S-PLACE GA for optimal water quality sensor locations in water distribution network for dual purpose: regular monitoring and early contamination detection – a software tool for academia and practitioner

Shweta Rathi

ABSTRACT

Security concerns about water distribution networks (WDNs) have led to increased interest in optimizing sensor locations in WDNs achieved through a calibrated hydraulic model. This paper presents a methodology, which consists of two stages. The first stage consists of calibration of a hydraulic model using a genetic algorithm (GA). A real-life network of one of the hydraulic zones of Nagpur city, India, is considered, which optimizes the settings of a throttled controlled valve at different timings for calibration. In this stage, a detailed case study, GA calibration model, methodology and results of calibrated models are discussed. The second stage consists of identifying optimal sensor locations using a newly developed software tool named 'S-PLACE GA' and its efficiency and effectiveness are discussed. It can be used for the dual purpose of routine monitoring of water quality and for early detection of contamination. The optimal locations are obtained considering two objective metrics, 'Demand Coverage' and 'Time-Constrained Detection Likelihood'. These two objectives are combined into a single objective by using weights. Key features, input data required for the software and their applications on (1) BWSN network 1 and result comparison with others and (2) calibrated model of the first stage are discussed. Results showed the effectiveness of S-PLACE GA for practical applications.

Key words | calibration, contamination detection, sensor placement, S-PLACE GA, TCV modeling, water distribution networks

Shweta Rathi
Former Research Scholar, Civil Engineering Department,
Visvesvaraya National Institute of Technology,
Nagpur,
Maharashtra,
India, 440010
E-mail: *shwetanabira12345@rediffmail.com*

HIGHLIGHTS

- Paper highlights the calibration of hydraulic model of water distribution system (WDS) using genetic algorithm performing the modelling of throttled control valve.
- The new software tool 'S-PLACE GA' is developed for dual purpose of routine monitoring of water quality in WDS and for early detection of contamination in case of accidental or intentional contamination simultaneously.
- The new formulation is suggested considering weighted objective function comprising two objective metrics, 'Demand Coverage' and 'Time-Constrained Detection Likelihood'. Application and result comparison are shown on benchmark problem.

doi: 10.2166/ws.2020.333

LIST OF ACRONYMS

a_{jp}	Binary variable, value is 1 or 0
b_{xp}	Binary variable, i.e if contamination event x during pattern p is detected/non-detected, 1 or 0
p	Pattern
q	Nodal demand
t_s^{in}	The start time of injection
t_d^s	Detection time for any contamination
BV	Butterfly valve
BWSN	Battle of sensor networks
CP	Critical point
DC	Demand coverage
DL	Detection likelihood
DS	Distribution systems
EPANET	United States Environmental Protection Agency Network
EPANET-MSX	Multi-species extension
ESR	Elevated service reservoir
GA	Genetic algorithm
IRA-WDS	Integrated risk assessment of water distribution systems
LOS	Level of service
MLC	Minor loss coefficient
MLD	Million liters per day
PE	Population exposed
PMPs	Pressure monitoring points
PTN	Total number of patterns
R	Reservoir/source
S-PLACE GA	Sensor placement genetic algorithm
TD	Time to detection
TCDL	Time-constrained detection likelihood
TCV	Throttled controlled valve
TEVA-SPOT	Threat ensemble vulnerability assessment and sensor placement optimization tool
WaterGEMS	Water distribution analysis geospatial engineering modeling systems
WDSA	Water distribution system analysis
VC	Volume of contaminated water consumed
W	Pre-specified weights assigned to the objective DC
Z	Objective function of sensor placement
ZC	Objective function of calibration
Z1	Expected time of detection
Z2	Expected population affected prior to detection
Z3	Expected consumption of contaminant water prior to detection
Z4	Expected detection likelihood
WDN	Water distribution network
WDS	Water distribution system
$F(i)$	The combined fitness value of chromosome i
$F(\bar{x})$	Minimization of ZC
$f(i, k)$	A fitness value of chromosome i for an individual objective k
$f^*(i, k)$	A normalized fitness value of chromosome i for an individual objective k
f_i	Fitness function
$f_{max}(k)$	The maximum fitness value for an individual objective k
$f_{min}(k)$	The minimum fitness value for an individual objective k
${}_iP_t^m$	Measured pressure head at test node i at time t;
${}_iP_t^s$	Simulated pressure or model predicted head at test node i at time t
$\bar{x}$	Trial solution containing valve settings

INTRODUCTION

Concerns about water quality and vulnerability of water distribution systems (WDS) to contamination have led to increased interest in on-line monitoring through sensors to have early warning against contamination events to mitigate its impact. Thus, water quality sensors basically serve two purposes. Firstly, they provide practical or real-time water quality monitoring to detect water quality degradation in the distribution systems (DS) caused by naturally occurring processes such as nitrification, iron corrosion, bacterial re-growth and so on. Secondly, they help in detecting contamination events to minimize public health exposures after release of a chemical contaminant or biological agent into the DS (in case of accidental or deliberate contamination). Installation, operation and maintenance of water quality sensors are costly, and therefore the number of sensors to be installed must be limited. Therefore, optimal locations for sensors in the network are required.

Lee & Deininger (1992) were the first to address the problem of optimal sensor placement in WDS based on the concept of 'demand coverage' (DC). Watson *et al.* (2004) were the first to introduce multi-objective formulation to sensor placement by employing a mixed integer linear programming model over a wide range of design objectives. Several researchers participated in the 'Battle of Sensor Networks' (BWSN) to locate sensors in two-example networks using their algorithms (Ostfeld *et al.* 2008). Ostfeld *et al.* (2008) compared solutions provided by several algorithms based on four objectives: (1) time of detection (TD) (Z1); (2) population exposed (PE) (Z2); (3) volume of contaminated water consumed (VC) (Z3); and (4) detection likelihood (DL) (Z4). The solutions from different algorithms provided a distinctive set of sensor locations. The main emphasis was thus given to early contaminant detection and minimizing the negative consequences. No importance was given to network coverage in terms of demand. Eliades & Polycarpou (2006) suggested including demand coverage (DC) as a fifth objective, Z5, with the consideration that it would be more useful in comparison when the only objective; that is, the detection likelihood, is the same for two or more solutions. They further obtained the values of DC for different solutions suggested in the literature for the BWSN networks. Hart & Murray (2010) and Rathi *et al.* (2015) provided a review of various methodologies on sensor placement and discussed several challenges and issues regarding methodologies and applications.

Some of the issues of concern with reference to application of methodologies to large WDNs are as follows. (1) *Skeletonized network* (Dorini *et al.* 2006; Huang *et al.* 2006; Wu & Walski 2006, Aral *et al.* 2010, Perelman *et al.* 2008; Perelman & Ostfeld 2012) – the skeletonization of the network should be such as to reduce the size the of the network without compromising the hydraulic and water quality equivalences between the whole and reduced network. Therefore, clustering of nodes and placing sensors cluster-wise can be considered the best solution for network reduction. (2) *Selection of contamination events* (Watson *et al.* 2009; Perelman & Ostfeld 2010, 2012; Shen & Mcbean 2011; Davis *et al.* 2014) – due to the uncertain nature of contamination events (i.e what, how much, when, where and for how long an event will occur) the size of the problem should be reduced either by restricting the number of injection locations and limiting injection time or by random sampling, considering importance-based sampling, or by designing a system based on consideration of high impact scenarios. (3) *Event modelling* (Ostfeld & Salomons 2006; Weickgenannt *et al.* 2010; Afshar & Marino 2012) – consideration of each and every scenario for simulation requires high computational cost in terms of computer memory and computational speed; hence, the computational requirement can be reduced by identifying a few important scenarios with respect to nodal demands, pump on and off situations and flow in and out conditions from tanks. Further, some have focused on conservative contaminants (Dorini *et al.* 2010; Weickgenannt *et al.* 2010), changes in the contaminant due to chemical reaction (Cozzolino *et al.* 2005) and EPANET-MSX for event modeling with contaminant reactions (Ohar *et al.* 2015).

(4) *Number of performance objectives* (Huang *et al.* 2006; Krause *et al.* 2008; Xu *et al.* 2008; Aral *et al.* 2010) – selection of objectives should be such that a balanced design can be obtained with respect to different objectives. A unified single objective formulation (Aral *et al.* 2010) jointly considers various parameters affecting sensor locations, which makes the problem formulation simple and provides better sensor locations compared to other methodologies.

(5) *Type of solution methodologies and software* – Preferred solution methodology is one which can prioritize selection with respect to different objectives considered in the sensor location problems. Prioritization selection helps in future extension of monitoring locations. Software like TEVA-SPOT (Berry *et al.* 2008, 2010; Hart *et al.* 2008), and the S-PLACE tool kit (Eliades *et al.* 2014), are required to handle large size networks, select flow distributions, restrict contamination events and limit objectives.

(6) *Uncertainty consideration* – Shastri and Diwekar (2006), Ostfeld & Salomons (2005), Ostfeld *et al.* 2004, Comboul & Ghanem (2013). The complexities in sensor network

design are due to various uncertain parameters, such as: flow reversal in some links (requiring consideration of either multiple or dynamic simulations), uncertainty in attack in terms of location, time, type, strength and duration of contaminant intrusion (requiring consideration of a large number of contamination scenarios and water quality simulations), uncertainty in nodal demands, sensor detection limits, delay by sensor in registering an event and raising alarm, imperfectness of sensors resulting in false alarms, imperfect mixing at nodes, uncertainty involved in quantification of impacts on human health and other impacts, etc. Sensor network design to select optimal locations of a limited number of sensors involving these complexities is computationally expensive. and further requires a calibrated hydraulic model for proper placement.

