Efficient k-means clustering and greedy selection-based reduction of nodal search space for optimization of sensor placement in the water distribution networks

Dinesh Kumar Gautam, Prakash Kotecha, Senthilmurugan Subbiah

Efficient k-means clustering and greedy selection-based reduction of nodal search space for optimization of sensor placement in the water distribution networks.

Dinesh Kumar Gautam, Prakash Kotecha, Senthilmurugan Subbiah

Department of Chemical Engineering, Indian Institute of Technology Guwahati, 781039,

Assam, India

Abstract

Monitoring of water distribution network (WDN) requires placement of sensors at strategic locations to detect maximum contamination events at the earliest. The multi-objective optimization (MOO) of sensor placement is a complicated problem owing to its combinatorial nature, interconnected and large WDN sizes, and temporal flows producing complex outcomes for a given set of contamination events. In this study, a new method is proposed to reduce the complexity of the problem by condensing the nodal search space. This method first segregates the nodes based on intrusion events detected using k-means clustering followed by selecting nodes from each group based on the improvement observed in the objectives, namely, contamination event detection, expected detection time, and affected population. The selected nodes formed the decision variable space for the MOO study. The developed strategy was tested on two benchmark networks: BWSN Network1 and C-town network, and its performance is compared with the traditional method in terms of hypervolume contribution rate (CR) indicator and the number of Pareto points. The optimal subset of nodes generated twice the number of Pareto points than the complete set of nodes set for placing 20 sensors and had 10% more than CR indicator than the traditional method. For the placement of 5 sensors, the proposed solutions were better at the higher detection likelihood values, which is required to achieve maximum detection. The proposed sensor placement algorithm can be easily scaled

to large WDNs. It is expected to provide a better optimal sensor placement solution irrespective of network size as compared to the traditional approach.

Key Words: sensor placement, water distribution network, multiobjective optimization, optimal sensor placement, BWSN, C-town

1 Introduction

The advancements in sensors for water quality measurements have enabled water supply assurance and steadily shifted the focus towards monitoring and maintaining the quality of supplied water in water distribution networks (WDN) (Storey et al., 2010). Monitoring of WDNs has been widely researched with the application of sensors in the network, concentrating on collecting crucial supply data to facilitate swift mitigation procedures (Creaco et al., 2019; Nurani et al., 2018). The placement of the sensors at niche locations is essential due to budgetary constraints, thus allowing only a limited number of sensors to be placed in a vast network (Zeng et al., 2018). This led to research in various sensor placement strategies aimed at monitoring of WDNs with a limited number of sensors. The main issue for placing sensors in a WDN arises in selecting the best solution from a vast number of feasible solutions available to utility managers. The efficiency of the 'best solution' has to be assessed and compared with other available solutions, and this 'efficiency' can be estimated based on the required goals or objectives for placing the sensors. Though the sensor placement problem (SPP) can be constructed as an optimization problem, finding the optimal solution is computationally expensive when deterministic methods are employed while stochastic approaches result in sub-optimal solutions (Hu et al., 2018a; Xu et al., 2013). The intricacies of optimizing the sensor placement in a WDN are further explained in the following paragraphs.

The main components of the sensor placement problem that have been extensively researched are: objectives, optimization algorithm, and decision variables. With the network flow configuration and nodal demands of a WDN, two crucial information can be estimated: flow and contaminant concentration at any node and the time taken for the contaminant to travel in the network. Based on this data, The Battle of Water Sensor Networks (BWSN) (Hart et al., 2007) design challenge was conducted to compare various approaches to solving SPP for

WDNs. The problem was formulated with four objectives: minimization of expected detection time, minimization of the expected population affected, minimization of expected amount of contaminated consumed, and maximization of detection likelihood. The study revealed that the first three objectives, 'affected population,' 'contaminated water consumed' and 'detection time,' conflicted with detection likelihood. Since then, many researchers have discussed the Multi-objective Optimization of SPP (MOSPP) for WDN, focusing on detecting events and either of the other three objectives (Hu et al., 2018b). The effect of contamination in WDN can be captured in pollution matrices, and it can be developed by simulating a series of contamination scenarios in EPANET 2.0 (https://www.epa.gov/water-research/epanet) (Rosmann, 2000) and recording the response from the network compactly in a matrix. The pollution matrices have number of rows equal to the number of nodes in the network and columns equal to the number of contamination events that are simulated. Each value represents the information like contamination detection, time taken for detection if contamination detected etc. by the corresponding node (row) for the corresponding event (column). So, each row provides values for the estimation of objective function if a sensor is placed at the corresponding node. Thus, by adopting similar procedure for all the nodes (rows), the values required for estimation of the objective function for each node can be stored efficiently. This step will reduce computation cost as during optimization, pollution matrix is utilized to estimate the objective function rather than simulating all the contamination events for estimating the objective function for each sensor solution. The recorded information in pollution matrices depicted contamination presence, time of detection, water consumed before detection, or the number of persons affected before detection, etc. (Chastain, 2006; Hart et al., 2007).

Various optimization algorithms have been implemented to solve the SPP and MOSPP, such as genetic algorithm (GA) and its variants (Eliades and Polycarpou, 2006; Hu et al., 2018a;

Preis and Ostfeld, 2007), greedy randomized adaptive search procedures (GRASP) techniques (J. Berry et al., 2006; Krause et al., 2008), heuristics (Aral et al., 2010; Guan et al., 2006), Particle swarm optimization (PSO) (Hu et al., 2015; Marlim and Kang, 2021) and constrained mixed-integer programming (MIP) formulations (J. W. Berry et al., 2006). Also, The TEVA-SPOT software tool was developed by the U.S. EPA to evaluate the contamination risks and design sensor monitoring stations for WDN. TEVA-SPOT allows the users to choose an objective among the following: a) number of failed contaminated detections, b) time of detection, c) estimated population exposed, d) estimation population contaminated, e) extent of contamination, f) mass of contaminated water consumed, g) length of pipe contaminated and h) volume of water contaminated. The objectives c, d, f, and h are correlated with the time of detection, while objectives e and g are correlated with objective a. It can be seen that minimization of objective a is complementary to the maximization of objective 'number (percentage) of successful detections', which is analogous to the detection likelihood objective considered in this study. However, TEVA-SPOT carries out single objective optimization of the above mentioned objectives and the conflicting objectives have to be provided as a constraint (Janke, 2018). Stochastic techniques generate near-optimal solutions with low memory requirements and CPU time for both SPP and MOSPP, while the MIP formulations consume huge memory to provide optimal solutions for SPP and constrained MOSPP. Nevertheless, irrespective of the methods applied for optimizing sensor locations, the complexity of WDN due to vastness, complex interflows, and temporal demands have kept the search for optimal sensor locations still a significant challenge (Adedoja et al., 2019; Hu et al., 2018b).

The third component represents the decision variables given by the set of nodes in the network. The problem size depends on the number of nodes considered for placing sensors as large number of nodes translate into large number of node-sensor pairs. This makes the problem

computationally and memory-wise expensive. The problem can be interpreted as a combinatorial problem of choosing a solution from ${}^{n}C_{k}$ combinations where 'n' is the number of nodes, and 'k' is the number of sensors (Blockeel et al., 2012). However, if a few nodes are deemed non-potential and only 'm' such nodes are considered for sensor placement, the number of combinations reduces to ${}^{m}C_{k}$, which is lesser than ${}^{n}C_{k}$. For instance, if a network has 100 nodes, but only 90 nodes are chosen for the placement of 5 sensors, then the total number of combinations reduces by more than 41%. This led to the following studies that focused on improving the nodal search space by contracting the WDN using graph trimming methods and topography-based clustering, designing contamination events based on heuristics, and decreasing decision variable (nodes) search space considered for optimization.

Klise et al. (Klise et al., 2013) pruned large WDN using graph trimming techniques followed by a two-tiered optimization procedure. The first step involved grouping similar nodes based on topography and flow into 'super nodes' and finding the optimal 'super nodes' for placing sensors, and then fine-tuning the sensor locations using the original nodes forming the optimal 'super nodes'. Xu et al. clustered the WDN into k regions for placing 'k' sensors, and from each cluster, one node was selected based on maximum connectivity with other nodes in the cluster (Xu et al., 2008). This selection of single node from 'k' clusters was improvised by greedily selecting 'k' nodes with better overall connectivity within the WDN, leading to better observability of the whole network (Di Nardo et al., 2018). The above methods oversimplified the SPP by overseeing the effects of temporal demands, reverse flows and contaminant transport velocity, as their applicability was tested only on a small number of contamination events.

Diao et al. implemented controllability analysis on BWSN Network1 to condense the decision variable search space by (Diao and Rauch, 2013) choosing nodes that facilitated maximum network observability in a given time frame. But, this prioritized the observability at the

expense of swiftness in detections when a set of nodes are considered together for sensor placement. In contrast, Khorshid el. al. (Khorshidi et al., 2018) developed a node selection strategy based on Value Of Information (VOI) and Transinformation Entropy (TE) techniques. The nodes were compared pair-wise in terms detection time and grouped according to the swiftness in detecting the given set of contamination events. The results of the VOI-TE method were compared with the results from TEVA-SPOT tool which uses GRASP-heuristics, but the comparison with the traditional multi-objective optimization is required to assess the efficiency of the VOI-TE method. Pierre et al., defined quality zones within the WDN by clustering the nodes based on time-varying concentration data (Mandel et al., 2015), but the study has not been extended to sensor placement problem. The quality zones were developed by tracing the water from the source nodes using EPANET2.0, but in a WDN the contaminant intrusion can occur at any node. Moreover, the simulations were carried out at steady-state conditions, which are generally not observed in real WDNs.

Thus, there is a need for a pre-selection procedure to reduce the nodal search space that encompasses the dynamic water quality simulations with a focus on the sensor placement objectives. The observations or pollution matrices acquired from the contamination transport simulations provide the required data on the objectives, detection of contaminant, time of arrival of contaminated water, and population affected at the nodes. Based on this data, the nodes from the WDN can be clustered in terms of similarity in detecting events followed by selecting nodes from these clusters with better objective function values.

