Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty

Ehsan Raei, M. Ehsan Shafiee, Mohammad Reza Nikoo and Emily Berglund

ABSTRACT

Large volumes of water are wasted through leakage in water distribution networks, and early detection of leakages is important to minimize lost water. Pressure sensors can be placed in a network to detect changes in pressure that indicate the presence of a new leak. This study presents a new approach for placing a set of pressure sensors by creating a list of candidate locations based on sensitivity to leaks that are simulated at all potential nodes in a network. The selection of a set of sensors is explored for two objectives, which are the minimization of the number of sensors and the time of detection. The non-dominated sorting genetic algorithm (NSGA-II) is used to explore trade-offs between these objectives. The effect of measurement uncertainty on the selection of sensor locations is explored by identifying alternative non-dominated fronts for different values for sensor error. The evolutionary algorithm-based approach is applied and demonstrated for the C-Town water network.

Key words | genetic algorithm, leak detection, NSGA-II, pressure sensors, uncertainty

Ehsan Raei
Mohammad Reza Nikoo
Department of Civil and Environmental Engineering,
Shiraz University,
Shiraz,
Iran

M. Ehsan Shafiee
Sensus USA Inc.,
639 Davis Dr., Morrisville, NC,
USA

Emily Berglund (corresponding author)
Department of Civil, Construction, and Environmental Engineering,
North Carolina State University,
2501 Stinson Dr., Mann Hall, Campus Box 7908, 27695, Raleigh, NC,
USA
E-mail: *emily_berglund@ncsu.edu*

INTRODUCTION

Water distribution networks are designed and operated to reliably deliver water to meet demands while maintaining pressures throughout a network. Leakages cause failures in delivering water reliably, and water loss reduction is a critical goal for managing infrastructure and water resources. In the UK, 3,281 mega liters (10^6) of water were reported as wasted due to leaks in pipelines during 2009–2011, and some utilities in the USA report 15% of water lost each year (Sadeghioon *et al.* 2014). Leakages contribute to failures in infrastructure and lead to economic impacts through lost revenue, excessive power consumption, and costs that are passed on to end users (Ponce *et al.* 2014). While the immediate effect of a new leak is the propagation of a transient wave in the pipe network, the transient wave disappears quickly after the event (Duan 2017a, 2017b), and the permanent effect of a leakage is a readjustment of pressures in nearby pipes (Filion & Karney 2002). The combined effect of multiple leakages is a substantial pressure drop in a water network, which causes consumer complaints and low water quality issues. To address loss of performance, utilities apply strategies to increase pressures in affected areas, such as managing pressure through tight control over isolated sub-sectors, or district metered areas (DMAs) (Perelman *et al.* 2015; Laucelli *et al.* 2017; Samir *et al.* 2017). Utilities may add booster pumps, tune pressure reducing valves to deliver higher pressures, close selected loops to deliver water along shorter paths, and reschedule pump operations to increase pressure in affected areas. Water utilities have also begun to invest in tools for detecting leakages to more effectively manage water losses (Mutikanga *et al.* 2012).

doi: 10.2166/hydro.2018.032

Acoustical listening leakage detection devices may not be practical for utilities due to the length of pipe that should be tested, limited municipal budgets, and a lack of available personnel; additionally, the deployment of these devices require field experience and training (Hamilton & Charalambous 2013). Computational methodologies that identify leakages may be more useful for utilities, because they are typically inexpensive to implement and can provide insight about the location of leakages and replacement of water system components (Vairavamoorthy & Lumbers 1998; Poulakis *et al.* 2003).

Many computational methodologies for identifying leakages are based on inverse methods, which require a large amount of data about system parameters, such as pressure and flows (Liggett & Chen 1994; Poulakis *et al.* 2003; Soares *et al.* 2011). Inverse methods compare measured data with the results of a numerical model to identify the location and size of leakages (Puust *et al.* 2010). A range of optimization methods have been applied to solve inverse problems for water distribution systems, including evolutionary computation, which provides a heuristic search that can be coupled with a hydraulic simulation model (Casillas *et al.* 2013; Cuguero-Escofet *et al.* 2017). Pressure sensors are the primary device to measure water pressure in a pipe network and are subject to measurement errors associated with any measuring device. Differences between measured and expected data can confirm leakages, but there is uncertainty in using these values due to the possible range of errors for these devices. The difference between measured data and the expected value for the data must exceed the measurement error to be considered an anomaly in the performance of the network. Small leaks may be difficult to detect and leaks may occur at any location in a network; therefore, the effective placement of pressure sensors is essential for collecting reliable data to identify leakages. Because the function of leak detection and localization is directly related to sensor location and sensor performance, the optimal position and number of sensors is affected by uncertainties. Sensor designs that are developed using assumptions of perfect sensors can perform sub-optimally when implemented (Casillas *et al.* 2013; Steffelbauer & Fuchs-Hanusch 2016).

In the design of pressure sensor networks, alternative solutions should be evaluated for a set of management objectives, including the practicality, feasibility, and reliability of a pressure sensor layout. The time of detecting a leakage is important as water distribution networks may reach a new equilibrium relatively quickly after a leak is introduced to the system. In addition, the presence of measurement error should be addressed as it introduces uncertainty to the estimate of the performance of a design. The research presented here develops a framework to model potential leakages that can occur in a network and to place sensors based on a set of performance goals. An evolutionary algorithm-based approach is developed to detect a pressure anomaly due to a leakage at any node in the network. The framework couples a hydraulic model with a multi-objective genetic algorithm-based methodology, the non-dominated sorting genetic algorithm-based (NSGA-II) approach (Deb 2000), to identify a set of nodes for placing pressure sensors. Uncertainty associated with the detection of leakages is addressed using a set of pressure sensor errors. Four scenarios are developed to simulate leakages that occur at varying times of day to capture the diurnal changes in demands, and every node is tested for introducing leaks. Different leakage scenarios paired with assumptions about measurement uncertainty lead to alternative pressure sensor layout designs for detecting a range of leakage magnitudes. The outcome of this approach is a set of pressure sensor networks that detect pressure anomalies due to leakages. The final locations for installing pressure sensors are identified from a large number of locations to minimize the number of sensors and the time of detection. The new framework is applied to illustrate the design of a pressure sensor network for the C-Town water network.

The manuscript is organized as follows. In the section immediately below, an overview of sensor placement for leak identification literature is presented. The problem formulation is then presented followed by the methodology that is developed in this research. The case study follows, along with modeling scenarios and assumptions. Results are then presented followed by discussion and finally conclusions.

BACKGROUND

Pressure sensors can be strategically located within water distribution networks through the use of systems analysis methods, including simulation and optimization. Leaks create both short-term and long-term signatures that can be used through analysis methods to identify leak locations. Short-term signatures may be noticeable if unsteady characteristics of the networks are monitored. Small leaks can create permanent or long-term effects, which may be detected based on pressure drops that are larger than expected.

A number of algorithms have been developed to place pressure sensors using these signatures. Early works established the use of the sensitivity matrix to design sensor networks. The sensitivity matrix is calculated by subtracting pressures of simulations with no leak from the pressures which are calculated at the same position under a leak scenario (Farley et al. 2008; Pérez et al. 2009). Blesa et al. (2016) use a clustering approach to place sensors, Sarrate et al. (2014) use a branch and bound approach, and Cuguero-Escofet et al. (2017) and Casillas et al. (2013) use gentic algorithm approaches to place sensors. Another methodology was developed to place sensors by creating a library of possible leakages, which is passed into a classifier algorithm to identify the most effective locations for placing sensors (Soldevila et al. 2018). Forconi et al. (2017) formulate the sensor network design problem as a ranking problem where the sensor locations are ranked based on the risk of non-detection of leaks. Blesa et al. (2016) formulated the sensor placement problem as a multi-objective problem to minimize leaks that are not detected and maximize the locatability of leaks, based on the magnitude of the residual in pressure as calculated for leak and no-leak scenarios.