(7) *Mobile sensors* – Mobile sensors, along with stationary sensors, are to be considered for applications in WDNs which improve effectiveness and help in reduction of cost (Perelman & Ostfeld 2013; Oliker & Ostfeld (2015)). Recently, Berglund *et al.* (2020) provided a compressive review for several challenges that the water industry faced for managing WDS security. This review provided a modeling framework that has been developed to assist in the management of emergencies and WDS security.

Calibration of the hydraulic model is the basic requirement for obtaining optimal sensor locations in WDS. However, calibration of a hydraulic model of a water distribution system is a challenging task as there are a large number of input parameters such as nodal elevations and demands, including water losses, pipe length, diameter and roughness coefficients, water level at the source and pump characteristics, valve settings etc. It is desirable that most of these parameters are measured accurately during field observations so that few uncertain parameters require adjustments during the calibration. The field survey provides reliable data regarding pipe length, ground levels, operating valve settings and status, and water level at the sources. In general, data regarding the nodal consumption and pipe roughness coefficients are less reliable and therefore may need adjustment during the calibration. Recently, Rathi *et al.* (2020) considered a case study of a real-life network of one of the hydraulic zones of Ramnagar Elevated Service Reservoir (ESR) of Nagpur city, Maharashtra State in India. Study area, complexity, challenges and steps for traditional calibration and initial calibration results are discussed in detail in Rathi *et al.* (2020). After detailed study, it was observed that there is a need to calibrate the model using optimization techniques because of uncertain data. Therefore, a hydraulic model is developed in EPANET consisting of 69 throttle control valves (TCVs) and openings are approximated by the minor loss coefficient (MLC), which replicates actual valve status in the field. After detailed analysis, it was observed that one of TCVs is critical and there is a need to calibrate the status (settings) of this identified TCV in the Ramnagar ESR premises zone (Figure 1), which is operated at different timings in a day. Therefore, in this study model, it is calibrated using a genetic algorithm (GA) which optimizes one of the throttled controlled valve statuses (settings) which are operated at different timings in a day.

Considering the two aspects of regular/routine water quality monitoring and contamination event detection two objective metrics, 'demand coverage' (DC) and 'time-constrained detection likelihood (TCDL)' are selected for optimal locations of sensors in WDNs. The demand coverage is a useful parameter for regular monitoring as it provides the importance of including maximum nodes on the paths from source to various sensor locations to cover maximum nodal demands. Thus, quality of water can be considered satisfactory at all the upstream nodes if the quality of water is found to be satisfactory at the sensor node. To maximize DC, any optimization algorithm tends to place sensors away from the source. The time constrained detection likelihood provides the importance to detection of contamination events within permissible time called as level of service (LOS). Thus, depending upon the LOS, location of sensors from source and the distance between them is optimized. This is a useful parameter for accidental or intentional contamination. Therefore, two objectives of sensor locations; that is, regular monitoring and contamination detection, were considered together in optimizing the sensor locations (Eliades & Polycarpou 2006; Rathi & Gupta 2015; Rathi & Gupta 2017a). Several solutions can be obtained by varying the weights of two objectives.

Based on this methodology, a software program called S-PLACE GA as a tool was developed in 'C' programming language to select optimal locations of a given number of water quality sensors using the GA for easy application to

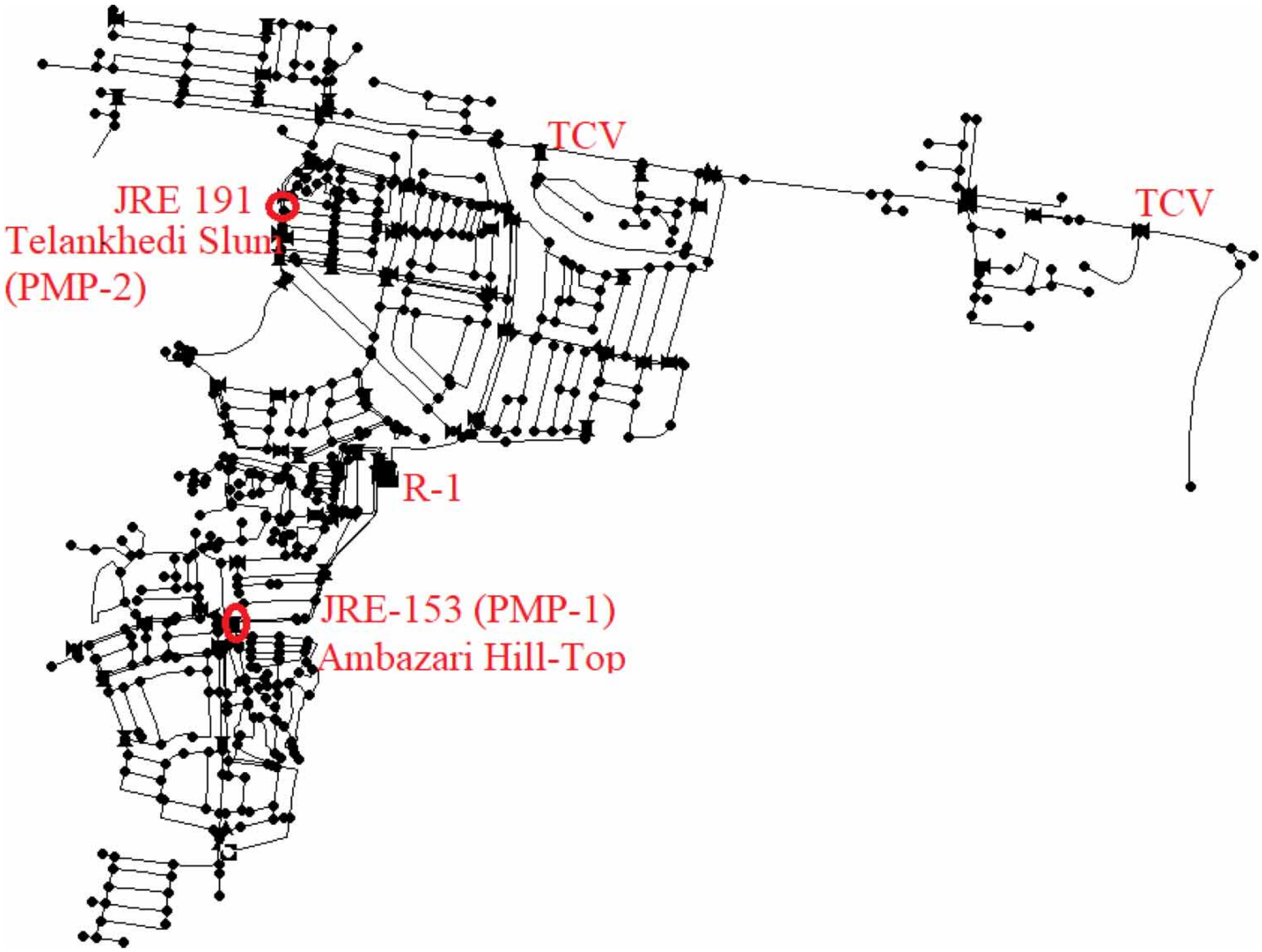

Figure 1 | Layout of water distribution network of Ramnagar ESR Nagpur, Maharashtra, India, consisting of 69 TCV (EPANET Model).

large WDS problems. S-PLACE GA is used for the dual purpose of regular monitoring of water quality and for early detection of contamination in WDN. Another important feature of the S-PLACE GA is that optimal sensor locations can be obtained under these conditions: (1) when all nodes of the network are considered as vulnerable points; (2) when some nodes are considered as vulnerable points of the system based on risk analysis with equal probability of occurrence; and (3) vulnerable nodes are considered with probability of occurrences based on the risk score.

The methodology of the manuscript is organized into two stages. (1) Stage 1 – discussed in detail about calibration of hydraulic model using GA using TCV modeling which includes about the study area and network details, traditional calibration and GA calibration model. (2) Stage 2 – discussed about new developed S-PLACE GA – a tool for optimal locations of water quality sensors for regular water quality monitoring and contamination event detection simultaneously in WDS. Key highlights of the software, objective function, input data required for the software, and GA process and operator are also discussed. Applications of S-PLACE GA software as a tool are shown through two problems, one a benchmark problem of BWSN network 1 and comparison of results with other methodologies are shown (Rathi & Gupta 2017b), and another is on a calibrated model of stage 1.

The novelty of the paper is that in the first stage, calibration of the hydraulic model is done using GA using throttle control valve (TCV) modeling, wherein one TCV valve's status and settings are optimized at different times of a day. Time is known from field operations; however, the settings are unknown. Therefore, the hydraulic model is calibrated using GA. In the second stage, S-PLACE GA – a software tool – and its efficiency and effectiveness are discussed. Key features, input data required for the software and its application on a large real-life calibrated model are discussed. Even though Rathi & Gupta (2017b) carried out a comparison of S-PLACE GA on BWSN network 1, its applicability on a real life calibrated model having 619 nodes and 733 pipes is shown here. On a calibrated network it is observed that only 21 sensors are required for 50% coverage in terms of demand and contamination detection and 32 sensors are required for covering 60% of coverage criteria. Further best 5, 20, 30 and 40 optimal sensor locations are also obtained covering 25%, 49, 58 and 65% of the coverage area respectively in the case study. This shows the effectiveness of S-PLACE GA for applications. After comparison on a BWSN network by Rathi & Gupta (2017b), it was observed that high demand coverage is achieved with a little compromise on other objectives. Discussion of results of stage 1 and stage 2 are discussed in section 4 and conclusions are given in section 5.

