing sensor detection likelihood ($f_1$) versus minimizing the percentage of affected active (consumer) nodes ($f_4$) are summarized.

To achieve the maximum detection likelihood, the sensors must be spread across the downstream nodes of the network, which infects a large number of network nodes. However, in the objective function 4, in order to achieve the minimum number of infected nodes, the sensors must be installed in locations close to where the pollution enters the network. Therefore, the optimal location of the sensors in the two objective functions is in conflict with each other.

Figure 6. b. shows the selected locations of the sensors in the network for the three solutions from the Pareto front. The location of the sensors according to solution 1.4.1 indicates the best percentage of detection likelihood equal to 77.4% but shows the worst percentage of infected nodes (12.77%). Solution 1.4.2 shows the location of the sensors according to the lowest (the best) percentage of infected nodes equal to 7.86%, but the percentage of infection detection is the lowest (58.2%). In solution 1.4.3, the method presented by Young (1993) is used to select the optimal point of Pareto front for locating the sensors in the distribution network. According to the Young's method, the optimal point has a detection likelihood equal to 70.7% and the percentage of affected nodes of 9.08%.

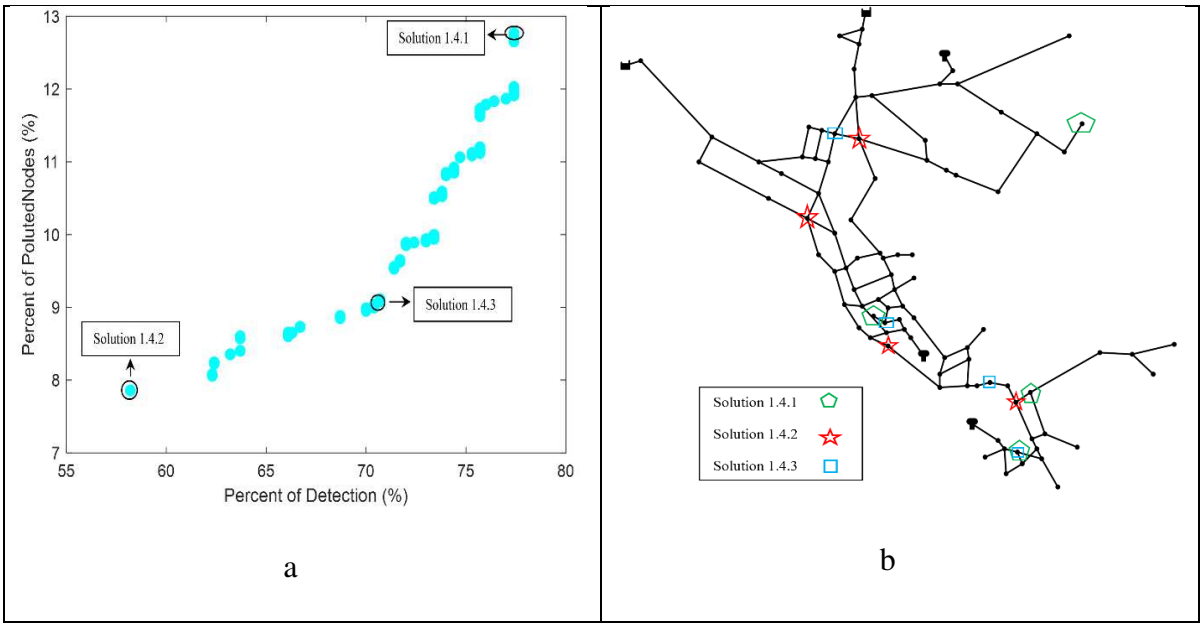


**Fig. 6.** Optimal Pareto front for (a) detection likelihood versus percentage of polluted nodes, and (b) selected locations of the sensors according to Pareto front

Figure. 7. a. optimal Pareto front for minimizing sensor expected detection time ($f_2$) versus maximizing sensor detection redundancy ($f_3$) are summarized.

In order to achieve the minimum sensor expected detection time in the distribution network, the sensors should be located in areas close to the entry point of contamination and in different parts of the network, but to achieve the optimal value according to objective function 3, the sensors should be located close to each other. Therefore, the optimal location of the sensors in the two objective functions is in conflict with each other.

Figure 7. b. shows the selected locations of the sensors in the network for the three solutions from the Pareto front. The location of the sensors according to solution 2.3.1 indicates the location of the sensors according to the minimum contamination detection time equal to 61.78 minutes, but the worst sensor detection redundancy (37.95 %). Solution 2.3.2 shows the location of the sensors according to the best redundancy value of 97.87%, but the worst expected detection time (111.1 min). In solution 2.3.3, the method presented by Young (1993) is used to select the optimal point of the Pareto front for locating the sensors in the distribution network. According to the Young's method, the optimal point has a contamination detection time equal to 70.14 minutes and the redundancy of 74.09%.

