Journal of Hydroinformatics

Journal of Hydroinformatics Vol 00 No 0, 1 doi: 10.2166/hydro.2023.057

Sensor placement in water distribution networks using centrality-guided multi-objective optimisation

Kegong Diao [a,*], Michael Emmerich[b,c], Jacob Lan[b], Iryna Yevseyeva[a] and Robert Sitzenfrei[d]

[a] Faculty of Computing, Engineering, and Media, De Montfort University, Leicester LE1 9BH, UK
[b] Faculty of Science, Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, Leiden, CA 2333, The Netherlands
[c] Faculty of Information Technology, University of Jyvaskyla, P.O. Box 35 (Agora), Jyvaskyla, FI 40014, Finland
[d] Unit of Environmental Engineering, Department of Infrastructure Engineering, Faculty of Engineering Sciences, University of Innsbruck, Technikerstrasse 13, Innsbruck 6020, Austria
*Corresponding author. E-mail: kegong.diao@dmu.ac.uk

KD, 0000-0003-2315-9455

ABSTRACT

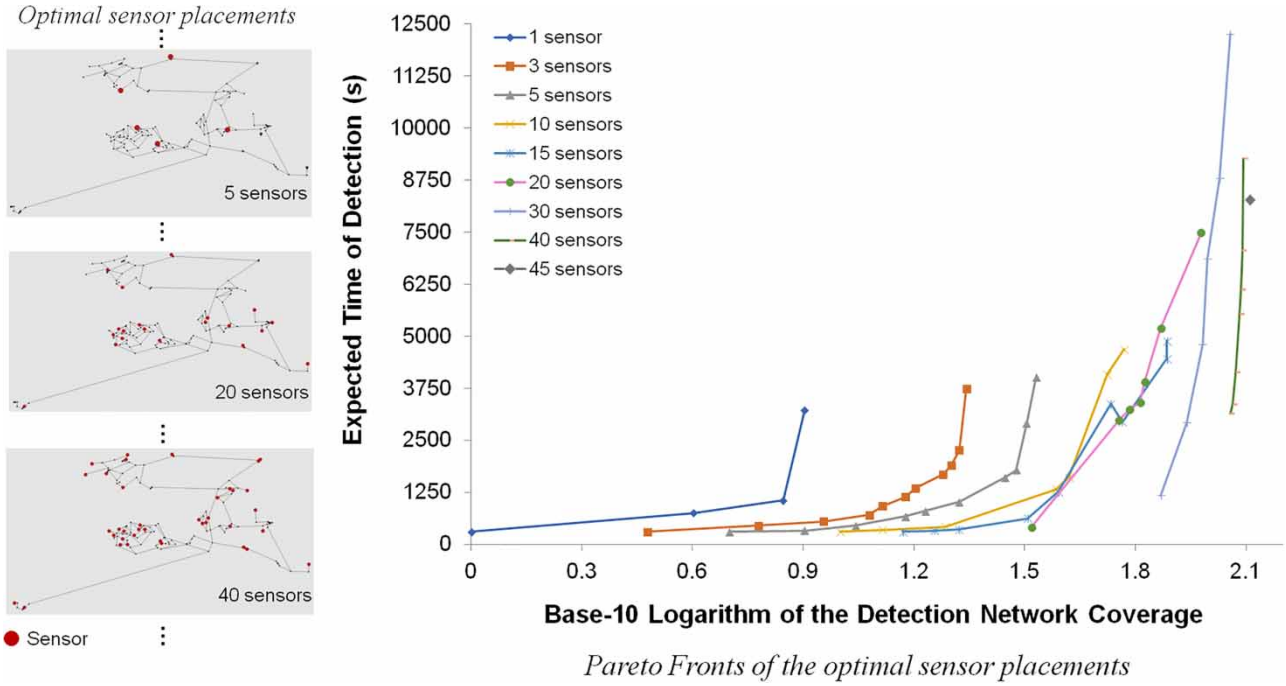
This paper introduces a multi-objective optimisation approach for the challenging problem of efficient sensor placement in water distribution networks for contamination detection. An important question is, how to identify the minimal number of required sensors without losing the capacity to monitor the system as a whole. In this study, we adapted the NSGA-II multi-objective optimisation method by applying centrality mutation. The approach, with two objectives, namely the minimisation of *Expected Time of Detection* and maximisation of *Detection Network Coverage* (which computes the number of detected water contamination events), is tested on a moderate-sized benchmark problem (129 nodes). The resulting Pareto front shows that detection network coverage can improve dramatically by deploying only a few sensors (e.g. increase from one sensor to three sensors). However, after reaching a certain number of sensors (e.g. 20 sensors), the effectiveness of further increasing the number of sensors is not apparent. Further, the results confirm that 40–45 sensors (i.e. 31 – 35% of the total number of nodes) will be sufficient for fully monitoring the benchmark network, i.e. for detection of any contaminant intrusion event no matter where it appears in the network.

Key words: centrality, contamination detection, early warning system, EPANET, optimisation, sensor, water distribution networks

HIGHLIGHTS

- It is possible to significantly reduce the number of undetected events by deploying only a few more sensors.
- Placing sensors on 31 – 35% of nodes is sufficient for full monitoring of the case study network.
- Maximising the opportunity to detect events prioritises the selection of nodes that neither have the highest centrality nor the lowest.
- Minimising the detection time of events prioritises nodes with centrality at/close to the extremes.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Water quality is essential to civilian daily life and therefore water distribution networks (WDNs) belong to the critical infrastructure of a region. In order to quickly and reliably detect contaminations, caused by leaks or malicious attacks, water quality sensors are installed at strategically important positions in the network. The optimisation of the sensor placement is of key importance when it comes to controlling the impact of contaminations. Yet the problem is difficult to solve, due to the high complexity and large size of city-level water distribution systems.

From a mathematical point of view, the placement of multiple sensors in WDNs is a combinatorial optimisation problem. The placement strategy aims to minimise the potential public health impact from any contamination intrusion with a limited number of sensors (Hart & Murray 2010). The challenge is to find an optimal subset of node locations for installing sensors in the WDN, which is represented as a set of interconnected nodes and edges. The subset selection problem faces a combinatorial explosion of possibilities with a growing number of nodes. It is certainly not always the best idea to merely place sensors at the most important locations, since the choice of any two sensors might be correlated. If one region is already covered by a sensor, another sensor might not have to be added to this region, even if the region as such is of high strategic importance. Another difficulty in the sensor placement problem is to consider the cost of sensors. It is a widely open question, i.e. how many sensors are required to monitor a network, and the related question is: how fast the effectiveness of the network monitoring decreases when fewer sensors are installed?