METHODOLOGY USED

Stage 1: calibration of hydraulic model using genetic algorithm

Study area and network details from Rathi *et al.* (2020)

Nagpur, a second capital of Maharashtra State, is situated in the center of India. The population of the city is 2.53 million (2015) spread over 217 sq. km. The total water supply to the city is over 650 MLD through 54 service reservoirs and a 3,200 km long distribution network covering 239,000 connections. The study area consists of a very complex hydraulic zone of Ramnagar ESR, located in the western part of Nagpur City and composed of a residential area with all types of dwellings, including slum areas, as well as institutional and commercial areas. During the conversion of mode from intermittent to continuous, all property connections were changed, faulty meters were replaced/repaired, illegal connections were removed/regularized, public stand posts were converted to grouped pipe connections with metering, and old pipelines (about one-third of the total pipelines) were replaced by new pipelines.

The water distribution network shown in Figure 1 has one source of supply (R-1), 619 nodes and 733 pipes. The network is organized in 33 subzones and consists of 12 wards. Out of 33 subzones, 26 are fully served by Ramnagar ESR and seven are partly served. The pressure and flow in these 33 subzones are controlled by 69 valves. Each of the subzones is controlled by one or more valves. Two predefined pressure monitoring points (PMPs), or critical points (CPs), are installed in the network, named as Ambazari Hill-Top Slum (PMP-1), which is on higher elevations, and the other PMP-2, which is named as Telankhedi Slum, is on lower elevation; these are shown by circle in Figure 1.

Traditional or manual calibration of the model

Rathi *et al.* (2020) discussed in detail several challenges in the study area and performed various stages of traditional calibration, which include the following. (1) *Collection of data* – this includes network data, demand data, operational data and source head data and supply from the source(s), which are collected from field survey and the water authority. (2) *Hydraulic model preparation and initial calibration results* – based on the available data, a hydraulic model is developed consisting of 69 partially closed valves. Most of those valves are adjusted at a particular fixed setting values; for example, percentage closer are shown in Table S1 (supplementary material) (Rathi *et al.* 2020). The valve can be operated and set at different degrees of opening, which will result in different pipe flows and junction pressures. These valves are considered as throttled control valves (TCVs) in EPANET during modeling and are shown in Figure 1. In hydraulic simulation models, the effect of the valve openings can be approximated by the minor loss coefficient (MLC). Results obtained using the initial calibration showed that there is a large pressure difference between observed and simulated pressure at both the PMPs shown in Figure S1 (supplementary material). (3) *Data validation, hydraulic model improvement, micro calibration, and results and conclusions* – detailed study of the network from field survey, physical verification and macro and micro level of calibration is performed. It was observed that there are two valves that are operated in the ESR premises, of which one is fixed at a particular valve setting value and the other is operated at different times of the day. A number of trials have been carried out for calibration of the model. Initially, the valve status of both the selected valves in ESR premises are considered as unknown parameters and calibration of the model using GA was tried. However, the complexity is greater in the model, and from the field visit it was known that one valve from ESR premises is operated at different times in a day. Therefore, in this study, the model is calibrated by identifying the valve status at different times in a day for the valve, which is operated in ESR premises and is shown by circle in Figure 2.

GA calibration model

The objective function (Equation (1)) is the minimization of the sum of the square of difference between the measured and simulated value of pressure heads at test nodes.

$$\min ZC = \sum_{i=1}^{n} \sum_{t=0}^{T} ({}_iP_t^m - {}_iP_t^s)^2 \tag{1}$$

where, ${}_iP_t^m$ = measured pressure head at test node i at time t;

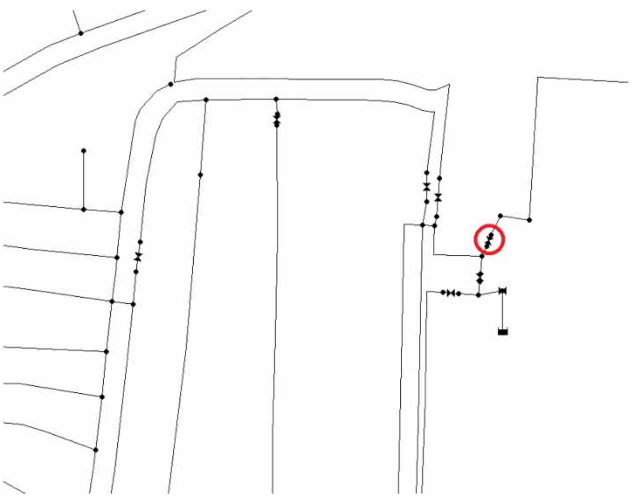

Figure 2 | TCV in ESR premises in EPANET model, which are operated at different times in a day.

${}_iP_t^s$ = simulated pressure or model predicted head at test node i at time t.

The decision variables are valve status (settings) identification at different timings for the identified critical valve (from field survey) in Ramnagar ESR premises, which is shown by circle in Figure 2. From a field survey, it was observed that the valve's operation schedule at different times in a day are: at 8 am, 8 pm, 8.25 pm, 10.25 pm, 12 am and so on. Therefore, the trial solution valve settings of the identified valve at different timings are considered as the decision variables. The valve setting variables are the values that represent open, closed and partially open. Let $\bar{x}$ represent the trial solution containing valve settings of a selected valve, the valve identification model is then formulated as a nonlinear optimization problem (Wu *et al.* 2012):

Search for $\bar{x}$

Minimize $F(\bar{x})$

$F(\bar{x})$ is the objective function given in Equation (1).

Thus, the valve settings of the unknown identified valve have multiple possible values at different times of a day; say, for example, at 8 am, 8 pm, 8.25 pm, 10.25 pm, 12 am etc (Table 1). The valve is operated and set at different degrees of opening, which will result in different pipe flows and junction pressures. In the EPANET hydraulic simulation model, the effect of the valve openings is approximated by the Minor Loss Coefficient (MLC). For an identified valve, random settings are selected from a range of minimum and maximum values of 0–3,500 are considered. The MLC value 0 represents fully open status and the value of infinity or a large MLC value in practice can be used to represent fully closed status. Examples of the trial solution are given in Table 1.

Measured values are obtained from field measurement devices using pressure loggers, which are installed at two

Table 1 | Examples of trial solutions

Trial Solution	0 to 8 am	8am to 8 pm	8 pm to 8.25 pm	8.25 pm to 10.25pm	10.25 pm to 24 am
Value	100.00	250.00	3,000.00	10.50	1,000.25

pre-defined places, and model pressure head values are obtained by running a hydraulic simulation toolkit such as EPANET with randomly generated solutions. Flows, heads and nodal demands are time dependent. Measured pressure head at test nodes and readings from supervisory control and data acquisition (SCADA) were taken at a time interval of every 15 minutes during an extended period simulation. The generalized diurnal demand pattern was created based on an hourly flow recorded by the flow meter installed on the outlet pipe of the Ramnagar ESR; that is, at the inlet of the distribution network. The observation of flow from the source, the water level in the reservoir and flow multipliers on the day of observation (31 October 2018) are given in Table S2 (supplementary material). The average flow during the day was 734.14 m^3/h.

The GA calibration model implemented has been written in C language and the flow chart of the algorithm is shown in Figure 3. Input for calibration of a model is collected from SCADA and field survey and observed pressures are measured at two pre-defined PMPs (Figure 1).

An initial population of chromosomes is randomly generated. Randomly generated valve statuses (openings or minor loss coefficients) selected from a range of minimum and maximum value of 0–3,500 are considered. This range is considered from the rough approximations from the field survey and observations of the case study from Rathi *et al.* (2020). This status is assigned to this identified critical valve, which is at different timings, and EPANET is subsequently called to simulate the steady-state hydraulic analysis of the system. Simulated pressures at the measurement locations are obtained and compared with their measured values through the calculation of the objective function. The fitness function for each member of the GA population is defined as the inverse of the objective function applied to define. Using GA operator's selection, crossover and mutation, new generations of populations

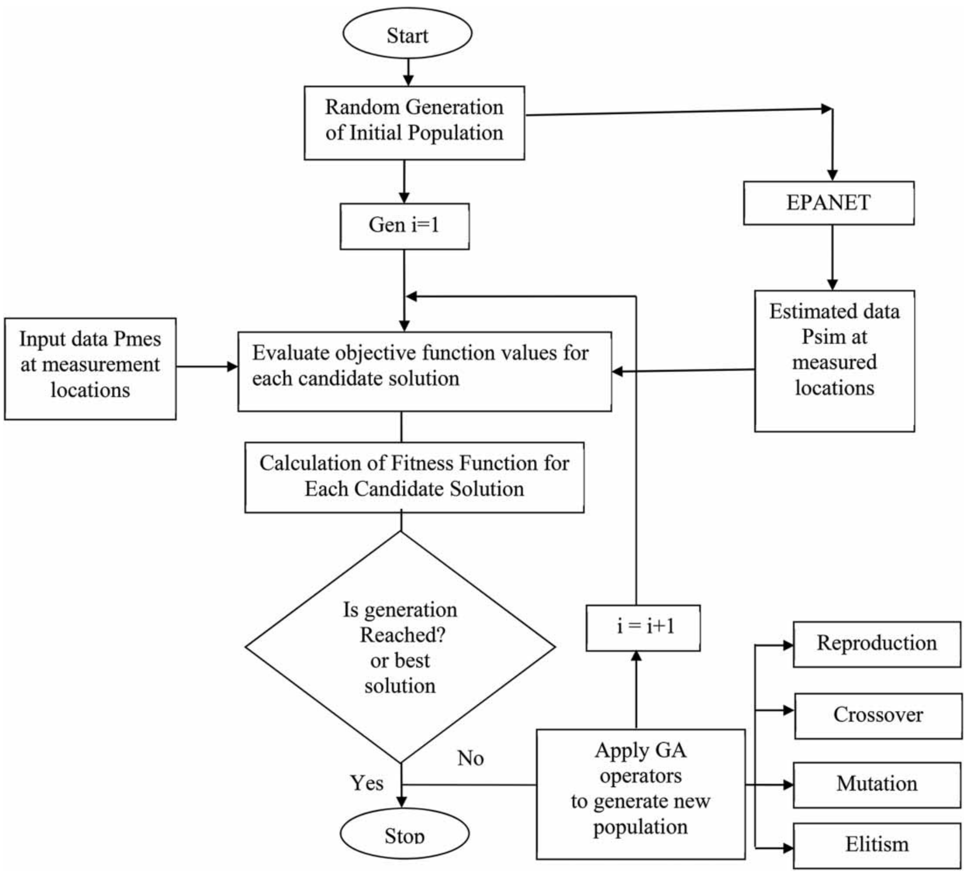

Figure 3 | Flow chart of hydraulic calibration model using GA.

are created and the calibration process is then repeated until the stopping criteria are met. In this study, a tournament selection, two-point crossover, and bitwise mutation have been applied in the GA calibration model to improve the performance of the GA model, as recommended by Goldberg & Deb (1991) for better convergence. Results of the GA model are given in Table 2 and in Figure 4(a) and 4(b).