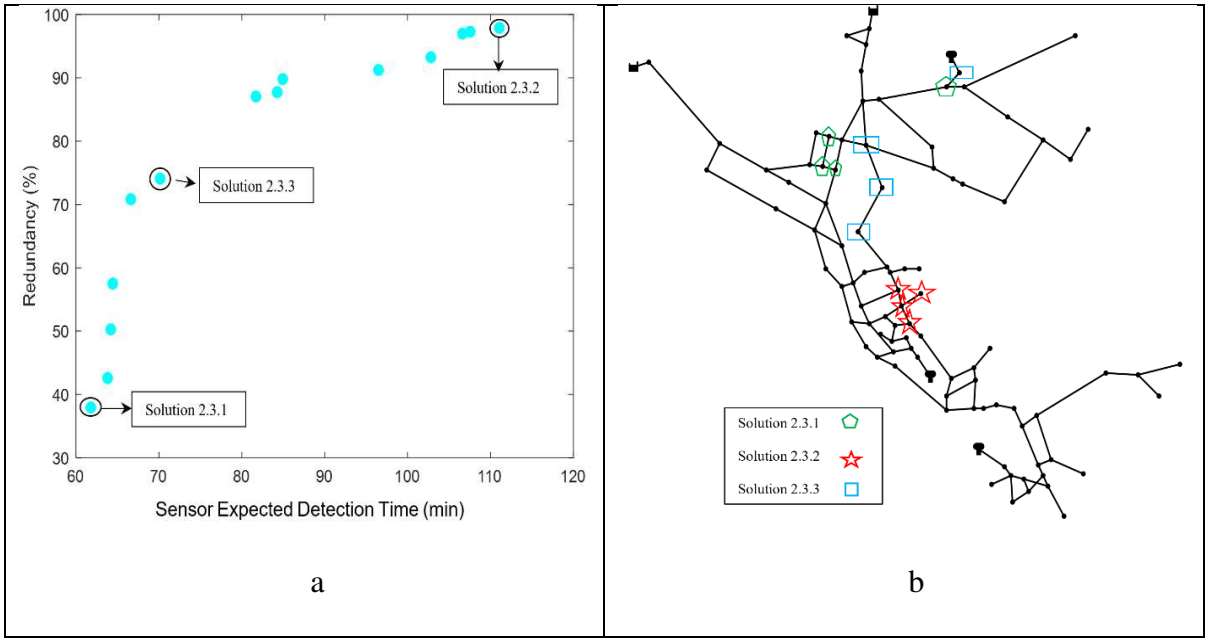



**Fig. 7.** Optimal Pareto front for (a) sensor expected detection time versus sensor detection redundancy, and (b) selected locations of the sensors according to Pareto front

Figure. 8. a. optimal Pareto front for minimizing sensor expected detection time ($f_2$) versus minimizing the percentage of affected consumer nodes ($f_4$) are summarized.

Figure 8. b. shows the selected locations of the sensors in the network for the three solutions from the Pareto front. The location of the sensors according to solution 2.4.1 indicates the location of the sensors according to the minimum contamination detection time equal to 1.9 minutes, but the worst percentage of active (consumer) nodes affected (22.87 %). Solution 2.4.2 shows the location of the sensors according to the best percentage of active (consumer) nodes affected 7.86%, but the worst expected detection time (115.8 min). In solution 2.4.3, the method presented by Young (1993) is used to select the optimal point of Pareto front for locating the sensors in the distribution network. According to the Young's method, the optimal point has a contamination detection time equal to 28.13 minutes and the percentage of affected nodes of 13.37%.

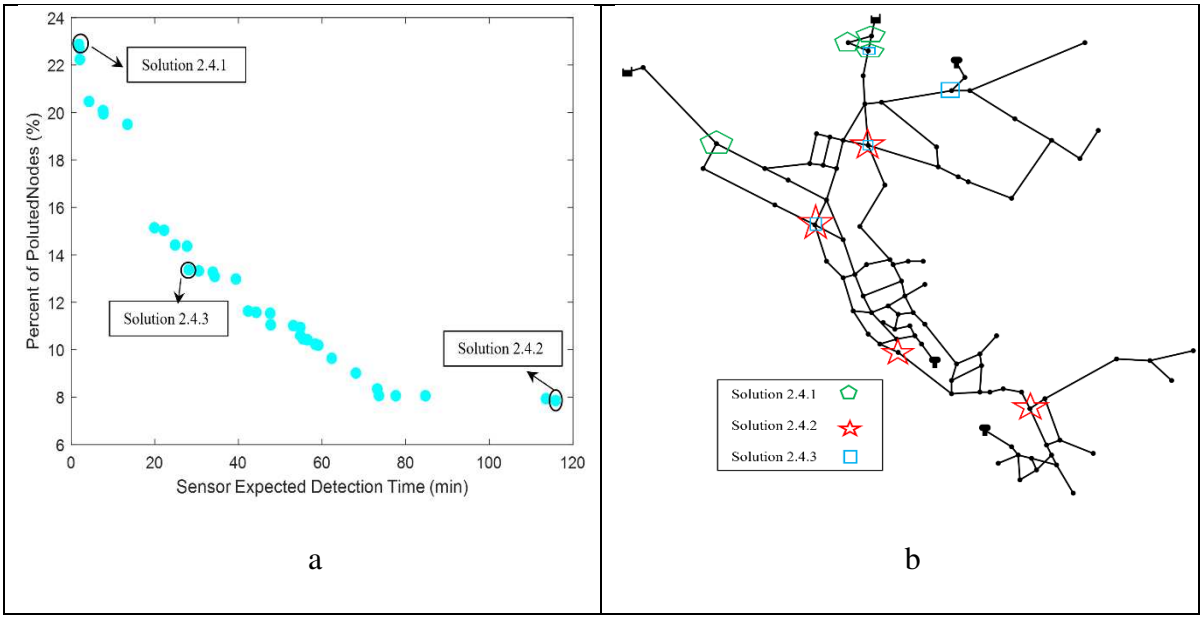



**Fig. 8.** Optimal Pareto front for (a) sensor expected detection time versus percentage of polluted nodes, and (b) selected locations of the sensors according to Pareto front

Figure. 9. a. optimal Pareto front for maximizing sensor detection redundancy ($f_3$) versus minimizing the percentage of affected active (consumer) nodes ($f_4$) are summarized.

