Sensor Placement Optimization in Sewer Networks: Machine Learning–Based Source Identification Approach

Aly K. Salem, S.M.ASCE[1]; and Ahmed A. Abokifa[2]

Abstract: Wastewater surveillance has recently emerged as a valuable tool for environmental and public health monitoring. By analyzing the constituents and biomarkers present in wastewater, stakeholders can gather critical information regarding contamination events and disease outbreaks. However, little attention has been given to the crucial question of where to collect water quality samples or place water quality sensors to maximize the usefulness of wastewater surveillance data. To address this gap, this study introduces a novel framework for sensor placement (SP) optimization in sewer networks. The objective of the optimization is to maximize both the observability and reliability of source identification (SI) under different scenarios. To achieve this objective, a machine learning–based SI model was integrated within the SP optimization framework. The SI model features a multilayer perceptron neural network model that was trained to forecast concentrations at various sensor locations, which were then propagated into a genetic algorithm that finds the optimal sensor network design that maximizes SI performance. The capabilities of the SP framework were demonstrated in a case study featuring a real-life, midsize sewer network. The SP framework was applied to multiple scenarios, including optimal design of a sensor network comprising one or more sensors, as well as optimal extension of existing sensor networks. The results showed that a clear trade-off exists between the sensor network's observability and reliability, highlighting the importance of considering both metrics for SP optimization. Overall, this study offers a practical approach for SP optimization to improve environmental and public health monitoring in a variety of contexts. **DOI: 10.1061/JWRMD5.WRENG-6430.**

Introduction

Monitoring the constituents of sewer systems is crucial because they affect the operations of wastewater treatment plants (WWTPs) and the quality of the recipient water bodies (Diaz-Fierros et al. 2002). In addition, residential wastewater has recently been found to reveal valuable epidemiological information, which has significant implications for public health (Calle et al. 2021; Lin et al. 2021). In general, sewer systems are routinely monitored to identify contaminants or biomarkers (Nourinejad et al. 2021; Sambito et al. 2020). The monitoring process has been supported by technological advancements in surveillance systems (e.g., real-time sensors), enabling so-called smart sewers and cities (Bourgeois et al. 2001; Edmondson et al. 2018; Tatiparthi et al. 2021). However, ubiquitous sewer system monitoring is still limited by the significant costs associated with the installation and operation of such sewer monitoring equipment (i.e., sensors) (Banik et al. 2015). Accordingly, finding the optimal location to place these sensors is a vital challenge that needs to be tackled (Banik et al. 2017a). This problem is formally known as the sensor placement (SP) optimization problem, which aims to find the optimal design for sensor networks to capture as much information as possible from the system.

SP optimization has been widely studied in the context of drinking water distribution networks (WDNs), which mainly aimed to minimize the impacts of potential contamination events (i.e., early warning systems). Several methodologies have been proposed, with different types and combinations of objectives, as has been summarized in multiple literature reviews (Adedoja et al. 2019; Hart and Murray 2010; Rathi and Gupta 2014). In these methodologies, the objectives varied from minimizing the time to detection, minimizing the impact and/or the extent of contamination, or maximizing the performance of the contamination source identification. To solve the SP problem in WDNs, numerous optimization approaches have been proposed, including single-objective (Aral et al. 2010; Preis and Ostfeld 2008a; Rathi and Gupta 2016), as well as multiobjective (Afshar and Mariño 2012; Brentan et al. 2021; Mu et al. 2022; Preis and Ostfeld 2008b) approaches.

In contrast, few studies were interested in SP optimization in sewer networks. Banik et al. (2015) introduced a methodology that utilizes the Non-dominated Sorting Genetic Algorithm (NSGA-II) to find the optimal sensor design that maximizes the information gain through entropy and minimizes the redundancy through total correlation. Banik et al. (2017b) then compared the solution of the previous study to a rank-based solution derived by the greedy algorithm and concluded that the greedy approach is computationally efficient but suboptimal. Later, Banik et al. (2017a) used the greedy algorithm to solve different formulations of the objective function (e.g., single and multiobjective formulations). Sambito et al. (2020) utilized a probabilistic approach (i.e., Bayesian decision network) to optimize sensor placement to isolate illicit intrusions. A similar approach was adopted by Sambito and Freni (2021) but with the consideration of organic reactive contaminants. Recently, Guadagno et al. (2023) applied a particle backtracking algorithm to optimize sensor placement based on impact coefficient evaluations.

[1]Ph.D. Student, Dept. of Civil, Materials, and Environmental Engineering, Univ. of Illinois Chicago, Chicago, IL 60607; Assistant Lecturer, Faculty of Engineering, Cairo Univ., Giza 12613, Egypt. ORCID: https://orcid.org/0000-0002-2295-9971

[2]Assistant Professor, Dept. of Civil, Materials, and Environmental Engineering, Univ. of Illinois Chicago, Chicago, IL 60607 (corresponding author). ORCID: https://orcid.org/0000-0002-2474-6670. Email: abokifa@uic.edu

Note. This manuscript was submitted on October 5, 2023; approved on June 3, 2024; published online on August 26, 2024. Discussion period open until January 26, 2025; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

One of the key uses of water quality data is source identification (SI), which is the process of identifying the characteristics of the sources that discharge certain species or constituents into the sewer network, such as the locations and concentrations of these sources. SI has significant implications in contamination detection and response, as well as wastewater-based epidemiology (WBE). Despite the importance of SI and its high reliance on water quality sensors, none of the aforementioned studies attempted to implement it as one of the critical criteria for SP optimization. Instead, SP optimization has typically been conducted based on various mathematical indicators (e.g., entropy and total correlation). Therefore, sensor designs produced by such indicators are not guaranteed to produce optimal SI performance.