In this study, a novel strategy is developed based on k-means clustering and cluster-wise greedy selection to reduce the nodal search space for multi-objective sensor placement study. This strategy is based on the transport of contaminant (both nodal observation and transportation time) in the WDN rather than just the connectivity. The resulting subset of nodes will be used for multi-objective optimization to generate multiple sensor design solutions with trade-offs

between objectives. The multi-objective optimization is carried out for two sets of conflicting objectives: (1) maximizing detection likelihood vs minimizing detection time and (2) maximizing detection likelihood vs minimizing the affected population. The pattern search algorithm (Custódio et al., 2011), which is based on direct search methods for optimization, is implemented for finding the Pareto points. The node selection strategy is tested on two benchmark networks, BWSN Network1 and C-town network. Finally, the Pareto Front generated using the optimal subset of nodes are compared with the traditional method of performing MOSPP.

2 Methodology

The work carried out in this study consists of three sections, as described in Figure 1. Firstly, the effect of contamination in the WDN is analysed by simulating various contamination

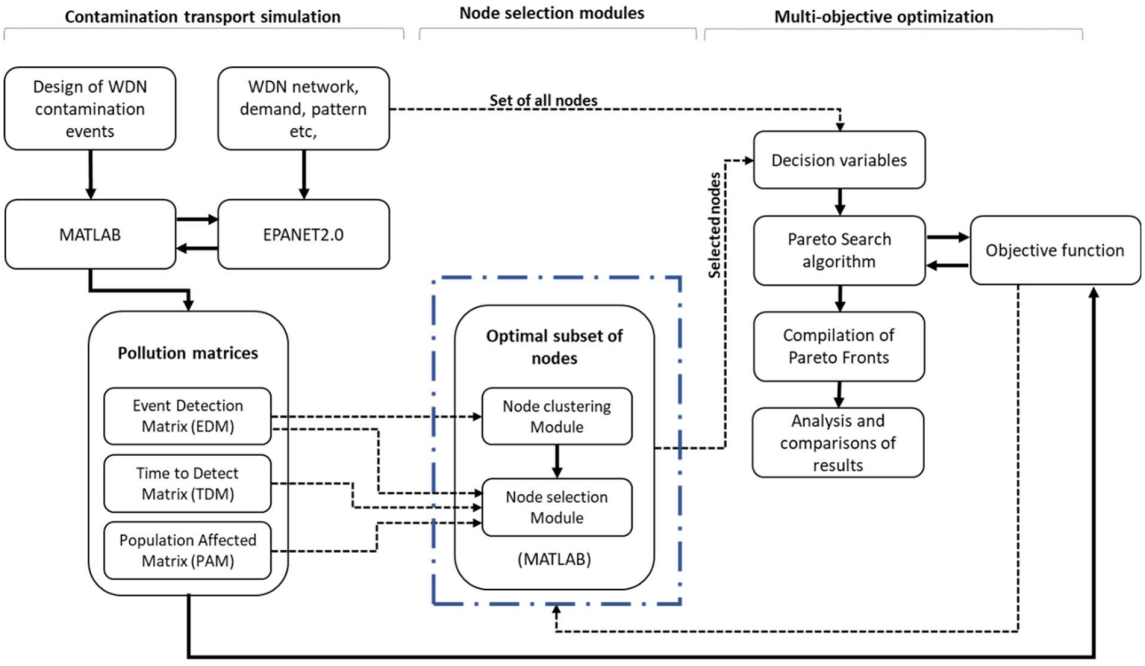


Figure 1 Overview of sensor placement problem for water distribution networks

events. The conditions of each event (time, node, etc.,) is provided by MATLAB to EPANET2.0 where extended hydraulic and water quality simulation of each event is carried

out. The parameters of the network (nodal concentration , time of contamination observance) are then fed back to MATLAB, for developing the pollution matrices. The second step is identifying a crucial set of nodes that provide higher detection likelihood and lower detection time. The nodes are first clustered in the node clustering module, followed by cluster-wise selection of nodes for the objectives considered in the node selection module. Both the modules are developed in MATLAB. The selected nodes are then used for multi-objective optimization study of sensor placement in the third step. The dotted lines represent the paths that will provide the decision search space for the multi-objective optimization. Generally, only the first and third steps are carried out to find the optimal Pareto, which suggests that all the nodes of WDN are potential locations for placing sensors. In order to compare with the proposed strategy, the traditional method of estimating the Pareto is also carried out.

2.1 Contaminant Transport Simulation and Pollution Matrix

The two benchmark networks considered as test cases for the developed node selection strategy: BWSN Network1 and C-town network are shown in Figure 2. The contamination scenario for BWSN Network 1 was a subset of BWSN Challenge Case A (Hart et al., 2007) with an intrusion at every half an hour interval at each node for two hours. The contamination event is defined by the parameters: injection flow rate = 125 L/h, contaminant concentration = 230,000 mg/L, and injection duration = 2 h. This translates to intrusion of 479167 mg/minute of contamination for two hours. It is assumed that the ingested contaminant remains inert throughout the network and it is consumed only at the demand nodes. While in C-town, a 15-minute interval for each contamination event was set to capture the intricacies of contaminant flow within the network. The number of sensors to be placed was fixed to 5 and 20, translating to 3.8% and 15% of nodes in BWSN Network1, respectively, and 1.2% and 5% of the C-town network. It is assumed that sensors have a sensitivity limit of 0.01 mg/L and there is no lag between detecting a contaminant by the sensor transmitting of

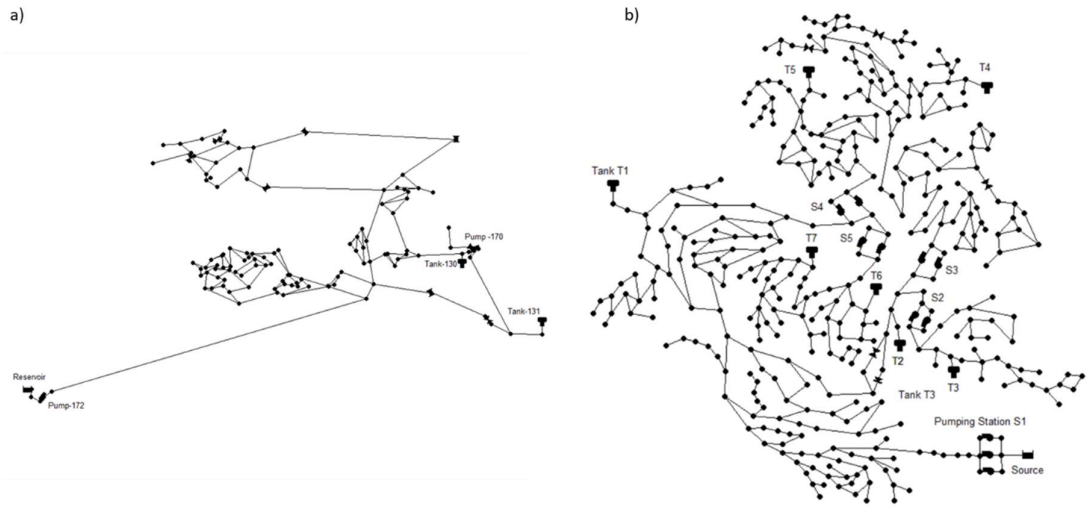


Figure 2 Water Distribution Network Layout a) BWSN Network1 b) C-town Network

information to the utility manager. Table 1 describes the contamination events design for both the WDNs.

The event simulations were carried out in EPANET2.0 integrated with MATLAB 2020b. The results of the contamination events were recorded in three pollution matrices, depicting the

Table 1 Contamination events design for BWSN Network1 and C-town Network

Features	**BWSN Network 1**	**C-town Network**
Nodal contamination	All nodes (129)	All nodes (396)
Time of intrusion	Every half an hour	Every 15 minutes
Intrusion duration	Half an hour	15 minutes
Contaminant mass injected	479167 mg/minute	10000 mg/minute
Detection delay	Real-time	Real-time
Detection sensitivity	0.01 mg/L	0.01 mg/L
Time period for extended simulation	96h	72h
Simulation time step	5 min	5 min
Number of sensors	5 and 20	5 and 20

detection of the event, time to detection, and population affected. The events detections matrix (EDM) is binary in nature, where '1' represented if the contamination was observed and '0' otherwise. The time to detection pollution matrix (TDM) stored the time delay in contaminant observation since the onset of contamination for each node for all the intrusion events. Similarly, the population affected matrix (PAM) provides the details about the number of people affected until the contamination is observed.

2.2 Generation of optimal subset of nodes

This section discusses the objectives for the sensor placement study and the objective function-based node selection strategy.

2.2.1 Definition of objective functions:

The objectives evaluated for the MOSPP study are: a) Detection likelihood, b) Expected detection time, and c) Population affected.

a) The detection likelihood (Z1) is defined as the percentage of contamination events detected for a given design of sensors.

$$Z1 = \frac{100}{N}\sum_{i=1}^{N} d_i \qquad i = 1,2,3....N \tag{1}$$

Where, i = 1,2,3…N refers to the 'N' contamination events simulated for the network and d_i = 1 if the contamination event 'i' was detected otherwise '0'.

b) The time of detection for a particular contamination event 'i', t_d, is the minimum time to detect among all sensors (j) present in the design,

$$t_{d,i} = min\,(t_{i,j}) \tag{2}$$

And, the Expected time of detection (Z2) is the expected value of t_d over N contaminations given by,

$$Z2 = E(t_{d,i}) \quad (3)$$

c) Similarly, the expected population affected prior to detection is the expected value of population affected (P_a) computed over the assumed probability distribution of 'i' contamination events.

