Optimal sensor placement for pipe burst detection in water distribution systems using cost–benefit analysis

Mengke Zhao, Chi Zhang, Haixing Liu, Guangtao Fu IWA and Yuntao Wang

ABSTRACT

Fast detection of pipe burst in water distribution systems (WDSs) could improve customer satisfaction, increase the profits of water supply and more importantly reduce the loss of water resources. Therefore, sensor placement for pipe burst detection in WDSs has been a crucial issue for researchers and practitioners. This paper presents an economic evaluation indicator named as net cost based on cost–benefit analysis to solve the optimal pressure sensor placement problem. The net cost is defined as the sum of the normalized optimal detection uncovering rate and investment cost of sensors. The optimal detection uncovering rate and the optimal set of sensor locations are determined through a single-objective optimization model that maximizes the detection coverage rate under a fixed number of sensors. The optimal number of sensors is then determined by analyzing the relationship between the net cost and the number of sensors. The proposed method is demonstrated to be effective in determining both the optimal number of sensors and their locations on a benchmark network Net3. Moreover, the sensor accuracy and pipe burst flow magnitude are shown to be key uncertainties in determining the optimal number of sensors.

Key words | burst detection, cost–benefit analysis, net cost, pressure sensor placement, water distribution systems

Mengke Zhao
Chi Zhang
Haixing Liu (corresponding author)
Yuntao Wang
School of Hydraulic Engineering,
Dalian University of Technology,
Dalian, Liaoning 116023,
China
E-mail: *hliu@dlut.edu.cn*

Guangtao Fu IWA
Center for Water Systems, College of Engineering, Mathematics, and Physical Sciences,
University of Exeter,
North Park Rd., Exeter EX4 4QF,
UK

INTRODUCTION

Water resource is one of the most important necessities for human survival and development. However, many countries in the world have been facing a severe shortage of water resources due to rapid population growth and urbanization, economic growth, environmental pollution and climate change (Huang *et al.* 2013). It is estimated by the United Nations that there will be an additional 3.5 billion people with most of the growth in developing countries that have been already suffering from water stress by 2050 (Salinas 2014). Water distribution systems (WDSs), as a critical infrastructure, play a vital role in safely delivering drinking water to customers. However, many networks have huge water losses, pipe bursts and leakage due to pipe aging, inadequate construction quality or lack of maintenance, which further exacerbates the water scarcity. Hence, developing effective and efficient detection methods of the anomaly events in WDSs is a proactive means to reduce water losses and improve water resources management. Moreover, it could also help the water utilities cut operating costs, improve water efficiency and customer satisfaction.

There are three main types of detection methods in pipe burst detection, i.e. equipment-based methods (Muntakim *et al.* 2017), transient-based methods (Ferrante *et al.* 2007; Sun *et al.* 2016) and data-driven methods (Bicik *et al.* 2010; Laucelli *et al.* 2013; Bakker *et al.* 2014; Jung *et al.* 2015; Laucelli *et al.* 2015). Due to the recent advances in sensor technology, information communication technology and cutting-edge artificial intelligence, data-driven methods are increasingly developed for anomaly detection of WDSs in recent years (Wu & Liu 2017). Data-driven methods highly rely on real-time monitoring hydraulic data, such as pipe flows and nodal pressures, from supervisory control and

doi: 10.2166/hydro.2020.158

data acquisition systems. These data follow periodically daily patterns under the normal demand pattern and abnormal events could cause changes different from the normal trend (Wu & Liu 2017). Therefore, machine learning methods such as artificial neural network and support vector machine are developed to predict pressures/flows, which are then compared with observed values to detect pipe bursts based on given thresholds (Mounce & Machell 2006; Mounce *et al.* 2010). Statistical process control methods are also used to detect pipe bursts by analyzing observed hydraulic values based on mean values and standard deviations of historical hydraulic values (Jung *et al.* 2015). Real-time pressure data from multiple sensors are fed into trained convolutional neural networks to detect and localize leaks (Zhou *et al.* 2019). Therefore, this raises a research question of how to design a sensor network that can effectively detect pipe burst events in WDSs within a given budget.

Various sensor placement models have been developed over the past two decades. Most sensor placement studies aimed for the purpose of contamination detection (Weickgenannt *et al.* 2009; Sankary & Ostfeld 2018) or hydraulic model calibration (de Schaetzen *et al.* 2000; Ostfeld *et al.* 2008; Behzadian *et al.* 2009), only a few aimed for the purpose of pipe burst detection (Farley *et al.* 2010; Raei *et al.* 2018). With regard to the purpose of contamination detection, the main objectives of sensor placement are the expected time of detection, expected population affected and expected volume of contaminated water consumed prior to detection and the detection likelihood (Ostfeld *et al.* 2008). With regard to the purpose of model calibration, most developed sensor placement methods aimed at finding the optimal set of sensor locations that are more sensitive to calibration parameters using a ranking (de Schaetzen *et al.* 2000) or optimization method (Behzadian *et al.* 2009). However, sensor placement for pipe burst detection is normally set up to find out leak events by identifying the variation of flow and/or pressure. A Jacobian sensitivity matrix of pressure change was used as a function of burst magnitude at given locations, and then, an optimization algorithm, such as the genetic algorithm and linear programming, was used to determine the optimal set of sensor locations (Casillas *et al.* 2013). Sarrate *et al.* (2014) proposed a depth-first search method combined with a *k*-means clustering algorithm to reduce the size and the complexity of the sensor placement optimization problem for the applicability to large-scale WDSs. Blesa *et al.* (2015) evaluated the robustness of sensor placement to different leak magnitudes and several operating points based on the proposed robustness percentage index and then established a multi-objective optimization model including maximization of both the mean and the worst leak locatability index according to the robustness analysis results. Hagos *et al.* (2016) proposed a single-objective optimal sensor placement model that maximizes the detection probability (DP) or minimizes the rate of false alarm (RF). The obtained result shows that pressure sensors cannot detect all the pipe bursts even if a large number of sensors are used and the optimal set of sensor locations differs substantially under the maximum DP and the minimum RF. However, it does not consider multiple simultaneous bursts for the generation of pipe burst events. Based on this trade-off between the DP and the RF, the total cost and mechanical reliability of the sensor network were also further considered to formulate a multi-objective optimal sensor placement model (Jung & Kim 2018).

However, all aforementioned studies have a common assumption that the total number of sensors is fixed in the sensor placement problem. There are only few studies focusing on the determination of the optimal number of sensors. Soroush & Abedini (2019) proposed a single-objective optimal sensor placement model to determine the optimal sensor locations by minimizing the variance of residuals between the estimated and true values of average pressure. The variation of the optimal variance of residuals against different numbers of sensors was analyzed to determine the optimal number of sensors. However, it does not take into account the cost of sensors. In addition, several studies introduce the cost of sensors to a multi-objective optimization model to determine the optimal number of sensors (Weickgenannt *et al.* 2009; Simone *et al.* 2016; Raei *et al.* 2018). Weickgenannt *et al.* (2009) formulated the water quality sensor placement problem as a twin-objective optimization problem including the minimization of risk of the contamination and sensor cost. The variables include the number of sensors and their locations. Based on the network topology and weights assigned to pipes, a novel multi-objective optimal sensor placement model was proposed to maximize the sampling-oriented modularity index and minimize the cost of pressure sensors (Simone *et al.* 2016). Raei *et al.* (2018) formulated the pressure sensor placement as a twin-objective optimization problem including the minimization of the number of sensors and the detection time of pipe leaks. Obviously, considering both the location and number of sensors is practical and valid in the sensor placement issues. Solving the bi-objective optimization problem of sensor placement derives a trade-off relationship between the cost and benefit of sensor placement. However, it is difficult to determine how many sensors are required given the complex trade-off and the large number of optimal solutions.