Motivated by the recent surge in WBE monitoring of sewer systems in the wake of the COVID-19 pandemic, several WBE studies explored the optimization of sample collection locations in sewer networks with the aim of pinpointing hot spots of COVID-19 outbreaks (Calle et al. 2021; Larson et al. 2020; Nourinejad et al. 2021; Wang et al. 2020). Although these studies attempted to include SI indicators in the optimization of monitoring locations, they mainly aimed to maximize the observance of the sewer network depending only on the geometric topology of the sewer network without conducting hydraulic and/or species transport simulations.

In this study, we developed an SP optimization framework that couples an SI model developed by Salem and Abokifa (2023) with genetic algorithm–based optimization. The SI model incorporates hydraulic and water quality simulations of the sewer network under various scenarios. To optimize the design of sensor networks, two SI-based indicators were developed for SP optimization, namely, the observability and reliability of the sensor network design. The observability of the sensor design reflects its ability to detect and identify the source characteristics of different injection and intrusion events, while reliability represents the accuracy of the identified source characteristics. Through this approach, we aim to (1) find the optimal sensor placement design that achieves the best SI performance under different design conditions, (2) analyze the relationship between the sensor location and its observability and reliability, and (3) investigate the trade-off between the sensor's observability and reliability.

Methodology

In this study, SP optimization is formulated as a nested optimization problem, where the optimal sensor design is obtained through performing two optimization processes, (1) a source identification (SI) optimization process within (2) a sensor placement (SP) optimization process. The SI aims to identify the injection characteristics corresponding to concentration data observed by one or more sensors (i.e., observation junctions). This process is conducted herein using an SI model developed by Salem and Abokifa (2023). The SP optimization process utilizes the results of the SI model to optimize the sensor design by connecting several modules to the SI model. The integration between the SI model and the various modules of the SP model is depicted in Fig. 1. In this section, a detailed explanation of the SP model is given, along with a brief explanation of the previously developed SI model.

SP Model

The SP model consists of several interconnected modules, each of which performs specific tasks. These modules are linked to the SI model to form the SP framework. In the SP framework, the sensor design alternatives are produced by the SP model, whereas the source identification results are produced by the SI model. In this subsection, the focus is on the SP modules and their interaction with the SI model, whereas the following subsection focuses on the SI model.

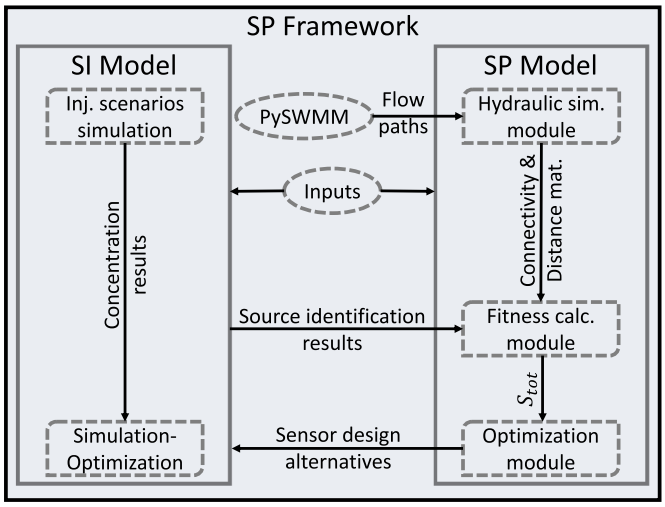

Fig. 1. Interaction between the SP model and the SI model in the proposed SP framework.

Model Formulation

The main objective of the SP model is to maximize the value of information (VOI) provided by a sensor design. The VOI represents the sensor design's ability to accurately identify the injection locations and concentrations of any species that is discharged into the network. The objective function to be maximized (i.e., the VOI of the sensor design) is determined through the calculation of two metrics, namely, the observability and the reliability metrics. The observability metric measures the sensor design's ability to observe the occurrence of an injection event and to estimate its source characteristics through a source identification process. The observability metric used in this paper is specifically tailored for sensor placement optimization and is different from the observability metric commonly used in control theory (Bartos and Kerkez 2021; Taha et al. 2021). The reliability metric evaluates the sensor design's reliability in identifying the correct source characteristics for the injection events it observes. An injection event is observed when all injection sources are located upstream of one or more sensors. The objective function of the SP framework was formed to maximize the weighted sum of these two metrics, which is represented mathematically as follows:

$$\text{Maximize: } S_{tot} = I_{obs} \times S_{obs} + I_{rel} \times S_{rel} \tag{1}$$

where S_{tot} = total score of a sensor design; and S_{obs}, S_{rel}, I_{obs}, and I_{rel} = observability and reliability scores and their weights, respectively. These weights are specified by the user to define the relative importance of the two scores according to the sensor design criteria. In Eq. (1), the observability metric is represented as a score where higher scores are associated with sensor designs that lead to accurate identification of injection characteristics, and vice versa. The observability score of a sensor design is calculated by averaging the inverses of the SI errors of all injection sources (N_s) across all injection events (N_e) as shown in Eq. (2)

$$S_{obs} = \frac{100}{N_e \times N_s} \sum_{e=1}^{N_e} \left\{ \sum_{s=1}^{N_s} \frac{1}{\text{SI}_s^{err}} \right\}_e \tag{2}$$

J. Water Resour. Plann. Manage., 2024, 150(11): 04024048

$$\mathrm{SI}^{err} = [1 + C^{err}] \times \left[1 + \frac{D}{D^{avg}}\right] \tag{3}$$

According to Eq. (3), the SI^{err} of an injection source (s) in an injection event (e) is controlled by (1) the absolute relative concentration error (C^{err}), and (2) the distance between the true and the identified injection locations (D) divided by the average distance between the true injection location and the sensors' locations (D^{avg}).