Thus far, three main approaches of sensor placement are described in the literature: empirical/empirically based methods (Bahadur *et al.* 2003; Berry *et al.* 2005; Ghimire & Barkdoll 2006; Trachtman 2006; Xu *et al.* 2008), topological methods (Deuerlein *et al.* 2010; Perelman & Ostfeld 2011; de Arruda et al. 2014; Di Nardo *et al.* 2018), and optimisation methods. Here, the empirically based methods refer to the ranking of potential sensor locations (Hart & Murray 2010) based on expert information (e.g. data from geographical information systems). The topological methods refer to using topology information of the WDNs to facilitate the identification of nodes suitable for sensor locations, e.g. by applying graph theory/complex network approaches or metrics (Santonastaso *et al.* 2021). Among these methods, optimisation methods are the most advocated ones today, given their capability to enable automated sensor placement based on hydraulic and water quality simulations. Thus, a sensor network that minimises contamination risks could be automatically planned using computational methods that perform an efficient search of potential solutions space.

In the last years, the problem of the optimal sensor location has been faced through single-objective (Lee & Deininger 1992; Kumar *et al.* 1997; Ostfeld & Salomons 2004; Propato & Piller 2006; Woo *et al.* 2006) and multi-objective (Dorini *et al.* 2006; Huang *et al.* 2006; Wu & Walski 2006; Preis & Ostfeld 2008; Yoo *et al.* 2015; He *et al.* 2018; Naserizade *et al.* 2018; Sankary & Ostfeld 2018; Brentan *et al.* 2021; Ponti *et al.* 2021; Gautam *et al.* 2022; Giudicianni *et al.* 2022; Shahra & Wu 2023) methodologies. Regarding methodologies, Ostfeld *et al.* (2008) compared 15 different approaches for optimal sensor placement in the battle of the water sensor networks (BWSN). The currently available optimisers are mainly based on integer programming (e.g. Lee & Deininger 1992), mixed-integer programming (e.g. Propato & Piller 2006; Zeng *et al.* 2018; Hooshmand *et al.* 2020), heuristic-based algorithms (e.g. Ghimire & Barkdoll 2006; Xu *et al.* 2013; Zhao *et al.* 2016), graph theory algorithms (e.g. Kessler *et al.* 1998; Diao & Rauch 2013; Nazempour *et al.* 2018; Taha *et al.* 2021), and genetic algorithm schemes (e.g. Ostfeld & Salomons 2004; Janke *et al.* 2012; Jafari *et al.* 2021). Although the optimisation methods are powerful tools for the placement of a fixed number of sensors, there are at least two research challenges for further exploration. The first challenge is to know the number of sensors required for satisfactory coverage of the network contamination detection, which is significantly under-researched. Shen & McBean (2011) assessed the impacts of changes in the number of sensors on the detection time of contaminant intrusion events and sensor detection redundancy. Optimal sensor placement solutions with increasing numbers of sensors (from 2 through 50) were obtained and analysed for the City of Guelph water distribution system that has 3,402 junctions. The results revealed that the performance improvement is the largest when increasing the number of sensors from 4 to 5 in the case of testing 2,912 intrusion events. However, this study does not reveal the most appropriate number of sensors for that WDN, and coverage of 1.5% (i.e. 50 sensors out of 3,402 nodes) is comparably small for fully revealing the sensor number-monitoring performance correlation (although the coverage rate is very reasonable since in practice it is restricted by the cost of sensors and their maintenance). Diao & Rauch (2013) applied controllability theory to identify the minimum number of sensors (30–40% of the total number of nodes) to ensure the detection of any contamination events no matter where it happens in the WDNs, i.e. fully monitoring. However, the resulting sensor layout is not guaranteed to be an optimal solution as the method is based on analysing directed graphs of the WDN without employing optimisation methods. Further, since 30–40% coverage is a tremendously high value, whether the method overestimated the needed number of sensors still needs verification. Hence, to answer the questions left from these two studies, it is worthwhile to work out optimal sensor placement solutions with increasing coverage rates, e.g. from a very small percentage to a percentage sufficient for fully monitoring of WDN. Such an extensive set of results shall reveal the trade-off between the number of sensors and the effectiveness of the monitoring. To obtain the results efficiently and also ensure the quality of the results, the optimisation method, which is usually computationally expensive, needs to be improved too.

The second challenge is the computational cost. There have been continuous efforts to improve the computational efficiency and quality of results for the optimisation of sensor placement. Several studies considered potential variations of nodal contamination probabilities in their multi-objective sensor placement (He *et al.* 2018; Naserizade *et al.* 2018; Cardoso *et al.* 2020; Hu *et al.* 2021; Ponti *et al.* 2021). Compared with traditional genetic algorithms, the improved method can increase the computational efficiency by approximately 10,000 times and also the detectability of contamination events by 2.6 times in achieved design solutions (He *et al.* 2018). Reduction of the nodal search space is another solution, i.e. selection of a subset of nodes as candidates for placing sensors. For instance, Giudicianni *et al.* (2022) restrained optimal sensor placement to nodes on hydraulic/topological-wise most important pipes and further narrowed the nodal search space by incorporating logistic/economic criteria. Gautam *et al.* (2022) reduced nodal search space by 34 and 45%, respectively, for two benchmarking WDNs by applying k-means clustering to select a subset of nodes prior to optimisation. Another alternative solution is to avoid the recourse to hydraulic modelling, since water quality simulation can be time-consuming. In this regard, techniques for topological analysis (e.g. graph theory/complex network analyses) are mostly used. Yoo *et al.* (2015) applied Betweenness Centrality that generates optimal sensor locations based on WDN's connectivity and found that the solution tends to select nodes close to water sources and water mains. Giudicianni *et al.* (2020) made a clustering of the WDN and next employed different topological centrality metrics to identify the most central nodes of each cluster for quality sensor placement. These studies significantly advanced the field, yet most of them focus on the development of preliminary analysis to assist optimisation or replace optimisation with other techniques. Hence, there is still a lack of exploration on how to improve the optimisation algorithms, e.g. by integrating cutting-edge techniques from graph theory/complex network analysis. According to the two challenges above, this paper aims to explore the trade-off between the number of sensors and the effectiveness of the monitoring by using a wide range of sensor coverage rates. The work is

done by applying a novel centrality-guided multi-objective genetic algorithm (CG-MOGA), which integrates the centrality metric into the mutation operator to guide the optimisation towards interesting areas of the search space, yet, in an unbiased way.. This technique combines principles from general-purpose metaheuristics for multi-objective optimisation, i.e. the non-dominated sorting genetic algorithm (NSGA-II) (Deb *et al.* 2002), with the tailored mutation operator. The centrality-guided mutation is particularly designed for node-subset selection problems in complex networks and has previously been used successfully for the multi-node immunisation of complex networks (Maulana 2017). Here it is used for the first time in the context of WDN sensor placement.