In order to achieve maximum redundancy, the sensors should be located close to each other, but in order to minimize the number of infected nodes, the sensors should be spread across different parts of the network and close to the entry point of contamination. Therefore, the optimal location of the sensors according to these two objective functions is in conflict with each other.

Figure 9. b. shows the selected locations of the sensors in the network for the three solutions from the Pareto front. The location of the sensors according to solution 3.4.1 indicates the location of the sensors according to the best redundancy value of 97.87%, but the worst percentage of active (consumer) nodes affected (14.68 %). Solution 3.4.2 shows the location of the sensors according to the best percentage of active (consumer) nodes affected 7.86%, but the worst redundancy (0 %). In solution 3.4.3, the method presented by Young (1993) is used to select the optimal point of Pareto front for locating the sensors in the distribution network.

According to the Young's method, the optimal point has a redundancy equal to 62.76% and the percentage of affected nodes of 10.48%.

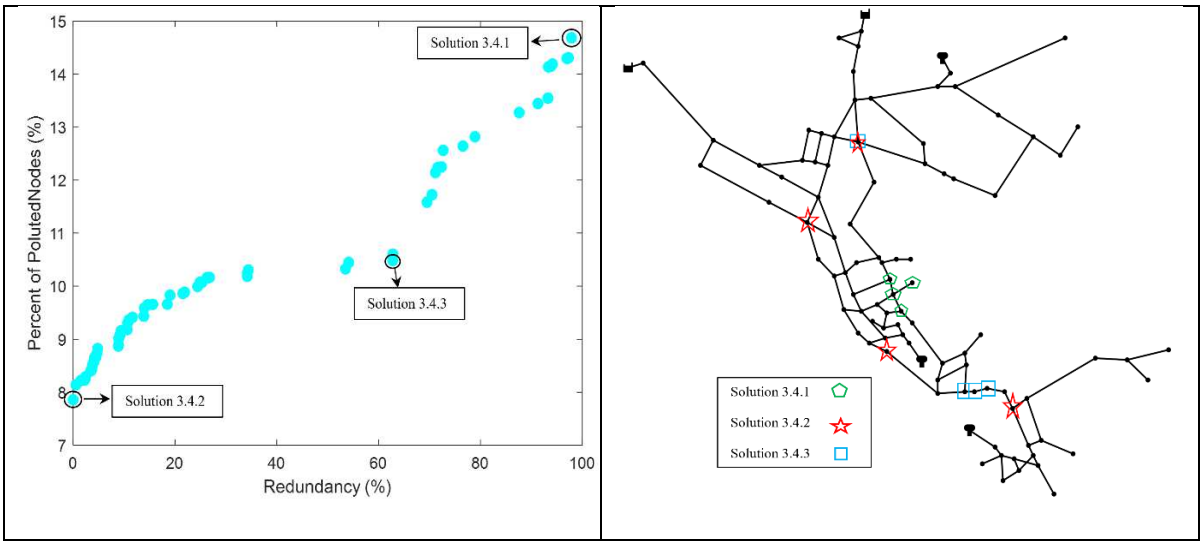
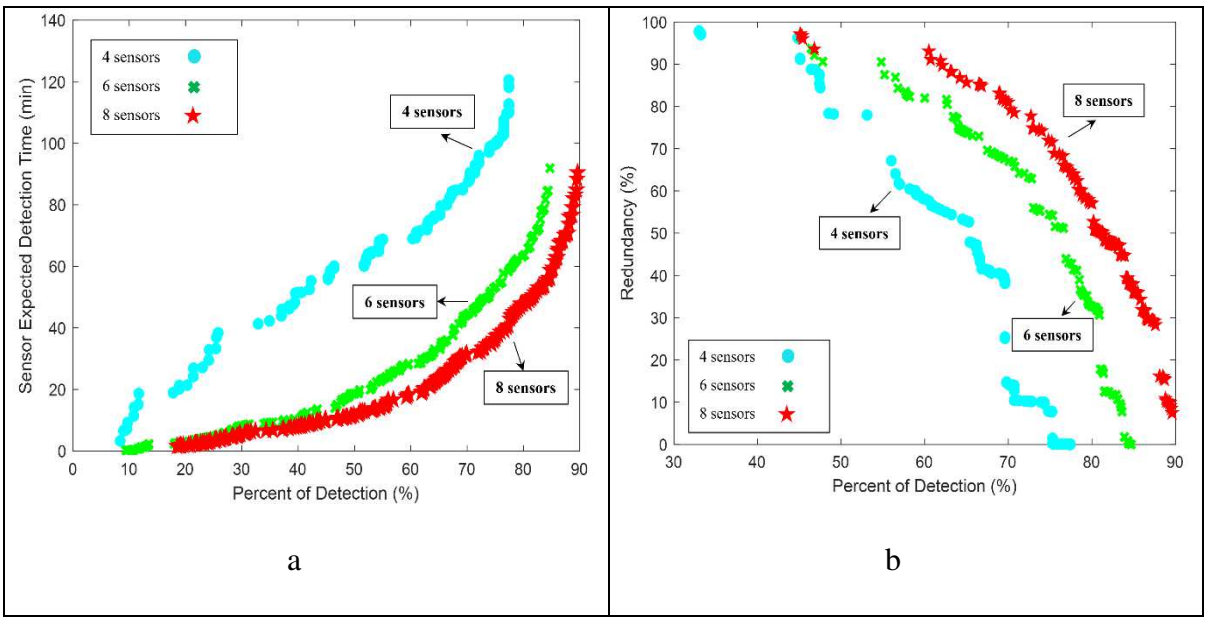