A set of studies explores the effects of uncertainty on the performance of sensor placement problems. Forconi et al. (2017) use a risk-based approach, where they account for the uncertainty in the performance of sensor networks, based on the measurement error of sensors. They represent the likelihood of detection as zero for locations where the pressure difference is lower than the accuracy of the sensor. Steffelbauer & Fuchs-Hanusch (2016) also address uncertainty in the placement of sensors. They account for the effect of demand uncertainty on modeled predictions of pressure, and use Monte Carlo simulation to calculate pressures for multiple realizations of nodal demands. Blesa et al. (2014) explore the effects of uncertainty in the magnitude of leaks and inflows to the network on the sensitivity matrix, and found that sensor positions are not sensitive to the size of leaks, but to the flows within the WDS.

The research presented here develops the pressure placement problem with measurement uncertainty as a multi-objective problem to explore trade-offs between the number of sensors and the time to detection. The effect of measurement errors on the placement of sensors is explicitly represented using a set of alternative expected pressure sensor errors. An evolutionary algorithm approach (Bäck et al. 1997) is used to search for locations to place sensors. Evolutionary algorithms are widely used to solve water distribution network problems. New solutions are created using probabilistic approaches to recombine and mutate a set of good solutions. Each solution is evaluated based on the fitness, which represents the satisfaction of a set of objectives. NSGA-II is an evolutionary algorithm that is designed to solve multi-objective problems and is applied here to explore trade-offs in the time of detection and the number of sensors.

PROBLEM STATEMENT

A multi-objective problem statement is formulated to identify a pressure sensor network. The multi-objective problem locates a set of pressure sensors to detect pressure differences and identify leakages by minimizing the number of sensors and the time of detection. The problem formulation is represented mathematically as follows:

$$\begin{aligned} &\mathit{minimize}\ Z_1 = k_s \\ &\mathit{minimize}\ Z_2 = \frac{\sum_i^n \sum_{t_l}^{T_L} t_{detection,i,t_l}(\bar{e}_s)}{N \times T_L} \\ &\mathit{subject\ to}\ m \le k_s \le m_s \\ &m_s \le N \end{aligned} \tag{1}$$

where the variables Z_j ($\forall\, j = 1\ \&\ 2$) are the objective functions; the variable k_s is the number of pressure sensors that is deployed for the water networks; and the variable m is the minimum number of sensors required by a utility. Sensor readings are expected to be used in a leak detection algorithm, and the number of sensors required may be determined by the requirements of the leak detection algorithm.

The variable T_L is the set of times, at which a leakage starts at the i^{th} node and represents the number of time periods within each diurnal pattern. The variable m_s is the number of nodes where sensors can be placed. The variable N is the number of nodes in the network. The variable $t_{detection,i,t_l}(\bar{e}_s)$ is the time at which a leakage at the i^{th} node is detected by one of the k_s sensors. The time of detection is the time at which one sensor out of the set of sensors registers a pressure difference that is larger than an error threshold, defined as $\bar{e}_s$. An approach for setting the threshold $\bar{e}_s$ is described as follows.

Capturing uncertainty in pressure reading for sensor placement

The error in pressure sensors creates some uncertainties in determining the difference in pressure that should be used to indicate a leak. Theoretically, a device can measure readings in a network with zero error; however, this perfect reading is practically impossible. The error associated with pressure readings is typically unknown, and it is difficult to explicitly separate the pressure difference and the error. For example, assume that e is the error of a pressure sensor that is reported by a sensor manufacturer, and the difference r between the expected pressure and the pressure reading is used to indicate that a leakage has occurred. The smallest value of r that could be used to identify that a leak is certainly present is $2e$. Here, we assume that the uncertainty in the calculation of expected pressure, or uncertainty associated with the simulation model, is neglected. To elaborate, if the expected value of pressure is p_{ex}, a pressure difference in the range of $p_{ex} \pm e$ should indicate that there is a leakage; that is, for a field pressure at p_{ex}, the range of potential pressure reading is $[p_{ex} - e, p_{ex} + e]$. A pressure difference larger than $2e$, therefore, indicates that a leakage has occurred.

To treat this uncertainty that affects the performance and best placement of pressure sensors, a range of error thresholds is used in assessing pressure differences and time of detection in Equation (1). This range of thresholds is determined based on the commercial pressure sensors available to be used in the market. Each threshold is a predefined input for Equation (1) to compare with the difference between the no leakage model and leakage model. By considering a range of thresholds, the effect of measurement uncertainty on the timing and location of detecting leakages can be determined. In addition, the sensor network, which is the result of this study, may vary as the sensor errors increase.

SIMULATION-OPTIMIZATION FRAMEWORK FOR IDENTIFYING PRESSURE SENSOR NETWORKS

A simulation-optimization framework is developed here to identify a pressure sensor layout (Figure 1). First, a hydraulic model is used to simulate a set of leakages that could occur at each node in a water network. The second step identifies a list of candidate nodes for placing sensors, based on locations where the change in pressure is relatively high when leaks are present. Finally, the list of candidate nodes

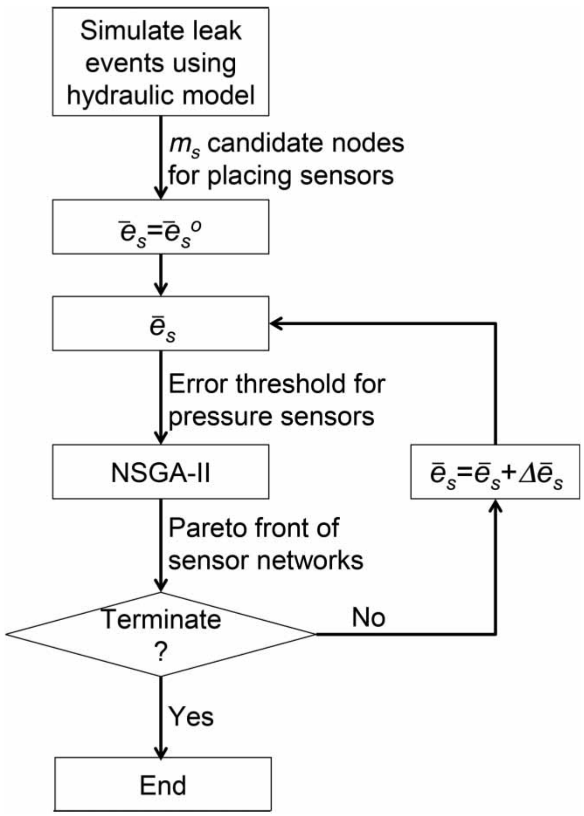

Figure 1 | Flowchart of the simulation-optimization framework for identifying a set of pressure sensors for water networks. The variable $\bar{e}_s$ is set at an initial value and updated by $\Delta\bar{e}_s$ for each execution of the evolutionary algorithm. This procedure results in multiple Pareto fronts, or sets of non-dominated solutions.

are used as decision variables for a multi-objective evolutionary algorithm-based approach, which selects nodes for placing pressure sensors. The multi-objective problem represented by Equation (1) is solved using NSGA-II, and results identify a trade-off, shown as a set of non-dominated solutions, between the number of sensors and the time to detect the leakage. As shown in Figure 1, the approach iterates through decreasing values of $\bar{e}_s$, which lower the threshold for the pressure error and assume the use of increasingly accurate sensors.

As described here, EPANET (Rossman 2000) is loosely coupled with the NSGA-II library using MATLAB software to develop the simulation-optimization framework. EPANET is run to develop a list of candidate nodes, but is not tightly integrated with NSGA-II. Each component of the framework is explained in detail as follows.

Hydraulic simulation of leakages in water distribution networks

Leakages can occur as various forms and types in a network. Experimental studies explored equations for alternative types of leakages such as longitudinal cracks, circumferential cracks, and round holes (Greyvenstein & Van Zyl 2007; Shafiee et al. 2016). Here, leakages are simulated as round holes using an orifice model. The emitter equation (Equation (2)) simulates the flow through leaks as:

$$q = cp^{\gamma} \tag{2}$$

where the variable q is the amount of flow through the leak, which varies as the function of the pressure, p. The variable c is the discharge coefficient, and the variable γ is the pressure exponent. The pressure exponent is 0.5 for an emitter modeled as an orifice (Rossman 2000). This is the same value used by Forconi et al. (2017) in their approach to place sensors for leak detection. Equation (2) can also be implemented within EPANET to represent alternative geometries of leaks.