Cost–benefit analysis (CBA) is a systematic approach to conduct economic analysis for engineering projects and has been widely applied to the WDSs management (Steffelbauer & Fuchs-Hanusch 2016; Creaco & Walski 2017). Creaco & Walski (2017) analyzed the economic effectiveness of two different pressure control solutions including conventional mechanical pressure reducing valves and remotely real-time pressure control valves for leakage and pipe burst reduction based on CBA. The CBA approach for the sensor placement problem was developed to estimate the relationships between the number of sensors as a surrogate of costs and the percentage of detected leak scenarios as a surrogate of benefits (Steffelbauer & Fuchs-Hanusch 2016), which allows the decision-maker to choose the number of required sensors given a specific level of leakage detection.

In this paper, we propose an economic evaluation indicator named as net cost based on CBA, defined as the sum of the normalized optimal detection uncovering rate and investment cost of sensors, to solve the problem of determining the optimal number of sensors and the set of sensor locations in WDSs. First, the optimal detection uncovering rate and the set of sensor locations are determined through a single-objective optimization model. The model maximizes the detection coverage rate under a fixed number of sensors and is solved by the fast messy Genetic Algorithm (fmGA). The optimization process is performed individually for different scenarios of different numbers of sensors. The net cost is then calculated by the sum of the normalized detection uncovering rate and investment cost of pressure sensors. The optimal number of sensors is determined by analyzing the relationship between the net cost and the number of sensors. The proposed net cost indicator is demonstrated on a benchmark network Net3 and used to investigate the impact of uncertainties of sensor accuracy and burst flow magnitude on the optimal number of sensors. This new method can provide informed decisions on finding the optimal number of sensors and their locations in a large network that can accurately and effectively detect pipe bursts should they occur.

METHODOLOGY

Problem formulation

Pipe burst event detection

Pipe burst can incur an unexpected demand increase at some nodes, and thus result in variations in pressure measurements. Assuming the WDSs installed with a given number (N_0) of pressure sensors at different locations, a pipe burst event can be detected by comparing the differences between the real-time measured pressures by sensors and the corresponding simulated pressures under the base diurnal variation pattern with detection thresholds. The difference between the measured and simulated pressures, i.e. residual, is calculated as below:

$$R_i(t) = P_i^{\mathrm{L}}(t) - P_i^{\mathrm{B}}(t) \quad i = 1, 2, \ldots, N_0 \tag{1}$$

where $R_i(t)$ is the pressure residual at ith sensor node at time step t; $P_i^{\mathrm{L}}(t)$ is the measured pressure after the event occurs; $P_i^{\mathrm{B}}(t)$ is the corresponding simulated pressure during an extend period simulation under a base diurnal variation pattern; N_0 is the number of sensors. Nodal demand always fluctuates on the basis of a base diurnal variation pattern under the normal operational state. The normal pressure difference between the normal operational state and the base diurnal variation pattern could be determined.

Here, the detection threshold at each sensor node for pipe burst events is defined as the sum of sensor accuracy and normal pressure difference, as shown below:

$$\Delta P_i^{\mathrm{threshold}}(t) = \delta_{\mathrm{sensor}} + |\Delta P_i^{\mathrm{N}}(t)| \tag{2}$$

where $\Delta P_i^{\mathrm{threshold}}(t)$ is the detection threshold at ith sensor node at time step t; δ_{sensor} is the sensor accuracy; $\Delta P_i^{\mathrm{N}}(t)$ is the normal pressure difference at ith sensor node at time step t. If the residual pressure by at least one sensor exceeds its detection threshold, it is considered to be a pipe burst.

In this paper, it is assumed that nodal demand always follows the same diurnal variation pattern on different days under the normal operational state. Thus, the normal pressure difference $\Delta P_i^{\mathrm{N}}(t)$ is equal to 0, which means that the detection threshold for the simulated pipe burst event is determined only by the sensor accuracy.

Sensor placement cost

In general, the investment cost (IC) of sensors consists of capital, maintenance and operational costs. The unit capital sensor cost $C(\delta)$ is related to the sensor accuracy. It is assumed that the overall investment of sensors ($\mathrm{IC_{sp}}$) is the unit capital sensor cost multiplied by the number of sensors (N) as below:

$$\mathrm{IC_{sp}} = C(\delta) \times N \tag{3}$$

where N is a variable representing the number of sensors. The other costs are not taken into account here due to the lack of the local price information.

Detection coverage rate

The performance of sensor placement with respect to pipe bursts is evaluated by the index of detection coverage rate, which reflects the effectiveness of a set of sensor locations. Here, the detection coverage rate is defined as the ratio of detectable events to the total simulated events, which is same as the DP proposed by Jung *et al.* (2015):

$$\mathrm{DCR} = \frac{\mathrm{NS_{detected}}}{\mathrm{NS_{total}}} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{\mathrm{NS_{total}}} D_{i,k}}{\mathrm{NS_{total}}} \tag{4}$$

where DCR is the detection coverage rate, which ranges from 0 to 1; $\mathrm{NS_{detected}}$ is the total number of detectable events by the set of sensors; i is the sensor index and k is the event index; $D_{i,k}$ is equal to 0 or 1 and it is calculated using Equation (5); $\mathrm{NS_{total}}$ is the total number of pipe burst events. Therefore, if the value of DCR is greater, the set of sensors can detect more pipe burst events and the deployment of the sensors is more effective:

$$D_{i,k} = \begin{cases} 1 & \text{if}(E_k \text{ is detected by } S_i \text{ but not by } \{S_1, \ldots, S_{i-1}\}) \\ 0 & \text{else} \end{cases} \tag{5}$$

where E_k is the kth event and S_i is the ith sensor. This ensures that the value of $\mathrm{NS_{detected}}$ is always no greater than the $\mathrm{NS_{total}}$.

Solution methods

An optimal sensor placement problem needs to determine the minimum number of sensors and their best locations with the purpose of detecting as many burst events as possible. This paper presents an economic indicator named as net cost based on the single-objective optimization model and CBA to solve the optimal pressure sensor placement problem. A two-stage decision-making framework based on the proposed indicator is illustrated in Figure 1.

Stage 1: optimization for the set of sensor locations using fmGA

When the number of sensors (N_0) is given, the optimal sensor placement is formulated as a single-objective optimization problem. The decision variables used here are the locations of sensors, denoted by $X_{N_0} = [x_1, x_2, \ldots, x_{N_0}]$, which should be the candidate node ID in WDSs. The objective function is the maximization of DCR given by Equation (4).