**Fig. 9.** Optimal Pareto front for (a) percentage of polluted nodes versus sensor detection redundancy, and (b) selected locations of the sensors according to Pareto front

### 3.2. Sensitivity analysis results

In this section, the effect of the number of sensors used in the network on the objective functions is investigated. The purpose is to check whether the increase in the number of sensors improves the optimal Pareto front s. For this purpose, the number of sensors was increased from four to six and eight sensors. Figures 10 describes the comparison between the optimal Pareto front s for the 4 sensors used in the base run and the Pareto front s for 6 and 8 sensors. As shown in Figures 10.a to 10.f, the Pareto front s for 6 sensors dominates the Pareto front s established with 4 sensors, and the Pareto front s for 8 sensors dominates the Pareto front s

established with 4 and 6 sensors. For example, according to Figure 10. a. When using four sensors, the best detection likelihood is equal to 77.4% and the worst solution for the detection time is 120.6 minutes. but with the increase in the number of sensors to six, the best detection likelihood by sensors increased by 7.3 % and reached 84.7%, On the other hand, the worst solution for the detection time decreased by 28.69 minutes (23.7%) and reached 91.91 minutes that This has improved the two objective functions. Then, with the increase in the number of sensors to eight, the best detection likelihood compared to four and six sensors increased by 12.2 and 4.9 percent, respectively, and reached 89.6 percent. Also, the worst solution for the detection time compared to four and six sensors decreased by 25 and 1.5 percent and reached 90.49 minutes.