Here, we model leaks using a demand-driven approach, and a new demand is calculated that varies based on the pressure at a node for each time step of the simulation. The new demand is added to the existing demand of the node. The new demand is calculated as follows:

$$d_{new,n,i} = d_{init,n,i} + \begin{cases} c\, p_{init,n,i}^{0.5}, & \text{if } i \geq t_l \\ 0, & \text{otherwise} \end{cases} \qquad \forall\ i \in 1, \ldots, T \,\&\, n \in 1, \ldots, N, \tag{3}$$

where the variable $d_{new,n,i}$ is the new demand for a leak modeled at node n at time step i of the simulation. The variable $d_{init,n,i}$ is the initial demand that is exerted at node n at time step i. The constant c is the emitter coefficient to simulate a leak at node n. The hydraulic model is executed without leaks present to calculate pressures at node n, and values for $p_{init,n,i}$ are reported as the pressure at node n and time step i. t_l is the time step at which the leak is initiated at node n, and T is the total simulation time, and in this work, the time step (*tstep*) is equal to 1 hour. N is the number of nodes in the network.

Leaks are modeled at every node in the water distribution network. Pressures at nodes vary during the day due to fluctuations in the water demand, which is typically represented as a diurnal pattern. The start time for leaks generates different pressure profiles at pressure sensors. To exhaustively identify the effect of leakages in a water network, leaks should be modeled at each location in the network for start times that vary in small increments (such as every 1-hour period). This approach would be computationally expensive. To capture the effects on pressure readings of leaks that may be initiated at various times in a day, we assume that leaks that are initiated during time periods when changes in demands are negligible create similar patterns in pressure readings. Periods of relatively constant demand patterns are determined from diurnal demand patterns that are used as input for the hydraulic simulation model. The start time for each leak is modeled at the beginning of each time period, and as a result, $T_L \times N$ leak events are created, where T_L is the number of time periods determined from the diurnal pattern. The hydraulic model is executed once to simulate each leak event.

Identifying candidate nodes for sensor placement

Candidate nodes are identified based on their pressure detection for all leak events that occur within time period t_{T_L}.

For one leak event, a matrix of pressure values is created by recording the pressure at every time step for each node. A matrix of pressure differences is calculated as the absolute value of the difference between the leak event and the baseline hydraulic model (without any leaks present) for each element of the pressure matrix. r_i is the difference between the pressure measurement (P_i) and its corresponding estimated value ($\wedge P_i$) obtained from the simulation model with no leak:

$$r_i = P_i - \wedge P_i \quad i \in 1, \ldots, N \tag{4}$$

The pressure difference (r_{ji}) is calculated for each node i at time step j, except for the node where the leak is initiated. The pressure difference matrix (L_S) is calculated for leak S:

$$L_S = \begin{bmatrix} r_{11} & \ldots & r_{1j} \\ \vdots & \ddots & \vdots \\ ri1 & \ldots & r_{ij} \end{bmatrix} \quad i \,\&\, S \in 1, \ldots, N, \quad j \in 1, \ldots, \frac{T}{tstep} \tag{5}$$

where L_S is the difference between the measured and the estimated pressure, affected by leak S among all nodes (columns of the matrix) throughout the simulation time steps (rows of the matrix).

For each column j in L_S, the nodes are ranked, where the highest and lowest values of r_{ij} receive the ranks of $N-1$ and 0, respectively. Nodes that have maximum pressure difference smaller than the error threshold are not included as potential nodes for detecting leakage for that leak event. These nodes receive a rank of zero in the ranking process.

To calculate a score (R) for each node, leaks are grouped according to the time of leak (t_{T_L}). A candidacy number, denoted as $R_S^{t_{T_L}}$, is the maximum rank over all time steps for leak S. By iterating over all leaks with the same time, each node receives $N-1$ candidacy numbers, and the candidacy numbers for each node are aggregated to assign a score for each node:

$$R = \sum [R_1^{t_{T_L}}, \ldots, R_S^{t_{T_L}}] \quad S \in 1, \ldots, N \tag{6}$$

The score can vary in the range of zero to $(N-1)^2$. The best score is $(N-1)^2$, representing the most effective location for placing a pressure sensor that detects leakages that occur at all other nodes in a network. For each time of leak setting, the top ranking m nodes are selected as candidates for placing pressure sensors. The size of the combined set of candidate nodes across the time of leak settings varies from m to $m \times T_L$.

Multi-objective optimization approach

A non-dominating sorting genetic algorithm is implemented to solve the multi-objective problem (Equation (1)). Evolutionary algorithms use a population of solutions and apply heuristic rules at each iteration to converge to a near-optimal solution(s). A solution is represented as an array of decision variables that may be specified as binary, real, or integer values. At each generation, selection is applied to select high performing solutions for crossover, which is applied to combine decision values across multiple solutions that perform well. Mutation is applied to make random changes to existing solutions. NSGA-II is widely used for generating non-dominated solutions for water resources planning and management problems (Nicklow *et al.* 2010). NSGA-II follows the steps for a genetic algorithm and ranks solutions based on dominance. To maintain diversity in the Pareto front, NSGA-II uses elitism and a crowding distance to select solutions without specifying additional algorithmic search parameters (Deb 2000).

A solution is represented as an array of integers, which shows sensor locations. The length of the array is the number of pressure sensors that will be placed in a network (Z_1). Using the hydraulic model, four leakages are modeled at every node with the various start times using Equation (3). The pressure difference is calculated for every node at each time step by taking the absolute difference between pressures reported by the no-leakage and the leakage hydraulic model. The time of detection ($t_{detection,i,t_l}(\bar{e}_s)$) for each leak event is the earliest time step that a sensor from the sensor list registers a pressure difference larger than the threshold. The variable Z_2, the time of detection, is calculated by averaging over leak events.