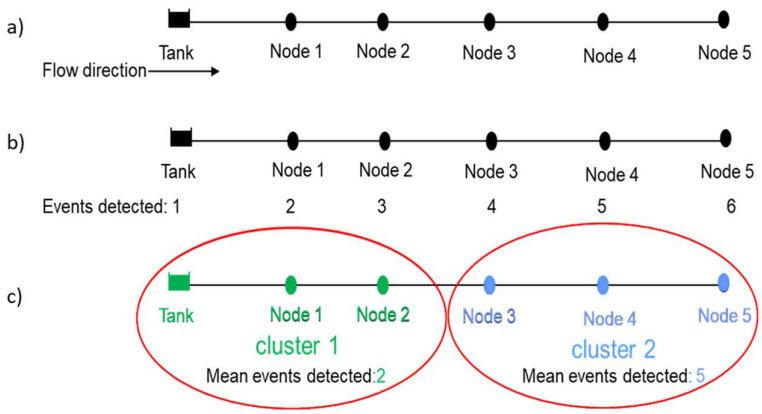
$$Z3 = E(P_{a,i}) \quad (4)$$

The estimation of $P_{a,i}$ is explained in section A.1. The objectives are taken 'as is' from BWSN challenge (Hart et al., 2007). In the following paragraphs, the expected time of detection will be referred to as detection time and expected population affected prior to detection as affected population for ease of discussion.

2.2.2 Generating optimal subset of nodes

The ideal case for complete observation of the WDN is to place sensors at all the nodes in the network, but it is not possible due to various constraints. The following best-case scenario is to design the sensor placement such that all the contamination events can be detected. But a guarantee of 100% detection is inefficient if the time to detect is large or large population gets affected before detection. It is necessary to assess the nodes based on swift detection of the contamination events. The objectives for improving the detections and detection time (or affected population) are competing in nature, thus placing sensors for better detections will deteriorate the detection time.

Consider the simple example network with six nodes in a straight line, as shown in Figure 3a. The example is explained for the MOO study of detection likelihood vs detection time but can also be easily extended to detection likelihood vs affected population. For placing only one sensor in the network, it can be observed that placing a sensor at node 5 can detect all the events, but the expected detection time will be maximum for this sensor placement design.



d) **Pollution Matrix of Node 2 and Node 4 for 6 contamination scenarios**

Contamination scenarios →	1	2	3	4	5	6
Node 2	1	1	1	0	0	0
Node 4	1	1	1	1	1	0

Figure 3 Clustering and node selection for an example network a) Example network b) Events detected at each node c) Cluster formation based on event detection d) Pollution matrix example

On the other hand, if the sensor is placed at the tank node, the expected detection time will be minimum, as the detection time is estimated based only on the detected events, and only those events will be detected that occur at the tank. Due to the competing nature, each objective requires different nodes for optimal value but at the cost of the other objective. However, if two sensors are to be placed in this simple network, then the first sensor can be dedicated to maximizing the detection likelihood (Node 5), and the second sensor that results in minimum detection time in combination with Node 5 can be selected. Note that the number of combinations for placing two sensors in nodes, ${}^{6}C_{2=}$ 15, is reduced to just 5 pairs of Node 5 and other nodes. This methodology can be extended to real WDNs by selecting all the nodes that improve the detections (predominantly dead-end nodes) followed by selecting nodes based that improve the detection time of previously selected nodes. The selection procedure is two phases as nodes will be first selected based on event detections followed by detection time or affected population. But real WDNs are much more complex in design than the above network in Figure

3, and this will lead to loss of important information when these nodes are considered for MOSPP. In real WDNs, multiple sources in the form of reservoirs and overhead tanks are installed for reliable water supply. The nodes that are connected to two sources might receive water from either of the sources. Or in the cases, where the nodes are present in between reservoir and overhead tanks, a change in water flow direction is observed whenever the source of water supply changes. So, it must be noted that these nodes and the nodes in its downstream will only be contaminated by the current source of water supply. These issues of changing water flow direction can be overcome by segregating the network into different groups or clusters based on the similarity in events detected and selecting the nodes from the group. In the following paragraphs, the clustering and greedy selection procedure are further explained.

2.2.2.1 k-means clustering and node selection

The nodes of the WDN were segregated into clusters using k-means clustering based on the dissimilarity of events detected by the nodes rather than the time-variant contamination concentration. In k-means clustering, the data is clustered into 'k' groups based on the 'distance metric' between the 'k' centroids and the observation data points (Lloyd, 1982). The difference in the detection ability of the nodes is translated into the 'distance metric' using the event detection matrix (EDM). The EDM is built of 0's and 1's, and thus, the 'Hamming distance metric' was used to estimate the distance between two nodal observations. The Hamming distance is calculated based on the XOR operation, and it estimates the number of bits that are different in each of the strings. For example, the distance between 101011 and 110011 is given by,

$$101011 \oplus 110011 = 011000;\ \textit{Hamming distance} = 2$$

When this metric is applied to the detection matrix, nodes with similar number of detections at the same bit locations (events) will be grouped into one. In other words, this calculates the

dissimilarity between the nodes in detecting a given set of contamination events. Also, if neither the nodes detect the contamination, it does not affect the distance metric. This clustering methodology prioritizes the transportation of contaminant to nodes during contamination events over connectivity or closeness of nodes, which is norm considered in previous literature.

The number of clusters is estimated by performing clustering with a multiple number of clusters and then fixing it based on statistical measures (Mandel et al., 2015), or they can be fixed based on the number of sensors to be placed (Xu et al., 2008). The former procedure requires multiple k-means clustering runs, while in the latter, only one node is chosen per cluster for sensor placement. These procedures result in larger computation time and neglect of many potential locations, respectively. In this study, a heuristics-based approach is developed by assuming that all the nodes in the WDN are selected by choosing 'k' nodes from 'k' clusters for each objective considered for MOO study. For two objectives in MOSPP, this results in the equation

$$2k^2 = number\ of\ nodes \tag{5}$$

Upon rearranging the above, the number of clusters can be calculated as,

$$k = round\left(\sqrt{\frac{number\ of\ nodes}{2}}\right) \tag{6}$$

The number of clusters was fixed on the Eq.6 and from each cluster the nodes were selected as discussed below.

2.2.2.2 Selection of nodes

The node selection procedure is two-phased and in each phase, the nodes are selected for each of the objectives in MOSPP. In the order of importance, first the nodes are chosen for event detection followed by the second competing objective, detection time (Z2) or affected population (Z3). The selection procedure relies on the pollution matrices for assessing the

importance of the nodes. In the first phase, nodes are selected from each cluster in the clustering module. The nodes belonging to each cluster are stored in sets G_i, where i=1,2..k, represents cluster 'i'. From each G_i set, nodes are selected in the order of maximum number of events detected (Z1) using the event detection matrix and stored in set $N_{G,i}$. The node selection for a cluster is terminated when no further improvements in Z1 is observed. In the example network (Figure 3), Node 2 from cluster 1 will have the maximum Z1 value, and no other nodes in the cluster can improve Z1 value. Similarly, only Node 5 from cluster 2 will be selected. Thus for a given WDN, $N_{G,i}$, i = 1,2…k, are generated that comprise the nodes selected from sets G_i, i = 1,2,…k, respectively. The nodes with higher sensitivity to detection likelihood (Z1) is given by the union of the sets $N_{G,i}$, denoted as $P_{DL} = \coprod_{i=1}^{k} N_{G,i}$.

In the second phase, the nodes are selected based on the improvement observed in detection time. The nodes are selected from each set G_i and their union will be defined as nodes with higher sensitivity to detection time. Firstly, sets $T_{G,i}$ are defined for each set G_i, and the nodes in P_{DL} are added to each of the $T_{G,i}$ sets. Inclusion of set P_{DL} in the set $T_{G,i}$ before selecting nodes ensures that all the contamination events are detected, and the selected node will improve the overall detection time. The detection times corresponding to the nodes were estimated using the time to detect matrix (TDM). Then from each set G_i nodes are appended one by one to set $T_{G,i}$ that results in the lowest value Z2. The nodes are appended till no further decrease in the detection time is observed or a maximum of 'k' nodes are selected. The nodes in a WDN will provide the lowest detection time when they are the source of contamination, and a sensor is placed at them. However, the trade-off in improving Z2 for each new node selected is very low when the number of nodes in set $T_{G,i}$ is large. Thus, to control the number of selected nodes, the second terminating criteria mandates choosing of a maximum of only 'k' nodes from each group. The set of nodes selected based on detection time is given by the union of set $T_{G,i}$ as $P_{DT} = \coprod_{i=1}^{k} T_{G,i}$.This is the optimal subset of nodes considered for MOO of Z1 vs Z2. Note that

the nodes for improving detection likelihood have already been added to sets $T_{G,i}$ and thus do not require to be added separately. During the implementation of the above selection procedure,