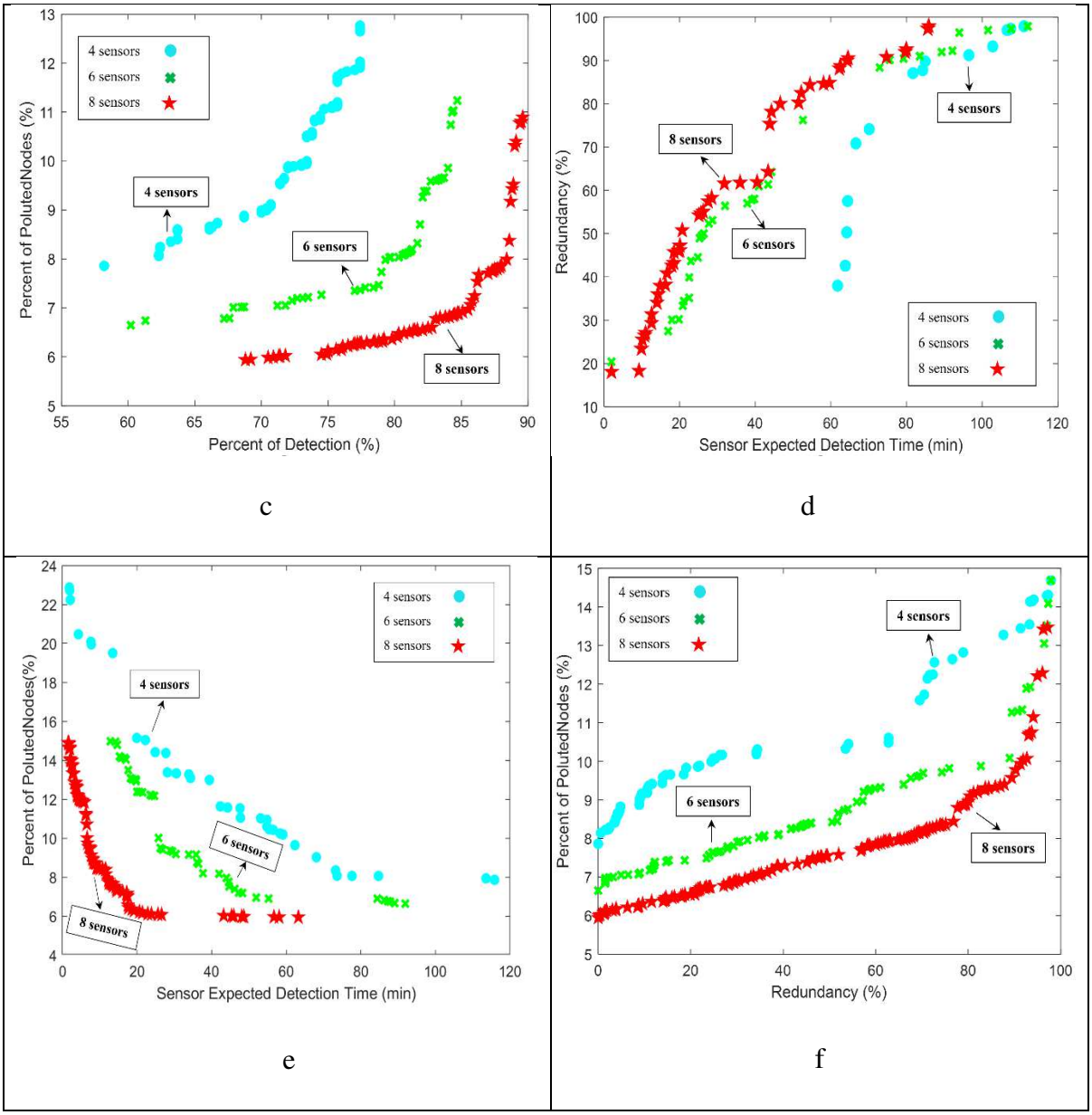



**Fig. 10.** Comparing the optimal Pareto fronts for 4 sensors used in the base run with those built with 6 and 8 sensors for different objective functions

In general, with increasing the number of sensors, the values of different objective functions have improved compared to the base run (4 sensors). Figure 11 shows the optimal Pareto front for the number of sensors used in the distribution network versus the best detection likelihood. According to this figure at first, when the number of sensors is lower the best detection likelihood meets a higher slope but when the number of sensors exceeds 14 sensors the best

detection likelihood increases with a lower slope. This means that for more than 14 sensors, an increase in the number of sensors does not have a noticeable impact on the Pareto fronts.

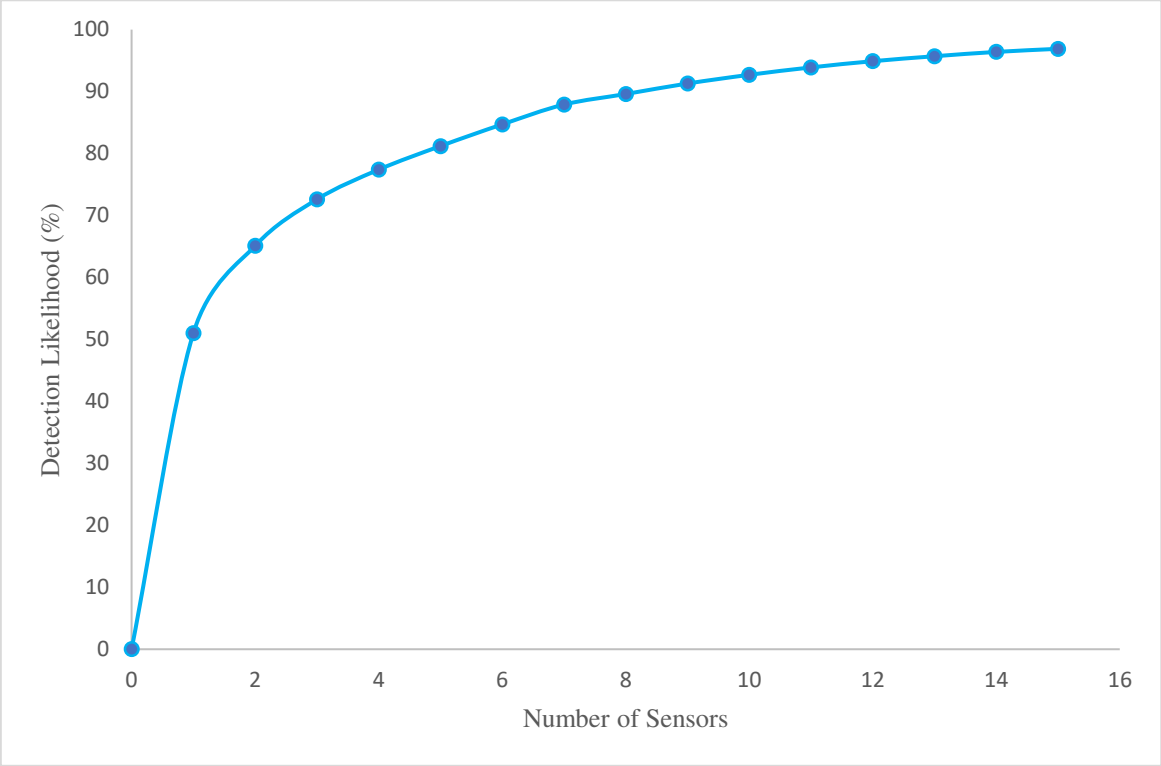



**Fig. 11.** Optimal Pareto front for number of sensors versus detection likelihood

## 4. Conclusions

Water distribution Networks (WDNs) are the main components of public infrastructure that distribute safe drinking water to billions of customers around the world. WDNs are vulnerable to intentional or accidental contamination events due to the complexity of their structures and a large number of access points. When contamination enters to WDNs, it spreads through the system and harms the customer health and the local economy. Therefore, protection of WDNs is an important issue and the security of drinking water should be increased. This leads to the use of sensors for identifying the pollutions in water distribution systems. A multi-objective scheme is developed for sensor network design in this study. For this purpose, EPANET example network No. 3 has been selected to optimally locate quality sensors in drinking water distribution networks. By selecting 1000 contaminations scenarios with the same