To evaluate the DCR of different sets of sensor locations, the database of the synthetic pipe burst events is generated initially. Due to the event being difficult to mimic in the real-world WDSs, the pressure measurements are simulated by a hydraulic model of the network. Here, pipe burst is simulated at a random node in the WDSs and concurrent multiple pipe bursts are taken into account in the study. All the nodes can be chosen as the potential burst locations except the reservoirs and tanks. The Monte Carlo method (Gasparini 1997) is employed to stochastically generate the number of burst events, the position of the increasing demand (i.e. node ID).

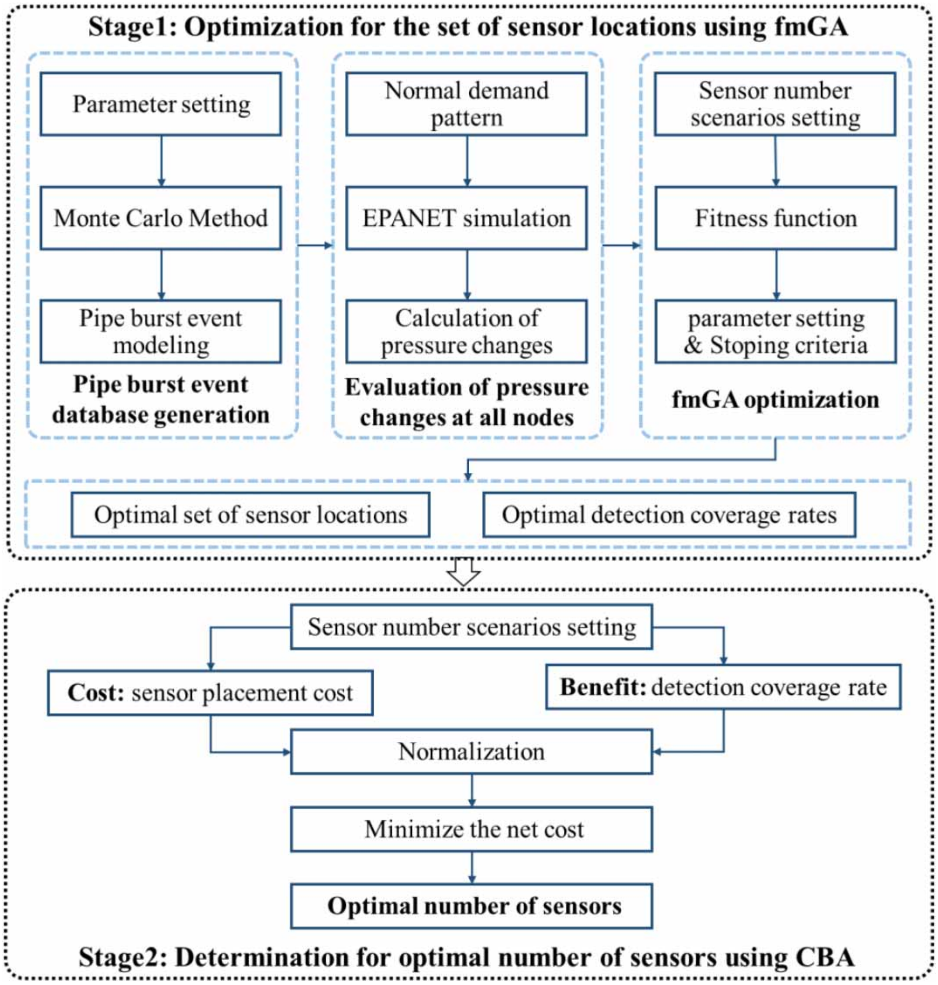

Figure 1 | Framework of the two-stage decision-making method for optimal sensor placement.

EPANET 2 is an open-source software developed by the United States Environmental Protection Agency and can be used for extended period simulation of hydraulic and water quality behaviors within pressurized pipe networks (Rossman 2000). Therefore, the EPANET 2 is employed to calculate the pressure changes at all nodes under the both normal demand pattern and synthetic pipe burst events.

For the given number of sensors, the fmGA is selected here to optimize the sensor placement model. As a variant of the simple Genetic Algorithm (sGA), the fmGA was developed by Goldberg *et al.* (1993) to solve the large-scale and high-dimensional optimization problem. In contrast to the uniform-length string representation of sGA, fmGA uses a variable-length string representation. The optimization process consists of an outer loop and an inner loop. Each outer loop corresponds to an inner loop, including initialization, a primordial phase and a juxtapositional phase. In comparison with the messy Genetic Algorithm (Goldberg *et al.* 1989), the probabilistically complete initialization method is used to substitute the enumeration method to generate the initial population in the initialization phase, which effectively reduces the size of the initial population and avoids the excessive memory use. In the primordial phase, the lengths of the initial population are periodically reduced using the build block filtering method to obtain the population which can be juxtaposed to obtain an optimal individual with high probability. In the juxtapositional phase, the cut-and-splice operator is used to reproduce offspring for the next generation. The key parameters for the fmGA are population size, generation number within the inner loop, cut rate and splice rate. Previous researches have proved that the fmGA outperforms the sGA when applying to the optimization of WDSs (Wu & Simpson 2001).

The optimal DCR and the corresponding set of sensor locations under different numbers of sensors are determined at Stage 1 and the results from Stage 1 are used to determine the optimal number of sensors at Stage 2.

Stage 2: determination for the optimal number of sensors using CBA

CBA is used to determine the optimal number of sensors at this stage. In this paper, the maximal value of sensor number is denoted by N_{max} and the minimal one is denoted by N_{min} within all the cases of sensor numbers. For a case of sensor number (N), the cost can be represented by the total IC of sensors, denoted by $IC_{sp}(\delta,N)$ in Equation (3), and the optimal DCR is employed to represent the benefit, denoted by $DCR(N)_{opt}$. Due to the difference of units between the $IC_{sp}(\delta,N)$ and the $DCR(N)_{opt}$, the cost and expected benefit are normalized using Equations (6) and (7):

$$\mathrm{IC}^{\mathrm{nor}}(\delta, N) = \frac{\mathrm{IC}(\delta, N) - \min \mathrm{IC}}{\max \mathrm{IC} - \min \mathrm{IC}} \tag{6}$$

$$\mathrm{DCR}^{\mathrm{nor}}(N)_{\mathrm{opt}} = \frac{\mathrm{DCR}(N)_{\mathrm{opt}} - \min \mathrm{DCR}_{\mathrm{opt}}}{\max \mathrm{DCR}_{\mathrm{opt}} - \min \mathrm{DCR}_{\mathrm{opt}}} \tag{7}$$

where $\mathrm{IC}^{\mathrm{nor}}(\delta, N)$ and $\mathrm{DCR}^{\mathrm{nor}}(N)_{\mathrm{opt}}$ are the normalized cost and benefit, respectively; max IC and max DCR_{opt} are the maximum cost and maximum benefit, respectively; min IC and min DCR_{opt} are the minimum cost and minimum benefit, respectively. Assumed that uniform sensors (i.e. uniform measurement accuracy) are employed for the sensor placement, Equation (6) is simplified to Equation (8) based on Equation (3):

$$\mathrm{IC}^{\mathrm{nor}}(\delta, N) = \frac{C(\delta) \times N - C(\delta) \times N_{\min}}{C(\delta) \times N_{\max} - C(\delta) \times N_{\min}} = \frac{N - N_{\min}}{N_{\max} - N_{\min}} \tag{8}$$

where N_{max} and N_{min} are the maximum and minimum number of sensors. The maximum number of sensors can be determined by analyzing the variation of optimal DCR with the increase in the number of sensors, which has an impact on the determination of the optimal number of sensors. When the optimal DCR does not show a significant improvement, the maximum number of sensors is determined. The maximum number of sensors has an impact on the determination of the optimal number of sensors.