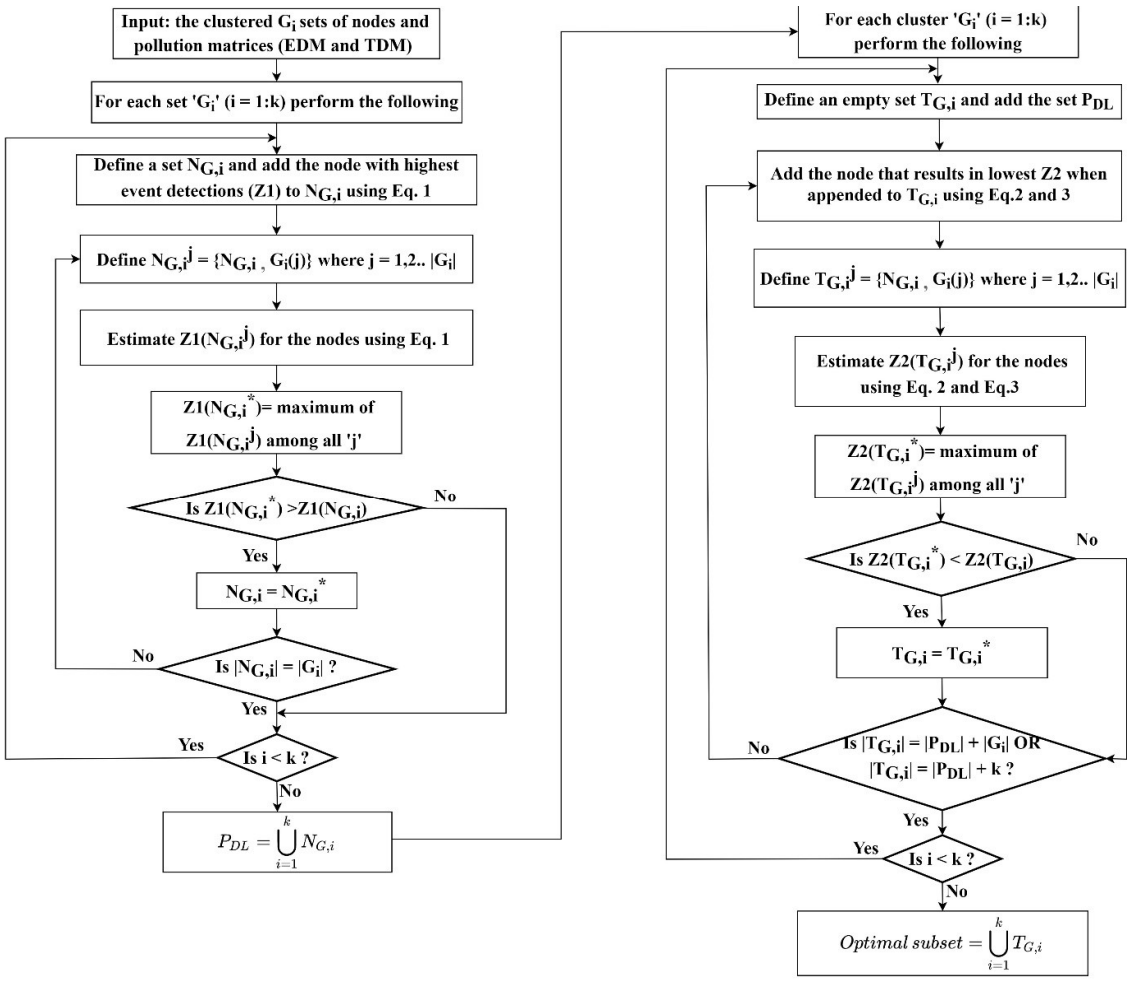


Figure 4 Objective based selection of nodes for multi-objective sensor placement study (Z1 vs Z2)

the nodes that were already part of $N_{G,i}$, P_{DL} and $T_{G,i}$ were removed from sets G_i to avoid repeated computations of already selected nodes. The cluster-wise greedy selection of nodes results in choosing nodes with lower detection time (which are generally located in the upstream of the cluster) as well as nodes with maximum observability (which are generally located in the downstream of the cluster). Thus, the combination of k-means clustering and greedy selection results in set of selected nodes that are targeted towards both objectives. In contrast, the previous studies only focused on the network connectivity, closeness to other nodes for trimming the

network or just maximum observability at the expense of detection time. The flowchart for node selection procedure is shown in Figure 4. An example of network clustering and node selection procedure is illustrated in Appendix A.2.

2.3 Multi-objective Optimization (MOO) and Contribution rate indicator

The pattern search algorithm from MATLAB2020b was implemented to find the Pareto front. This algorithm uses a derivative-free methodology by extending direct search methods to multi-objective problems (Custódio et al., 2011). In the node selection procedure, the objective functions would be evaluated for selecting nodes. Therefore, to provide a fair comparison, the number of objective function calls made during the selection procedure is reduced from the maximum number of function counts for the proposed method.

Pareto fronts of the traditional and proposed methods were compared in terms of contribution rate (CR) indicator based on hypervolume, developed by (Cao et al., 2015). The Pareto fronts (PF_i) to be compared are first compiled to generate a surrogate true Pareto Front (PF_s). Then the contribution from each of the PF_i's in PF_s is estimated using hypervolume. Hypervolume measures the size of the space enclosed by all points on the Pareto front and a user-defined reference point, and it is indicated as I_H(PF,r) for a given Pareto front (PF) and a reference point (r). After the evaluation of the reference point estimated based on (Cao et al., 2015), CR indicator of Pareto Fronts from each method was calculated. The contribution rate (CR) associated with each of the Pareto Front are estimated as:

$$CR_i = \frac{I_H\left(PF'_i, r\right)}{I_H\left(PF_s, r\right)} \tag{7}$$

where PF_i' is the set of non-dominating points of Pareto Front 'i' present in the surrogate true Pareto PF_s, 'r' refers to the reference point and 'I_H' refers to hypervolume. The procedure for estimation is explained in Appendix A.3. The value of CR_i close to '1' means that PF_i is closest

to replicating the Surrogate true Pareto front. Compared to other PFs, a higher CR indicator translates into higher hypervolume dominance.

3 Results and Discussion

3.1 BWSN Network1

The contamination event simulations for BWSN Network1 yielded three pollution matrices of dimensions 129x6192, as 48 events per node were simulated for each of the 129 nodes. It was observed that 622 events remained undetected due to zero flow conditions at the nodes during contaminant intrusion. The nodes of BWSN Network1 were clustered into 8 groups based on Eq.6 and 21 nodes were selected based on EDM, 51 based on TDM and 48 based on the PAM. Most of the dead ends have been selected (except Junction 13 and Junction 36, due to zero flow), ensuring that all the 'observable events' will be detected. The 8 clusters with the average number of detections are depicted in Figure 5a. The average number of detections for clusters ranged from 203.4 (3.6% of observable events) at the network entrance nodes and low flow nodes to 2893.0 (51.94% of observable events) at the downstream nodes. The 72 nodes selected for detection likelihood vs detection time and 69 nodes selected for detection likelihood vs affected population are shown in Figure 5b and Figure 5c, respectively. The first cluster had the lowest mean detections and consisted of Junctions 1, 7, 38, 109, 124, 128, and reservoir node (129). Junction 1, 7, 38 and 124 were selected during first phase of selection, leaving junctions 109, 128 and 129 as no improvement in detections were observed on appending these nodes.

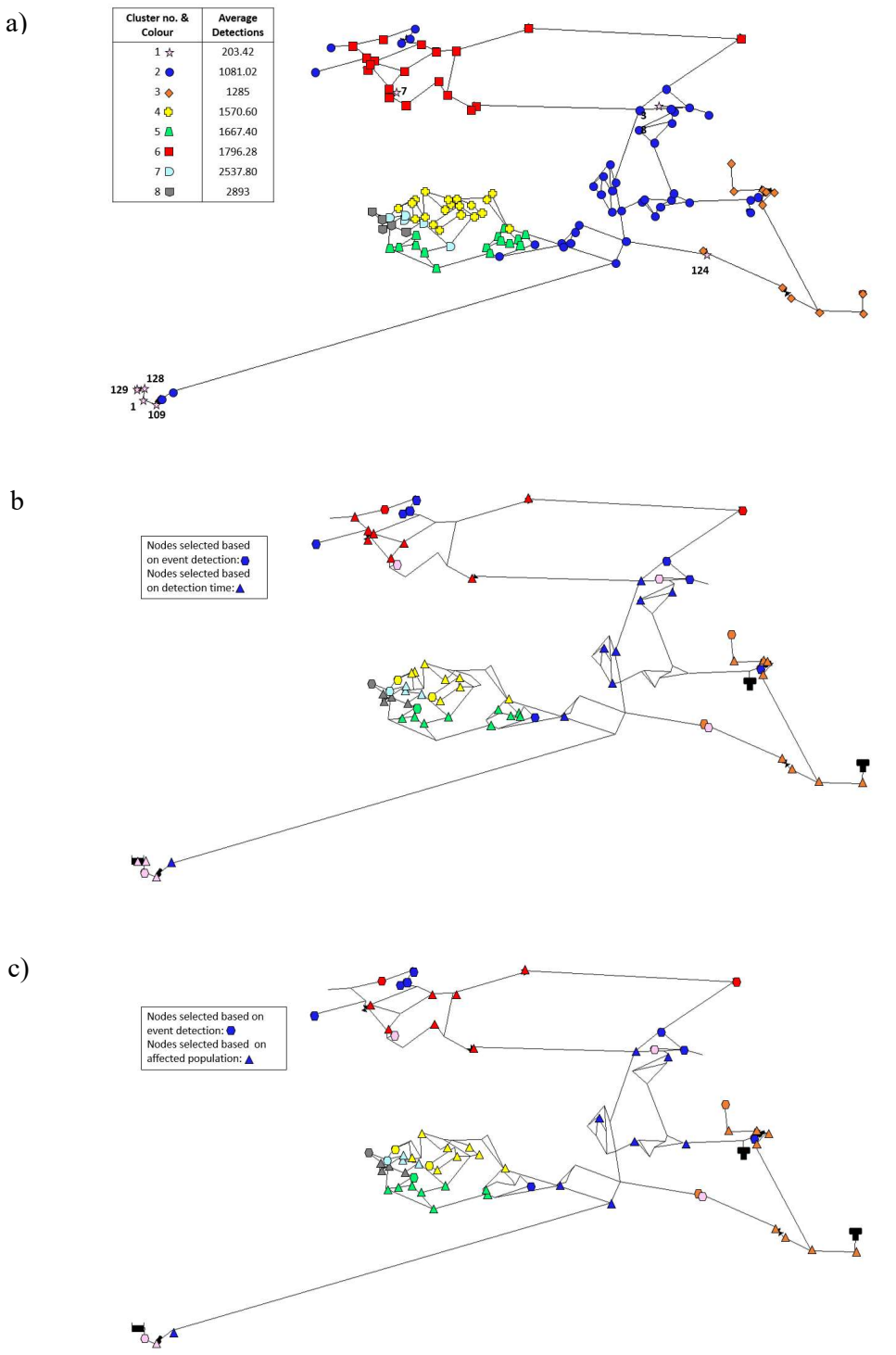


Figure 5 a) Clusters formed for BWSN Network1 b) Nodes selected for Z1 Vs Z2 MOSPP study c) Nodes selected for Z1 vs Z3 MOSPP study (Please go through the web version of this article for interpretation of the colours given in figure legend)

In the second phase of selection, all three nodes improved detection time and were thus added to the optimal subset of nodes, whereas only Junction 109 was selected for the affected population. It must be noted that all three nodes have zero base demand, but as the algorithm chooses at least one node from each group, the node at downstream was selected. Out of 48 nodes selected in the affected population, 13 had zero base demands. These nodes were selected either due to their position between two nodes with non-zero base demands (like Junction 91) or due to condition that makes the algorithm select at least one node from each cluster, as observed for Junction 109. It can be observed that five nodes in cluster 2 are disjointed from cluster 2 and are separated by cluster 6. This is because, though the disjointed nodes are located closer to and at the downstream of cluster 6, the events detected by these nodes are more similar to the events detected by nodes of cluster 6. This is verified by the average number of events detected by the disjointed nodes =1181.8, which is closer to average number of events detected of cluster 2 (=1081.02) than cluster 6 (=1796.28).