characteristics and effects as all possible contamination combinations, a contamination matrix is formed using the genetic algorithm. The optimal location of four sensors in the network was analyzed using a simulation-optimization approach. EPANET and NSGA-II were applied as a hydraulic-quality simulator and an optimizer approach, respectively. Sensor locations have been selected based on the following four objectives: 1- sensor detection likelihood, 2- sensor expected detection time, 3- sensor detection redundancy, and 4- the affected population before detection. The importance of contamination and also the importance of network nodes are not considered the same by considering two important coefficients based on the amount of damage caused by contamination in this study. According to the important coefficient 1, the contaminations that cause more damage to the distribution network (affect more people) are more important and should be identified in less time and also according to the important coefficient 2, nodes with more demands are more important and sensors must be installed in places to detect contamination before entering these nodes. Then, in the second step, the sensitivity analysis related to the number of sensors used in the network on different objective functions was investigated so that the number of sensors increased from four to six and eight. The results showed that with an increasing number of sensors the Pareto fronts for 6 sensors dominate the Pareto fronts established with 4 sensors, and the Pareto fronts for 8 sensors dominate the Pareto fronts established with 4 and 6 sensors. Finally, the optimal Pareto front for the number of sensors used in the distribution network versus the best detection likelihood was depicted. The results demonstrated that when the number of sensors is small, with increment the number of sensors, the best detection likelihood changes rapidly; But when the number of sensors exceeds 14, the best detection likelihood changes slowly. This means that for cases where the number of sensors in the network is more than 14, increasing the number of sensors does not have a significant effect on the optimal Pareto front.

## Statements & Declarations

### Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

### Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

### Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Siroos harif], [Mohsen Dehghani Darmian], [Gholamreza Azizyan] and [Mohammad Givehchi]. The first draft of the manuscript was written by [Siroos harif] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

### Ethical Approval

The authors declare that all ethical points in this manuscript are observed.

### Data transparency

All data and materials comply with field standards.

### Consent to Participate

All authors consent to participate.

### Consent to Publish

All authors consent to publish the manuscript.

**References**


Adedoja, O. S., Hamam, Y., Khalaf, B., & Sadiku, R. (2018a). A state-of-the-art review of an optimal sensor placement for contaminant warning system in a water distribution network. *Urban Water Journal*, *15*(10), 985-1000.

Adedoja, O. S., Hamam, Y., Khalaf, B., & Sadiku, R. (2018b). Towards development of an optimization model to identify contamination source in a water distribution network. *Water*, *10*(5), 579.

Afshar, A., & Miri Khombi, S. M. (2015). MULTIOBJECTIVE OPTIMIZATION OF SENSOR PLACEMENT IN WATER DISTRIBUTION NETWORKS DUAL USE BENEFIT APPROACH. *Iran University of Science & Technology*, *5*(3), 315-331.

Ahmed, T., Pervez, A., Mehtab, M., & Sherwani, S. K. (2015). Assessment of drinking water quality and its potential health impacts in academic institutions of Abbottabad (Pakistan). *Desalination and Water Treatment*, *54*(7), 1819-1828.

Al-Zahrani, M. A., & Moeid, K. (2001). Locating optimum water quality monitoring stations in water distribution system. In *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges* (pp. 1-9).

Bazargan-Lari, M. R. (2014). An evidential reasoning approach to optimal monitoring of drinking water distribution systems for detecting deliberate contamination events. *Journal of cleaner production*, *78*, 1-14.

Bekele, E. G., & Nicklow, J. W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology*, *341*(3-4), 165-176.

Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., & Watson, J. P. (2006). Sensor placement in municipal water networks with temporal integer programming models. *Journal of water resources planning and management*, *132*(4), 218-224.

Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., & Watson, J. P. (2009). Designing contamination warning systems for municipal water networks using imperfect sensors. *Journal of Water Resources Planning and Management*, *135*(4), 253-263.

Bush, G. W. 2002. The National Security Strategy of the United States of America. Tech. Rep. Washington, DC: Executive Office of the President.

Cheifetz, N., Sandraz, A. C., Féliers, C., Gilbert, D., Piller, O., & Lang, A. (2015). An incremental sensor placement optimization in a large real-world water system. *Procedia Engineering*, *119*, 947-952.

ChinaNews, 2016. http://www.chinanews.com/sh/2016/05-25/7883161.shtml

Comboul, M., & Ghanem, R. (2013). Value of information in the design of resilient water distribution sensor networks. *Journal of Water Resources Planning and Management*, *139*(4), 449-455.

Corso, P. S., Kramer, M. H., Blair, K. A., Addiss, D. G., Davis, J. P., & Haddix, A. C. (2003). Costs of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin. *Emerging infectious diseases*, *9*(4), 426.

Costa, D. M., Melo, L. F., & Martins, F. G. (2013). Localization of contamination sources in drinking water distribution systems: a method based on successive positive readings of sensors. *Water resources management*, *27*(13), 4623-4635.