The value of C^{err} is calculated using Eq. (4), where C and $\hat{C}$ are the true and identified injection concentrations, respectively; whereas D/D^{avg} is calculated using Eq. (5), where Z, $\hat{Z}$, and Z_m are the true injection location, identified injection location, and the mth sensor location, l_i is the pipe length, and P_{AB} is a pipe on the shortest flow path from A to B

$$C^{err} = \frac{C - \hat{C}}{C} \tag{4}$$

$$\frac{D}{D^{avg}} = \frac{\sum_{i \in P_{Z\hat{Z}}}^{I} l_i}{\frac{1}{M}\sum_{m=1}^{M}\sum_{i \in P_{ZZ_m}}^{I} l_i}, \qquad P_{\mathrm{AB}} = \{i | F_i = \mathrm{A} \leftrightarrow \mathrm{B}\} \tag{5}$$

For cases where the injection source is not observable by any of the sensors, the SI^{err} for this injection source is set to 2. A higher observability score means a higher number of observed injection sources or events and/or better estimated injection concentrations. The reliability metric is also represented as a score in Eq. (1) and is calculated as the ratio between the correctly identified injection events to the total number of injection events detected by the sensor design

$$S_{rel} = \frac{\sum_{e=1}^{N_e} S_c}{\sum_{e=1}^{N_e} S_s} \tag{6}$$

where $S_c = 1$ when the injection sources characteristics of an event (i.e., location and concentration) are correctly identified, and $S_c = 0$ otherwise; $S_s = 1$ when all sources are observed (i.e., injection sources are located upstream of one or more sensors), and $S_s = 0$ otherwise. A higher reliability score means a more unique signal a sensor design is receiving, and hence a higher accuracy in identifying the characteristics of the injection events. The score calculations are conducted in the fitness calculation module, which is explained in detail subsequently.

SP Framework Inputs

The parameters governing the performance of the SP framework have to be defined by the framework's user. These parameters are divided based on their role into two groups: (1) design parameters, and (2) optimization parameters. The design parameters correspond to the data informing the decisions taken by the SP framework, such as the total number of sensors, the number of injection events, and the scores' weights. The optimization parameters control the optimization modules within the SP framework (e.g., SP and SI optimizations), such as the number of individuals and generations. Other parameters are revealed subsequently in conjunction with their implementation.

Hydraulic Simulation Module

The calculation of the through-pipe distance (P_{AB}) and the hydraulic connectivity status (S_s) requires the determination of the flow paths between every two junctions in the network. This is done in this module by conducting a hydraulic simulation of the network so that the simulation results are then used to generate two matrices: (1) the sensors' connectivity matrix, and (2) the junctions' distance matrix. The sensors' connectivity matrix is a binary matrix showing the hydraulic connectivity status between the network junctions and the potential sensors (PSs) representing the junctions at which sensors can be placed. The junctions' distance matrix is a square matrix demonstrating the longitudinal through-pipe distance between every two junctions in the network. Three different cases can exist during the calculation of the distance matrix: (1) junctions are on the same flow path, in which the actual flow distance is assigned; (2) junctions are not on the same flow path, in which the longest through-pipe distance in the network is assigned; and (3) junctions are duplicated (both junctions are the same), in which zero is assigned.

PySWMM

In this study, PySWMM, a Python interface of the Storm Water Management Model (SWMM), is used to perform the hydraulic simulations (McDonnell et al. 2020). To identify the flow paths, a conservative tracer was introduced at each junction in the network. Then the system elements containing this tracer (i.e., the pipes and junctions the tracer passes through) are reported by PySWMM. Then these results are processed to calculate the sensors' connectivity and junctions' distance matrices. These matrices are used in the score calculation module as explained subsequently.

Fitness Calculation Module

In this module, the S_{total} (i.e., fitness) of a sensor design is calculated based on the results of the SI model and the matrices produced in the hydraulic simulation module. The fitness calculation involves the generation of multiple injection events, which the SI model will attempt to identify for several sensor design alternatives. Because the futuristic injection event is unknown, the injection events matrix is generated randomly by applying the Latin hypercube sampling method, considering a user-defined number of events. The latter method was used due to its ability to produce more diverse samples to represent the full domain of potential injection events by a small subset of events (Huntington and Lyrintzis 1998). For each event, a user-defined number of injection locations are selected from the potential injection junctions (IJs), along with an injection concentration between user-defined lower and upper bounds. An injection vector is then formed by merging N_s injection locations (L_{N_s}) with N_s injection concentrations (C_{N_s}) as shown in Eq. (7), and then all vectors are stacked to produce the injection events matrix

$$[L_1, C_1, L_2, C_2, \ldots, L_{N_s}, C_{N_s}] \tag{7}$$

To calculate the SI error, the true injection locations are compared to the injection locations identified by the SI model. Relative to the true injection locations, the identified injection locations can be either perfectly matching or partially or not matching. According to Eq. (6), the first case is the only applicable case in the S_{rel} calculations. However, the two cases apply to the S_{obs} calculations.

For the case of two injection sources, if the true and identified injection locations match, one set of two SI errors can be directly calculated by Eq. (3), where D will equal zero for the two errors. However, in case of partial or no matching between the true and identified injection locations, multiple sets of SI errors will be possible. In other words, different combinations between the true and identified injection sources should be considered. For example, if Junctions 1 and 2 are the true injection locations and Junctions 3 and 4 are the identified injection locations, four SI errors should be calculated. The first set is by considering Junction 1 to be identified

as Junction 3, and Junction 2 to be identified as Junction 4; the second is by considering Junction 1 to be identified as Junction 4, and Junction 2 to be identified as Junction 3. In this case, both sets are calculated, and the set with the smallest sum of SI errors is used. A similar approach can be used for more than two injection sources. Once the S_{obs} and the S_{rel} are calculated, the weights are applied and the S_{tot} of the sensor design is calculated by Eq. (1) and sent back to the genetic algorithm (GA).