The matching rate is defined as the percentage of nodes from BWSN results appearing in the pre-selected nodes, and it is provided in Table 2. The sensor locations (nodes) in BWSN challenge were given by 14 different research groups using different methods to optimize these locations. The efficiency of the proposed method is tested by comparing the optimal subset of nodes with the possible sensor nodes provided in BWSN. The matching rate of the BWSN 5 SPP solution nodes with the optimal subset of nodes (set OSN) was 78.8% and 75.6% for Z1 vs Z2 and Z1 vs Z3 study, respectively, while only 26.9% matching rate was observed in the controllability analysis method (set CAS), described in (Diao and Rauch, 2013). For BWSN 20 SPP, 64.71% and 67.06% of the BWSN solution nodes matched with set OSN compared to 57.4% for set CAS. The lower matching rate of set CAS might be due to excluding the cumulative effect of selected nodes in improvising the detection time. Meanwhile, the proposed methodology

Table 2 Matching rate of selected nodes with BWSN sensor nodes

MOO Study	Matching rate % (Diao et al)	Matching rate (This study) Z1 Vs Z2	Matching rate (This study) Z1 Vs Z3
5 sensors	23.9	78.8	75.76
20 sensors	57.4	64.71	67.06

weighs the nodes in terms of their individual significance in detecting events and reducing detection time or exposed population. The nodes of set OSN and BWSN results are enumerated in Appendix A.4.

The percentage of events detected with respect to mean detection time for set OSN and set CAS (Diao and Rauch, 2013) are shown in Figure 6. About 66.16 % of events were detected by OSN, while CAS detected only 41.36% within 5 minutes of contamination. Also, the set CAS observed a maximum of 94.5% of events while OSN observed 100% of the events. This shows that the selected nodes have quicker reaction time to the contamination events than DN nodes. For comparison purposes, the sensor solutions from BWSN 5 sensor (set B5) and BWSN 20 sensor (set B20) were also plotted (Figure 6). The order of improvement of event detection percentage with respect to detection time was B5, CAS, OSN, and finally B20 sets. The set OSN was swift in detecting events than the set B5 and set CAS but lagged behind set B20. It must be noted that only the combination of nodes from set OSN is required for placing the sensor, and thus, multi-objective optimization was carried out to verify the efficiency of selected nodes.

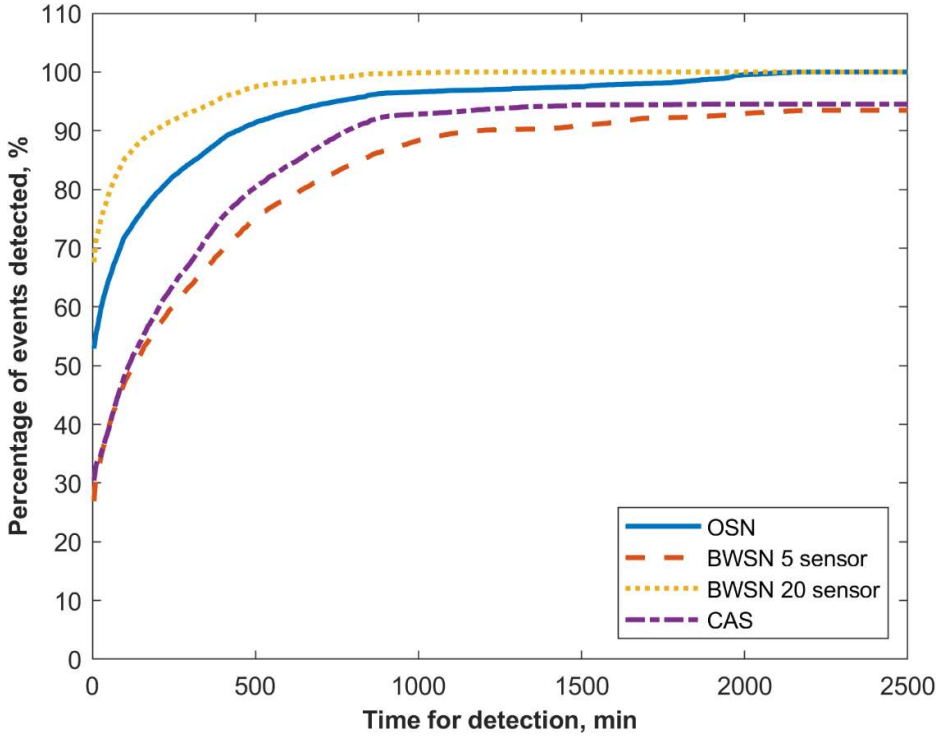


Figure 6 Percentage of events detected with respect to mean detection time for BWSN Network1

The parameters for multi-objective optimization study were: number of runs=30, maximum function count =50000, bounds = [1 number of nodes], tolerance = $1*10^{-6}$, final Pareto Front = Compiled non-dominating set of all 30 runs. The 'paretosearch' tool of MATLAB2020b that incorporates pattern search technique was used for generating Pareto fronts. The legend 'Traditional' refers to the generic method where all the nodes in WDN are deemed as plausible sensor locations while the legend 'Proposed' refers to the method developed in this study where only the optimal subset of nodes (OSN) are considered for placing sensors. Figure 7 represents the process to estimate the surrogate true Pareto front for placing 20 sensors for objectives Z1 and Z2. Figure 7a depicts the complete set of Pareto points obtained from 30 runs of the optimization. A clear separation of the Pareto points between traditional and proposed procedures was observed, with the latter dominating the former in the range 50% to 90%

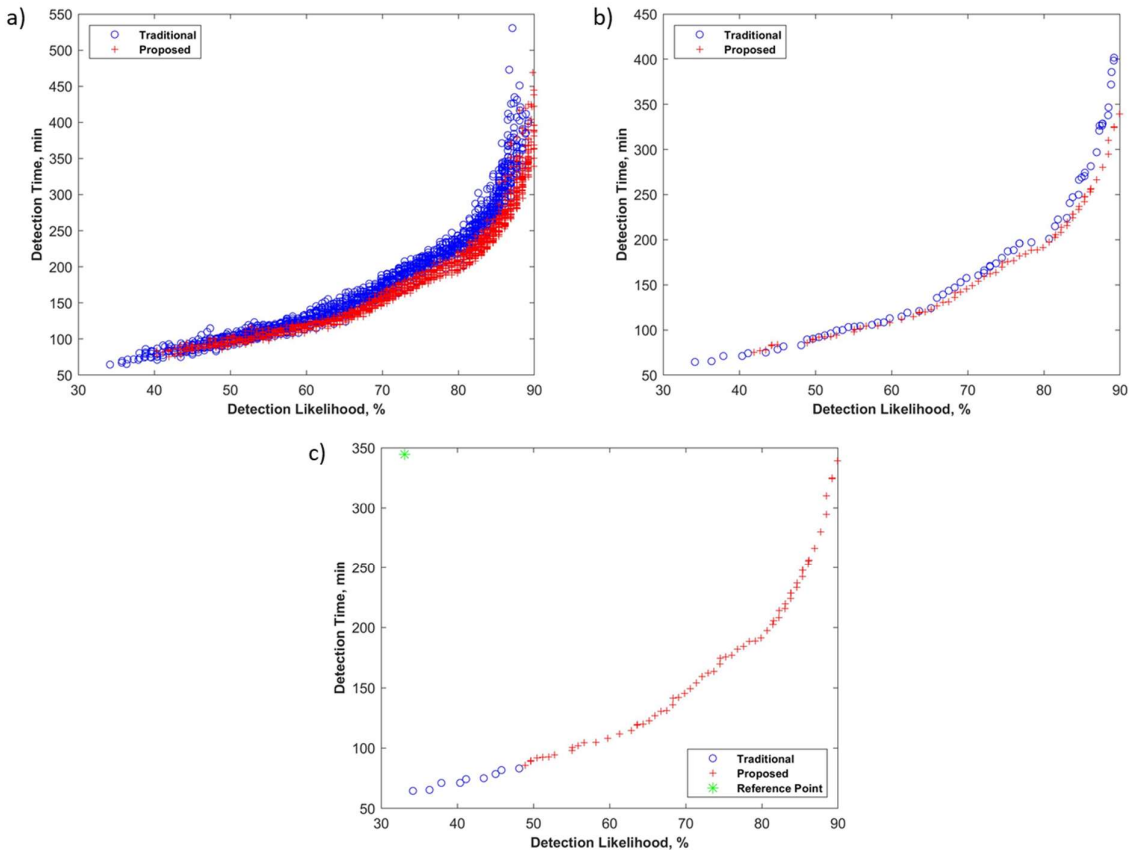


Figure 7 Generation of surrogate true Pareto front for BWSN Network1 - 20 sensors problem a) Pareto fronts of all 30 runs b) Compiled final Pareto c) Surrogate true Pareto front with reference point