In order to determine the optimal number of sensors, the optimal detection uncovering rate (UCR), denoted by $UCR(N)_{opt}$, is introduced here:

$$\mathrm{UCR}(N)_{\mathrm{opt}} = 1 - \mathrm{DCR}(N)_{\mathrm{opt}} \tag{9}$$

Similar to Equation (7), the normalized UCR, $\mathrm{UCR}^{\mathrm{nor}}(N)_{\mathrm{opt}}$ can be obtained.

The net cost (NC) of sensor placement is the difference between the normalized cost and benefit, which can be also denoted by the sum of the normalized UCR and IC, as shown below:

$$\begin{aligned} \mathrm{NC}(N) &= \mathrm{IC}^{\mathrm{nor}}(\delta, N) - \mathrm{DCR}^{\mathrm{nor}}(N)_{\mathrm{opt}} \\ &= \mathrm{IC}^{\mathrm{nor}}(\delta, N) + \mathrm{UCR}^{\mathrm{nor}}(N)_{\mathrm{opt}} \end{aligned} \tag{10}$$

where NC(N) is the NC of sensor placement. The optimal number of sensors, denoted by N^*, can be determined by analyzing the relationship between the NC and the number of sensors.

CASE STUDY

The proposed two-stage decision-making method of sensor placement is applied to a well-known example network from the literature, i.e. Net3, as shown in Figure 2 (Diao et al. 2016). It consists of 92 junctions, 117 pipes, 2 reservoirs (one lake and one river) and 3 tanks. The total mean demand for the Net3 network is approximately 10,950 GPM.

In this case network, some simplifications are made in the simulation of pipe burst: (1) here, all the burst events are assumed to be located at nodes of WDSs, which might result in the loss of accuracy and information at the pipe level (Hagos et al. 2016; Raei et al. 2018); (2) all of the nodes are considered to be the possible leak nodes except the reservoirs and tanks; and (3) pipe burst is simulated by adding an extra demand at the target node (Casillas et al. 2013; Li et al. 2019). The minimum and maximum of the

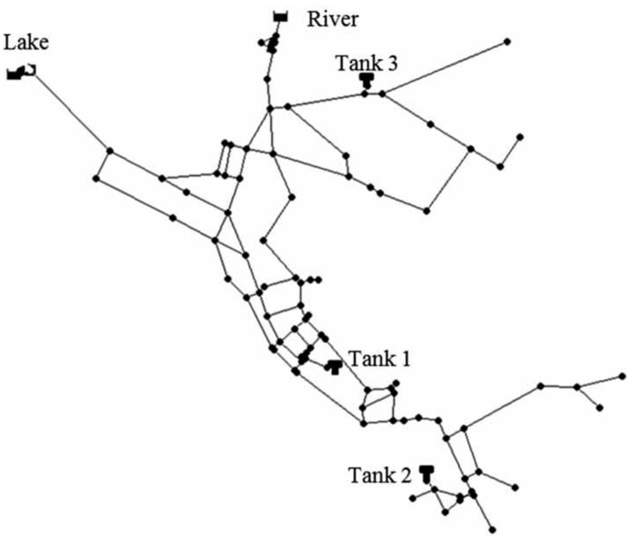

Figure 2 | Layout of the Net3 network.

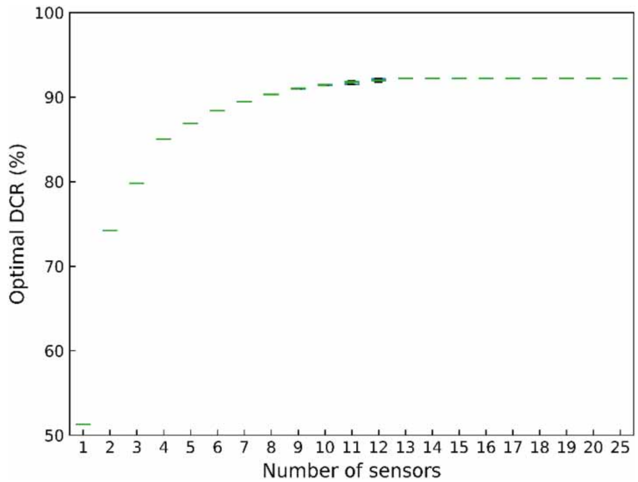

Figure 3 | Variation of the optimal DCR with the number of sensors.

extra demands, which are used to represent the variation range of burst flow, are set to 50 and 100 GPM, respectively. The resulting burst magnitudes range from 0.46% to 0.92% of the total mean demand. Only single or double leaks are considered in this study since more than two simultaneous bursts are unlikely to occur in a network in practice. In total, 1,000 synthetic pipe burst events are generated using the Monte Carlo method, i.e. by randomly sampling the number of leak nodes (i.e. 1 or 2 leaks), their locations and their burst flows. The sensor accuracy is initially set to 0.05 psi by which the pressure changes can be evaluated in comparison with the normal operation mode of pressures, and other sensor accuracies are considered in the impact analysis.

RESULTS AND DISCUSSION

Optimal number of sensors and set of sensor locations

To analyze the variation of optimal DCR with an increasing number of sensors, Figure 3 shows the boxplot from 10 random runs. It shows that for a given number of sensors, the optimal DCRs obtained by 10 random runs are almost the same, which indicates that the optimization process is converged well with the fmGA. Moreover, the optimal DCR increases with the increase in the number of sensors, but increase slowly after 12 sensors. The optimal DCR is approximately 92.23% after 25 sensors; thus, there is no benefit in deploying more sensors. This is because some of the pipe burst events cannot be detected by the deployed sensors due to the sensor accuracy and the burst locations. This result agrees with that obtained by Hagos *et al.* (2016). However, with the improvement of the sensor accuracy, the optimal DCR can be improved further as illustrated in the next section.

Figure 4 shows how frequent a node is chosen to be the location for sensor deployment, based on the optimal solutions obtained. Recall that there are 21 cases with respect to the number of sensors and for each case 10 runs are conducted with the fmGA. Therefore, in total 210 solutions of sensor placement are obtained. There are 11 sensor locations which have been chosen over 60 times. This demonstrates that these nodes are critical locations for sensor placement which are more sensitive to simulated pipe burst events and are highlighted in the topology of Net3 in Figure 5. It demonstrates that the sensors tend to be placed at the center of the WDSs. From the perspective of a topological analysis, the sensors at the topological center of the network can reach each boundary of the network with the shortest paths and access to the variation of pressures at nodes. Therefore, the sensors deployed at these locations can cover more pipe burst events.