For the following sections, the problem formulation for the multi-objective sensor placement will be introduced, followed by the application of the CG-MOGA on a moderate-sized WDN model. Next, the results are discussed highlighting the key findings and conjectures. Finally, we summarise our main conclusions and propose directions for extended work and future research.

2. METHOD

Overall, the methodology for the optimal sensor placement is to use multi-objective optimisation to minimise the expected time of detection of contaminant events (F_1 in Equation (1)) and maximise the detection network coverage (F_2 in Equation (2)). Further, to reduce the number of decision variables for the reduction of the computational burden, a subset of nodes is selected as sensor candidates based on using centrality (see section 2.2) in the mutation operator of the optimisation algorithm.

2.1. Problem formulation

The sensor placement problem is formulated as a multi-objective optimisation as described by the following equations.

$$\text{minimise } F_1 = \frac{\sum_{i=1}^{m} t_{\mathrm{d},i}}{m}, \quad t_{\mathrm{d}} \in [0, t_{\mathrm{s}}] \tag{1}$$

$$\text{maximise } F_2 = \sum_{i=1}^{m} S_i \tag{2}$$

where t_{d} is the detection time of contamination event i, $t_{\mathrm{d}} = \min_j t_j,\ j = [1, n]$, where t_j denotes the first time of detection at the jth sensor location, and n denotes the total number of sensor candidates; t_{s} denotes the simulation duration; Please note that if $t_{\mathrm{d}} > t_{\mathrm{s}}$, the contamination event is regarded as undetected. m is the total number of contamination scenarios. In Equation (2), $S = 1$ if a contamination event is detected at any sensor and otherwise 0; thus, $\sum_{i=1}^{m} S$ denotes the total number of detected contamination events, named as the detection network coverage.

2.2. Centrality-guided multi-objective optimisation

We adopted the NSGA-II (Maulana *et al.* 2017) algorithm to execute the multi-objective optimisation problem formulation but made an improvement in the algorithm's mutation operator in order to elevate the intensity of search in parts of the network that likely impacts the system's performance. A general guideline in genetic algorithms is that the mutation operator should be unbiased. When searching large search spaces (with many possible locations for placing sensors), it might be impossible to find interesting solutions merely by chance. However, intensification of the search in relevant regions without biasing the mutation operator can be achieved by the proposed *centrality-guided mutation*.

We represent the sensor selection by means of a bit-vector x with $x_i = 1$ if node v_i is selected. i.e. a sensor is placed at this location, and $x_i = 0$, otherwise. The genetic algorithm maintains a population of individuals. Non-dominated sorting is used in combination with crowding distance for each layer of equal dominance rank in the selection (Maulana *et al.* 2017). For the mutation operator, each node v_i in the network is assigned an individual mutation probability p_i. This value is proportional to its centrality in the network. Centrality is a metric in graph theory and network analysis, which gives an estimation on how important a node or edge is for the connectivity or the information flow of the network (EMBL-EBI Training 2022). Here we consider the eigenvector centrality as a non-local centrality measure. It is non-local and thus the importance of a node for the network is not only measured based on its direct neighbours but with respect to its influence on the entire network. The eigenvector centrality of a network can be computed by standard linear algebra operations. Let A denote the adjacency matrix of

the network, and let $(\lambda_{\max}, u_{\max}) = \arg\max_{\lambda,u}\{\lambda | Au = \lambda u\}$ denotes the principal eigenvalue and eigenvector of the network. Then, the eigenvector centrality is given by the components u_i, $i = 1, \ldots, m$ of $u_{\max}$. The resolving process starts with an approximation or a random vector u^0, followed by an iteration of $u^{k+1} = Au^k$ and $u^{k+1} = u^{k+1}/m(u^{k+1})$ until the desired precision is achieved, i.e. u^k converges to the dominant eigenvector of A and $m(u^{k+1})$ converges to the dominant eigenvalue of A (Collins *et al.* 2021). The $m(u^{k+1})$ denotes the signed component of the maximal magnitude of vector u^{k+1}, which is introduced for normalisation of u^k to avoid data overflow. A good example of the computation process can be found in Meghanathan (2015). Since the matrix A of WDN is sparse, each vector-matrix product can be performed in linear time in the size of the graph.

The described mutation operator intensifies the search in more promising regions proportional to the centrality of nodes while still maintaining the possibility to access all points in the search space by means of mutation. More specifically, when the mutation process in the genetic algorithm happens, different weights are assigned for each potential sensor candidate according to their eigen-centrality value. The sensor candidates with higher weights are more likely to be selected in the mutation process. The process of mutation is as follows:

1. Let $x_i = 1$ if v_i is selected and $x_i = 0$ otherwise.
2. Let $s := \sum (1 - x_i)u_i$ // Sum of centralities of non-selected nodes
3. Let $p_i = (1 - x_i)u_i/s$ // only non-selected nodes have a positive probability
4. Choose sensor node v_i proportional to p_i
5. Change state of sensor node v_i to selected by setting $x_i = 1$
6. Set $x_i = 0$ for randomly chosen node v_i with $x_i = 1$ // Makes sure that the total number of selected nodes remains constant (that is equal to n)

For details of the centrality mutation, the reader is referred to Lan *et al.* (2018).

The optimisation results in Pareto fronts for different numbers of installed sensors, while the objectives are maximising network coverage and minimising the time of contamination event detection.

3. CASE STUDY

The case study focuses on the battle of the water sensors network1 (BWSN1) (Ostfeld *et al.* 2008). The system (Figure 1), composed of 126 junctions, 168 pipes, 1 constant head source, 2 tanks, 2 pumps, and 8 valves, is subject to a varying demand pattern of 96 h.