Optimization Module

The optimization module is the core of the SP model because it utilizes the outputs of the SP modules and the SI model to maximize the objective function given in Eq. (1) by applying GA, which is a heuristic optimization technique that follows the mechanisms of natural selection to come up with a powerful population starting with weak parents (Elbeltagi et al. 2005). GA was used as the optimization technique due to its ability to handle complicated discontinuous functions (Haupt and Haupt 2003) because it does not require the derivatives of the objective function to be known. In addition, GA has been extensively used by researchers interested in sensor placement optimization (Banik et al. 2015) and water network optimization in general (Pan and Kao 2009; Preis and Ostfeld 2008a; Xuesong et al. 2017).

In the optimization module, GA randomly generates sensor design alternatives (i.e., individuals), which are sent along with the injection events matrix to the SI model. Then the optimization module receives the S_{tot} of each sensor design alternative from the fitness calculation module. The previous process is repeated several times (i.e., generations). In each generation, GA produces new design alternatives by applying three different operators: (1) selection, (2) crossover, and (3) mutation. The best sensor designs from one generation form a portion of the next generation, whereas the remaining portion is formed by applying crossover and mutation operators. The crossover operator pairs the best sensor designs and the mutation operator alters some of the genes of these sensor designs to increase the diversity and to avoid being stuck in local minimums. The PyGAD package introduced by Gad (2024) is used to implement the GA. The genetic algorithm optimization parameters used in this study are listed in Table S1.

SI Model

As mentioned previously, the SI model is used to assess the ability of a sensor design to be employed in identifying the characteristics of futuristic injection events. In other words, the SI model evaluates the accuracy of source identification based on the concentration data observed by the sensors. A previous study proposed a machine learning–based SI model that proved to be computationally efficient and showed to be capable of identifying multiple simultaneous sources of reactive constituents (Salem and Abokifa 2023). Hence, this SI model was implemented in this study.

The SI model proposed by Salem and Abokifa (2023) follows a simulation-optimization approach to locate the injection source(s) and their injection pattern based on the concentration of a certain constituent observed by one or more sensors. The SI model combines a surrogate model constructed by a multilayer perceptron neural network (MLP-NN) with GA. The objective of the SI model is to minimize the difference between the time-series concentration of the simulated and observed injection event at the location of the sensors

$$\text{Minimize: } n\text{RMSE} = \sum_{j=1}^{J} \frac{[\sum_{t=1}^{T} (c_{jt}^{obs} - c_{jt}^{sim})^2 / T]^{1/2}}{(\sum_{t=1}^{T} c_{jt}^{obs})/T} \quad (8)$$

where C^{obs} and C^{sim} = concentration observed and simulated at the sensor location; j = index of the sensor; J = total number of sensors; t = index of the time step; and T = total number of time steps.

For a given sensor design, the SI model starts by simulating each injection event and extracting the concentration at the location of all sensors forming the sensor design. Then, for each injection event, the SI model utilizes the previously extracted concentration data to inversely find the characteristics of the injection sources (e.g., location and concentration). The structure of the MLP-NN model is listed in Table S2, and detailed information about the SI model can be found in Salem and Abokifa (2023).

Case Study

To test the performance of the developed SP framework, we applied it to a real medium-sized combined sewer network located in the United States (the exact location is confidential). The case study network is a branched network with several loops extending from east to west and covers an area of 308,350 m^2 [Fig. 2(a)]. The network consists of 84 conduits connecting 80 junctions, collecting wastewater from 59 residential sewer sheds. The dry weather flow (DWF) pattern was considered to match the typical DWF pattern developed by Butler et al. (2018) [Fig. 2(b)], resulting in a peak discharge of 42.8 m^3/h at the network outfall.

Injection Events Generation

As mentioned previously, the Latin hypercube sampling method is used to generate the injection events matrix. In this case study, 150 random injection events were considered, wherein each event, a reactive constituent (e.g., a contaminant) was assumed to be injected at two simultaneous junctions out of the 53 drainage junctions (i.e., receiving DWF). A uniform distribution was used to allow each junction to have an equal selection frequency as the injection location. Similarly, a uniformly distributed injection concentration was considered between 1 and 100 mg/L.

Sensor Placement Scenarios

Four sensor placement scenarios were tested to investigate the performance of the SP framework under various design conditions. In the first scenario (S1), the framework was applied to find the optimal location to place one sensor considering two different score weights. The aim was to gain insight into the significance and contribution of the observability and reliability scores, and the effect of the weights in determining the optimal design. In Scenarios S2.1 and S2.2, the objective was to find the optimal location to place two sensors, considering two constituents with different decay rates, assuming that no sensors already exist in the network. In Scenario S3, it was considered that the two sensors in Scenario S2.1 were already placed, and the objective was to find the optimal location to place an additional sensor to enhance the sensor network performance. In the four scenarios, a high decay rate of 25 day^{-1} was used, except for S2.2, where a low decay rate of 1.4 day^{-1} was used to represent the decay of SARS-CoV-2 in sewer networks (Bivins et al. 2020). In addition, only the junctions detecting (i.e., hydraulically connected to) at least four drainage junctions were considered as PSs (i.e., valid locations to place sensors). PSs are depicted by diamonds in Fig. 3.