To determine the optimal number of sensors, the maximum number of sensors, N_{max}, is set to 25 and the minimum one, N_{min}, is 1. The normalized UCR and the

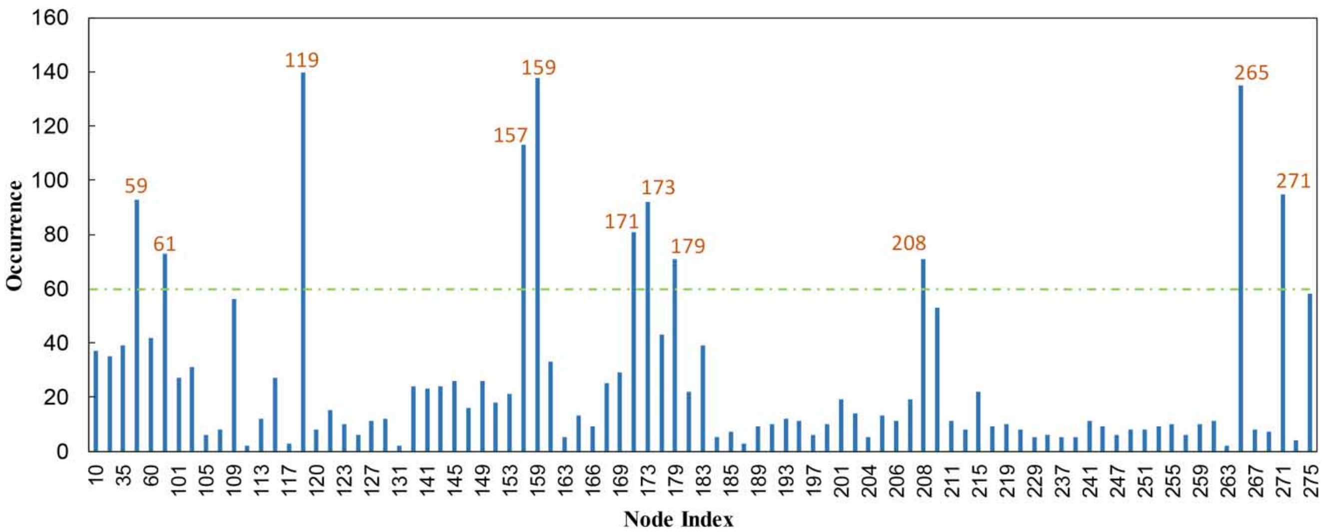

Figure 4 | Occurrences of the optimal pressure sensor locations.

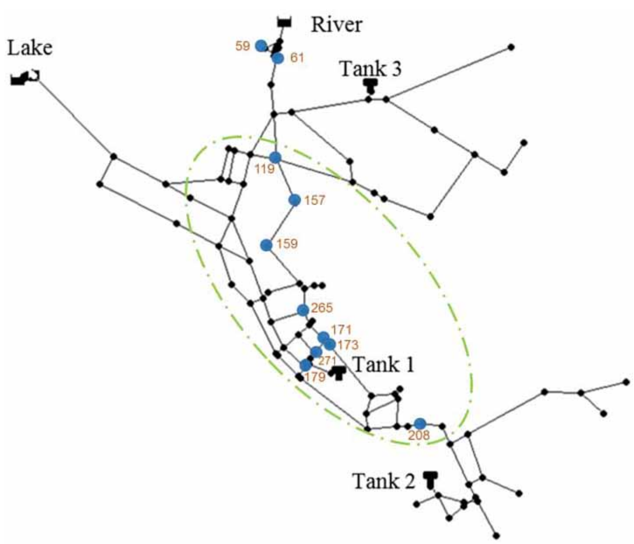

Figure 5 | Critical sensor locations in Net3.

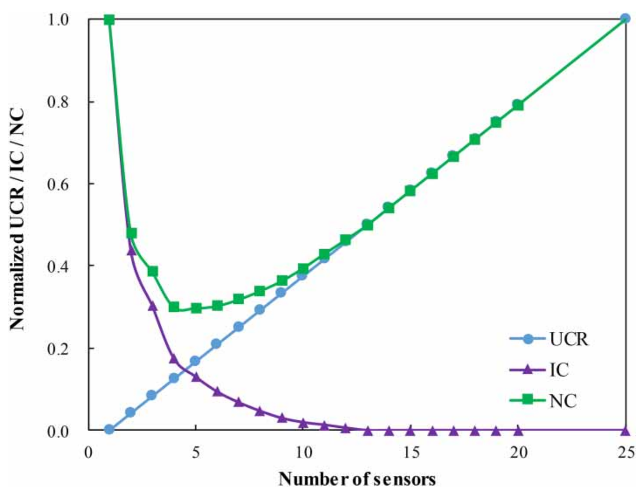

Figure 6 | Determination of the optimal number of sensors using CBA.

normalized IC are calculated by Equations (9) and (10) for different numbers of sensors. Figure 6 shows the variation of the normalized UCR, the normalized IC and the normalized NC when the sensor accuracy is 0.05 psi. The normalized IC is represented by the circular symbols. It increases linearly with the number of sensors because the overall investment of sensors is assumed here to be the unit sensor cost multiplied by the number of sensors. The normalized NC (i.e. the sum of the normalized UCR and investment IC) is represented by the squared symbols in Figure 6. The lowest point of the green curve represents the optimal number of sensors, which is 5 in this case.

Table 1 shows the node ID of optimal sets of sensor locations and corresponding DCRs under 10 random runs of fmGA when the optimal number of sensors is 5. Results show that all of 10 solutions could achieve the equal DCR, which has been demonstrated in Figure 3. However, the optimal set of sensor locations vary slightly with different solutions. Only Node 208 and 265 are chosen to be the optimal sensor locations for all of the optimal sets of sensor locations. It means that sensors at these two nodes are more sensitive to the simulated pipe burst events.

Table 1 | Node ID of optimal sets of sensor locations for five sensors

Solution	1	2	3	4	5	6	7	8	9	10
Loc. 1	10	101	10	109	10	10	115	101	10	101
Loc. 2	149	147	153	147	147	143	147	145	143	143
Loc. 3	171	171	173	171	171	171	171	171	173	173
Loc. 4	208	208	208	208	208	208	208	208	208	208
Loc. 5	265	265	265	265	265	265	265	265	265	265
DCR (%)	86.86	86.86	86.86	86.86	86.86	86.86	86.86	86.86	86.86	86.86

To analyze the topological relation in Net3 among different solutions, the optimal sensor locations of solution 2 (represented by triangles) and other alternative sensor locations (represented by circles) are shown in Figure 7. As can be seen, the optimal sensor locations are evenly distributed in the whole network. This finding is different from the findings obtained in Hagos *et al.* (2016), in which results show that the best locations for sensors are located at the end of the case network. The main reason is the topology difference between two case networks. Comparing with Hagos *et al.*'s case network including only one reservoir, there exist two reservoirs and three tanks in network Net3. They are located in different areas of the Net3, which could reduce the impact of pipe burst events on nodal pressures at the end of the network. In addition, alternative sensor locations obtained by other solutions are located near to the sensor locations of solution 2, which illustrates sensors in these areas could achieve a nearly equal DCR.