Daud, M. K., Nafees, M., Ali, S., Rizwan, M., Bajwa, R. A., Shakoor, M. B., ... & Zhu, S. J. (2017). Drinking water quality status and contamination in Pakistan. *BioMed research international*, *2017*.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, *6*(2), 182-197.

de Winter, C., Palleti, V. R., Worm, D., & Kooij, R. (2019). Optimal placement of imperfect water quality sensors in water distribution networks. *Computers & Chemical Engineering*, *121*, 200-211.

Dorini, G., Jonkergouw, P., Kapelan, Z., Di Pierro, F., Khu, S. T., & Savic, D. (2008). An efficient algorithm for sensor placement in water distribution systems. In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-13).

Eliades, D., & Polycarpou, M. (2008). Iterative deepening of Pareto solutions in water sensor networks. In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-19).

Falco, R., & Williams, S. I. (2009). Waterborne Salmonella Outbreak in Alamosa, Colorado March and April 2008. *Outbreak identification response, and investigation, Safe Drinking Water Program, Water Quality Control Division, Colorado Department of Public Health and Environment*.

Gong, W., Cai, Z., & Liang, D. (2014). Adaptive ranking mutation operator based differential evolution for constrained optimization. *IEEE transactions on cybernetics*, *45*(4), 716-727.

Gong, W., Zhou, A., & Cai, Z. (2015). A multioperator search strategy based on cheap surrogate models for evolutionary optimization. *IEEE transactions on Evolutionary Computation*, *19*(5), 746-758.

Grayman, W. M., Murray, R., Savic, A. A., & Farmani, R. (2016). Redesign of water distribution systems for passive containment of contamination. *Journal-American Water Works Association*, *108*(7), E381-E391.

Gueli, R. (2008). Predator-prey model for discrete sensor placement. In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-9).

Hart, W. E., & Murray, R. (2010). Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *Journal of Water Resources Planning and Management*, *136*(6), 611-619.

He, G., Zhang, T., Zheng, F., & Zhang, Q. (2018). An efficient multi-objective optimization method for water quality sensor placement within water distribution systems considering contamination probability variations. *Water research*, *143*, 165-175.

Hrudey, S. E., Payment, P., Huck, P. M., Gillham, R. W., & Hrudey, E. J. (2003). A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. *Water Science and Technology*, *47*(3), 7-14.

Hu, C., Ren, G., Liu, C., Li, M., & Jie, W. (2017). A Spark-based genetic algorithm for sensor placement in large scale drinking water distribution systems. *Cluster Computing*, *20*(2), 1089-1099.

Huang, J. J., McBean, E. A., & James, W. (2008). Multi-objective optimization for monitoring sensor placement in water distribution systems. In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-14).

Huang, J. J., & McBean, E. A. (2009). Data mining to identify contaminant event locations in water distribution systems. *Journal of Water Resources Planning and Management*, 135(6), 466-474.

Kalyanmoy, D. (2001). Multi-Objective Optimization Using Evolutionary Algorithms John Wiley & Sons.

Kannan, S., Baskar, S., McCalley, J. D., & Murugan, P. (2008). Application of NSGA-II algorithm to generation expansion planning. *IEEE Transactions on Power systems*, *24*(1), 454-461.

Keramati, H., Miri, A., Baghaei, M., Rahimizadeh, A., Ghorbani, R., Fakhri, Y., ... & Khaneghah, A. M. (2019). Fluoride in Iranian drinking water resources: a systematic review, meta-analysis and non-carcinogenic risk assessment. *Biological trace element research*, *188*(2), 261-273.

Kessler, A., Ostfeld, A., & Sinai, G. (1998). Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management*, *124*(4), 192-198.

Khorshidi, M. S., Nikoo, M. R., Ebrahimi, E., & Sadegh, M. (2019a). A robust decision support leader-follower framework for design of contamination warning system in water distribution network. *Journal of cleaner production*, *214*, 666-673.

Khorshidi, M. S., Nikoo, M. R., Sadegh, M., & Nematollahi, B. (2019b). A multi-objective risk-based game theoretic approach to reservoir operation policy in potential future drought condition. *Water Resources Management*, *33*(6), 1999-2014.

Kou, J. Drinking water reservoirs were poisoned in Ruyang City of Henan Province. http://news.china.com/zh_cn/social/1007/20031004/11549967.html.

Kumar, A., Kansal, M. L., & Arora, G. (1997). Identification of monitoring stations in water distribution system. *Journal of Environmental Engineering*, *123*(8), 746-752.

Lee, B. H., & Deininger, R. A. (1992). Optimal locations of monitoring stations in water distribution system. *Journal of Environmental Engineering*, *118*(1), 4-16.

Massoudinejad, M., Shahsavani, A., Kamarehie, B., Jafari, A., Ghaderpoori, M., Amini, M. M., & Ghaderpoury, A. (2018). Highly efficient adsorption of fluoride from aqueous solutions by metal organic frameworks: modeling, isotherms, and kinetics. *Fluoride*, *51*(4).

McKenna, S. A., Hart, D. B., & Yarrington, L. (2006). Impact of sensor detection limits on protecting water distribution systems from contamination events. *Journal of Water Resources Planning and Management*, *132*(4), 305-309.