detection likelihood. This effect is reiterated when compiled Pareto points from 30 runs are compared in Figure 7b. Based on these Pareto fronts, the surrogate true Pareto front and reference point were obtained and depicted in Figure 7c. The Surrogate true Pareto fronts with the reference points for 5 and 20 sensors placement for objectives Z1 vs Z2 and Z1 vs Z3 study are shown in Figure 8. An overview of Pareto points for Z1 vs Z2 study revealed that the OSN performs better than the complete set. Whereas in Z1 vs Z3 study for placing 5 sensors, the optimal subset generated fewer Pareto points than traditional method, and most of these Pareto points had higher detection likelihood values. This could be due to the exclusion of nodes with very low base demand that were not selected, and when these combined with zero demand nodes with higher detection ability, it resulted in a low number of people affected at higher Z1

value. This trend was again pronounced in 20 sensor results shown in Figure 8d, where all the Pareto points at Z1<65% were obtained from the traditional method and at Z1 > 75% were from the proposed method. The comparisons of both methods in terms of CR indicator is shown in Table 3. The CR indicator of both methods for 5SPP was same in Z1 vs Z2 study, indicating that both procedures dominate a significant portion of hypervolume at the same level. Whereas in Z1 vs Z3 study, the traditional method was higher by 4.16%. The CR indicator for 20 SPP was 190% and 22.7% more than the traditional method in Z1 vs Z2 and Z1 vs Z3 study, respectively. This indicated that the optimal subset provided better Pareto points in terms of hypervolume dominance. Since the solutions with lower detection likelihood cannot be considered for placing sensors, a cut-off criterion was maintained to assess the number of Pareto points contributed by both methods. Only those solutions whose Z1 value was more than 50% of maximum observed events (50% of max(Z1)) were considered as potential solutions. The number of these Pareto points observed for 5 sensors were higher for the traditional method, while for 20 sensors proposed method provided a higher number of solutions. The time taken for estimating the final Pareto Front using the proposed method for 5SPP was 25% (by average) less than the traditional method while it was less by only 0.7% (by average) for 20SPP. Both traditional and proposed methods generated better Pareto points in comparison to BWSN results. Only 4 Pareto points from BWSN were coinciding with Pareto Fronts obtained in this study (illustrated in Appendix A.5).

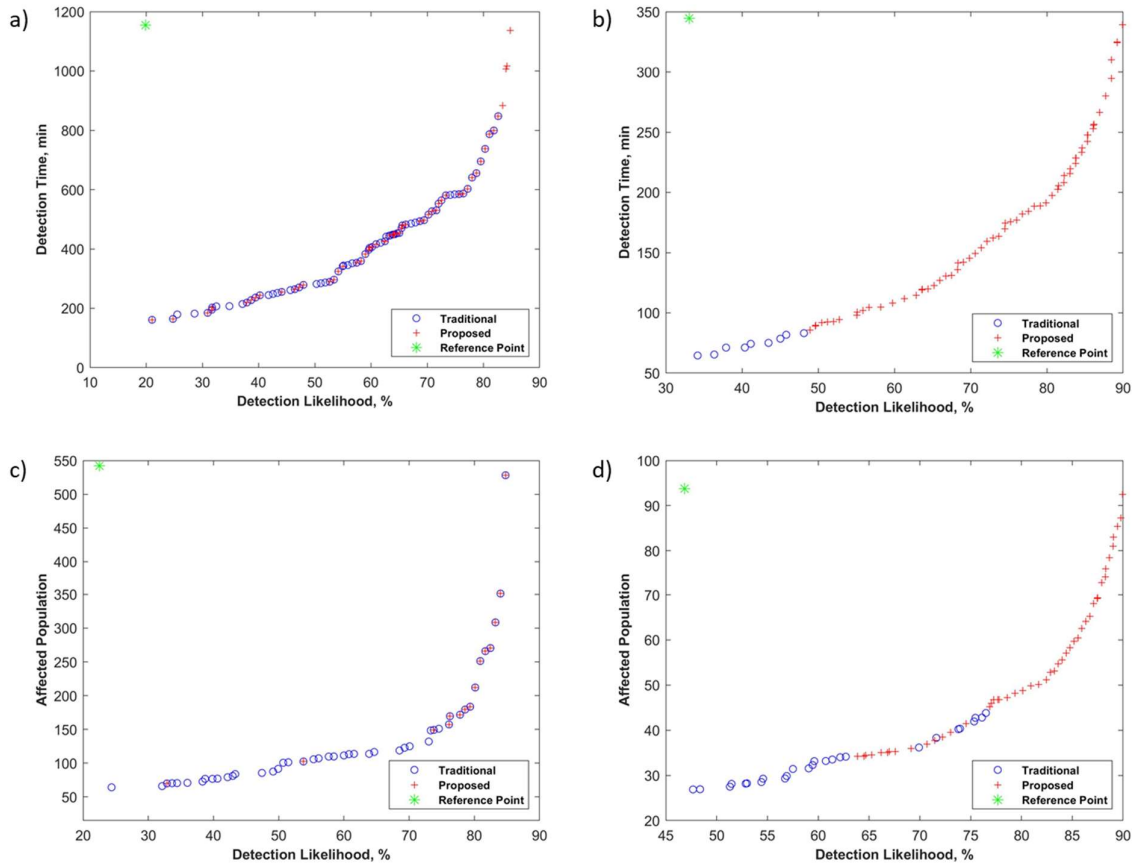


Figure 8 Surrogate true Pareto fronts and Reference point for BWSN Network1 a) Z1 vs Z2 5 sensors b) Z1 vs Z2 20 sensors c) Z1 vs Z3 5 sensors d) Z1 vs Z3 20 sensors

Table 3 BWSN Network comparison of the traditional and proposed method

	Detection likelihood Vs Detection Time				**Detection likelihood Vs Affected Population**			
	5 Sensors		**20 Sensors**		**5 sensors**		**20 sensors**	
Parameters	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**
Number of nodes	129	72	129	72	129	69	129	69
Contribution rate Indicator	0.99	0.99	0.21	0.60	1	0.96	0.44	0.54
Pareto points in PF_s (Z1>50%)	59	47	3	65	36	14	26	46
Time taken for Final Pareto, sec	433.10	318.84	5426.26	5415.68	679.49	418.45	6402.20	6324.27

3.2 C-town Network

The pollution matrices for the C-town network had 38016 contamination events for 396 nodes yielding a pollution matrix of dimension 396x38016. Three nodes in the network did not detect any event and their respective contamination events were undetected due to zero flow. The C-town network was segregated into 14 groups, and 263 nodes were selected for Z1 vs Z2 study and 261 nodes were selected for Z1 vs Z3 study (illustrated in A4.7). The surrogate true Pareto

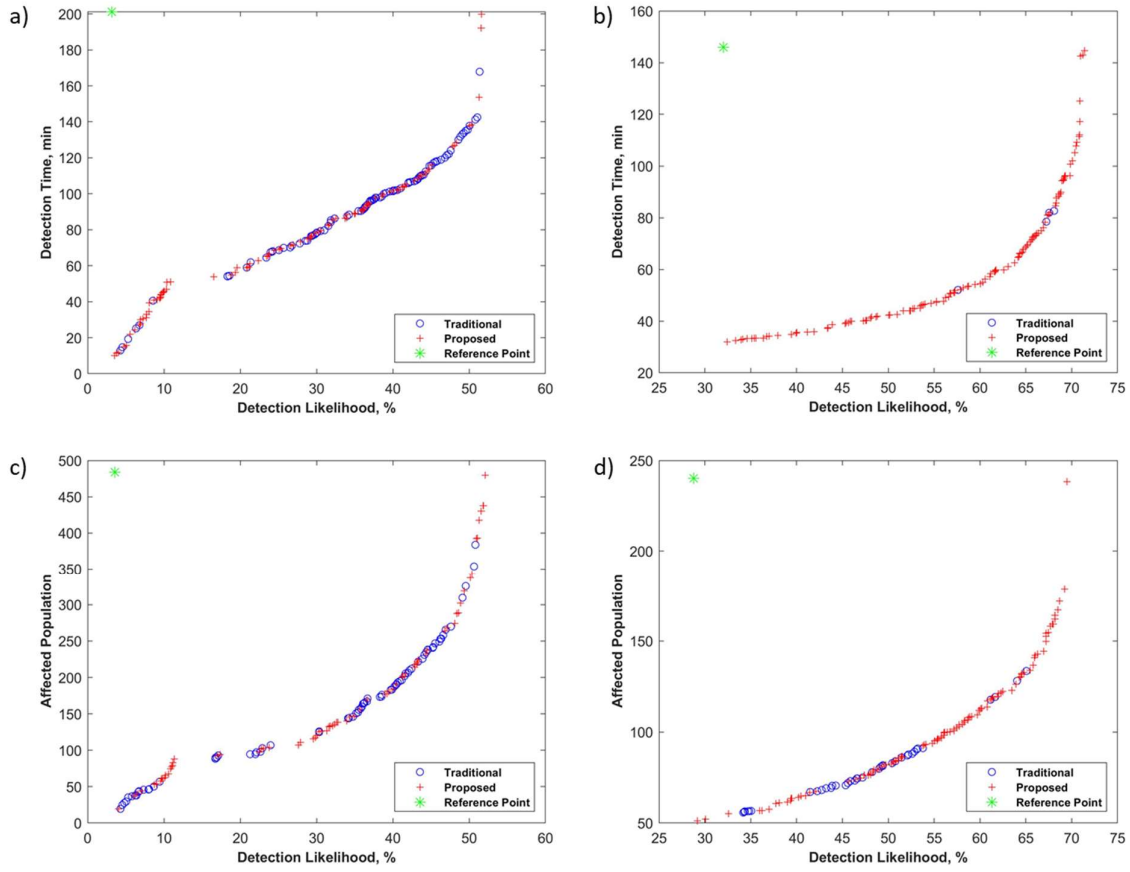


Figure 9 Surrogate true Pareto fronts and Reference point for C-Town Network a) Z1 vs Z2 5 sensors b) Z1 vs Z2 20 sensors c) Z1 vs Z3 5 sensors d) Z1 vs Z3 20 sensors