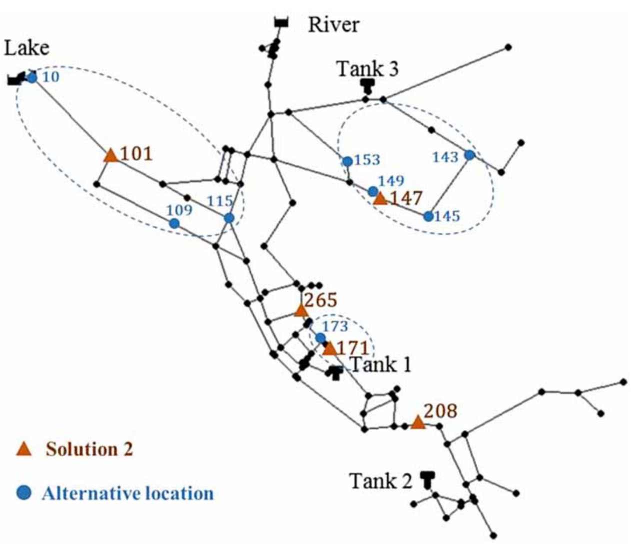

Figure 7 | Optimal sets of sensor locations for five sensors in Net3.

To determine the best set of sensor locations for five sensors, all the optimal solutions obtained by the fmGA under the number of sensors (ranging from 5 to 10) are analyzed to find the nested solutions, as shown in Table 2. When the number of sensors is increased from 5 to 9, the optimal solution for N sensors is always a superset of that for $N-1$ sensors. It means that increasing the number of sensors will not result in a complete variation of the proposed sensor network but only its expansion. Comparing with the solution for nine sensors, the difference of the solution for 10 sensors is that IDs of sensor locations including 149, 171 and 181 are replaced by the IDs of 147, 173 and 183, respectively. According to the topological analysis, they are located near to each other in Net3. Thus, the flexibility of sensor placement can be guaranteed if water utilities have an increase in the budget or network expansion.

Impact of sensor accuracy

According to Equations (3) and (4), the DCR and the IC of the sensors are significantly influenced by the sensor accuracy. To investigate the impact of the sensor accuracies on the determination of the optimal number of sensors, four-sensor accuracies, including 0.03, 0.05, 0.07 and 0.09 psi, are used to test the variation of pressures and analyze the DCRs and the optimal number of sensors. Here, both the single leak and double leak are considered to analyze the impact of sensor accuracy. The minimum and maximum of the extra demands at one leak node are set to 50 and 100 GPM, respectively, for single leak condition. The minimum and maximum of the extra demands at one leak node are set to 25 and 50 GPM, respectively, for double leak condition.

Table 2 | Best sensor locations with different numbers of sensors

Number	ID of sensor location										DCR (%)
5	10	149	171	208	265						86.86
6	10	149	171	208	265	59					88.36
7	10	149	171	208	265	59	159				89.43
8	10	149	171	208	265	59	159	119			90.29
9	10	149	171	208	265	59	159	119	181		90.94
10	10	147	173	208	265	59	159	119	183	271	91.47

Figure 8 shows the optimal DCRs for different sensor accuracies under the single leak condition. It shows that the maximal DCR can approximately reach 93.1% with the sensor accuracy of 0.03 psi, while the maximal one is only 64.7% with the sensor accuracy of 0.09 psi. Under the double leak condition, the optimal DCRs for different sensor accuracies are shown in Figure 9. The results show that the maximal DCR can approximately reach 94.5% with the sensor accuracy of 0.03 psi, while the maximal one is only 54.0% with the sensor accuracy of 0.09 psi. Under both single leak and double leak conditions, the higher the sensor accuracy, the larger the optimal DCR is. For all sensor accuracies, the best DCR appears to have an upper limit. It indicates that the sensor accuracy has a significant impact on the detection coverage effectiveness. To achieve the higher DCR, increasing the sensor accuracy is more effective than increasing the sensor number.

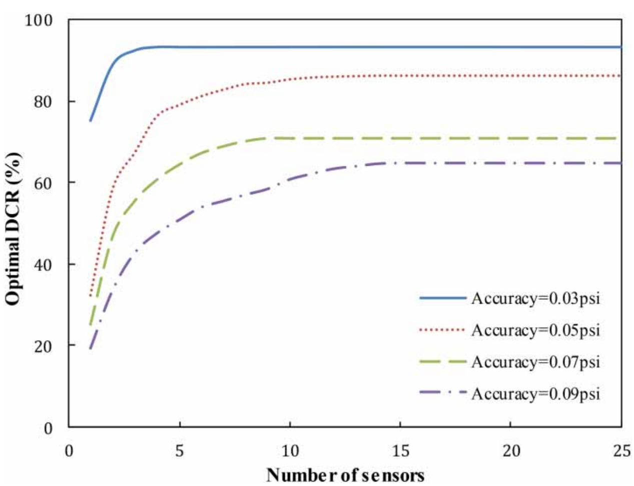

Figure 8 | Optimal DCRs for different sensor accuracies (Single leak condition).

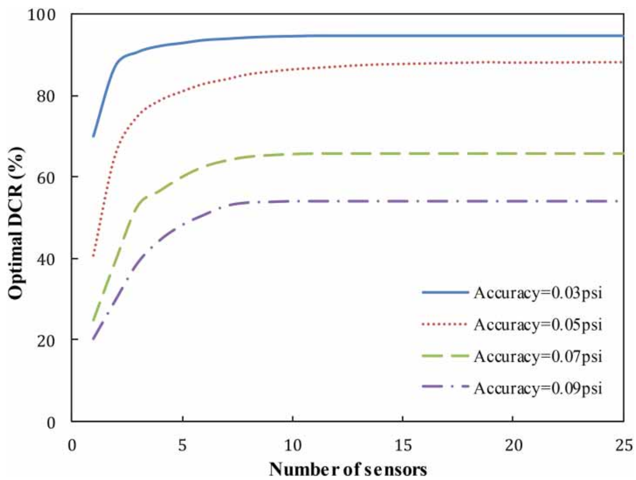

Figure 9 | Optimal DCRs for different sensor accuracies (Double leak condition).

With different sensor accuracies, the normalized NCs (i.e. the sum of the normalized UCR and normalized IC) under the single leak and double leak conditions are shown in Figures 10 and 11, respectively. The results reveal that the optimal number of sensors gradually decreases from 6 to 4 under the single leak condition, and the optimal number of sensors gradually decreases from 6 to 3 under the double leak condition when the sensor accuracy improves from 0.09 to 0.03 psi. The variation is consistent with the practical experience that the sensors with higher accuracy are more sensitive to the variation of pressure and easily detect the occurrence of the pipe burst event. However, high accuracy would incur the extra cost for the sensor network construction. Thus, both the sensor accuracy and the number of sensors should be considered in the sensor placement, and the CBA can illustrate the optimal sensor number in the trade-off relationship between the sensor cost and the optimal DCR.