Table 2 | Final solutions obtained using GA for calibration

Trial solution	0 to 8 am	8am to 8 pm	8 pm to 8.25 pm	8.25 pm to 10.25pm	10.25 pm to 24 am
Value	5.95603	929.72033	1,758.04126	2,964.26977	5.95603

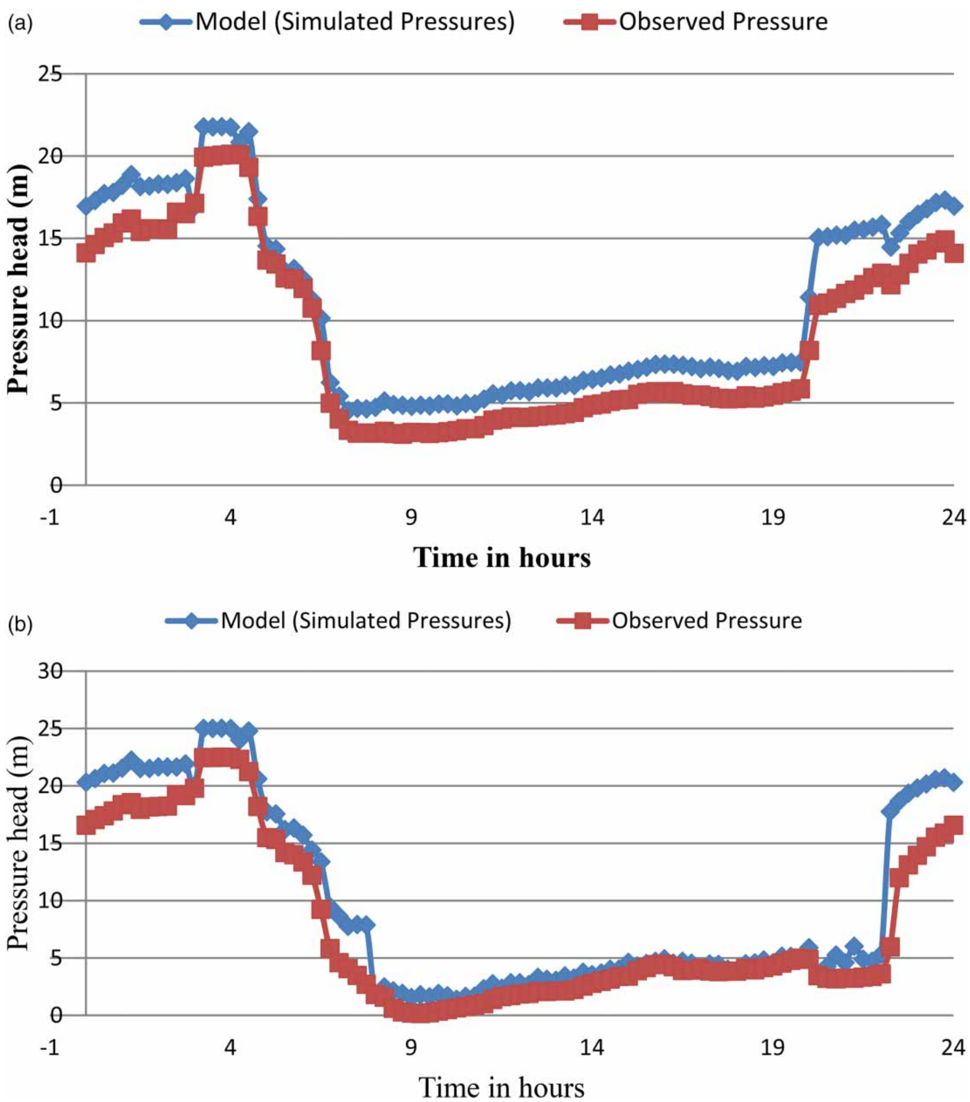

Figure 4 | Comparison of computed/simulated pressure with observed pressure at both CPs. (a) Ambazari Hill-Top Slum (PMP-1/CP1) (b) Telankhedi Slum (PMP-2/CP2).

Stage 2: S-PLACE GA – a tool for optimal locations of water quality sensors in water distribution network

Key highlights of the software

- A newly developed software tool named S-PLACE Genetic Algorithm (GA) is used to select optimal locations of a given number of water quality sensors in WDN using the methodology of Rathi & Gupta (2015) for easy application to a large WDS network.
- The optimal locations are obtained considering two objective metrics, 'demand coverage' (DC) and 'time-constrained detection likelihood (TCDL)'. These two objectives are combined into a single objective by using weights. This code is written in C language.
- It can be used for regular water quality monitoring and also in case of accidental or intentional contamination in WDSs.
- This tool is simple, easy to use and is verified on a benchmark problem; that is, BWSN network 1, and the results are compared with other studies. It is observed that high demand coverage is achieved with a little compromise on other objectives (Rathi & Gupta 2017b). Further, its applications are shown on a calibrated real-life network. The best 5, 20, 30 and 40 optimal sensor locations are also obtained covering 25, 49, 58 and 65% of the coverage area respectively in the case study.
- The genetic algorithm requires hydraulic simulation results from software like EPANET, or WaterGEMS or WaterCAD or others.
- Another important feature of GA is that robust sensor network design is obtained under various conditions: (1) when all nodes of the network are considered as vulnerable points; (2) when some nodes are considered as vulnerable points of the system based on risk analysis with equal probability of occurrence; and (3) when vulnerable nodes are considered with probability of occurrences based on the risk score.

Objective function

The objectives of regular monitoring and contamination event detection are considered through (1) maximization of demand coverage (DC) and (2) maximization of time-constrained detection (TCDL) likelihood respectively.

The demand coverage (Equation (3)) is defined as the percentage of network demand monitored by a particular sensor or set of sensors located at different nodes in a WDS (Lee & Deininger 1992; Rathi & Gupta 2014). Kumar *et al.* (1999) defined detection likelihood (DL) considering only those events in the list of detected events that have time of detection (TD) less than some acceptable time and are termed as time-constrained DL (TCDL) (Equation (4)). Thus, any event detected beyond a specific time interval is not included with computing detection likelihood. Maximization of TCDL would restrict sensor locations to move away from the source, but within the considered TD. Thus, TCDL combines the impact of two objectives, TD and DL.

DC is maximized to assure the quality of delivered water to consumers during regular monitoring; TCDL is used for early detection of a greater number of contamination events. Maximization of TCDL minimizes the undetected events and their negative consequences indirectly in terms of PE, VC, EC and risk and assuring early detection of contamination events.

These two objectives are combined by assigning different weights to them, and determination of DC and TCDL are given by Rathi & Gupta (2015).

$$\text{Max } Z = W \times DC + (1 - W) \times TCDL \tag{2}$$

W = pre-specified weights assigned to the objective DC.

The objective of regular monitoring through maximization of DC is given by f_1:

$$DC = f_1 = \frac{\sum_{p=1}^{PTN} \sum_{j=1}^{J} a_{jp} \times q_{jp}}{\sum_{p=1}^{PTN} \sum_{j=1}^{J} q_{jp}} \tag{3}$$

where, $a_{jp} = 1$, if node j is covered by any of the sensor nodes during pattern p, else $a_{jp} = 0$; q = nodal demand. J = number of covered nodes, PTN = number of flow patterns.

The objective of contamination detection through maximization of TCDL is expressed by f_2

$$TCDL = f_2 = \sum_{x=1}^{J} \sum_{p=1}^{PTN} \sum_{t=t_s^{in}}^{t_d^s} \alpha_{xp} b_{xp} \tag{4}$$

where, α_{xp} = probability of event x during pattern p and it is calculated as per the expert's knowledge of network vulnerability; $b_{xp.} = 1$, if contamination event x is detected during the pattern PTN. t_s^{in} is the start time of injection of contamination. t_d^s is the maximum detection time for any contamination event, which is pre-defined or pre-specified or it is the maximum level of damage prior to a pre-defined detection time. These two objectives are combined by assigning weights based on the priority of these objectives, as also done by Krause *et al.* (2008) and Xu *et al.* (2008).

Since the objective function consists of two different parameters with different units, a normalization method for these fitness functions is necessary. The following describes the normalization method and the same used herein, as suggested by Huang *et al.* (2006). Let $f(i, k)$ = a fitness value of chromosome i for an individual objective k, where $i = 1, \ldots, P$ and P is the population size; $f^*(i, k)$ = a normalized fitness value of chromosome i for an individual objective k as given in Equation (3); $f_{\max}(k)$ = the maximum fitness value for an individual objective k; $f_{\min}(k)$ = the minimum fitness value for an individual objective k; and $F(i)$ = the combined fitness value of chromosome i as given in Equation (4).

$$f*(i, k) = \frac{f(i, k) - f_{\min}(k)}{f_{\max}(k) - f_{\min}(k)} \tag{5}$$

$$F(i) = \sum_{k=1}^{K} w \times f*(i, k) \tag{6}$$

The objective function needs to be maximized; w = pre-specified weights assigned to the objective DC. $F(i)$ = combined fitness value of chromosome i.