front and reference point for CR indicator estimation for C-town network is displayed in Figure 9. For 5SPP, comparable Pareto points are generated by both the methods for both the MOSPP studies. In the case of Z1 vs Z3, the proposed methodology generated better Pareto points for higher detection likelihood similar to the results observed for BWSN Network1. But for 20SPP

most of the Pareto points in PF_s were generated from the proposed method. These observations have been reflected in the CR value of the corresponding problem provided in Table 4. The set OSN yielded 96.6% and 74.6% of Pareto points in PF_s for Z1 vs Z2 and Z1 vs Z3 respectively for 20SPP. The time taken for estimating the final Pareto Front using the proposed method for Z1 vs Z2 study was 3.85% (by average) less than the traditional method, whereas it was less by 9.75% (by average) for Z1 vs Z2 study.

Table 4 C Town Network comparison of the traditional and proposed method

	Detection likelihood Vs Detection Time				**Detection likelihood Vs Affected Population**			
	5 Sensors		**20 Sensors**		**5 sensors**		**20 sensors**	
Parameters	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**
Number of nodes	396	263	396	263	396	261	396	261
Contribution rate (CR) Indicator	0.615	0.617	0.543	0.650	0.691	0.697	0.576	0.628
Pareto points in PF_s (Z1 > 50% Z1)	77	55	4	117	50	49	36	106
Time taken for Final Pareto, sec	3012.18	3185.59	3799.45	3877.15	3678.44	3167.42	3629.90	3425.35

From the results of both the methods, it was observed that the pre-selection procedure performs well when the number of combinations is huge, as in 20SPP ($^{129}C_5 << {}^{129}C_{20}$ and $^{396}C_5 << {}^{396}C_{20}$). But the deterioration in performance was only observed for the lower detection likelihood solutions, which are generally omitted.

3.3 Comparison of sensor placement solutions based on detection likelihood cut-off

The expected time for detection for various percentages of detections is described in Table 5. For BWSN 5SPP detection, the maximum Z1 value from traditional method was 84.01% and 84.78% for traditional method and proposed method, respectively and for 20 sensors it was

89.21% and 89.95%. On the other hand, the detection time at the maximum Z1 values were lower in proposed method. But in order to provide a viable comparison, Z2 and Z3 are compared at values at four fixed values of Z1, starting at a minimum level of detecting 50% of

Table 5 Comparison of objectives Z2 and Z3 at different Z1 cut-off values

Detection likelihood (Z1)	Detection Time (Z2)				Affected Population (Z3)			
	BWSN 5 Sensors		BWSN 20 Sensors		BWSN 5 Sensors		BWSN 20 Sensors	
	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed
50%	281.96	289.23	92.10	91.86	100.61	102.46	27.49	32.08
60%	406.31	406.31	114.91	111.70	113.1	118.61	33.18	34.03
70%	515.91	515.91	160.38	149.25	124.97	133.79	37.03	36.88
80%	737.38	737.38	201.02	197.38	212.07	212.07	50.32	48.80

the events and then increasing in the order of 10%. It was observed that there is no significant difference in the traditional and proposed method for 5 SPP, while for 20 SPP the expected time of detection reduces by 2.8%, 6.94% and 1.8% for 60%, 70%, and 80% detections, respectively. The affected population metric was higher in the proposed method for 5 sensor design while it was lower for 20 sensor design at 60%, 70% and 80% detection likelihood. The lowest Z1 value Pareto point for 20SPP was 58% in proposed method and thus at cut-off of 50%, the affected population has 18% higher value.

As shown in Figure 10, the sensor locations for 70% detection for Z1 vs Z3 study for 5SPP have 3 locations in common, and one node each solution had zero base demand. While 12 sensor locations were similar for 20 SPP, with 5 and 7 zero demand nodes for traditional and proposed solutions, respectively. The sensor locations of 5 SPP are not repeated in 20 SPP, which indicates that solutions of a smaller number of sensors cannot be replicated for a larger number of sensors, and for each sensor design, optimization is required to estimate the best

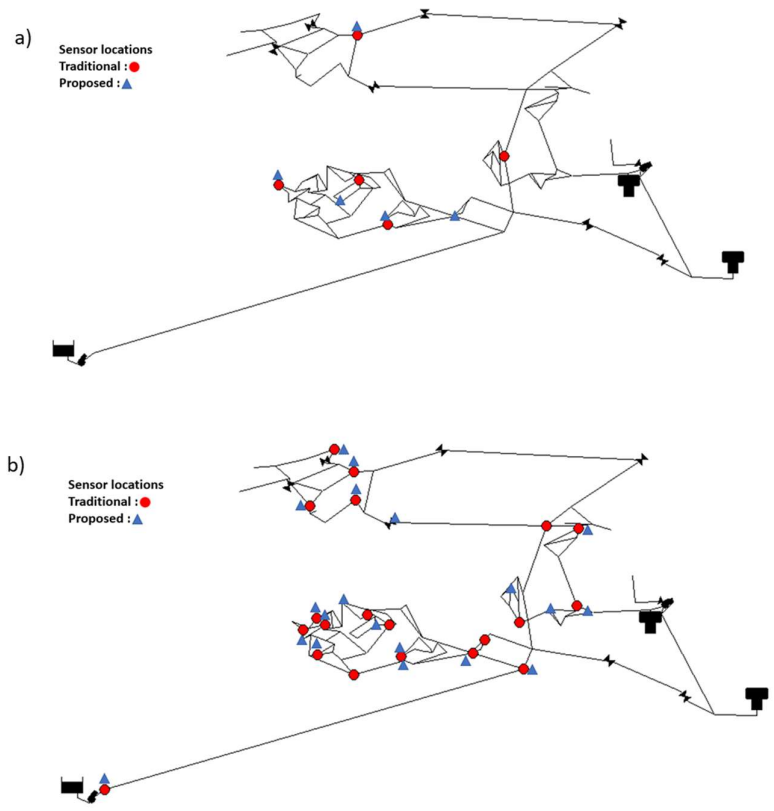


Figure 10 Sensor node location for 5 and 20 sensors in BWSN Network from Z1 vs Z3 study at Z1 = 70%

location. Similarly, Z2 and Z3 are compared at Z1 values of 40%, 50%, 60% and 65%. This range of Z1 was fixed owing to the maximum Z1 value of 53% and 71% observed for placing 5 and 20 sensors. The expected time for detection for both 5 SPP and 20 SPP from the proposed

Table 6 Comparison of objectives Z2 and Z3 at different Z1 cut-off values for C-town Network

Detection	**Detection Time (Z2)**				**Affected Population (Z3)**			
Likelihood	**C-town 5 Sensors**		**C-town 20 Sensors**		**C-town 5 Sensors**		**C-town 20 Sensors**	
(Z1)	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**
40%	101.37	101.91	38.40	35.64	187.43	188.71	66.35	64.01
50%	137.81	137.81	44.68	42.22	342.15	337.30	82.76	82.30
60%	-	-	57.00	54.49	-	-	113.78	112.86
65%	-	-	116.60	102.03	-	-	133.57	133.97

method is comparable to values observed in the traditional method Table 6. All the objective values obtained from the proposed method for objectives Z2 and Z3 (except at Z3 at 50% detection likelihood, where it is almost equal) are lower than the traditional method.

The sensor nodes at 50% detections is displayed in Figure 11, where 3 out 5 sensor locations are at the same nodes for 5 sensors (Figure 11a) and in the case of 20 sensors, only 6 nodes are common in solutions provided by both methods (Figure 11b). The sensor nodes from the proposed method are evenly distributed in the WDN, such that the nodes are placed at the farthest ends to obtain maximum detections. The set of nodes obtained from the proposed (Figure 11) are also present in the decision space of the traditional method, but due to the combinatorial nature of the problem, that particular solution was not obtained in the traditional method during the optimization process. The focused optimal sub-space developed in this study made it easy for the optimizer to determine a better combination of significant nodes, as observed in the results of proposed method.

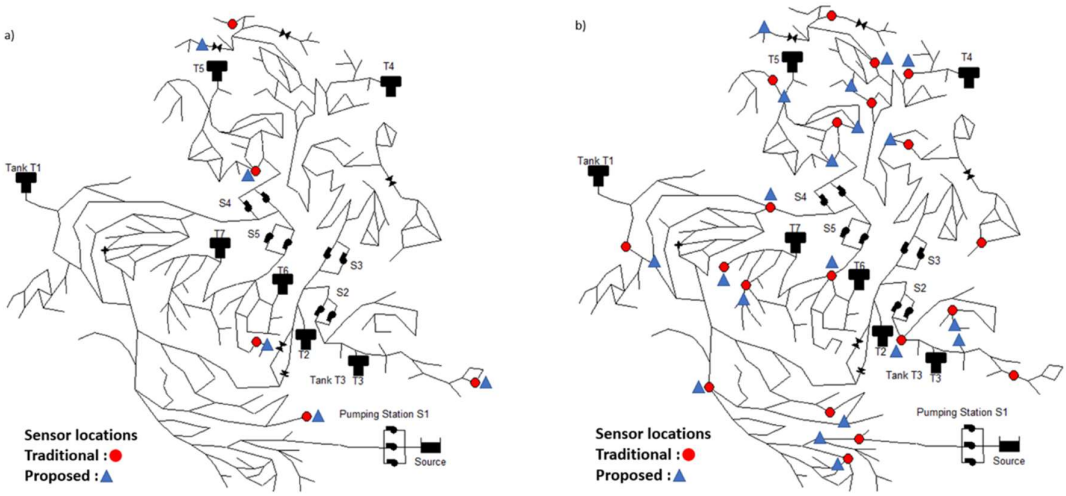


Figure 11 C town Sensor node location for 5 and 20 sensors for Z1 vs Z3 study at Z1 = 50%

The repeatability of the pre-selection strategy was tested out by running the selection module 100 times and estimating the percentage of the nodes from the final non-dominating Pareto

Fronts selected in all the 100 runs. On average, 80% of the nodes from Pareto Fronts were selected in all the runs, while 90.5% of nodes were selected in at least 85 of runs (Appendix A.7). The average time for finding the optimal subset for BWSN was 3.01 secs and for C-town network 156.57 secs.