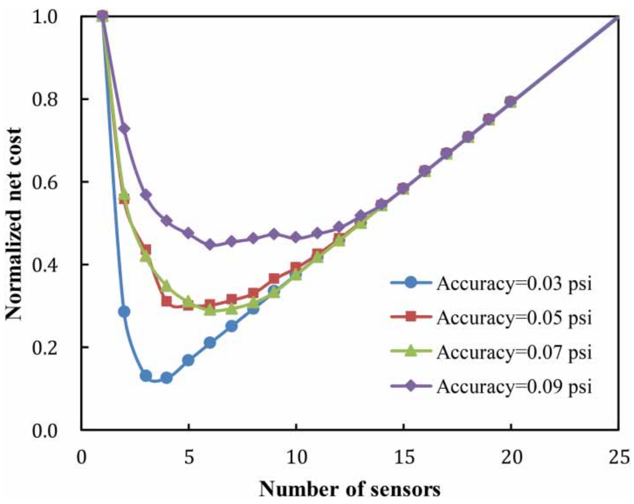

Figure 10 | Optimal number of sensors for different sensor accuracies (Single leak condition).

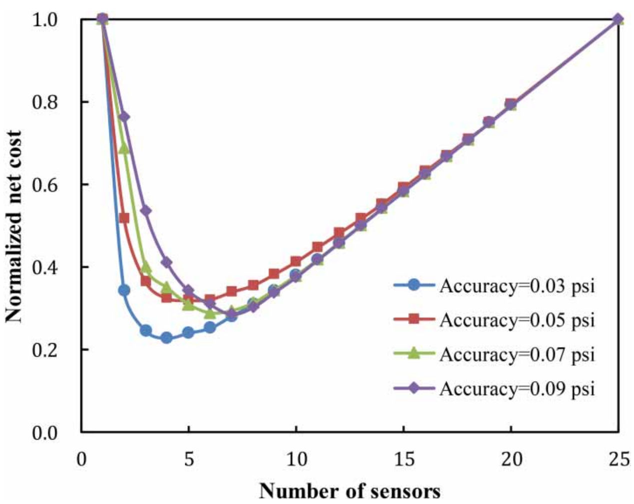

Figure 11 | Optimal number of sensors for different sensor accuracies (Double leak condition).

Impact of the uncertainty of burst flow magnitude

Due to the effect of pipe pressure and burst characteristics (such as cracking size), the burst flow varies significantly. The large burst incidents can severely change the pressures of nodes in WDSs. To analyze the impact of the uncertainty of burst flow on the optimal sensor placement, the uncertainty of burst flow is investigated here. The single leak and double leak are also considered here to analyze the impact of the uncertainty of burst flow magnitude. Four variation ranges of total pipe burst flow are examined, including S1 (30–40 GPM), S2 (50–60 GPM), S3 (70–80 GPM) and S4 (90–100 GPM). The pipe burst flow is sampled from these four ranges. The sensor accuracy is fixed to 0.05 psi. The optimization and CBA are conducted using the same parameters with the section of the case study.

Figures 12 and 13 show the cost–benefit curves for four-pipe burst flow cases under the single leak and the double leak conditions, respectively. Both figures indicate that the optimal number of sensors is reduced with the increase in the pipe burst flow magnitude. If the large bursts frequently happen or only the large burst flows are concerned with, the

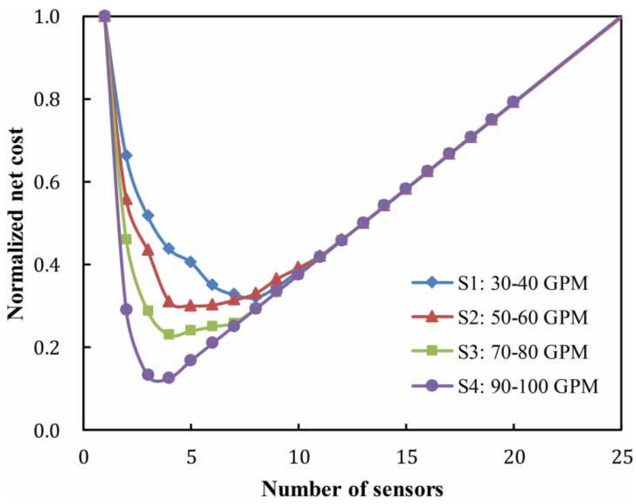

Figure 12 | Optimal number of sensors under different pipe burst flow magnitudes (Single leak condition).

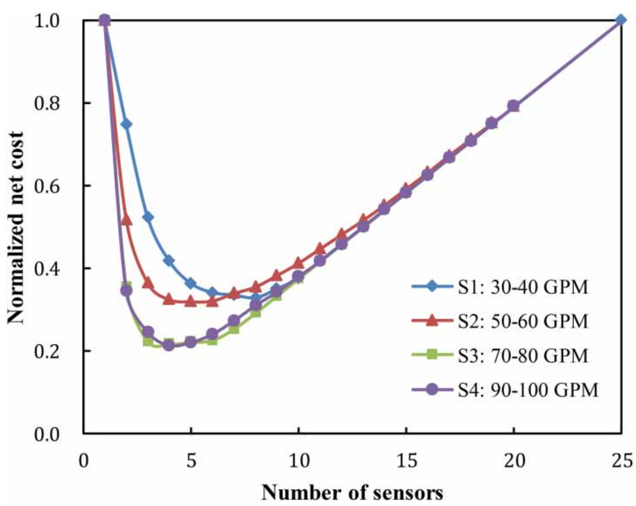

Figure 13 | Optimal number of sensors under different pipe burst flow magnitudes (Double leak condition).

number of sensors to achieve the same detection level can be reduced. By contrast, to detect the smaller burst events, more sensors should be deployed in the network, resulting in the cost increase. Thus, in the practice of the sensor placement, factors such as infrastructure deterioration, external conditions and pipe pressure should be taken into account, which can impact the burst flow magnitude.

CONCLUSIONS

This paper proposes an economic evaluation indicator named net cost, defined as the sum of the normalized optimal detection uncovering rate and investment cost of sensor based on CBA, to determine the optimal number of sensors and optimal set of sensor locations. The benchmark network Net3 is used to evaluate the performance of the proposed economic evaluation indicator. Moreover, the impacts of sensor accuracy and pipe burst flow uncertainty on the optimal sensor placement under both single leak and double leak conditions are analyzed. According to the results obtained from the case study, the key conclusions are drawn as follows:

1. The proposed net cost indicator based on the single-objective optimization model and the CBA method has been proven to be effective in determining the optimal number of sensors by analyzing the relationship between the NC (i.e. the sum of the normalized UCR and IC of sensors) and the number of sensors. Moreover, the critical sensor locations tend to be placed at the center of the WDSs.
2. The sensor accuracy has a significant impact on the detection coverage effectiveness. The optimal number of sensors deployed in WDSs reduces with the improvement of the sensor accuracy. To achieve the higher detection coverage rate, improving the sensor accuracy is more effective than increasing the sensor number.
3. Pipe burst flow magnitude is shown to be a key uncertainty in determining the optimal number of sensors. More sensors are needed to be deployed in effective places to detect small leaks in WDSs. Thus, the burst flow magnitude should be considered in practice for sensor placement.