Input data required for the software

S-PLACE GA requires the hydraulic analysis data obtained from any hydraulic simulation software like EPANET, WaterGEMS, WaterCAD or others. Pipe data; that is, pipe number, upstream node, downstream node, length, travel time and flow (Table 3) and node data; that is, node number and nodal demand, are required (shown in Table 3). Hydraulic simulation of a network can be used for obtaining the data of travel times and flows in different pipes.

S-PLACE GA can be used for robust sensor network design. Although intrusion may take place at any point in the network, only the nodes that are in risk-prone areas can be considered to limit the number of contamination events. Value of 1 indicates the node is selected as a possible location of contaminant intrusion and 0 indicates no risk at that node, and based on the probability of contamination of any node risk is assigned to that node. This is a useful feature to restrict the number of nodes as possible contamination events for simulation. Restricted nodes for objective function evaluation and their probability of contamination are given in Table 3.

Using any risk analysis model or software, such as the improved risk assessment of water distribution systems (IRA-WDS) developed by Vairavamoorthy *et al.* (2007) and used by Rathi *et al.* (2016) or Naserizade's model (Naserizade *et al.* 2017) or methodology, vulnerable nodes are identified. Once the vulnerable nodes are identified, then based on the probability of their occurrence, the fitness function is evaluated. Thus, S-PLACE GA has a special feature that it can model those contamination events that are at risk and different probabilities of contamination can be added to those nodes. Accordingly, sensor network design is obtained.

GA process and operator

For fitness function calculation, three parameters are needed: T-hour LOS - for example, 5 hours; TCDL multiplier - for example, 0.5; DC weightage - for example, 0.5, crossover rate; for example, 0.95, mutation rate; for

Table 3 | Pipe and node data for S-PLACE GA

Pipe no.	Upstream node	Downstream node	Length (ft)	Flow (GPM)	Travel time (Hour)	Node no.	Nodal demand (GPM)	Probability of contamination
1	0	17	359	2,964.26977	0.554	1	3.83	0.5

example, 0.05; population size - for example, 100; generation size - for example, 100; selection type - for example, roulette wheel selection, tournament selection, crossover type – one point crossover, two point crossover, uniform crossover, number of sensors, mode: maximization, elitism: true, crossover rate: 0.95 (e.g.), mutation rate: 0.05 (e.g.), population size: 100 (e.g.), generation size: 100 (e.g.).

Required number of generations, best solutions, worst solution, maximum, minimum and average fitness values are obtained at the end. GA is used to solve the optimization problem in Equations (2) and (6).

A simple GA started with selection of a random population of candidate solutions. Each member in the population is given a chance to improve in the next generation based on its fitness. Three basic GA operators — reproduction, cross-over and mutation — are used. Elitism is also used to retain one or a few good solutions in the next generation without change. The process of improvement continued for a pre-selected number of generations and the best solution at the end is considered as the optimal solution (Deb *et al.* 2000, Ostfeld & Salomons 2004). The chart of S-PLACE GA is given in Figure 5.

APPLICATIONS OF S-PLACE GA ON EXAMPLE NETWORKS

Application to BWSN network 1

'The Battle of Sensor Network (BWSN): A Design Challenge for Engineers and Algorithms' was held as a part of the Eighth Annual Water Distribution System Analysis (WDSA) symposium in Cincinnati, Ohio, in 2006. This was prepared to discuss the most efficient and effective solution for optimizing sensor deployment locations to minimize the impact of contamination events (Ostfeld *et al.* 2008). The BWSN organizer asked the participants to place 5 and 20 sensor locations using their algorithm or methodology for the two networks of BWSN, network 1 and network 2 of increasing size for base case A and three derived cases B, C and D (Ostfeld *et al.* 2008). In contrast to network 1, network 2 is a large WDN and with greater complexity. Four test scenarios, A, B, C and D, are explained in detailed in Ostfeld *et al.* (2008).

Rathi & Gupta (2017b) considered Network 1 of BWSN, which has 126 nodes, 1 source, 2 tanks, 168 pipes, 2 pumps, and 8 valves having four variable demand patterns shown in Figure S2 (supplementary material), which were considered for obtaining an optimal solution using S-PLACE GA. The EPANET input files of the network are available on the website http://www.exeter.ac.uk/cws/bwsn (Ostfeld *et al.* 2008).

GA and its parameters

The unconstrained optimization problem (Equation (6)) is solved using simple GA by pre-defining a number of sensor locations on BWSN Network 1. Considering all nodes as a possible contamination event (equal probability in Equation (4)), the fitness function is evaluated. A contamination event is considered detected if its time of detection is found to be less than the selected level of service. Level of service is the permissible time within which any contamination event is desired to be detected.

Results from Rathi & Gupta (2017b)

The network was simulated for 96 hours to acquire the information of the dynamic flow pattern. The dominating flow patterns and average discharges and velocities in the pipe lines of the network are calculated and used for location of sensors. Initially, the weights are varied for both the objectives. The best 5 and 20 sensor locations are obtained by considering LOS given by a time of detection of 3 hrs. Thus, a contamination event is considered detected if is detected within 3 hours by any of the sensors. The GA program is run five times and solutions with maximum objective function values in five runs are reported for each combination of weights. GA parameters used for this network are given in section 2.2.4. The results of the best five and 20 locations are provided in Tables S3 and S4 (Supplementary Material) and some observations are given in the supplementary material (Rathi & Gupta 2017b).

Comparison of results (Rathi & Gupta 2017b)

Various optimization or heuristic models with a total of 14 different algorithms were discussed, compared, and applied to determine the optimal locations of sensors with respect to

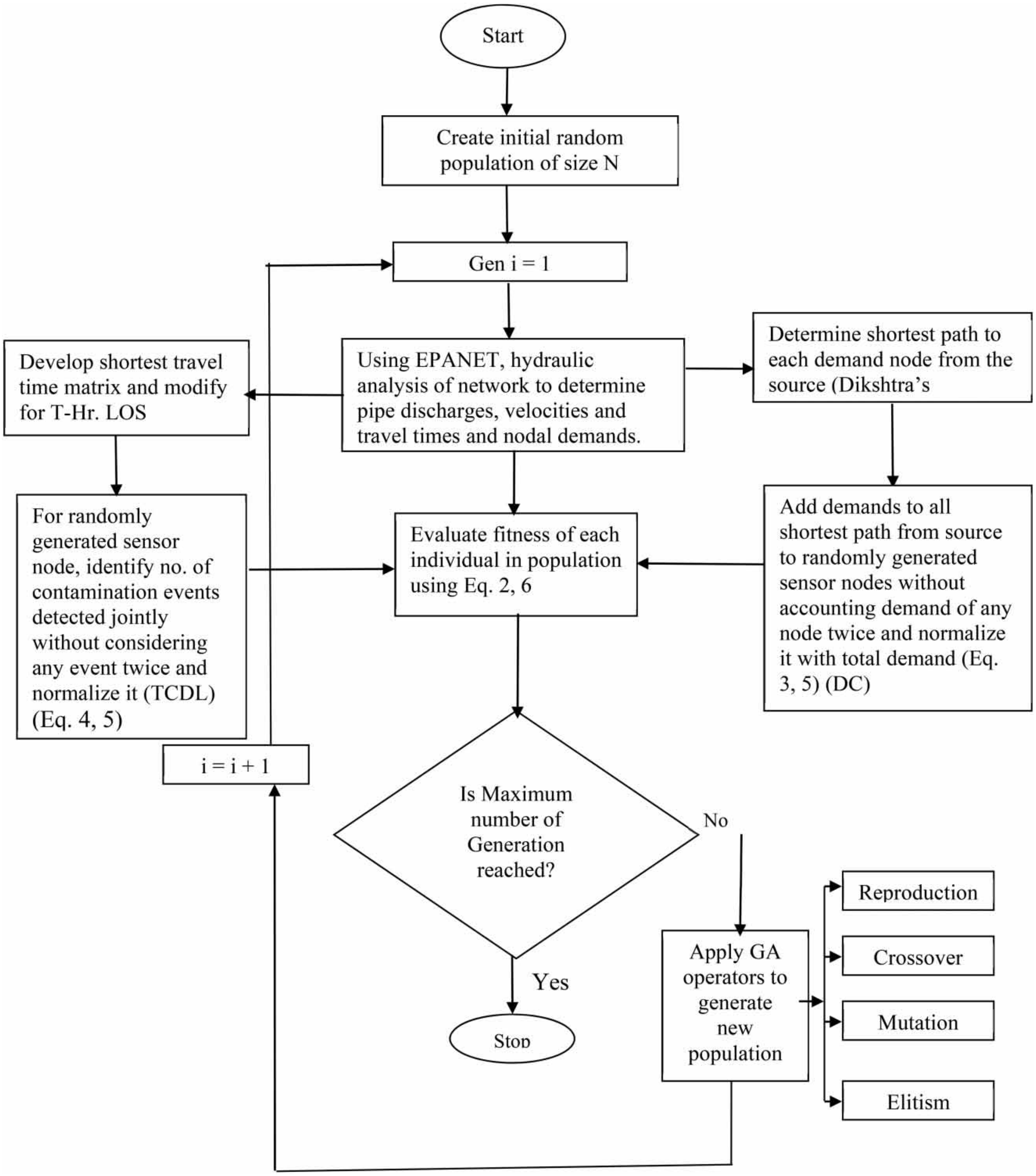

Figure 5 | Flow chart of S-PLACE GA.