J. Water Resour. Plann. Manage., 2024, 150(11): 04024048

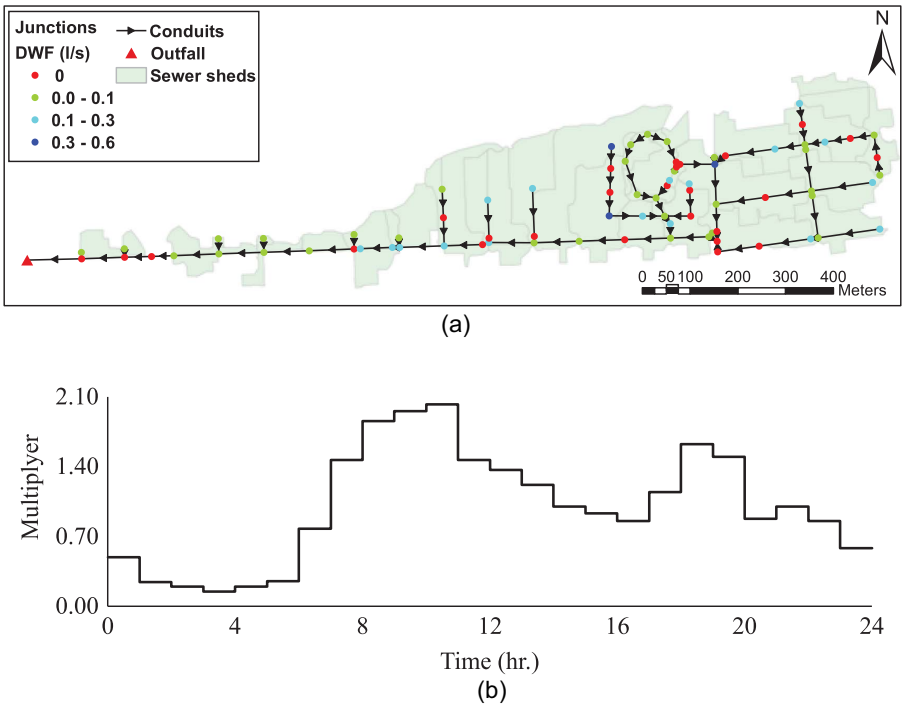

Fig. 2. Case study network: (a) layout; and (b) DWF pattern.

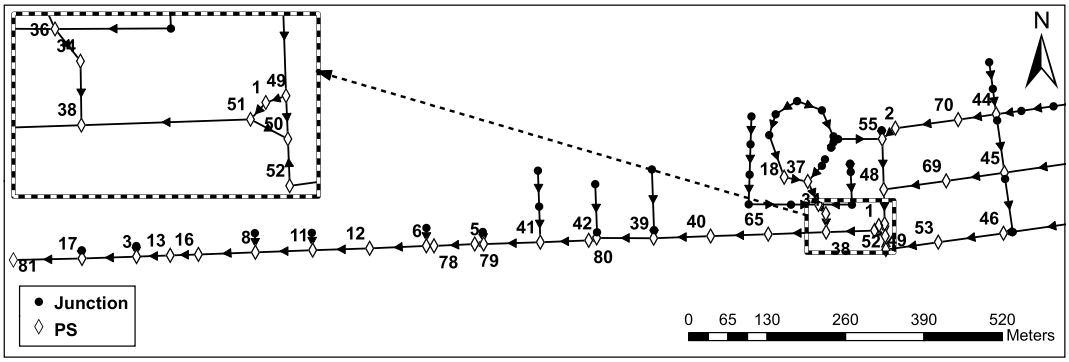

Fig. 3. Location of the potential sensors (PSs).

Results and Discussion

Placement of a Single Sensor (Scenario S1)

In Scenario S1, the observability and reliability scores (S_{obs} and S_{rel}) were calculated for all PSs. Because this analysis aims to find the optimal location of only one sensor, an enumeration approach was adopted instead of an optimization approach so the two scores can be compared for all potential sensors.

Equal Score Weights

At first, equal score weights were considered for the two scores (I_{obs}:I_{rel} = 1:1), and the total score of each sensor (S_{tot}) was calculated according to Eq. (1). Both scores are shown in Fig. 4(a), where the x-axis represents the junction's ID, and the y-axis shows the junction's S_{obs} and S_{rel} scores. The junctions are sorted by their total score so that the highest S_{tot} is to the left, and the junction rank in terms of each individual score is displayed.

Fig. 4(a) shows that sensors placed at the downstream junctions tend to have high observability scores with relatively low reliability scores, and vice versa. For instance, the three most downstream junctions (17, 81, and 3) are the three top-ranked junctions for observability score, while their reliability rankings are in the lower 40th percentile. On the other hand, the three top-ranked junctions for reliability score (Junctions 45, 69, and 52) are all located at the upstream portion of the network and showed very low observability scores (all in the lower 20th percentile). Nevertheless, some of the upstream junctions still had very low reliability scores [Fig. 4(a)]. This is because they could not detect both injection sources in most (or all) of the 150 injection events because they can only detect a very small section of the network. These results highlight the significant role played by the location of the sensor in the trade-off between observability and reliability.

Moreover, when both scores were assigned equal weights, the reliability score appeared to dominate the observability score in the