Thus, the simulation period is 96 h, with time steps 1 h for hydraulic and water quality equal to 5 min. Once the contaminant reaches any node (i.e. the concentration of the contaminant at the node is >0), the event is regarded as being detected, and the time, until detection is recorded as t_j, and used to compute objective functions. Otherwise, the event would be labelled as undetected. These steps are repeated until all nodes have been considered as the contamination source (i.e. 129 contamination events in total). This process is a simplification given the fact that the precision of the sensors do have limitations as contaminants with very low concentration may not be detectable. However, this is out of the scope of this study yet an important question for future research. The nodal injection of the contaminant is simulated by imposing a mass booster source (set as 3,000 mg/L) for the duration of 2 h from the beginning of the simulation period. Regarding the dynamics of the pollutant, the nonlinear dynamics of the contaminant in the network are considered. Accordingly, the settings in the EPANET software are chosen as bulk reaction order = 1.5; global bulk coefficient = −1.0; limiting concentration = 0.01 (using this parameter ensures that the contaminant would not vanish in the system). Additionally, the dead-end nodes (the nodes at the downstream end of a branch pipe, i.e. here nodes 13, 16, 36, 38, and 125) are assigned a base demand of 2.0 gallons/min (0.45 m^3/h).

Regarding coverage, if based on the evaluation of a node a contamination event is detected, this incidence is marked. After the entire set of contamination events has been considered as the contamination source, we check the number of marked events in each node as the detection network coverage of these potential sensors. In total, there are 45 nodes which are selected as candidates for placing sensors. A full list of the candidate nodes is provided in Table 1. In the table, the longest detection time is 87,900 s. This is roughly 24.4 h which is smaller than 96 h (the simulation period). Therefore, data from the simulation are all valid for optimisation.

Since there is a lack of guidance on sensor numbers, we proposed a clear comparison over a different number of sensors' capability in dealing with the two objective functions defined in Section 2.1. Therefore, this case study needs to work through

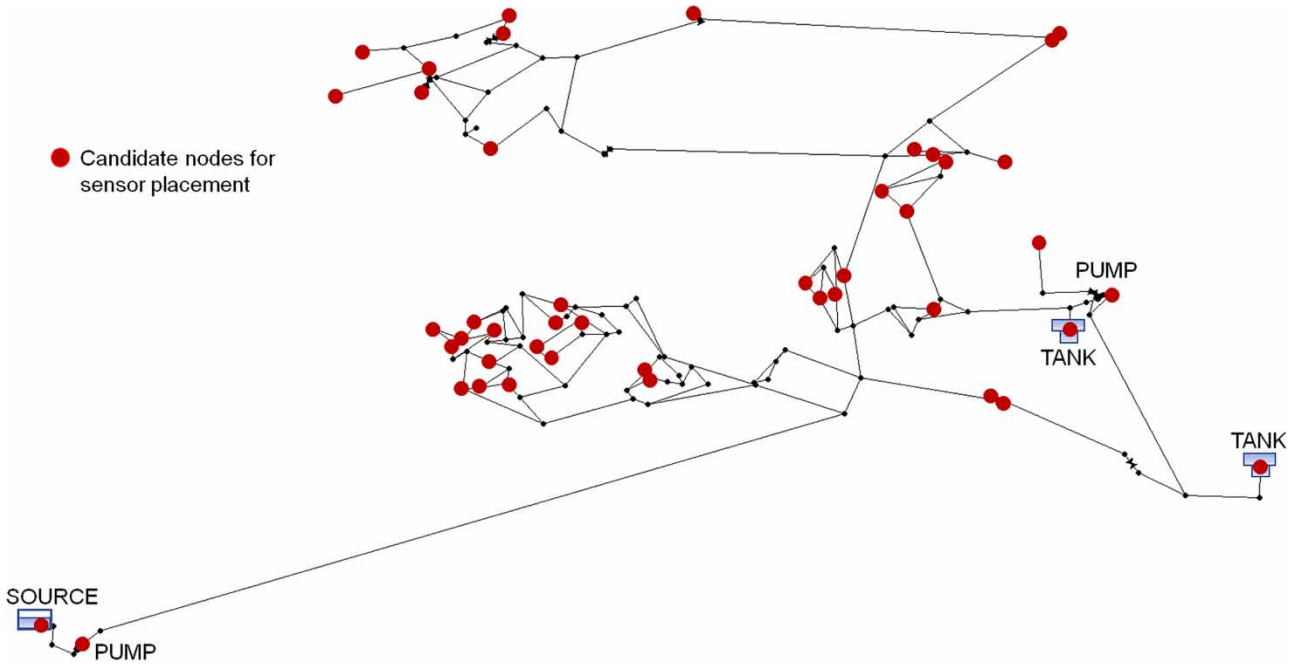


Figure 1 | Layout of BWSN network 1 (Ostfeld *et al.* 2008).

nine experiments with the improved algorithms. The number of sensors suggested for each experiment is 1, 3, 5, 10, 15, 20, 30, 40, 45.

For the optimisation based on NSGA-II, the parameter settings are population size = 90, mutation rate = 5%, crossover rate = 30%, and generation number = 3,000. These parameter values are decided based on pre-testing. For example, the mutation rate is kept relatively low to avoid unnecessary delay in reaching convergence. In addition, the previous Pareto front is used to replace the 'random' initialised population for new optimisations. For instance, for optimisation of the five-sensor case, the Pareto front from three-sensor optimisation is used to generate the initial population.

4. RESULTS AND DISCUSSIONS

Figure 2 shows the comparison of all Pareto Fronts obtained using a different number of sensors. To avoid the Pareto fronts being very close to each other, the Base-10 logarithm of the Detection Network Coverage is used for the horizontal axis for better illustration. It can be seen that the logarithm of detection network coverage increases dramatically (i.e. significant improvement of detection network coverage) by deploying only a few more sensors, e.g. increase from one sensor to three sensors. This observation is similar to the study from (Shen & McBean 2011), which reveals the Pareto front performance improvement is the largest when increasing the number of sensors from 4 to 5 in their case study testing 2,912 intrusion events. After reaching a certain number of sensors (e.g. 15 sensors), the effectiveness of further increasing the number of sensors is not apparent, as the Pareto fronts are very close to each other while the detection time fluctuates.