a pre-defined (five and 20) number of sensors for BWSN network 1. Four design objectives, including the expected time of detection (Z1), the expected population affected prior to detection (Z2), the expected consumption of contaminant water prior to detection (Z3), and the detection likelihood (Z4) were used to evaluate various algorithms in response to events that can occur due to intentional attack from terrorists. The optimal solution, say SGA1; that is, the best five nodes obtained using the S-PLACE GA for sensor locations for 10-hour LOS are 58, 83, 101, 118, and 124, shown in Table 4. To compare the performance of the S-PLACE GA against these existing 14 optimization and

Table 4 | Comparison of results obtained from proposed GA with other optimization and heuristic algorithms for five sensor solutions (Rathi & Gupta 2017b)

Methodology (1)	Sensor location (nodes) (2)	Z1 (min) (3)	Z2 (People) (4)	Z3 (gallons) (5)	Z4 (DL) (%) (6)	Z5 (DC) (%) (7)
Berry *et al.* (2006)	17,21,68,79,122	542	140	2,459	0.609	25.23
Dorini *et al.* (2006)	10,31,45,83,118	1,068	258	7,983	**0.801 (1)**	30.37
Eliades & Polycarpou (2006)	17,31,45,83,126	912	221	7,862	**0.763 (3)**	34.39
Ghimire & Barkdoll (2006a)	126,30,118,102,24	432	357	4,287	0.367	68.06
Ghimire & Barkdoll (2006b)	126,30,102,118,58	424	331	3,995	0.402	71.40
Guan *et al.* (2006)	17,31,81,98,102	642	159	2,811	0.663	22.88
Gueli (2006)	112,118,109,100,84	794	403	10,309	0.699	31.00
Huang *et al.* (2006)	68,81,82,97,118	541	280	4,465	0.676	33.44
Krause *et al.* (2006)	17,83,122,31,45	842	181	3,992	**0.756 (4)**	24.02
Ostfeld & Salomons (2006)	117,71,98,68,82	461	250	4,499	0.622	22.76
Preis and Ostfeld (2006)	68,101,116,22,46	439	151	7,109	0.477	12.63
Propato & Piller (2006)	17,22,68,83,123	711	164	3,148	**0.725 (5)**	27.45
Trachtman (2006)	1,29,102,30,20	391	142	2,504	0.237	26.41
Wu & Walski (2006)	45,68,83,100,118	704	303	8,406	**0.787 (2)**	37.34
Chang *et al.* (2011)	47,68,76,97,118	479	209	2,824	0.575	31.83
Rathi & Gupta (2015) (SGA1)	58, 83,101,118,124	724	297	4,510	**0.725 (5)**	43.85

heuristics models, four designed objectives, denoted by Z1–Z4, were used as the performance criteria applied for BWSN networks.

BWSN software utilities developed by Salomons (2006) were used to achieve the comparisons. Utility 1, which is 'Build injection data', allows the user to create the data needed to evaluate the fitness function for a given sensor layout design, and utility 2; that is, 'Calculate fitness', allows the user to calculate the fitness function for a given sensor layout design. After running both utility 1 and utility 2, respectively, the four design objectives from Z1 to Z4 can be generated based on the sensor deployment locations analyzed by S-PLACE GA. The four performance objectives were evaluated by using the BWSN utility software are (1) Z1 = 724 minutes, (2) Z2 = 297 persons, (3) Z3 = 4,510 gallons, and (4) Z4 = 0.725 (i.e. 72.5%). These values are observed to be comparable with solutions provided by other researchers. The same comparison is provided by Rathi & Gupta (2017b).

An analysis program is also developed herein based on Equation (2) to obtain the values of DC or TCDL for solution (shown in Table 4) SGA1. The DC (Z5) obtained for solution SGA1 is 0.439 (43.9%). DC for all the 14 models and algorithms is further evaluated using the same software. This general analysis program can further be used to evaluate DC and TCDL values for known sensor locations, as suggested by the methodology of Rathi *et al.* (2015).

Application of S-PLACE GA on calibrated model of case study

Application of S-PLACE GA is further shown on a calibrated model of the case study of Nagpur city, India (stage 1). The optimal sensor locations are also obtained with the combined objective of maximizing DC and TCDL, which is given by Equation (2) using S-PLACE GA. The best five, 20, 30 and 40 sensor locations are obtained for equal weights of 50% for each objective by considering 1-hour LOS. The results are provided in Table S4 (supplementary material). It is observed that the objective function values vary from 0.2569 to 0.6596 for five sensor locations to 40 sensor locations. It is observed that 21 sensor locations covered a combined objective function value of 0.5077 (that is, 50.77% coverage in terms of demand and contamination detection) and 32 sensor locations covered a combined objective function value of 0.6046. Figure 6 shows 32 sensor locations with 60% coverage criteria.

DISCUSSION

Discussion of results of stage 1

Results of hydraulic calibration of the model showed that the difference between observed and simulated pressure is small for Ambazari Hill-Top Slum (PMP-1/CP1) located at a higher elevation, say 1–2 meters for all timings and for one of the 15 minute slots difference between observed and simulated pressure goes to 4 m. However, 96 values (after every 15 minutes) are considered for simulation and calibration. For another critical point, such as for Telankhedi Slum (PMP-2/CP2) located at a lower elevation, from 12 am to 6 am, the difference is less; that is, up to 2.5 m and from 6 am to 8am, ranges from 4 m to 6 m, for timings 8 am to 9 pm and ranges from 0.8 m to 2 m. In timings, 9 pm to 10 pm, values lie between 2 m and 3 m; however, this difference is high during the period from 10 pm to 11 pm and the values range between 6 m and 7 m. Looking to this result, it is recommended that a greater number of pressure loggers should be installed in the field for this zone since it is a very complex and large network

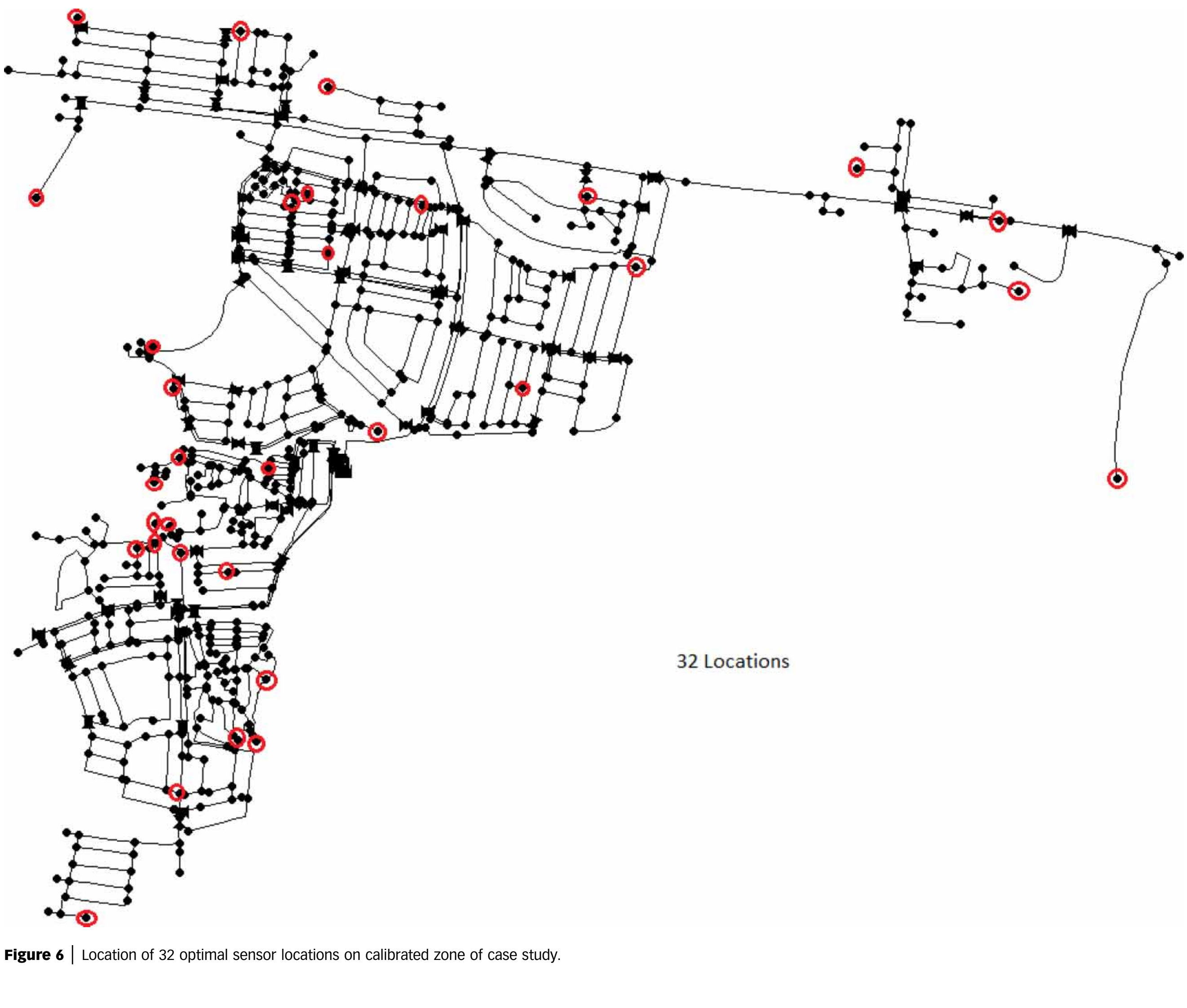

Figure 6 | Location of 32 optimal sensor locations on calibrated zone of case study.

for calibration. However, one can use this model for various purposes for design, analysis, and operation of water distribution systems.