In future work, the DCR indicator could be further modified by taking into account the number of nodes covered by sensors, which might evaluate the effectiveness of sensor locations more appropriately. Real WDSs cases should be investigated. In terms of pipe burst events, they should be simulated on a specific location of the pipe, which could represent the relationship of burst flow with pipe pressure and the characteristics of the crack. It will be more appropriate than simulating burst at nodes. Although demand-driven analysis has been selected here to calculate variations of nodal pressures under small burst, the pressure-driven analysis must be conducted to calculate more accurate nodal pressures under large burst in real WDSs cases. Burst probabilities of different pipes could be considered by analyzing historical pipe burst data and other factors impacting the pipe burst, such as infrastructure deterioration, external conditions and operating pressure.

ACKNOWLEDGEMENTS

This study is supported by the National Key Research and Development Program (2017YFC0406005) and the National Natural Science Foundation of China (91647201, 91747102 and 51708086) and the UK Royal Society (IEC\NSFC\170249 and IF160108). The third author is funded by the DUT startup grant (DUT18RC(3)072).

REFERENCES

Bakker, M., Vreeburg, J. H. G., Van De Roer, M. & Rietveld, L. C. 2014 Heuristic burst detection method using flow and pressure measurements. *Journal of Hydroinformatics* **16** (5), 1194–1209.

Behzadian, K., Kapelan, Z., Savic, D. & Ardeshir, A. 2009 Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. *Environmental Modelling & Software* **24** (4), 530–541.

Bicik, J., Kapelan, Z., Makropoulos, C. & Savić, D. A. 2010 Pipe burst diagnostics using evidence theory. *Journal of Hydroinformatics* **13** (4), 596–608.

Blesa, J., Nejjari, F. & Sarrate, R. 2015 Robust sensor placement for leak location: analysis and design. *Journal of Hydroinformatics* **18** (1), 136–148.

Casillas, M., Puig, V., Garza-Castañón, L. & Rosich, A. 2013 Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **13** (11), 14984.

Creaco, E. & Walski, T. 2017 Economic analysis of pressure control for leakage and pipe burst reduction. *Journal of Water Resources Planning and Management* **143** (12), 04017074.

de Schaetzen, W. B. F., Walters, G. A. & Savic, D. A. 2000 Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms. *Urban Water* **2** (2), 141–152.

Diao, K., Sweetapple, C., Farmani, R., Fu, G., Ward, S. & Butler, D. 2016 Global resilience analysis of water distribution systems. *Water Research* **106**, 383–393.

Farley, B., Mounce, S. R. & Boxall, J. B. 2010 Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal* **7** (6), 345–356.

Ferrante, M., Brunone, B. & Meniconi, S. 2007 Wavelets for the analysis of transient pressure signals for leak detection. *Journal of Hydraulic Engineering* **133** (11), 1274–1282.

Gasparini, M. 1997 Markov chain Monte Carlo in practice. *Technometrics* **39** (3), 338–338.

Goldberg, D. E., Korb, B. & Deb, K. 1989 Messy genetic algorithms: motivation, analysis, and first results. *Complex Systems* **3** (3), 493–530.

Goldberg, D. E., Deb, K., Kargupta, H. & Harik, G. R. 1993 Rapid Accurate Optimization of Difficult Problems Using Fast Messy Genetic Algorithms. Proceedings of the Fifth International Conference on Genetic Algorithms, pp. 56–64.

Hagos, M., Jung, D. & Lansey, K. E. 2016 Optimal meter placement for pipe burst detection in water distribution systems. *Journal of Hydroinformatics* **18** (4), 741–756.

Huang, Y., Chen, J., Zeng, S., Sun, F. & Dong, X. 2013 A stochastic optimization approach for integrated urban water resource planning. *Water Science and Technology* **67** (7), 1634–1641.

Jung, D. & Kim, J. H. 2018 Using mechanical reliability in multiobjective optimal meter placement for pipe burst detection. *Journal of Water Resources Planning and Management* **144** (7), 04018031.

Jung, D., Kang, D., Liu, J. & Lansey, K. 2015 Improving the rapidity of responses to pipe burst in water distribution systems: a comparison of statistical process control methods. *Journal of Hydroinformatics* **17** (2), 307–328.

Laucelli, D., Rajani, B., Kleiner, Y. & Giustolisi, O. 2013 Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. *Journal of Hydroinformatics* **16** (4), 743–757.

Laucelli, D., Romano, M., Savić, D. & Giustolisi, O. 2015 Detecting anomalies in water distribution networks using EPR modelling paradigm. *Journal of Hydroinformatics* **18** (3), 409–427.

Li, J., Wang, C., Qian, Z. & Lu, C. 2019 Optimal sensor placement for leak localization in water distribution networks based on a novel semi-supervised strategy. *Journal of Process Control* **82**, 13–21.

Mounce, S. R. & Machell, J. 2006 Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal* **3** (1), 21–31.

Mounce, S. R., Mounce, R. B. & Boxall, J. B. 2010 Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics* **13** (4), 672–686.

Muntakim, A. H., Dhar, A. S. & Dey, R. 2017 Interpretation of acoustic field data for leak detection in ductile iron and copper water-distribution pipes. *Journal of Pipeline Systems Engineering and Practice* **8** (3), 05017001.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., Pierro, F. d., Khu, S.-T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., 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. B., Wu, Z. Y. & 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.

Raei, E., Shafiee, M. E., Nikoo, M. R. & Berglund, E. 2018 Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty. *Journal of Hydroinformatics* **21** (2), 223–239.

Rossman, L. A. 2000 EPANET2: Users Manual. https://www.epa.gov/water-research/epanet (accessed 11 September 2000).

Salinas, A. 2014 Integrated water resources management under a new global scenario. *Water Supply* **15** (2), 215–223.

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.

Sarrate, R., Blesa, J., Nejjari, F. & Quevedo, J. 2014 Sensor placement for leak detection and location in water distribution networks. *Water Science and Technology* **14** (5), 795–803.

Simone, A., Giustolisi, O. & Laucelli, D. B. 2016 A proposal of optimal sampling design using a modularity strategy. *Water Resources Research* **52**, 6171–6185.

Soroush, F. & Abedini, M. J. 2019 Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm. *Journal of Hydroinformatics* **21** (6), 1030–1047.

Steffelbauer, D. B. & Fuchs-Hanusch, D. 2016 Efficient sensor placement for leak localization considering uncertainties. *Water Resources Management* **30** (14), 5517–5533.

Sun, J., Wang, R. & Duan, H.-F. 2016 Multiple-fault detection in water pipelines using transient-based time-frequency analysis. *Journal of Hydroinformatics* **18** (6), 975–989.

Weickgenannt, M., Kapelan, Z., Blokker, M. & Savic, D. A. 2009 Optimal Sensor Placement for the Efficient Contaminant Detection in Water Distribution Systems. Water Distribution Systems Analysis 2008, pp. 1–10.

Wu, Y. & Liu, S. 2017 A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal* **14** (9), 972–983.

Wu, Z. Y. & Simpson, A. R. 2001 Competent genetic-evolutionary optimization of water distribution systems. *Journal of Computing in Civil Engineering* **15** (2), 89–101.

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K. & Fu, G. 2019 Deep learning identifies accurate burst locations in water distribution networks. *Water Research* **166**, 115058.

First received 1 September 2019; accepted in revised form 6 February 2020. Available online 16 March 2020