As Figure 2 illustrates, the logarithm of detection network coverage increases dramatically from around 0.9 (1 sensor) to 2.06 (30 sensors), denoting an increase in the number of detected events from 8 to 114, with the expected time of detection fluctuating in a range of approximately. 3,500 to 12,000 s. After reaching 40 sensors, both the logarithm of detection network coverage and expected time of detection have only negligible differences, i.e. from about 2.09 (i.e. 124 detected events, 40 sensors) to about 2.11 (129 detected events, 45 sensors) for the former, and from about 9,500 s (40 sensors) to 8,300 s (45 sensors) (about 20 min difference) for later. This observation confirms that 40–45 sensors (i.e. 31 – 35% of the total number of nodes) will be sufficient for fully monitoring the benchmark network. This result is consistent with that from Diao & Rauch (2013), who conclude that the full monitoring of the same network requires sensor placement on 36.4% of nodes. Hence, this research verifies that solving optimal sensor placement problems, even when two objectives are considered, results in a similar minimum number of required sensors to the graph-based method (Diao & Rauch 2013). It is

Table 1 | Full list of the 45 candidate nodes for sensor placement

Sensor candidate node	Frequency (coverage)	Detection time (s)	Eigenvector centrality
26	8	3,225	0.389607
118	7	6,900	0.264835
64	7	1,056	0.074546
91	6	6,350	0.144328
8	6	2,500	0.123821
80	5	10,800	0.374543
106	5	1,380	0.232671
101	5	4,440	0.199984
112	5	9,360	0.075092
76	4	41,775	0.655913
21	4	2,475	0.606180
130	4	3,900	0.301279
73	4	16,575	0.206232
131	4	1,600	0.091494
45	4	7,800	0.082226
123	4	10,650	0.082226
110	4	750	0.025722
36	3	62,100	0.672393
84	3	1,900	0.657461
85	3	12,700	0.425350
48	3	3,500	0.184193
39	2	10,800	0.438871
52	2	1,500	0.415769
10	2	87,900	0.328839
72	2	7,350	0.185071
93	2	1,800	0.136205
66	2	7,950	0.074522
13	2	37,950	0.000000
83	1	300	1.000000
37	1	300	0.752721
38	1	300	0.672393
82	1	300	0.646136
126	1	300	0.445138
114	1	300	0.419577
74	1	300	0.342580
100	1	300	0.337818
125	1	300	0.309490
124	1	300	0.290348
42	1	300	0.251222
99	1	300	0.199984
41	1	300	0.198965
50	1	300	0.184193

(*Continued.*)

Table 1 | Continued

Sensor candidate node	Frequency (coverage)	Detection time (s)	Eigenvector centrality
14	1	300	0.075469
16	1	300	0.000000
129	1	300	0.000000

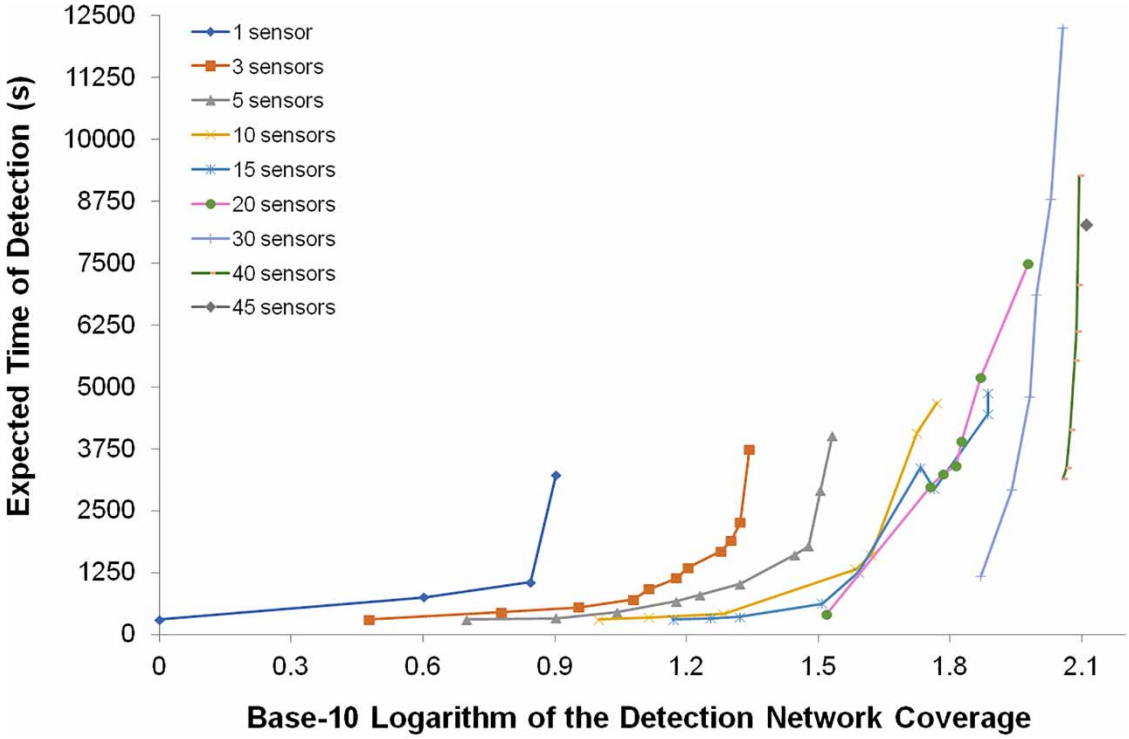


Figure 2 | All Pareto Fronts for different sensors number(s) (Sensor numbers:1, 3, 5, 10, 15, 20, 30, 40, 45).

thus safe to use the graph-based method to quickly identify the required minimum number of sensors without employing a computationally expensive optimisation process. However, future research is still needed to systematically compare the optimal solutions with graph-based solutions in order to check the possibility of replacing optimal sensor placement completely with the graph-based method.

To further understand the sensor node selection, analysis of the design solutions reveals that maximising detection network coverage reasonably results in the selection of nodes with high coverage (see as Table 1 for the values of the coverage), and minimising the expected time of detection leads to the selection of nodes with short detection times (see as Table 1), which are reasonable too. In terms of centrality, the nodes with high coverage (i.e. the ability to detect more contamination events) are the nodes that neither have the highest eigenvector centrality nor the lowest eigenvector centrality, as shown in Figure 3 – the left column for instance. Figure 4 plots the eigenvector centrality of all sensor candidate nodes and sensors in each set of solutions. Contrarily, the nodes with short detection times have a wide range of eigenvector centrality values including both the highest (i.e. 1) and lowest (i.e. 0). Interestingly, minimisation of the expected time of detection tends to select nodes with their eigenvector centrality at/close to the extremes (i.e. the highest and lowest value), as shown in Figure 4 – the right column.

However, one common fact for both cases (maximising the Detection Network Coverage or minimising the expected time of detection) is that, for any number of sensors, the solutions select sensor nodes with a wide range of centralities, rather than only focus on selecting high/low centrality nodes. Figure 4 shows an example of the physical locations of the placed sensors.