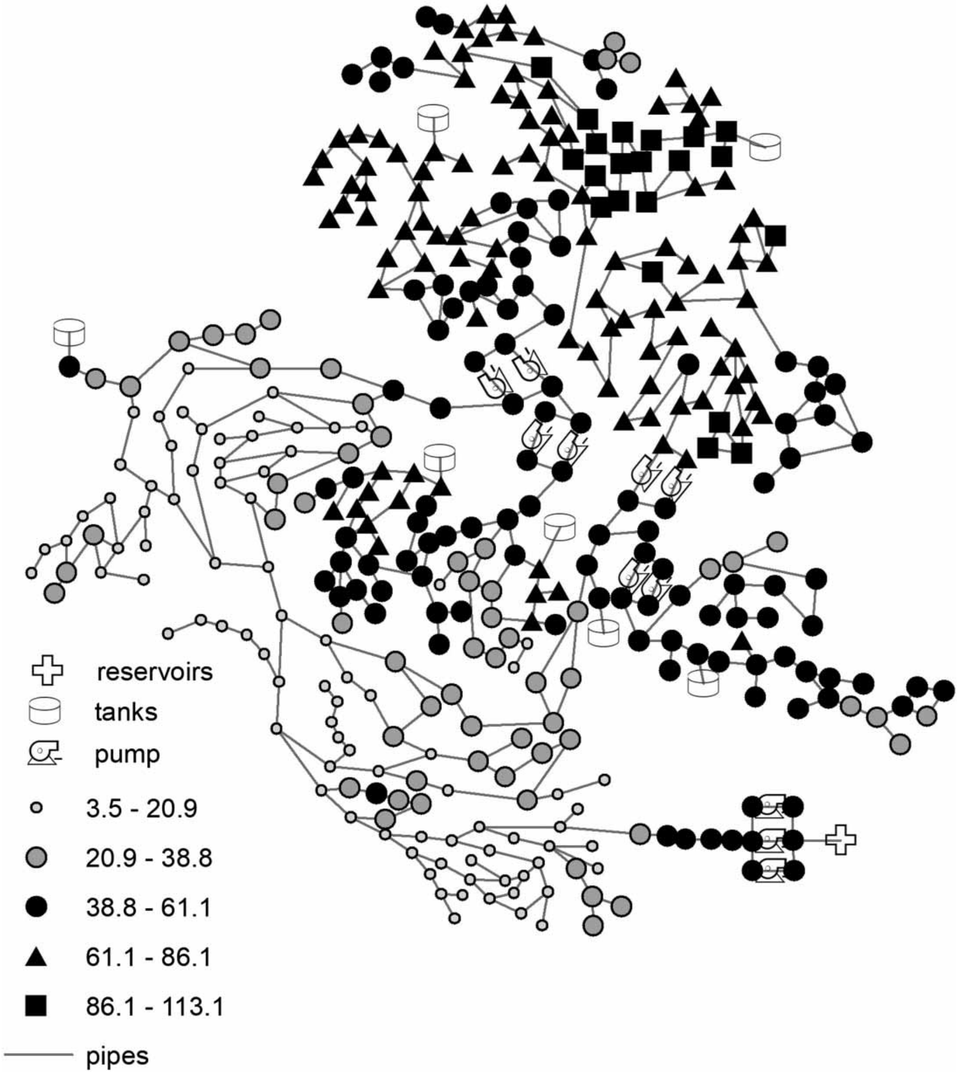

Figure 2 | C-Town water distribution network.

ILLUSTRATIVE CASE STUDY: C-TOWN

The pressure sensor framework is demonstrated here to simulate and identify leaks for the virtual city of C-Town (Marchi *et al.* 2014). C-Town has five distinct pressure zones, which are distinguished from each other by varying topographical regions (Figure 2). The original C-Town network has 388 nodes, seven water tanks, one reservoir, and 11 pumps. The lowest and highest elevations among nodes are 3.48 and 113.08 meters, respectively. Leakages are introduced at all 388 nodes, except the nodes that are located after each pump. The simulation time is 96 hours.

This study is based on a numerical experiment, and field data are not used for validation of the model. It is assumed the hydraulic model is calibrated and represents the conditions of the C-Town system, which is a virtual city.

Simulation scenarios and model settings

Four time periods are modeled to capture the effect of diurnal patterns on the volume of water that is wasted. Diurnal patterns are divided into four periods based on the consumption behaviors of end users (Figure 3). These

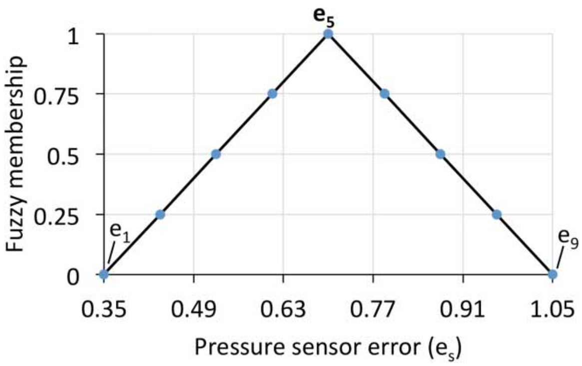

Figure 3 | The average diurnal pattern of C-Town water network over five residential and commercial diurnal patterns, which are defined in the C-Town water model.

periods consider ascending, descending, and flat changes. The first period (Scenario 1) lasts from midnight to 6am, during which the demands are at the minimum. The demands increase after 6am until 12pm, and maximum demands are observed at lunchtime (Scenario 2). During the afternoon period (12pm–6pm) (Scenario 3), demands remain at the largest value for the commercial and industrial demands and at the smallest values for residential demands. There is an increase in the residential demands towards the end of the day as consumers arrive at residential nodes. This shift in the value of demands begins at 6pm and lasts until 12am (Scenario 4), when most consumers are asleep. For each scenario, leakages are introduced at the beginning of the time period, or 12am, 6am, 12pm, and 6pm ($T_L = 4$ in Equation (1)).

We use values for the c coefficient to generate leaks at a rate of 0.2–0.5 $\frac{lt}{sec}$. This range of leakages was tested to ensure that a minimum level of change occurs in the pressure in the network. These tests were conducted to check that the addition of leakage demand to the water network did not violate the minimum pressure at nodes. If the minimum pressure is not satisfied, the Wagner function can be used instead of Equation (3).

The set of error thresholds is generated and used to generate the rank matrix. These thresholds determine the score of nodes for each leakage event and are tested at nine equidistant values within the range of 0.35 psi and 1.05 psi (0.35, 0.4375, 0.525, 0.6125, 0.7, 0.7875, 0.875, 0.9625, and 1.05 psi).

The minimum number of sensors (m) required by the leakage detection algorithm is set to three ($m = 3$ in Equation (1)). The trade-off that is identified by the algorithm provides a range of sizes for the sensor network that a manager can evaluate to make a final decision.

RESULTS

NSGA-II is executed for each error threshold. The settings for NSGA-II are a population size of 400, 100 generations, a crossover rate of 0.8, and a mutation rate of 0.2. The MATLAB toolbox is used to calculate the objective functions Z_1 and Z_2. Due to stochasticity in the convergence of NSGA-II, three runs of NSGA-II are executed for each error threshold, and the best non-dominated set is selected as the Pareto front for this error. The best Pareto front is determined by calculating the hypervolume for each run. The results of the 27 NSGA-II runs are used to identify nine trade-off fronts between the objectives. Each of these Pareto fronts corresponds to an error threshold.

Candidate nodes for placing pressure sensors

Figure 4 shows the candidate nodes that are identified as potential nodes for sensor placement. In total, 110 nodes are found as the most sensitive nodes to all leak events. As some nodes are sensitive to two or more events, the original list of potential sensors was reduced by 40 by removing repeated nodes. All potential nodes have the same likelihood to be selected by the NSGA-II algorithm for a solution. Only a few nodes that are located at an elevation lower than 30 meters are selected as part of the candidate list.

Effect of error threshold: trade-offs between number of sensors and time of detection

Figure 5 shows the best (based on the calculated hypervolume) trade-offs between the number of sensors and the average time of detection for each setting of $(\bar{e})_s$. The average earliest time of detection is 18.3 hours (66,000 seconds) when 20 or more sensors are deployed and the devices work with a small error value. The average time of detection increases to 21.4 hours with 16 pressure sensors that perform with low precision (error is equal to 1.05 psi).