Nabeela, F., Azizullah, A., Bibi, R., Uzma, S., Murad, W., Shakir, S. K., ... & Häder, D. P. (2014). Microbial contamination of drinking water in Pakistan—a review. *Environmental Science and Pollution Research*, *21*(24), 13929-13942.

Naserizade, S. S., Nikoo, M. R., & Montaseri, H. (2018). A risk-based multi-objective model for optimal placement of sensors in water distribution system. *Journal of hydrology*, *557*, 147-159.

Nazempour, R., Monfared, M. A. S., & Zio, E. (2018). A complex network theory approach for optimizing contamination warning sensor location in water distribution networks. *International Journal of Disaster Risk Reduction*, *30*, 225-234.

Ohar, Z., Lahav, O., & Ostfeld, A. (2015). Optimal sensor placement for detecting organophosphate intrusions into water distribution systems. *Water research*, *73*, 193-203.

Ostfeld, A., & Salomons, E. (2004). Optimal layout of early warning detection stations for water distribution systems security. *Journal of water resources planning and management*, *130*(5), 377-385.

Ostfeld, A., & Salomons, E. (2005). Securing water distribution systems using online contamination monitoring. *Journal of Water Resources Planning and Management*, *131*(5), 402-405.

Ostfeld, A., & Salomons, E. (2008). Sensor network design proposal for the battle of the water sensor networks (BWSN). In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-16).

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., ... & Walski, T. (2008). The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. *Journal of Water Resources Planning and Management*, *134*(6), 556-568.

Perelman, L., Arad, J., Housh, M., & Ostfeld, A. (2012). Event detection in water distribution systems from multivariate water quality time series. *Environmental science & technology*, *46*(15), 8212-8219.

Preis, A., & Ostfeld, A. (2008). Multiobjective contaminant sensor network design for water distribution systems. *Journal of Water Resources Planning and Management*, *134*(4), 366-377.

Propato, M. (2006). Contamination warning in water networks: General mixed-integer linear models for sensor location design. *Journal of water resources planning and management*, *132*(4), 225-233.

Propato, M., & Piller, O. (2008). Battle of the water sensor networks. In *Water Distribution Systems Analysis Symposium 2006* (pp. 1-8).

Rathi, S., & Gupta, R. (2016). A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks. *KSCE Journal of Civil Engineering*, *20*(2), 597-608.

Rosen, J. S., Whelton, A. J., McGuire, M. J., Clancy, J. L., Bartrand, T., Eaton, A., ... & Adams, C. (2014). The crude MCHM chemical spill in Charleston, W. Va. *Journal-American Water Works Association*, *106*(9), 65-74.

Rossman, L. A. (2000). EPANET 2: Users manual. Cincinnati US Environmental Protection Agency National Risk Management Research Laboratory 38 (September): 200.

Schwartz, R., Lahav, O., & Ostfeld, A. (2014). Optimal sensor placement in water distribution systems for injection of chlorpyrifos. In *World Environmental and Water Resources Congress 2014* (pp. 485-494).

Shastri, Y., & Diwekar, U. (2006). Sensor placement in water networks: A stochastic programming approach. *Journal of water resources planning and management*, *132*(3), 192-203.

Srinivas, N., & Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, *2*(3), 221-248.

Tinelli, S., Creaco, E., & Ciaponi, C. (2017). Sampling significant contamination events for optimal sensor placement in water distribution systems. *Journal of Water Resources Planning and Management*, *143*(9), 04017058.

UNESCO (2003) Water for people water for life. The United Nations world water development report. United Nations Educational, Scientific and Cultural Organization

UNISEF and WHO (2012) Progress on drinking water and sanitation update 2012. Retrieved from; http://www.unicef.org/media/files/JMPreport2012.pdf

Wang, C., & Zhou, S. (2017). Contamination source identification based on sequential Bayesian approach for water distribution network with stochastic demands. *IISE Transactions*, *49*(9), 899-910.

Woo, H. M., Yoon, J. H., & Choi, D. Y. (2001). Optimal monitoring sites based on water quality and quantity in water distribution systems. In *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges* (pp. 1-9).

World Health Organization (2010) Right to Water.

World Health Organization. (2015). Improving environment and health in Europe: how far have we gotten?

Wu, Z. Y., and Walski, T. (2006). Multiobjective optimization of sensor placement in water distribution systems. Proc., *8th Annual Water Distribution System Analysis Symp*., ASCE, Reston, VA.

Yang, X., & Boccelli, D. L. (2016). Model-based event detection for contaminant warning systems. *Journal of Water Resources Planning and Management*, *142*(11), 04016048.

Yazdi, J. (2018). Water quality monitoring network design for urban drainage systems, an entropy method. *Urban Water Journal*, *15*(3), 227-233.

Yusoff, Y., Ngadiman, M. S., & Zain, A. M. (2011). Overview of NSGA-II for optimizing machining process parameters. *Procedia Engineering*, *15*, 3978-3983.

Yousefi, M., Ghoochani, M., & Mahvi, A. H. (2018). Health risk assessment to fluoride in drinking water of rural residents living in the Poldasht city, Northwest of Iran. *Ecotoxicology and environmental safety*, *148*, 426-430.