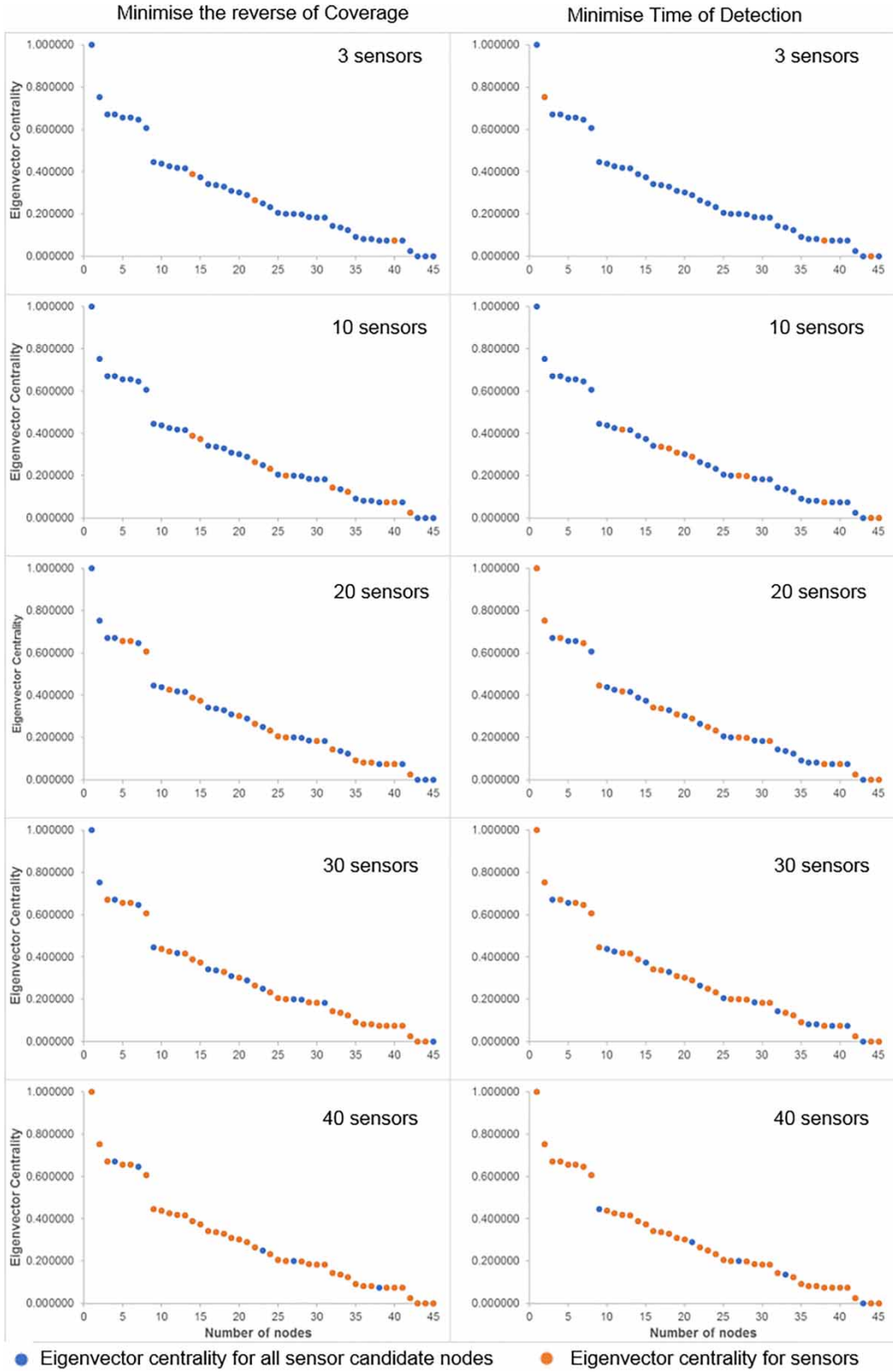


Figure 3 | Eigenvector centrality of sensors in each set of solutions and all sensor candidate nodes.

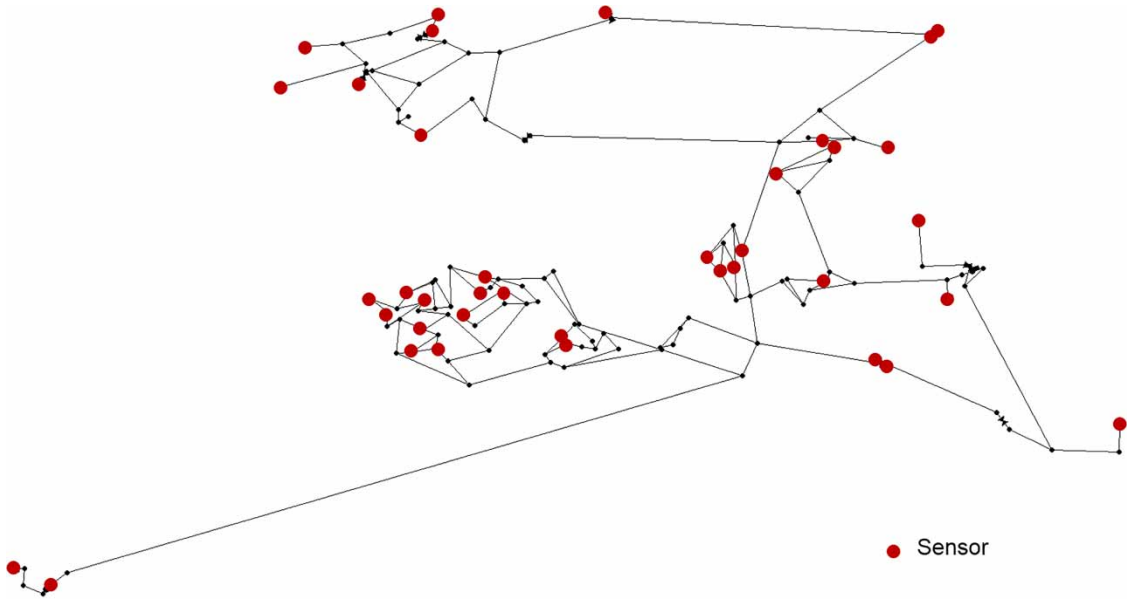


Figure 4 | Sensor layout of the 40 sensors solution with network detection coverage maximised.

It can be seen that the sensors are spread over the network, which covers the highly looped areas in the centre of the network and many dead ends. More sensors are deployed in the highly looped regions due to the more complicated topology and flow patterns (e.g. bi-directional flows that change over time), yet fewer sensors are required for branched pipelines. The wide range of the centrality values may result from the diversified roles and positions of the selected sensor nodes, which however definitely need future research to further explore the correlations.