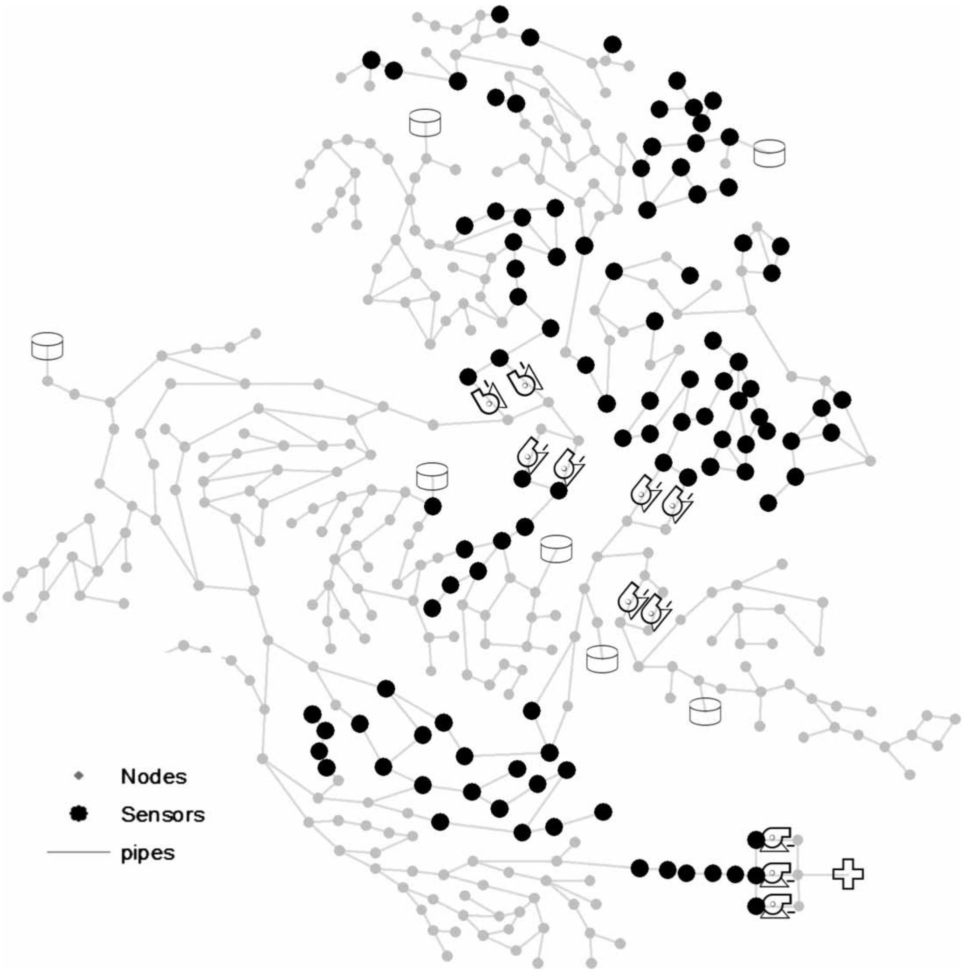

Figure 4 | Candidate nodes that are identified as potential nodes for sensor placement. The results of four scenarios are combined, and 110 nodes are selected as input for NSGA-II.

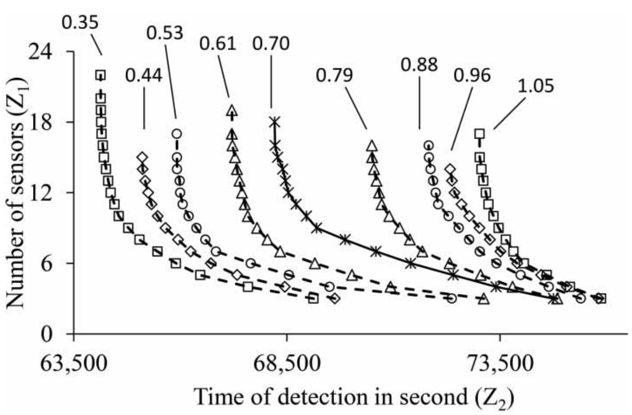

Figure 5 | Trade-offs between the number of sensors, Z_1, that is deployed in the network and the average time of detection, Z_2. Each trade-off curve is generated corresponding to an error threshold. From left to right, the trade-offs are identified for error thresholds of 0.35, 0.4375, 0.525, 0.6125, 0.7, 0.7875, 0.875, 0.9625, and 1.05 psi, respectively.

Using three sensors, however, increases the average detection time by 3.6 hours with the same precision. With the deployment of three sensors, utilities can detect a leakage within 20 to 25 hours after it started. With only three sensors, however, the amount of pressure data may not be sufficient to use in inverse methods to identify the leak location. Deploying a larger number of pressure sensors has a small influence on the average time, but it provides more data points for leakage identification algorithms, which can increase the uniqueness of solutions that are identified by inverse approaches (Zechman & Ranjithan 2009; Zechman *et al.* 2013). There is a negligible increase in the average detection time for six to ten pressure sensors, and the decision to select a number of sensors from within this range can be further explored based on

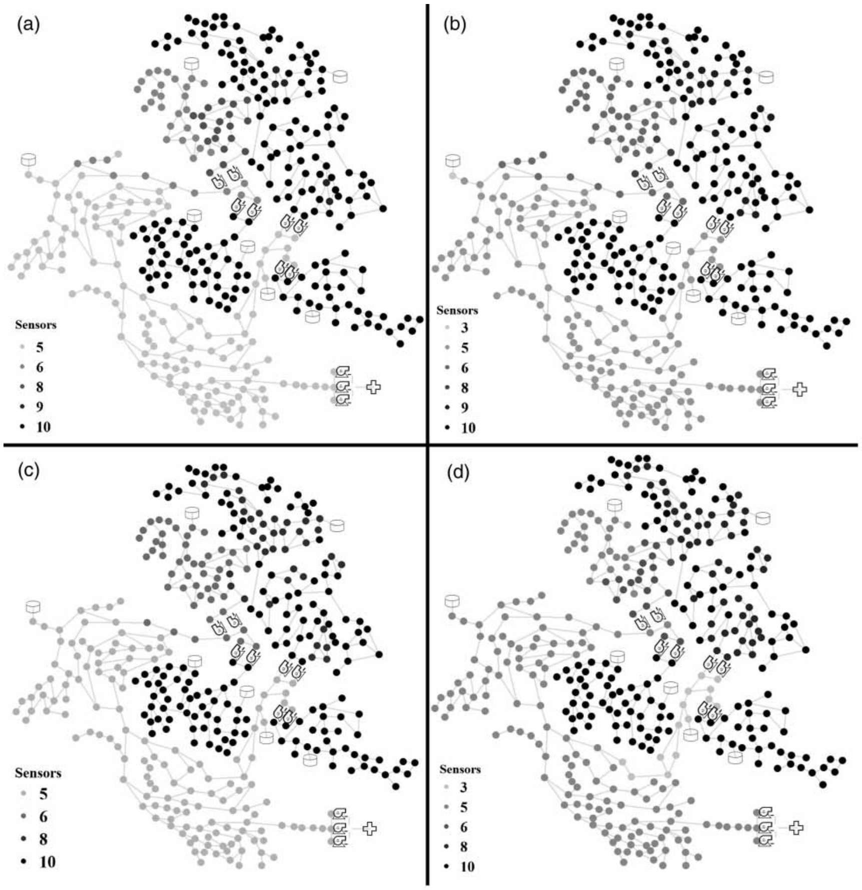

Figure 6 | Leak event coverage using the ensemble of ten pressure sensors, which are selected from the trade-off for 0.7 psi error threshold when the leak event start time is (a) Scenario 1, (b) Scenario 2, (c) Scenario 3, and (d) Scenario 4. The map shows the number of sensors that detects a leak at that node. The shade represents the number of sensors (out of ten) that detects the leak event that is started at a node.

budget limitations and the use of data for leakage detection algorithms.

Leak event coverage using an ensemble of pressure sensors

Figure 6 shows the coverage of an ensemble of pressure sensors for detecting leakage events for C-town for the sensor network shown in Figure 7. The coverage is 100% for this network, because ten sensors are sufficient to cover all the leakage events that are simulated. The coverage rate may be lower for fewer sensors or networks that are made up of isolated sub-systems, such as district metered areas. For systems where lower coverage of leakage events is likely, the optimization model can be reformulated to maximize coverage explicitly, rather than minimizing time of detection.

Effect of device error on sensor placement

Figure 7 shows the effect of the device error on the sensor placement for networks of three, six, eight, and 15 sensors. For small sensor networks, sensors are located near the central region of the network, and additional sensors are placed on the periphery of the network for larger sensor networks. There are a few nodes in the network that are included in the solution for nearly every size of sensor networks. These nodes are effective at detecting pressure changes and, thus, are included in most solutions. Repeated nodes are advantageous in selecting the number of nodes for utilities, as these nodes may be considered as robust components of a network for multiple sensor designs and sensor error. These solutions can also provide guidance for a phased deployment of a large sensors network. Sensors that are

Figure 7 | The effect of device error on the location of pressure sensors. The solutions that have the ensemble size of (a) 3, (b) 6, (c) 8, (d) 15 sensors are selected from Figure 5. The number next to each sensor location corresponds to the error code. Error codes of 1, 2, …, 9 represent 0.35, 0.4375, …, 1.05 psi error thresholds, respectively.

selected for both the small and large networks can be deployed first, followed by sensors that are included in the large sensor networks.