J. Water Resour. Plann. Manage., 2024, 150(11): 04024048

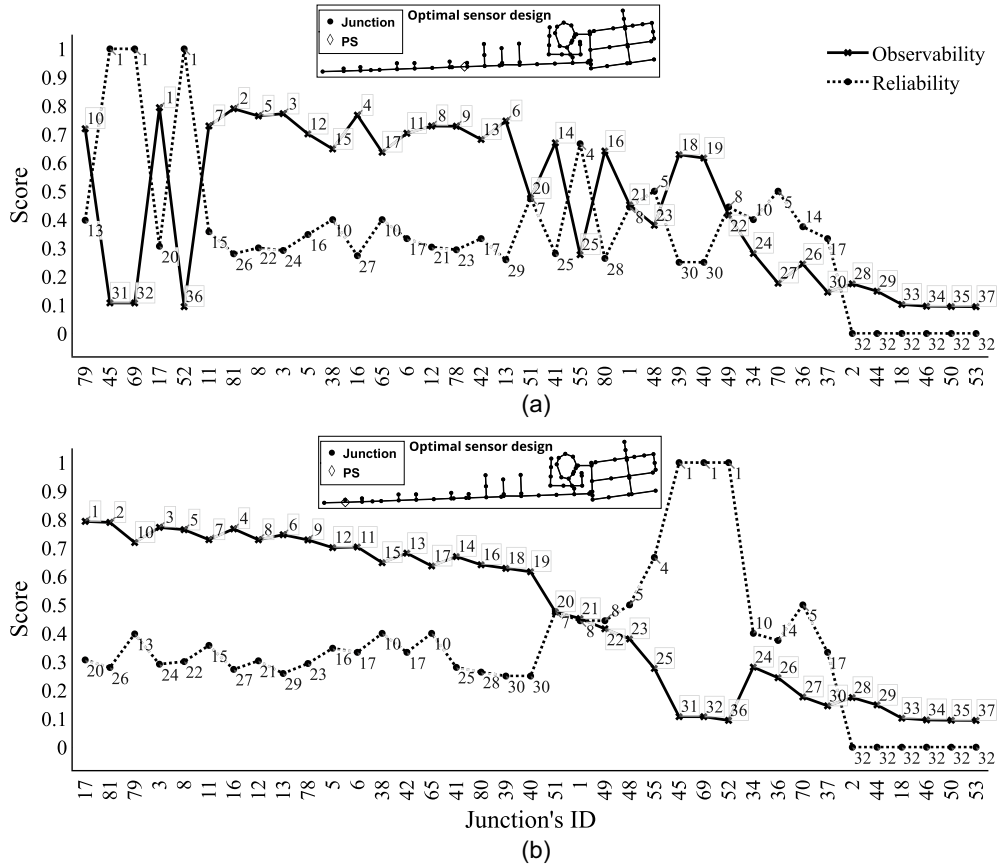

Fig. 4. Scenario S1 results considering (a) equal weights; and (b) unequal weights. The junctions are sorted based on the total score, highest to the left. The inset shows the optimal sensor design projected on the network layout.

determination of the optimal sensor location. Fig. 4(a) shows that the top five junctions in terms of the total score had an average reliability score of 0.74 and an observability score of 0.36. Moreover, the top three junctions in terms of the reliability score (Junctions 45, 69, and 52) were ranked second, third, and fifth in terms of the total score, while the top three junctions in terms of the observability score (Junctions 17, 81, and 3) were ranked fourth, seventh, and ninth in terms of the total score.

As can be seen from Fig. 4(a), Junction 79 was the optimal location to place the sensor, which, according to the inset, is a central junction. Placing the sensor at Junction 79 allowed for achieving a good balance between observability and reliability. Furthermore, this balance can be noted by the junction's rank in terms of individual scores because it was ranked 10th and 13th in terms of the observability and reliability scores, respectively [Fig. 4(a)]. However, this balance was accomplished at the expense of not observing a considerably large section of the network (the section downstream of Junction 79). For that reason, observing the whole network is typically prioritized if only one sensor is to be placed (e.g., at the WWTP). Thus, a more practical design can be retrieved by assigning a higher weight for the observability score, which is explored in the following scenario.

Unequal Score Weights

To test the effect of score weights and to employ practicality in the sensor network design, the observability score was prioritized over the reliability score by assigning I_{obs}:I_{rel} = 2:1. As shown in Fig. 4(b), four of the five highest junctions in terms of the observability score appeared in the highest five junctions in terms of the total score, which means the observability score dominates the reliability score when assigned a greater weight. Furthermore, Junction 17 appeared as the optimal sensor location instead of Junction 79, which dropped to the third rank in terms of the total score. These changes highlight the relative significance of the score weights in determining the optimal sensor design. Junction 17 achieved a higher total score than Junction 81 (i.e., the outfall) because it lies closer to the other junctions as shown in the inset, and is thus receiving a clearer signal, allowing it to achieve a higher reliability score.

Placement of Multiple Sensors (Scenario S2)

High Decay Rate (Scenario S2.1)

Increasing the number of sensors to two in Scenario S2 gave the SP framework higher flexibility in finding an optimal sensor design that balances observability and reliability scores. This can be seen in Fig. 5, which demonstrates the best 10 sensor designs produced by the GA in Scenario S2.1.

In Fig. 5, the x-axis represents the junctions' ID of the sensors' locations sorted by the total score, and the y-axis shows the observability and reliability scores of the sensor designs. There appears to be no significant variability in the observability and reliability scores between the best 10 sensor designs in Scenario S2.1. Moreover, unlike Scenario S1 where a clear trade-off existed between the reliability and observability scores of the sensor designs, the 10 designs in Scenario S2.1 had both high observability and reliability scores. This can be attributed to the fact that the sensor designs in Scenario S2.1 typically featured one sensor placed at the downstream portion of the network and another sensor at a central or

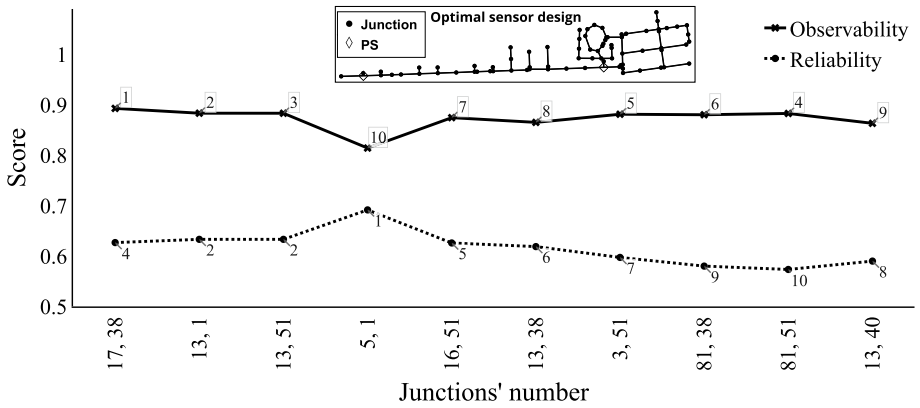

Fig. 5. Best 10 sensor designs in Scenario S2.1. The designs are sorted based on the total score, highest to the left. The inset shows the optimal sensor design projected on the network layout.

upstream portion of the network. For instance, in the optimal sensor design (Junctions 17 and 38), the SP framework achieved a high observability score by placing a sensor in the downstream portion (i.e., Junction 17) while also achieving a high reliability score by placing the other sensor at a central junction (i.e., Junction 38). The same observation can be made for most of the top designs, except for the fourth best design, which featured one junction in a central location (Junction 5) in addition to one junction in the upstream portion (Junction 1), resulting in higher reliability with lower observability scores compared to the other designs. More importantly, a synergistic effect was noticed in the top 10 sensor designs, where the observability and reliability scores of the two-sensor design were higher than the individual scores of each of the two sensors calculated in Scenario S1.