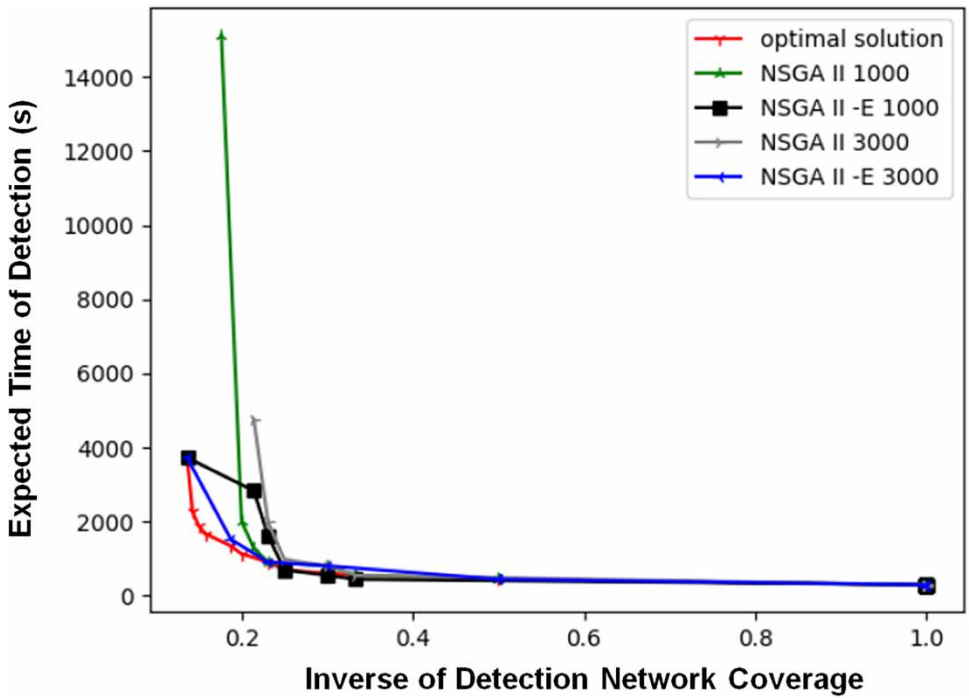


Figure 5 | Comparison between classic NSGA II and the centrality-guided NSGA II (NSGA-II-CG).

Lastly, a comparison between the centrality-guided NSGA II (NSGA-II-CG) and classic NSGA II is made. For benchmarking, the optimal solution of a three-sensor design is generated by using the 45 nodes in Table 1 as the candidate pool and enumerating all possible combinations of three-sensor solutions ($C(45,\ 3) = 14,190$). Next, the NSGA-II and NSGA-II-CG are implemented for the same optimisation, respectively. It is found that the centrality-guided algorithm does not increase the computational load and also leads to solutions that are much closer to the optimal solution (Figure 5). Although increasing the Generation Number (e.g. from 1,000 to 3,000) helps to improve solutions from both the NSGA-II and NSGA-II-CG methods, the improvement from the classic NSGA-II is not as significant as the NSGA-II-CG, reflected by the distance to the optimal solution.

5. CONCLUSIONS

This study proposed a novel centrality-guided multi-objective optimisation method for optimal sensor placement in WDNs. The method uses two objectives, minimising the detection time of contaminant events and maximising the number of detectable events, respectively. Based on a case study on the BWSN1 network, the following conclusions are reached:

- It is possible to significantly reduce the number of undetected events by deploying only a few more sensors, e.g. to increase the number of sensors from one sensor to three sensors.
- Placing sensors on 31 – 35% of the total number of nodes will be sufficient for full monitoring of the case study network. This result is consistent with the graph-based method. Further, this study proves that any further increase in the number of sensors will have marginal effects on both objectives.
- The centrality-guided multi-objective optimisation method reveals that maximising the opportunity to detect contamination events prioritises the selection of nodes that neither have the highest eigenvector centrality nor the lowest eigenvector centrality; yet minimising detection time of contamination events prioritises the selection of nodes with their eigenvector centrality at/close to the extremes (i.e. the highest and lowest value).

ACKNOWLEDGEMENTS

The contribution of the University of Innsbruck was supported by the project RESIST (FO999886338) which is funded by the Austrian security research programme KIRAS of the Federal Ministry of Agriculture, Regions and Tourism (BMLRT).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Bahadur, R., Samuels, W. B., Grayman, W., Amstutz, D. & Pickus, J. 2003 PipelineNet: A model for monitoring introduced contaminants in a distribution system. In: *Proceedings of the World Water and Environmental Resources Congress 2003 and Related Symposium*, Reston, VA, ASCE.

Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G. & Walski, T. M. 2005 Water quality sensor placement in water networks with budget constraints. In: *Proceedings of the World Water and Environment Resources Conference*, Reston, VA, ASCE.

Brentan, B., Carpitella, S., Barros, D., Meirelles, G., Certa, A. & Izquierdo, J. 2021 Water quality sensor placement: A multi-objective and multi-criteria approach. *Water Resources Management* **35**, 225–241.

Cardoso, S. M., Barros, D. B., Oliveira, E., Brentan, B. & Ribeiro, L. 2020 Optimal sensor placement for contamination detection: A multi-objective and probabilistic approach. *Environmental Modelling & Software* **135**, 104896.

Collins, A., Engströmb, C., Magero, J., Kasumbaa, H., Silvestrov, S. & Abola, B. 2021 Eigenvector centrality and uniform dominant eigenvalue of graph components. *arXiv:2107.09137*.

de Arruda, G. F., Barbieri, A. L., Rodríguez, P. M., Rodrigues, F. A., Moreno, Y. & Costa, L. F. 2014 Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E*. doi:10.1103/PhysRevE.90.032812.

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.

Deuerlein, J. W., Wolters, A., Meyer-Harries, L. & Simpson, A. R. 2010 Graph decomposition in risk analysis and sensor placement for water distribution network security. In: *Water Distribution System Analysis 2010 – WDSA2010*, September 12–15, 2010, Tucson, AZ, USA.

Diao, K. & Rauch, W. 2013 Controllability analysis as a Pre-Selection method for sensor placement in water distribution systems. *Water Research* **47** (16), 6097–6108.

Di Nardo, A., Giudicianni, C., Greco, R., Herrera, M. & Santonastaso, G. F. 2018 Applications of graph spectral techniques to water distribution network management. *Water* **10** (1).

Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S. T. & Savic, D. 2006 An efficient algorithm for sensor placement in water distribution systems. In: *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*, Reston, VA, ASCE.

EMBL-EBI Training 2022 *Centrality Analysis in Network Analysis of Protein Interaction Data*. Available from: https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/building-and-analysing-ppins/topological-ppin-analysis/centrality-analysis/ (accessed 8 March 2023).