Figure 8 shows the relationship between the average time of detection and the volume of water that is wasted. The volume of wasted water is between 20 to 26 thousand

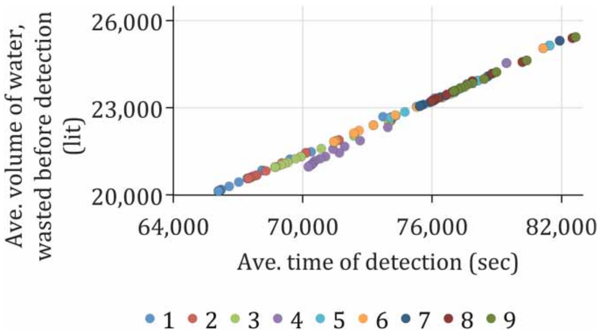

Figure 8 | The average volume of water that is wasted at the average time of detection. Each dot represents an ensemble of sensors as identified in Figure 5. The e_i is the error threshold that is used to determine the threshold for detecting leakages, where e_1 is 0.35 psi, and e_9 is 1.05 psi.

of liters for the set of sensor networks. As expected, lower times of detection correspond to lower volumes of wasted water. The lowest average detection time results in the minimum average volume of wasted water, corresponding to an ensemble of 19 sensors for the error threshold of 0.35 psi. In contrast, the maximum average wasted water corresponds to the deployment of three sensors for the 1.05 psi error threshold. Because the detection time and wasted water are presented as averages over all leak events in C-Town, there is a strong linear relationship that is observed between these two parameters. As previously shown in Figure 5, there is some overlap in the performance of sensor networks of various sizes.

Effect of leakage start time on the performance of pressure sensors

Figure 9 shows the effect of leakage start time on the time of detection by plotting the performance (detection time) of an ensemble of sensors for all 388 leakages that are simulated in C-Town. The sensor ensembles are of different sizes as identified for the trade-off for 0.7 psi error threshold shown in Figure 5. The start time of leakages is plotted separately for four scenarios. In general, pressure sensors perform better for leaks that start during high demand hours (Scenarios 2, 3, and 4). This is because the total demand is high throughout the network (Figure 3), and any leakage produces a significant disturbance in pressures. During peak hours, many components of the network perform together to deliver water, and, therefore, an interruption (or leak) creates a significant deviation from normal operations. As total demands decrease in Scenario 1, pressure sensors require more time to register a pressure difference. In addition, sensors perform poorly when demands are at the lowest, from 12am to 6pm.

As shown in Figure 9, higher numbers of sensors are most effective in reducing time to detection for Scenario 1. The first quantile of Scenario 1 decreases significantly by using five sensors instead of four sensors. The median of Scenarios 2, 3, and 4 remains unchanged as the number of sensors increases, and the performances for Scenarios 2, 3, and 4 are similar when a large number of sensors are used. This is because a common set of sensors that are located at the middle of the network are used for Scenarios 2, 3, and 4. As more sensors are deployed, additional sensors are located on the perimeter of the network for Scenario 1 (Figure 7) to reduce the detection time (see Pareto front for 0.7 psi error threshold in Figure 5). Large sizes of sensor networks reduce the first quantile in Scenarios 1 and 2, and the third quantile is not changed for varying sizes of sensor networks. As shown above, the performance of pressure sensor networks is similar for scenarios of high demands.

Figure 10 shows the effect of the leakage start time on the time of detection for each node. For each node, the detection times are shown for leakages that are introduced using the four scenarios when 11 sensors (Figure 11(b)) are deployed and the error threshold is 0.7 psi. As shown in Figure 9, ten-sensor networks perform poorly in detecting leakages for Scenario 4. Figure 10 shows that the sensor layout is able to detect leakages introduced at some nodes significantly earlier than the other nodes. For example, the sensor layout detects leakages that occur at the nodes with the index of 100–150 more quickly. These nodes are located near the center of the network.

Effect of the range of device error on the time of detection

To evaluate the effect of device error on the time of detection, an experiment is performed by selecting three solutions that have ten sensors from the 0.35, 0.7, and 1.05 error threshold trade-offs in Figure 5. The error threshold may have a negligible effect on the location of sensors.

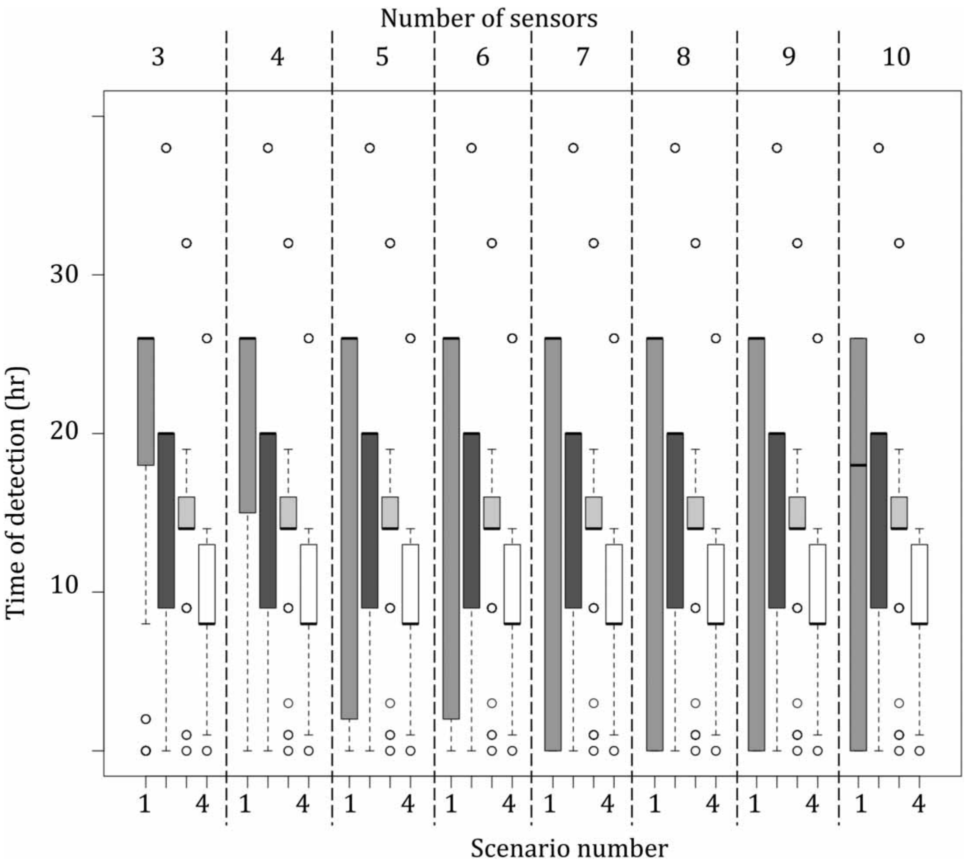

Figure 9 | Effect of leakage start time on the time of detection. Boxplots represent the detection time for leak events simulated at 388 nodes in the C-Town water network for four start times. Each boxplot is depicted for a sensor ensemble with size of 3–10, using an error threshold of 0.7 psi.

Figure 11 demonstrates the placement of three ten-sensor networks, corresponding to error thresholds of 0.35, 0.7, and 1.05. The layouts of these solutions are similar, as some sensors are located in the center of the network and a few sensors near the upper part of the network. Figure 12 shows the delays in detecting a leakage, where the delay is calculated as the difference in the time to detection for an event initiated at one node for two error threshold values. The delay shows that for a majority of leakages, the time of detection is not dependent on the range of device errors. For 30.5% of leakages, however, the range of errors can lead to up to 26 hours' delay in detecting leakages.

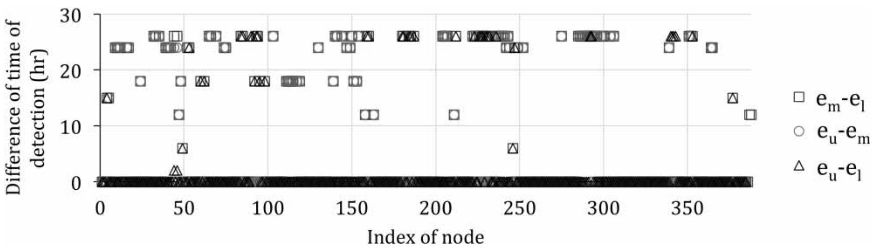

Figure 10 | Effect of leakage start time on time of detection for each node, at which a leakage is modeled. The error threshold is assumed at 0.7 psi, and ten sensors are deployed (Figure 11(b)). The location of sensors is determined using the 0.7 psi error threshold trade-off in Figure 5.