By comparing Scenario S1 to Scenario S2.1, it can be concluded that Scenario S1's results cannot be used to simply derive the optimal design of Scenario S2.1. For instance, the junctions forming the best design in Scenario S2.1 (17 and 38) were ranked 4th and 11th in Scenario S1 in terms of the total score [Fig. 4(a)]. Moreover, Junctions 13 and 51 were the most frequently featured in the top 10 designs in Scenario S2.1 (Fig. 5) despite individually ranking sixth and seventh in terms of the observability and reliability scores in Scenario S1 [Fig. 4(a)].

Low Decay Rate (Scenario S2.2)

The 10 best sensor designs produced by the SP optimization model in Scenario S2.2 (Fig. 6) featured a trade-off between the observability and reliability scores with relatively lower observability scores than Scenario S2.1. Fig. 6 also shows that, unlike Scenario S2.1, the observability score no longer dominates the best sensor designs; in fact, the reliability score dominates the best two sensor designs because they were ranked 10th and 5th in terms of reliability compared to 12th and 38th in terms of observability. More importantly, the sensors of the optimal sensor design (1 and 79) were placed at the middle of the network compared to one sensor in the upstream and downstream portion of the network in Scenario S2.1.

The differences in sensor placement results are primarily attributed to the reduction in the decay rate from 25 day^{-1} to 1.4 day^{-1}. As previously stated by Salem and Abokifa (2023), reducing the decay rate elevates the complexity of the SI problem by diminishing signal uniqueness. In instances of low-decay-rate constituents, signals from injections at specific junctions become less distinct from those at upstream and downstream junctions. Accordingly, the SI model struggled to accurately identify the injection locations and concentrations in Scenario S2.2, resulting in lower observability scores. Furthermore, accurately identifying injection characteristics becomes even more challenging for far downstream sensors, making them less advantageous from the SP model's perspective.

Existing Sensor Network Expansion (Scenario S3)

In Scenario S3, two sensors were considered to exist at Junctions 17 and 38, which is the optimal sensor design achieved in Scenario S2.1, and the objective was to find the optimal location to place a third sensor. The best 10 sensor designs identified in Scenario S3 are shown in Fig. 7. Similar to Scenario S2.1, Fig. 7 shows that the trade-off between the observability and reliability scores is

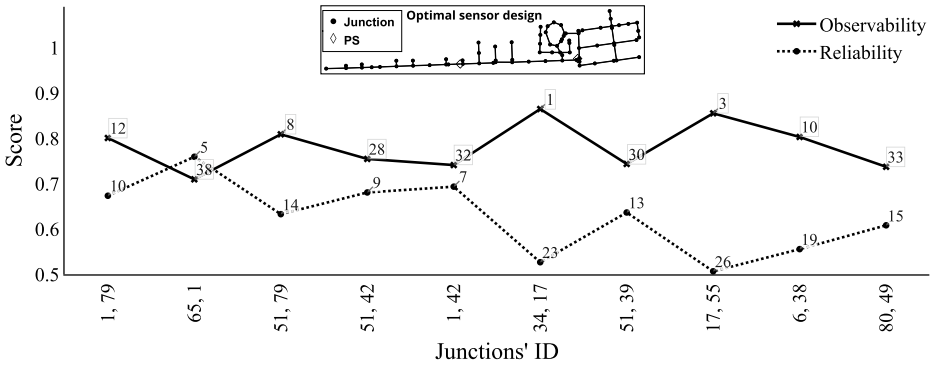

Fig. 6. Best 10 sensor designs in Scenario S2.2. The designs are sorted based on the total score, highest to the left. The inset shows the optimal sensor design projected on the network layout.

J. Water Resour. Plann. Manage., 2024, 150(11): 04024048

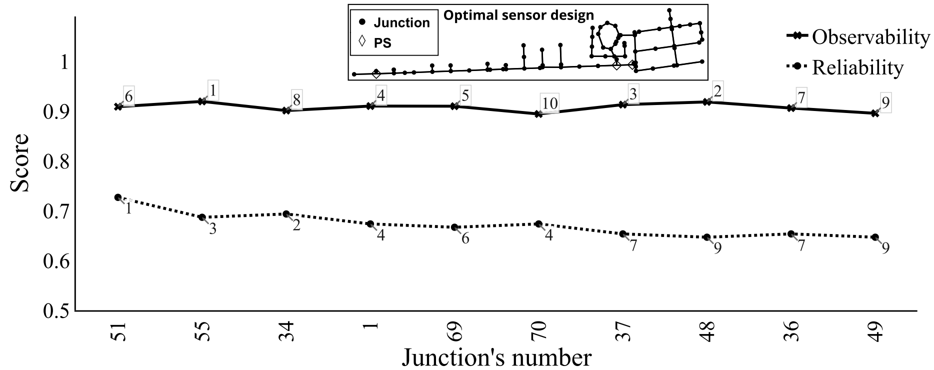

Fig. 7. Best 10 sensor designs in Scenario S3 in a network that already includes Junctions 17 and 38. The junctions are sorted based on the total score, highest to the left. The inset shows the optimal sensor design projected on the network layout.

reduced compared to Scenario S1. More importantly, Fig. 7 shows that the 10 best sensor designs were all located at the upstream portion of the network, highlighting a potential improvement in the reliability score. For example, placing the third sensor at Junction 51 increased the reliability score by 16%, compared to only 1.8% in the observability score. Taken together, Scenario S3's results showed that adding more sensors to the network would generally result in better reliability with no significant enhancement to observability. This can be attributed to the fact that the sensor placed at the network downstream already provides high observability. However, the reliability of that sensor is restively low because it does not receive a clear signal from the network's junction. Hence, adding more sensors at the upstream portion of the network enhanced the signal, and accordingly the sensors' reliability.