Gautam, D. K., Kotecha, P. & Subbiah, S. 2022 Efficient k-means clustering and greedy selection-based reduction of nodal search space for optimization of sensor placement in the water distribution networks. *Water Research* **220**, 118666.

Ghimire, S. R. & Barkdoll, B. D. 2006 A heuristic method for water quality sensor location in a municipal water distribution system: Mass-released based approach. In: *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*, Reston, VA, ASCE.

Giudicianni, C., Herrera, M., Di Nardo, A., Greco, R., Creaco, E. & Scala, A. 2020 Topological placement of quality sensors in water-distribution networks without the recourse to hydraulic modelling. *Journal of Water Resources Planning and Management* **146** (6), 04020030.

Giudicianni, C., Herrera, M., Di Nardo, A., Creaco, E. & Greco, R. 2022 Multi-criteria method for the realistic placement of water quality sensors on pipes of water distribution systems. *Environmental Modelling & Software* **152**, 105405.

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.

Hooshmand, F., Amerehi, F. & MirHassani, S. A. 2020 Logic-based benders decomposition algorithm for contamination detection problem in water networks. *Computers & Operations Research* **115**, 104840.

Hu, Z., Chen, W., Shen, D., Chen, B., Ye, S. & Tan, D. 2021 Optimal sensor placement for contamination identification in water distribution system considering contamination probability variations. *Computers & Chemical Engineering* 107404.

Huang, J., McBean, E. A. & James, W. 2006 Multiobjective optimisation for monitoring sensor placement in water distribution systems. In: *8th Annual Water Distribution System Analysis Symposium*, August 27–30, 2006, University of Cincinnatti, Cincinnati, Ohio USA.

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 of important nodes contamination. *Water Supply* **22** (1), 928–944.

Janke, R., Murray, R., Haxton, T. M., Taxon, T., Bahadur, R., Samuels, W. & Uber, J. 2012 Threat ensemble vulnerability assessment-sensor placement optimization tool (TEVA-SPOT) graphical user interface user's manual. *US EPA National Homeland Security Research Center (NHSRC)* 1–109.

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.

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

Lan, J., Emmerich, M. & Diao, K. 2018 *Critical Water Infrastructure Sensor Placement Optimisation. Master Thesis*, LIACS, Leiden University, The Netherlands.

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

Maulana, A., Kefalas, M. & Emmerich, M. T. 2017 Immunization of networks using genetic algorithms and multiobjective metaheuristics. In: *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, November 27 to December 1, 2017, Honolulu, Hawaii, IEEE, pp. 1–8.

Meghanathan, N. 2015 Use of Eigenvector Centrality to Detect Graph Isomorphism. In: *Proceedings of the Fourth International Conference on Advanced Information Technologies and Applications (ICAITA)*, November 6–7, Dubai, UAE, pp. 1–9.

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.

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., Uber, J., Salomons, E., Berry, J., Hart, W., Phillips, C., Watson, J., Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S., Barkdoll, B., Gueli, R., Huang, J., McBean, E., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G., Wu, Z. & 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. & Ostfeld, A. 2011 Topological clustering for water distribution systems analysis. *Environmental Modelling & Software* **26** (7), 969–972.

Ponti, A., Candelieri, A. & Archetti, F. 2021 A new evolutionary approach to optimal sensor placement in water distribution networks. *Water* **13**, 1625.

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

Propato, M. & Piller, O. 2006 Battle of the water sensor networks. In: *8th Annual Water Distribution System Analysis Symposium*, August 27–30, 2006. University of Cincinnatti, Cincinnati, Ohio, USA.

Sankary, N. & Ostfeld, A. 2018 Analyzing multi-variate water quality signals for water quality monitoring station placement in water distribution systems. *Journal of Hydroinformatics* **20** (6), 1323–1342.

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

Shahra, E. Q. & Wu, W. 2023 Water contaminants detection using sensor placement approach in smart water networks. *Journal of Ambient Intelligence and Humanized Computing* **14**, 4971–4986.

Shen, H. & McBean, E. 2011 Diminishing marginal returns for sensor networks in a water distribution system. *Journal of Water Supply: Research and Technology-Aqua* **60** (5), 286–293.

Taha, A. F., Wang, S., Guo, Y., Summers, T. H., Gatsis, N., Giacomoni, M. H. & Abokifa, A. A. 2021 Revisiting the water quality sensor placement problem: Optimizing network observability and state estimation metrics. *Journal of Water Resources Planning and Management* **147** (7), 04021040.

Trachtman, G. 2006 A 'strawman' common sense approach for waterquality sensor site selection. In: *Proceedings of 8th Annual Water Distribution Systems Analysis Symposium*, Reston, VA, ASCE.

Woo, H. M., Yoon, J. H. & Choi, D. Y. 2006 Optimal monitoring sites based on water quality and quantity in water distribution systems. In: *World Water and Environmental Resources Congress, CD-ROM*, Reston, VA, ASCE.

Wu, Z. Y. & Walski, T. 2006 Multiobjective optimisation of sensor placement in water distribution systems. In: *8th Annual Water Distribution System Analysis Symposium*, August 27–30, 2006. University of Cincinnatti, Cincinnati, Ohio, USA.

Xu, J. H., Fischbeck, P. S., Small, M. J., VanBriesen, J. M. & Casman, E. 2008 Identifying sets of key nodes for placing sensors in dynamic water distribution networks. *Journal of Water Resources Planning and Management* **134** (4), 378–385.

Xu, X., Lu, Y., Huang, S., Xiao, Y. & Wang, W. 2013 Incremental sensor placement optimization on water network. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, September 2013. Springer, Berlin, Heidelberg, pp. 467–482.

Yoo, D. G., Chung, G., Sadollah, A. & Kim, J. H. 2015 Applications of network analysis and multi-objective genetic algorithm for selecting optimal water quality sensor locations in water distribution networks. *KSCE Journal of Civil Engineering* **19**, 2333–2344.

Zeng, D., Zhang, S., Gu, L., Yu, S. & Fu, Z. 2018 Quality-of-sensing aware budget constrained contaminant detection sensor deployment in water distribution system. *Journal of Network and Computer Applications* **103**, 274–279.

Zhao, Y., Schwartz, R., Salomons, E., Ostfeld, A. & Poor, H. V. 2016 New formulation and optimization methods for water sensor placement. *Environmental Modelling & Software* **76**, 128–136.

First received 11 March 2023; accepted in revised form 29 September 2023. Available online 24 October 2023