Discussion of results of stage 2

Rathi & Gupta (2017b) compared the result (SGA1) with others. On comparison of the solution SGA1 with others (Table 4) in BWSN network 1, it is observed that S-PLACE GA provided a detection likelihood (Z4) of 0.725. Now, the solutions that have a higher detection likelihood are considered and used for comparison (Table 4). The reason for consideration is that it is assumed that the solutions that have higher detection likelihood (DL) reduce negative consequences through volume consumed (VC), population exposed (PE) and extent of contamination (EC) as discussed in Ostfeld *et al.* (2008). Thus, herein one of the objectives considered is maximization of TCDL, which assures early detection of a greater number of contamination events. Therefore, TCDL assures that all the detected events have a time of detection within the required level of service (LOS). Therefore, solutions are arranged based on descending order of DL (Table 5). On comparing solution SGA1 with others, it is further observed that solution SGA1 is ranked 3rd, 5th and 3rd in order in terms of Z1, Z2 and Z3, respectively. The higher value of DL shows that the risk of contamination is less. The DC is maximum for the proposed solution in comparison to others. Thus, S-PLACE GA is not only giving importance to detecting a higher number of contamination events but also covering higher consumption nodes. Even though the proposed methodology has not resulted in the best solution considering the objectives Z1–Z4, the solution obtained by GA with the formulation of Rathi & Gupta (2015) is comparable with those obtained through other methodologies. Further, S-PLACE GA is an efficient and effective tool for applicability on a large network having 619 nodes and 733 pipes. Further, only 21 sensors are required for 50% coverage in terms of demand and contamination detection and 32 sensors are required for 60% coverage criteria. Results are shown in Table 6. This shows the effectiveness of S-PLACE GA in application.

CONCLUSIONS

Water distribution system are vulnerable to various types of contamination events, natural, accidental and intentional. Delivering safe water through water distribution systems (WDSs) is a challenging task that has led to increased interest in on-line monitoring through sensors to have early warning against contamination events to mitigate their impact. For accurate identification of optimal sensor locations in the water distribution network, there is a need for a calibrated hydraulic model. However, calibration of a hydraulic model is a challenging task for large, real water distribution systems but is useful for accurate determination of optimal locations of water quality sensors in the water distribution network. Therefore, in this paper, a methodology is presented which consists of two stages: (1) calibration of a hydraulic model using GA using TCV modeling and (2) development of a new software tool named S-PLACE GA; its effectiveness by comparing with others and applicability to a large sized network' are discussed.

In the first stage, a case study of a real-life network of one of the hydraulic zones of Ramnagar ESR of Nagpur

Table 5 | Best solutions as compared to Detection Likelihood (Rathi & Gupta 2017b)

Methodology (1)	Sensor location (nodes) (2)	Z1 (min) (3)	Z2 (People) (4)	Z3 (gallons) (5)	Z4 (DL) (%) (6)	Z5 (DC) (%) (7)
Dorini *et al.* (2006)	10,31,45,83,118	1,068 (6)	258 (4)	7,983 (5)	0.801	30.37 (4)
Wu & Walski (2006)	45,68,83,100,118	704 (1)	303 (6)	8,406 (6)	0.787	37.34 (2)
Eliades & Polycarpou (2006)	17,31,45,83,126	912 (5)	221 (3)	7,862 (4)	0.763	34.39 (3)
Krause *et al.*(2006)	17,83,122,31,45	842 (4)	181 (2)	3,992 (2)	0.756	24.02 (6)
Propato & Piller (2006)	17,22,68,83,123	711 (2)	164 (1)	3,148 (1)	**0.725**	27.45 (5)
Rathi & Gupta (2015) (SGA1)	58, 83,101,118,124	724 (**3**)	297 (**5**)	4,510 (**3**)	**0.725**	43.85 (1)

Table 6 | Best sensor locations using GA

Solution	Sensor location at nodes	DC	TCDL	Z
	For 5 best sensor locations			
S_1	904,346,482,151,402	0.3625	0.1509	0.2569
	For 20 best sensor locations			
S_2	904,346,482,151,402,307,223,39,136,298,422,47,283,462,329,197,271,43,98,115	0.5945	0.3996	0.497
	For 30 sensor locations			
S_3	904,346,482,151,402,307,223,39,136,298,422,47,283,462,329,197,271,43,98,115,439,102,329,304,123,98,257,27,103,385	0.6761	0.5036	0.5897
	For 40 sensor locations			
S_4	904,346,482,151,402,307,223,39,136,298,422,47,283,462,329,197,271,43,98,115,439,102,329,304, 123,98,257,27,103,385,323,178,8,299,436,458,416,147,372,468	0.7225	0.5971	0.6596

city, India, is considered. A hydraulic model is developed in EPANET consisting of 69 TCV and openings are approximated by the minor loss coefficient (MLC), which replicates actual valve status in the field. Rathi *et al.* (2020) carried out detailed analysis, field verification and micro calibration of the network and showed that one of TCV (in Ramnagar ESR premises) is critical and there is a need to calibrate the status (settings) of identified TCV that are operated at different times in a day. Therefore, this study model is calibrated using a genetic algorithm (GA), which optimizes identified TCV status (settings) at different times in a day. Results of stage 1 are given in the discussion section. Results showed that for both PMPs, at some of the 15 minutes slots, the difference between simulated and observed pressure is high and for the rest of the timings this difference is less. The identified reason for the high pressure difference between simulated and observed pressures is that this network is very complex, large and there are too many challenges for calibration (Rathi *et al.* 2020). In future, this difference can be minimized by installing a greater number of pressure loggers in the case study at lower and higher elevation areas. However, one can use this model for various purposes for design, analysis, and operation of water distribution systems. Therefore, for identifying the optimal sensor network layout, this calibrated model is used in the second stage.

The calibrated model of the first stage is then used for the second stage. The second stage consists of application of a software tool named S-PLACE Genetic Algorithm (GA) and its efficiency and effectiveness are discussed. It can be used for the dual purpose of routine monitoring of water quality and for early detection of contamination in WDN based on the methodology of Rathi *et al.* (2015). The optimal locations are obtained considering two objective metrics, 'demand coverage' and 'time-constrained detection likelihood'. These two objectives are combined into a single objective by using weights. Key features, input data required for the software and its application in large real-life calibrated models are discussed. Even though Rathi & Gupta (2017b) carried out a comparison of S-PLACE GA on BWSN network 1, its applicability to a real life calibrated model with 619 nodes and 733 pipes is shown here. On a calibrated network, it is observed that only 21 sensors are required for 50% coverage in terms of demand and contamination detection and 32 sensors are required for 60% coverage of the network. A further best 5, 20, 30 and 40 optimal sensor locations are also obtained covering 25, 49, 58 and 65% of the coverage area respectively in the case study. Application on BWSN network 1 (Rathi & Gupta 2017b) showed that high demand coverage is achieved with a little compromise on other objectives using S-PLACE GA. S-PLACE GA is not only giving the importance to detect a higher number of contamination events but also to cover higher consumption nodes. Even though it is not best in evaluating the objectives of Z1–Z4, the results are however comparable. Application in these two examples and comparison in benchmark problems shows the effectiveness of S-PLACE GA for applications in practice.

ACKNOWLEDGEMENTS

I am thankful to my Ph.D supervisor Dr Rajesh Gupta to guide me during my Ph.D work. This has helped me to carry out my further research work.

DATA AVAILABILITY

Software S-PLACE GA can be made available based on the requirements of Editor and Reviewer.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Afshar, A. & Marino, M. A. 2012 Multi-objective coverage-based ACO model for quality monitoring in large water networks. *J. Water Resour. Plann. Manage.* **26**, 2159–2176.

Aral, M. M., Guan, J. & Maslia, L. M. 2010 Optimal design of sensor placement in water distribution networks. *J. Water Resour. Plann. Manage.* **136** (1), 5–18. doi: 10.1061/ (ASCE) WR.1943-5452.0000001.

Berglund, E. Z., Pesantez, J. E., Rasekh, A., Shafiee, M. E., Sela, L. & Haxton, T. 2020 Review of modeling methodologies for managing water distribution security. *J. Water Resour. Plann. Manage.* **146** (8), 03120001–23.

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

Berry, J. W., Boman, E., Riesen, L. A., Hart, W. E., Phillips, C. A. & Watson, J. P. 2008 *User's Manual: TEVA-SPOT Toolkit 2.0*. Sandia National Laboratories, Albuquerque, NM.

Berry, J. W., Boman, E., Riesen, L. A., Hart, W. E., Phillips, C. A. & Watson, J. P. 2010 *User's Manual: TEVA-SPOT Graphical User Interface User's Manual. EPA-600-R-08-147*. U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.

Chang, N. B., Pongsanone, N. P. & Ernest, A. 2011 Comparisons between a rule-based expert system and optimization models for sensor deployment in a small drinking water network. *Expert Syst. Appl.* **38**, 10685–10695.

Comboul, M. & Ghanem, R. 2013 Value of information in the design of resilient water distribution sensor networks. *J. Water Resour. Plann. Manage.* **139** (4), 449–455.

Cozzolino, L., Mucherino, C., Pianese, D. & Pirozzi, F. 2005 Optimal allocation of monitoring stations aiming at an early detection of intentional contamination of water supply systems. *In: Proceedings of CCWI 2005 Conference*, University of Exeter, Exeter, UK.

Davis, M., Janke, R. & Phillips, C. A. 2014 Robustness of designs for drinking-water contamination warning systems under uncertain conditions. *J. Water Resour. Plann. Manage.* **140** (10). doi:04014028.

Deb, K., Agarwal, S., Pratap, A. & Meyarivan, T. 2000 A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *International Conference on Parallel Problem Solving from Nature (PPSN) 2000: Parallel Problem Solving from Nature PPSN VI*, pp. 849–858.