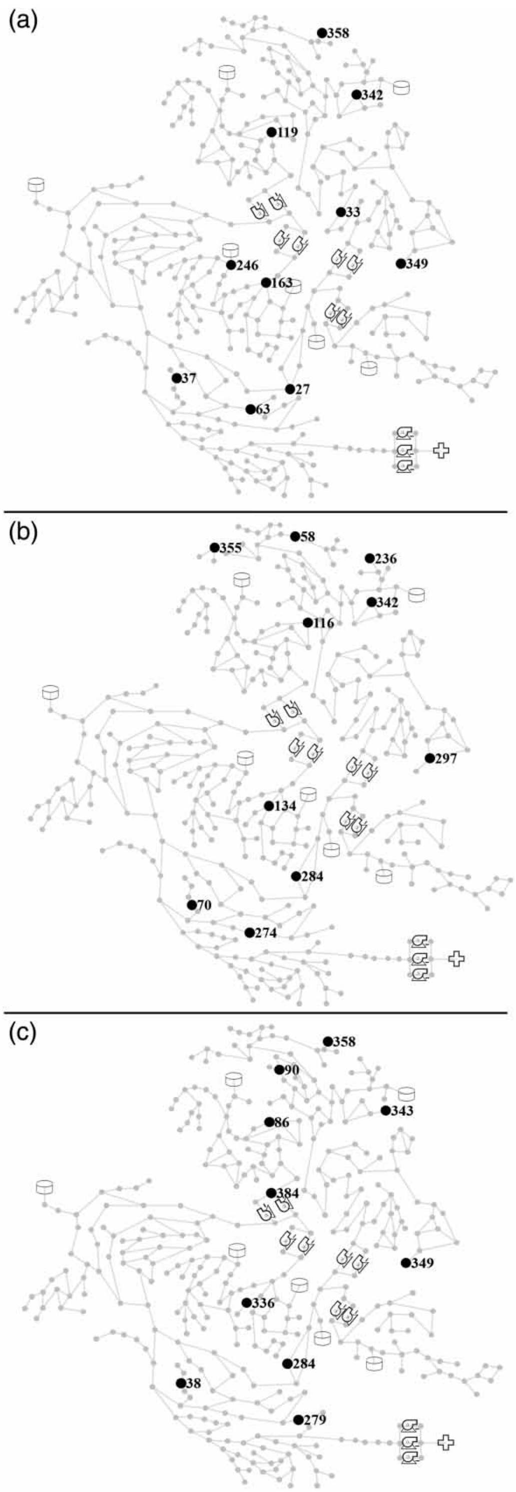

Figure 11 | The location of sensors when the error threshold is (a) 0.35, (b) 0.7, and (c) 1.05 and the size of sensor network is ten. The number is the index of nodes in the hydraulic model file.

DISCUSSION

This study uses a demand-driven model for leaks. Specifically, Equation (3) is an expansion of the implementation of the orifice model in EPANET (Rossman 2000). Braun *et al.* (2017) revisited the assumptions which are made for solving hydraulics of a water network and reviewed the advantages and disadvantages of both demand-driven and pressure-driven solution approaches and suggested improvements to find more realistic modeling results. In this study, we use a demand-driven approach and enforce a minimum pressure constraint for nodes. The pressure constraint ensures that the water pressure at every node is sufficient for firefighting purposes. In preliminary work, we tested an iterative approach to adjust $d_{new,n,i}$ based on the pressure values that are produced after leaks have been introduced to the network. The experimental results showed only an incremental change in the demand, $d_{new,n,i}$, when pressures were changed to reflect leaks. Therefore, neglecting the effect of pressure variation on the flow of leaks is reasonable for this methodology. Shafiee & Berglund (2017) demonstrated that demand changes in response to contaminated water can significantly affect the movement of a contaminant in a network. Similarly, pressure variation is affected by the amount of water that is extracted from the networks and drives the volume of leaks that is lost through compromised pipes. As the layout of a network influences the hydraulics of the network, this assumption concerning pressure variation should be re-evaluated for application to real-world networks. Therefore, for real networks, an analysis of the pressure variation is important before implementation of this methodology to place sensors. In addition, this research approach does not solve for the size of leaks, but places sensors to detect pressure variations. The use of pressure-driven modeling may more significantly affect methodologies that detect the location and size of leaks using optimized sensors.

The hydraulic solver is an extended-time hydraulic model (EPANET), which is a quasi-steady state model. The fundamental assumption of EPANET is that flows and pressures in a system change slowly. Considering slow changes in flows and pressures enables the solver to simulate the state of networks at each time step by assuming a linear change between each state. These assumptions, however, may be

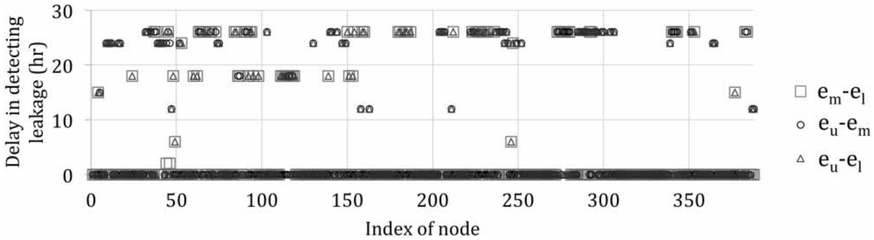

Figure 12 | The delay between the time of detection for Scenario 1 leakages when an ensemble of ten sensors is used. The error threshold is modeled at discrete values of 0.35, 0.70, and 1.05 psi. The sensor configuration is presented in Figure 11, respectively.

violated through manipulation of pumps and valves in the network. Improvements in hydraulic simulation capabilities may affect the results of the framework developed here. As shown, results are only slightly sensitive to the error threshold. For example, in Figure 11, the location of ten sensors are similar for three different error thresholds. Future research can test the sensitivity of this method to alternative hydraulic models.

The methodology that is developed here relies on the discretization of the demand pattern into four clusters to represent different starting times. The number of starting times was selected to capture variations in demand patterns and their effects on hydraulics within the network. Simulating leaks at every time step (e.g., every 1-hour period) would create a large computational burden, as leaks are simulated at every node in the network. The required number of simulations may be tractable for high performance computing cluster, although for typical desktop computers, reducing the number of simulations required to complete the methodology presents an advantage for encouraging its use in practice. Further research can explore the effects of using higher numbers of simulations on sensor design.

CONCLUSIONS

This study focuses on the placement of pressure sensors to detect water leakages at nodes within a water distribution network. A new framework is developed that uses an enumeration approach to simulate leakages that can occur at nodes in the network and different times of day. A set of nodes is selected as candidate locations for placing pressure sensors. Once a list of candidate nodes is created, the list is passed into a multi-objective optimization algorithm, NSGA-II, which selects a smaller set of nodes that can be used as an ensemble of pressure sensors. Because pressure sensors read pressure measurements with some error, the uncertainty that arises from the device error is captured by considering a range of pressure sensor error. The device error is modeled as a threshold for detecting leakage events, and, for each device error, a set of solutions is obtained using NSGA-II. The developed approach is applied to the illustrative virtual city of C-Town.

The new approach that is developed here is able to identify a set of locations for placing pressure sensors. Considering a range of device errors quantifies the threshold in the nodal ranking mechanism for improving the robustness of pressure sensors in detecting leakages. Results demonstrate that a large ensemble of pressure sensors is needed to register pressure readings within the range of error variation. The analysis of the sensor networks shows that the deployment of a large ensemble can increase confidence in detecting leakages by 30.5%. It is expected that the size of a sensor network would increase with the level of complexity for a water network. A more complex water network is likely to have various independent pressure zones. Within these complex networks, the effect of a leakage is not transmitted to other zones as zones are separated by valves and pump stations.