4 Conclusion

This study developed a strategy for the pre-selection of nodes to reduce the decision variable space for optimal sensor placement. The network was split into clusters based on a novel heuristics-based approach that incorporates the nodal disparity in detecting a set of contamination events. The selection procedure was two-phased; first, a set of nodes are selected based on their ability to detect events, followed by selecting another set of nodes that improvised the detection time or affected population in combination with the previously selected nodes. The proposed algorithm was tested on BWSN Network1 and C-town network and compared based on the quality of Pareto front produced after MOO to optimize detection likelihood and expected time of detection. The search space was successfully reduced by 45% and 34% for the BWSN and C-town network, respectively and the Pareto front obtained from the optimal subset of nodes was better in terms of CR indicator than the complete set of nodes. The time taken for generating the final Pareto Front was better than the traditional method in 6 of 8 MOO studies carried out in the study. These cases are BWSN Network: 5SPP and 20 SPP for both Z1 vs Z2 and Z1 vs Z3 study, and C-town network: 5SPP and 20 SPP for Z1 vs Z3 study.

The proposed method is robust and can be applied to any network irrespective of its complexity, different design of contamination events, and different objectives with appropriate modifications. The only prerequisites required are the pollution matrices corresponding to the objectives considered for placing sensors.

Further, the selection strategy also guarantees that a major set of significant nodes (i.e., >70%) are always selected. Thus, it can be integrated with the decision support system to monitor and maintain larger WDN. In future studies, the effect of modification and/or improvisation in optimization algorithm specifically to MOSPP by methods like the inclusion of initial guesses, tweaking of optimization parameters etc., will be analysed to enhance the quality of Pareto fronts.

Acknowledgements: Partial financial support was received from the Department of Science and Technology and the European Union. We thank the project partners for their valuable suggestions provided for successful completion of this study.

Funding: This work is funded by DST and EU through the LOTUS Project No: DST/IMRCD/India EU/LOTUS/208/(G) and EU Grant No 820881

References

Adedoja, O.S., Hamam, Y., Khalaf, B., Sadiku, R., 2019. A state-of-the-art review of an optimal sensor placement for contaminant warning system in a water distribution network. Urban Water J. 15, 985–1000. https://doi.org/10.1080/1573062X.2019.1597378

Aral, M.M., Guan, J., Maslia, M.L., 2010. Optimal design of sensor placement in water distribution networks. J. Water Resour. Plan. Manag. 136, 5–18. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000001

Berry, J., Hart, W.E., Philips, C.A., Uber, J.G., Watson, J.P., 2006. Sensor placement in municipal water networks with temporal integer programming models. J. Water Resour. Plan. Manag. 132, 218–224. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:4(218)

Berry, J.W., Hart, W.E., Phillips, C.A., Watson, J., 2006. A facility location approach to sensor placement optimization., in: 8th Annual Water Distribution Systems Analysis Symposium. pp. 1–4.

Blockeel, H., Kersting, K., Nijssen, S., Zelezny, F., 2012. A Revised Publication Model for ECML PKDD.

Cao, Y., Smucker, B.J., Robinson, T.J., 2015. On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design. J. Stat. Plan. Inference 160, 60–74. https://doi.org/10.1016/j.jspi.2014.12.004

Chastain, J.R., 2006. Methodology for Locating Monitoring Stations to Detect Contamination in Potable Water Distribution Systems. J. Infrastruct. Syst. 12, 252–259. https://doi.org/10.1061/(ASCE)1076-0342(2006)12:4(252)

Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P.R., Walski, T., 2019. Real time control of water distribution networks : A state-of-the-art review. Water Res. 161, 517–530. https://doi.org/10.1016/j.watres.2019.06.025

Custódio, A.L., Madeira, J.F.A., Vaz, A.I.F., Vicente, L.N., 2011. Direct multisearch for multiobjective optimization. SIAM J. Optim. 21, 1109–1140. https://doi.org/10.1137/10079731X

Di Nardo, A., Giudicianni, C., Greco, R., Herrera, M., Santonastaso, G.F., Scala, A., 2018. Sensor Placement in Water Distribution Networks based on Spectral Algorithms 3, 593–584. https://doi.org/10.29007/whzr

Diao, K., Rauch, W., 2013. Controllability analysis as a pre-selection method for sensor placement in water distribution systems. Water Res. 47, 6097–6108. https://doi.org/10.1016/j.watres.2013.07.026

Eliades, D., Polycarpou, M., 2006. Iterative deepening of Pareto solutions in water sensor networks, in: 8th Annual Water Distribution Systems Analysis Symposium. pp. 1–19.

Guan, J., Aral, M.M., Maslia, M.L., Grayman, W.M., 2006. Optimization model and algorithms for design of water sensor placement in water distribution dystems. J. Water Resour. Plan. Manag. 1–16.

Hart, W., Berry, J., Phillips, C., Watson, J.-P., Avi, O., Uber, J., Salomons, E., 2007. The Battle of the Water Sensor Networks (BWSN): A design challenge for engineers and algorithms. J. Water Resour. Plan. Manag. 134, 556–568.

Hu, C., Dai, L., Yan, X., Gong, W., Liu, X., Wang, L., 2018a. Modified NSGA-III for sensor placement in water distribution system. Inf. Sci. (Ny). 0, 1–13. https://doi.org/10.1016/j.ins.2018.06.055

Hu, C., Li, M., Zeng, D., Guo, S., 2018b. A survey on sensor placement for contamination detection in water distribution systems. Wirel. Networks 24, 647–661. https://doi.org/10.1007/s11276-016-1358-0

Hu, C.Y., Tian, D.J., Liu, C., Yan, X., 2015. Sensors placement in water distribution systems based on co-evolutionary optimization algorithm. Proc. 2015 1st Int. Conf. Ind. Networks Intell. Syst. INISCom 2015 7–11. https://doi.org/10.4108/icst.iniscom.2015.258402

Janke, R., 2018. Threat Ensemble Vulnerability Assessment-Sensor Placement Optimization-Sensor Placement Optimization Tool (TEVA-SPOT) Graphical User Interface Threat Ensemble Vulnerability Assessment-Sensor Placement Optimization-Sensor Placement Optimization Tool (T. https://doi.org/10.13140/RG.2.2.18849.71521

Khorshidi, M.S., Nikoo, M.R., Sadegh, M., 2018. Optimal and objective placement of sensors in water distribution systems using information theory. Water Res. 143, 218–228. https://doi.org/10.1016/j.watres.2018.06.050

Klise, K.A., Phillips, C.A., Janke, R.J., 2013. Two-Tiered Sensor Placement for Large Water Distribution Network Models. J. Infrastruct. Syst. 19, 465–473. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000156

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., Faloutsos, C., 2008. Efficient sensor placement optimization for securing large water distribution networks. J. Water Resour. Plan. Manag. 134, 516–526. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:6(516)

Lloyd, S.P., 1982. Least Squares Quantization in PCM. IEEE Trans. Inf. Theory 28, 129–137. https://doi.org/10.1109/TIT.1982.1056489

Mandel, P., Maurel, M., Chenu, D., 2015. Better understanding of water quality evolution in water distribution networks using data clustering. Water Res. 87, 69–78. https://doi.org/10.1016/j.watres.2015.08.061

Marlim, M.S., Kang, D., 2021. Optimal water quality sensor placement by accounting for possible contamination events in water distribution networks. Water (Switzerland) 13. https://doi.org/10.3390/w13151999

Nurani, S., Abdul, H., Lau, W., 2018. Sensors and Actuators B : Chemical Detection of contaminants in water supply : A review on state-of-the-art monitoring technologies and their applications. Sensors Actuators B. Chem. 255, 2657–2689. https://doi.org/10.1016/j.snb.2017.09.078

Preis, A., Ostfeld, A., 2007. Multiobjective sensor design for water distribution systems security. 8th Annu. Water Distrib. Syst. Anal. Symp. 2006 107. https://doi.org/10.1061/40941(247)107

Rosmann, L., 2000. EPANET 2 user's manual, [WWW Document].

Storey, M. V, Gaag, B. Van Der, Burns, B.P., 2010. Advances in on-line drinking water quality monitoring and early warning systems. Water Res. 45, 741–747. https://doi.org/10.1016/j.watres.2010.08.049

Xu, J., 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. J. Water Resour. Plan. Manag. 134, 378–385. https://doi.org/10.1061/(asce)0733-9496(2008)134:4(378)

Xu, X., Lu, Y., Huang, S., Xiao, Y., Wang, W., 2013. Incremental sensor placement optimization on water network. Lect. Notes Comput. Sci. (including Subser. Lect. Notes

Artif. Intell. Lect. Notes Bioinformatics) 8190 LNAI, 467–482. https://doi.org/10.1007/978-3-642-40994-3_30

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. J. Netw. Comput. Appl. 103, 274–279. https://doi.org/10.1016/j.jnca.2017.10.018