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

Dorini, G., Jonkergouw, P., Kapelan, Z. & Savic, D. 2010 SLOTS: Effective algorithm for sensor placement in water distribution systems. *J. Water Resour. Plann. Manage.* **136** (6), 620–628.

Eliades, D. & Polycarpou, M. 2006 Iterative deepening of Pareto solutions in water sensor networks. In: *Proc., 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA.

Eliades, D. G., Kyriakou, M. & Polycarpou, M. M. 2014 Sensor placement in water distribution systems using the S-PLACE Toolkit. In: *Proc. 12th International Conference on Computing and Control for the Water Industry*. CCWI2013, Procedia Engineering 70, pp. 602–611.

Ghimire, S. R. & Barkdoll, B. D. 2006a Heuristic method for the battle of the water network sensors: demand- based approach. In: *Proceedings, 8th Annual Water Distribution System Analysis Symposium*. ASCE, Reston, VA.

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

Goldberg, D. E. & Deb, K. 1991 A comparative analysis of selection schemes used in genetic algorithms. In Proceedings of Foundations of Genetic Algorithms 1 (FOGA-1), pp. 69–93.

Guan, J., Aral, M. M., Maslia, M. L. & Grayman, W. M. 2006 Identification of contaminant sources in water distribution systems using simulation–optimization method: case study. *J. Water Resour. Plann. Manage.* **132** (4), 252–262.

Gueli, R. 2006 Predator-prey model for discrete sensor placement. In: *Proc. 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA.

Hart, W. E. & Murray, R. 2010 Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *J. Water Resour. Plan. Manage.* **136** (6), 611–619.

Hart, W. E., Berry, J. W., Boman, E. G., Murray, R., Phillips, C. A., Riesen, L. A. & Watson, J-P. 2008 The TEVA-SPOT toolkit for drinking water contaminant warning system design. In: *Proc. World Environmental and Water Resources Congress*. ASCE, Reston, VA.

Huang, J. J., McBean, E. A. & James, W. 2006 Multi-objective optimization for monitoring sensor placement in water distribution systems. In: *Proc. 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. & Faloutsos, C. 2006 Optimizing sensor placements in water distribution systems using submodular function maximization. In: *Proc., 8th Annual Water Distribution Systems Analysis Symp.*, August 27–30, 2006. ASCE, Cincinnati, Ohio, USA, pp. 1–17.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. & Faloutsos, C. 2008 Efficient sensor placement optimization for securing large water distribution networks. *J. Water Resour. Plan. Manage.* **134** (6), 516–526.

Kumar, A., Kansal, M. L. & Arora, G. 1999 Detecting accidental contaminations in municipal water networks. *J. Water Resour. Plan. Manage.* **125** (5), 308–310.

Lee, B. H. & Deininger, R. A. 1992 Optimal locations of monitoring stations in water distribution system. *J. Environ. Eng.* **118** (1), 4–16.

Naserizade, S. S., Nikoo, M. R. & Montaseri, H. 2017 A risk-based multi-objective model for optimal placement of sensors in water distribution system. *J. Hydrol.* https://doi.org/10.1016/j.jhydrol.2017.12.028.

Ohar, Z., Lahav, O. & Ostfeld, A. 2015 Optimal sensor placement for detecting organophosphate intrusion into water distribution systems. *J. Water Res.* **73**, 193–203.

Oliker, N. & Ostfeld, A. 2015 Inclusion of mobile sensors in water distribution system monitoring operations. *J. Water Resour. Plann. Manage.* **142** (1), 04015044.

Ostfeld, A. & Salomons, E. 2004 Optimal layout of early warning detection stations for water distribution systems security. *J. Water Resour. Plann. Manage.* 130 (5), 377–385.

Ostfeld, A. & Salomons, E. 2005 Securing water distribution systems using online contamination monitoring. *J. Water Resour. Plann. Manage.* 131 (5), 402–405.

Ostfeld, A. & Salomons, E. 2006 Sensor network design proposal for the battle of the water sensor networks BWSN. In: *Proc., 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Cincinnati, VA.

Ostfeld, A., Uber, J., Salomons, E., Berry, J., Hart, W., Phillips, C., Watson, J., Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S., Barkdoll, B., Gueli, R., Huang, J., McBean, E., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G., Wu, Z. & Walski, T. 2008 The battle of the water sensor networks BWSN: a design challenge for engineers and algorithms. *J. Water Resour. Plann. Manage.* **134** (6), 555–568. doi:10.1061/(ASCE)0733-9496(2008)134:6(556).

Perelman, L. & Ostfeld, A. 2010 Extreme impact contamination events sampling for water distribution systems security. *J. Water Resour. Plann. Manage.* **136** (1), 80–86.

Perelman, L. & Ostfeld, A. 2012 Water distribution systems simplifications through clustering. *J. Water Resour. Plann. Manage.* **138** (3), 218–229. doi: 10.1061/(ASCE)WR.1943-5452.0000173.

Perelman, L. & Ostfeld, A. 2013 Operation of remote mobile sensors for security of drinking water distribution systems. *Water Res.* **47** (13), 4217–4226.

Perelman, L., Maslia, M. L., Ostfeld, A. & Sautner, J. B. 2008 Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies. *J. Am. Water Works Assn.* **100**, 122–133.

Preis, A. & Ostfeld, A. 2006 Multiobjective sensor design for water distribution systems security. *Proceedings, 8th Annual Water Distribution Systems Analysis Symposium*, American Society of Civil Engineers (ASCE), Reston, VA.

Propato, M. & Piller, O. 2006 Battle of the water sensor networks. In: *Proc., 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA, pp. 1–10.

Rathi, S. & Gupta, R. 2014 Location of sampling stations for water quality monitoring in water distribution networks. *J. Environ. Sci. Eng.* **56** (2), 165–174.

Rathi, S. & Gupta, R. 2015 A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks. *J. Korean Soc. Civ. Eng. (KSCE)* doi:10.1007/s12205-015-0024-x.

Rathi, S. & Gupta, R. 2017a Optimal sensor locations for contamination detection in pressure-deficient water distribution networks using Genetic Algorithm. *Urban Water J.* **14** (2), 160–172. doi:10.1080/1573062X.2015. 1080736.

Rathi, S. & Gupta, R. 2017b Optimal sensor locations with dual objectives of regular monitoring and contamination detection in water distribution network of BWSN. *J. Indian Water Work Assoc.* **3**, 212–218.

Rathi, S., Gupta, R. & Ormsbee, L. 2015 A review of sensor placement objective metrics for contamination detection in water distribution networks. *Water Sci. Technol. Water Supply* **15** (5), 898–917. doi:10.2166/ws.2015.077.

Rathi, S., Gupta, R., Kamble, S. & Sargaonkar, A. 2016 Risk based analysis for contamination event selection and optimal sensor placement for intermittent water distribution network security. *Water Resour. Manage.* **30** (8), 2671–2685. doi:10.1007/s11269-016-1309-7.

Rathi, S., Gupta, R., Labhsetar, P. & Nagarnayak, P. 2020 Challenges in calibration of water distribution network: a case study of Ramnagar elevated service reservoir command area in Nagpur city, India. *Water Sci. Technol. Water Supply* 20 (4), 1294–1312. https://doi.org/10.2166/ws.2020.047.

Salomons, E. 2006 *BWSN-Software Utilities*. Available from: http://www.watersimulation.com/wsp/bwsn.

Shastri, Y. & Diwekar, U. 2006 Sensor placement in water networks: a stochastic programming approach. *J. Water Resour. Plann. Manage.* **132** (3), 192–203.

Shen, H. & McBean, E. 2011 Pareto optimality for sensor placements in a water distribution system. *J. Water Resour. Plann. Manage.* **137** (3), 243–248. doi:10.1061/(ASCE)WR.1943-5452.0000111.

Trachtman, G. B. 2006 A strawman common sense approach for water quality sensor site selection. *Proceedings 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.

Vairavamoorthy, K., Yan, J., Galgale, H. M. & Gorantiwar, S. D. 2007 IRA-WDS: A GIS-based risk analysis tool for water distribution systems. *Environmental Modelling & Software* **22**, 951–965.

Watson, J. P., Greenberg, H. J. & Hart, W. E. 2004 A multiple objective analysis of sensor placement optimization in water networks. In: *Proc., World Water and Environmental Resources Congress*. ASCE/EWRI, Reston, VA, pp. 456–465. doi:10.1061/40737(2004)456.

Watson, J. P., Murray, R. & Hart, W. E. 2009 Formulation and optimization of robust sensor placement problems for drinking water contamination warning systems. *J. Infrastruct. Syst., ASCE* **15** (4), 330–339. doi:10.1061/(ASCE)1076-0342(2009),15:4(330).

Weickgenannt, M., Kapelan, Z., Blokker, M. & Savic, D. A. 2010 Risk based sensor placement for contaminant detection in water distribution systems. *J. Water Resour. Plann. Manage.* **136** (6), 629–636. doi:10.1061/(ASCE)WR.1943-5452.0000073.

Wu, Z. Y. & Walski, T. 2006 Multi objective optimization of sensor placement in water distribution systems. In: *Proc. 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA, pp. 1–11, doi:10.1061/40941(247)105.

Wu, Z. Y., Song, Y. & Syed, J. L. 2012 Optimization model for identifying unknown valve statuses and settings. In: *Proceedings of the 14th Water Distribution Systems Analysis*, 24–27, Adelaide, Australia.

Xu, J., Fischbeck, P., Small, M., VanBriesen, J. & Casman, E. 2008 Identifying sets of key nodes for placing sensors in dynamic water distribution networks. *J. Water Resour. Plann. Manage.* **134** (4), 378–385.

First received 31 March 2020; accepted in revised form 9 November 2020. Available online 24 November 2020