Future research can explore the effect of the characteristics of a water network on the number of sensors required and the performance of sensor networks. The C-Town water network is similar to real water networks; however, it is a skeletonized system, and a more descriptive model may create complications in convergence of the algorithmic approaches used here. The same approach should be evaluated for a mesh or looped water network. In an ongoing study, we explore the use of this approach for a larger, more complex network. The use of flow meters is integrated with an ensemble of pressure sensor and explored to improve certainty in detecting leakages for alternative types of leaks.

REFERENCES

Bäck, T., Fogel, D. B. & Michalewicz, Z. 1997 *Handbook of Evolutionary Computation*. IOP Publishing Ltd, Bristol, UK.

Blesa, J., Nejjar, F. & Sarrate, R. 2014 Robustness analysis of sensor placement for leak detection and location under uncertain operating conditions. *Procedia Engineering* **89**, 1553–1560.

Blesa, J., Nejjari, F. & Sarrate, R. 2016 Robust sensor placement for leak location: analysis and design. *Journal of Hydroinformatics* **18** (1), 136–148.

Braun, M., Piller, O., Deuerlein, J. & Mortazavi, I. 2017 Limitations of demand- and pressure-driven modeling for large deficient networks. *Drinking Water Engineering and Science* **10**, 93–98.

Casillas, M. V., Puig, V., Garza-Castanón, L. E. & Rosich, A. 2013 Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **13** (11), 14984–15005.

Cuguero-Escofet, M., Puig, V. & Quevedo, J. 2017 Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: application to the Barcelona water network. *Control Engineering Practice* **63**, 1–12.

Deb, K. 2000 *Multi-objective Optimization Using Evolutionary Algorithm*. John Wiley and Sons, Chichester, UK.

Duan, H.-F. 2017a Transient frequency response based leak detection in water supply pipeline systems with branched and looped junctions. *Journal of Hydroinformatics* **19** (1), 17–30.

Duan, H.-F. 2017b Transient wave scattering and its influence on transient analysis and leak detection in urban water supply systems: theoretical analysis and numerical validation. *Water* **9**, 789.

Farley, B., Mounce, S. & Boxall, J. 2008 Optimal locations of pressure meters for burst detection. In: *Water Distribution Systems Analysis 2008*. American Society of Civil Engineering, Kruger National Park, South Africa, pp. 1093–1102.

Filion, Y. R. & Karney, B. W. 2002 Extended-period analysis with a transient model. *Journal of Hydraulic Engineering* **10**, 616–624.

Forconi, E., Kapelan, Z., Ferrante, M., Mahmoud, H. & Capponi, C. 2017 Risk-based sensor placement methods for burst/leak detection in water distribution systems. *Water Science & Technology* **17** (6), 1663–1672.

Greyvenstein, B. & Van Zyl, J. E. 2007 An experimental investigation into the pressure-leakage relationship of some failed water pipes. *Journal of Water Supply: Research and Technology-AQUA* **56** (2), 117–124.

Hamilton, S. & Charalambous, B. 2013 *Leak Detection: Technology and Implementation*. IWA Publishing, London, UK.

Laucelli, Z. B., Simone, A., Berardi, L. & Giustolisi, O. 2017 Optimal design of district metering areas for the reduction of leakages. *Journal of Water Resources Planning and Management* **143** (6), 04017017.

Liggett, J. A. & Chen, L.-C. 1994 Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering* **120** (8), 934–955.

Marchi, A., Salomons, E., Ostfeld, A., Kapelan, Z., Simpson, A., Zecchin, A., Maier, H., Wu, Z., Elsayed, S., Song, Y., Walski, T., Stokes, C., Wu, W., Dandy, G., Alvisi, S., Creaco, E., Franchini, M., Saldarriaga, J., Páez, D., Hernández, D., Bohórquez, J., Bent, R., Coffrin, C., Judi, D., McPherson, T., van Hentenryck, P., Matos, J., Monteiro, A., Matias, N., Yoo, D., Lee, H., Kim, J., Iglesias-Rey, P., Martínez-Solano, F., Mora-Meliá, D., Ribelles-Aguilar, J., Guidolin, M., Fu, G., Reed, P., Wang, Q., Liu, H., McClymont, K., Johns, M., Keedwell, E., Kandiah, V., Jasper, M., Drake, K., Shafiee, E., Barandouzi, M., Berglund, A., Brill, D., Mahinthakumar, G., Ranjithan, R., Zechman, E. M., Morley, M. S., Tricarico, C., Tolson, B. A., Khedr, A. & Asadzadeh, M. 2014 The battle of the water networks II (BWN-II). *Journal of Water Resources Planning and Management* **140** (7), 04014009.

Mutikanga, H. E., Sharma, S. K. & Vairavamoorthy, K. 2012 Methods and tools for managing losses in water distribution systems. *Journal of Water Resources Planning and Management* **139** (2), 166–174.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. & Zechman, E. 2010 State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management* **136** (4), 412–432.

Perelman, L. S., Allen, M., Preis, A., Iqbal, M. & Whittle, A. J. 2015 Automated sub-zoning of water distribution systems. *Environmental Modelling & Software* **65**, 1–14.

Pérez, R., Puig, V., Pascual, J., Peralta, A., Landeros, E. & Jordanas, L. 2009 Pressure sensor distribution for leak detection in Barcelona water distribution network. *Water Science and Technology: Water Supply* **9** (6), 715–721.

Ponce, M. V. C., Castañón, L. E. G. & Cayuela, V. P. 2014 Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. *Journal of Hydroinformatics* **16** (3), 649–670.

Poulakis, Z., Valougeorgis, D. & Papadimitriou, C. 2003 Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probabilistic Engineering Mechanics* **18** (4), 315–327.

Puust, R., Kapelan, Z., Savic, D. A. & Koppel, T. 2010 A review of methods for leakage management in pipe networks. *Urban Water Journal* **7** (1), 25–45.

Rossman, L. 2000 *EPANET User's Manual*. U.S. Environmental Protection Agency, Cincinnati, OH, USA.

Sadeghioon, A. M., Metje, N., Chapman, D. N. & Anthony, C. J. 2014 Smartpipes: smart wireless sensor networks for leak detection in water pipelines. *Journal of Sensor and Actuator Networks* **3** (1), 64–78.

Samir, N., Kansoh, R., Elbarki, W. & Fleifle, A. M. R. 2017 Pressure control for minimizing leakage in water distribution systems. *Alexandria Engineering Journal* **4**, 601–612.

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

Shafiee, M. E. & Berglund, E. Z. 2017 Complex adaptive systems framework to simulate the performance of hydrant flushing rules and broadcasts during a water distribution system contamination event. *Journal of Water Resources Planning and Management* **143** (4), 04017001.

Shafiee, M. E., Berglund, A., Berglund, E. Z., Brill Jr., E. D. & Mahinthakumar, G. 2016 Parallel evolutionary algorithm for designing water distribution networks to minimize background leakage. *Journal of Water Resources Planning and Management* **142** (5), C4015007.

Soares, A. K., Covas, D. I. & Reis, L. F. R. 2011 Leak detection by inverse transient analysis in an experimental PVC pipe system. *Journal of Hydroinformatics* **13** (2), 153–166.

Soldevila, A., Blesa, J., Tornil-Sin, S., Fernandez-Canti, R. M. & Puig, V. 2018 Sensor placement for classifier-based leak localization in water distribution networks using hybrid feature selection. *Computers & Chemical Engineering* **108**, 152–162.

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

Vairavamoorthy, K. & Lumbers, J. 1998 Leakage reduction in water distribution systems: optimal valve control. *Journal of Hydraulic Engineering* **124** (11), 1146–1154.

Zechman, E. M. & Ranjithan, S. R. 2009 Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. *Water Resources Planning and Management* **135** (5), 334–343.

Zechman, E. M., Giacomoni, M. H. & Shafiee, M. E. 2013 An evolutionary algorithm approach to generate distinct sets of non-dominated solutions for wicked problems. *Engineering Applications of Artificial Intelligence* **26** (5–6), 1442–1457.

First received 12 April 2018; accepted in revised form 13 October 2018. Available online 24 December 2018