Conclusions

In this study, we introduced an optimization framework for placing water quality sensors in sewer networks to enhance source identification performance. The goal of the presented framework was to find the optimal sensor design that yields the best SI performance, and accordingly helps in identifying the injection characteristics of futuristic injection events. To achieve this objective, we incorporated a previously developed machine learning–based SI model within the SP optimization. The optimization process was carried out by a genetic algorithm to maximize the value of the information provided by the sensor design, represented by two performance indicators, observability and reliability. The effectiveness of our proposed SP framework was examined by applying it to a real-life sewer network considering different decay rates and design conditions.

The SP framework results revealed a trade-off between the observability and reliability of the sensor network. This highlights the importance of incorporating both metrics within the objective function because it is crucial to have sensors that can accurately observe the occurrence of injection events in the sewer system and also provide reliable estimates of the injection event characteristics. Moreover, the results showed that the observability and reliability of the sensors depend on their location within the network. In general, downstream sensors displayed high observability scores and low reliability scores because they observed a larger portion of the network. However, because they receive a diluted signal from the far junctions, they are less accurate in estimating the injection characteristics. On the other hand, upstream sensors showed high reliability scores and low observability scores because they received a concentrated signal from a limited number of junctions, allowing for a more accurate estimation of the injection characteristics. Furthermore, the weight of each score seemed to play an important role in determining the optimal sensor location.

Nevertheless, the results showed that the trade-off was less pronounced when more than one sensor was considered in the SP framework compared to the case where only one sensor was placed. For the case of high decay rate, the optimal design featured sensors placed in both the upstream and downstream sections of the network, thus simultaneously providing high observability and reliability. More importantly, as the number of sensors increases, the SP framework favored the placement of additional sensors at the upstream portion of the network to increase the reliability of the sensor network. In contrast, the increased complexity of the SI problem caused by the diminishing signal uniqueness associated with low decay rates pushed the SP framework to place the sensors in the middle of the network. In general, the proposed SP framework can be applied to a wide range of applications to minimize the risk of pollution and to protect public health. This includes optimizing the placement of water quality sensors to reveal the source characteristics of the species not intended to exist in the sewer systems (i.e., early warning systems) in addition to its application in choosing water sampling locations for epidemiological purposes (i.e., sampling location).

Although integrating the machine learning–based SI model in the proposed sensor placement model enhances computational efficiency, it also introduces certain limitations. These limitations arise from the SI model training on the simulations performed by SWMM, which only accounts for first-order decay and single-species water quality dynamics. Future studies are encouraged to apply more advanced water quality models to overcome these limitations. Furthermore, exploring alternative optimization techniques, such as Bayesian optimization, is also recommended to expand the applicability of the proposed framework to more complex networks.

Data Availability Statement

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

Funding by the National Science Foundation under Award No. 2134747 is gratefully acknowledged.

Author contributions: Aly K. Salem: conceptualization; data curation; formal analysis; software; investigation; methodology; validation; visualization; and writing–original draft. Ahmed Abokifa: conceptualization; funding acquisition; methodology; project administration; resources; supervision; and writing–review and editing.

Supplemental Materials

Tables S1 and S2 are available online in the ASCE Library (www.ascelibrary.org).

References

Adedoja, O. S., Y. Hamam, B. Khalaf, and R. Sadiku. 2019. "Sensor placement strategies for contamination identification in water distribution networks: A review." *WIT Trans. Ecol. Environ.* 229 (Jul): 79–90. https://doi.org/10.2495/WRM190081.

Afshar, A., and M. A. Mariño. 2012. "Multiobjective coverage-based ACO model for quality monitoring in large water networks." *Water Resour. Manage.* 26 (Jun): 2159–2176. https://doi.org/10.1007/s-012-0008-2.

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

Banik, B. K., L. Alfonso, C. Di Cristo, and A. Leopardi. 2017a. "Greedy algorithms for sensor location in sewer systems." *Water* 9 (11): 856. https://doi.org/10.3390/w9110856.

Banik, B. K., L. Alfonso, C. Di Cristo, A. Leopardi, and A. Mynett. 2017b. "Evaluation of different formulations to optimally locate sensors in sewer systems." *J. Water Resour. Plann. Manage.* 143 (7): 04017026. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000778.

Banik, B. K., L. Alfonso, A. S. Torres, A. Mynett, C. Di Cristo, and A. Leopardi. 2015. "Optimal placement of water quality monitoring stations in sewer systems: An information theory approach." *Procedia Eng.* 119 (Mar): 1308–1317. https://doi.org/10.1016/j.proeng.2015.08.956.

Bartos, M., and B. Kerkez. 2021. "Observability-based sensor placement improves contaminant tracing in river networks." *Water Resour. Res.* 57 (7): 1–20. https://doi.org/10.1029/2020WR029551.

Bivins, A., J. Greaves, R. Fischer, K. C. Yinda, W. Ahmed, M. Kitajima, V. J. Munster, and K. Bibby. 2020. "Persistence of SARS-CoV-2 in water and wastewater." *Environ. Sci. Technol. Lett.* 7 (12): 937–942. https://doi.org/10.1021/acs.estlett.0c00730.

Bourgeois, W., J. E. Burgess, and R. M. Stuetz. 2001. "On-line monitoring of wastewater quality: A review." *J. Chem. Technol. Biotechnol.* 76 (Jul): 337–348. https://doi.org/10.1002/jctb.393.

Brentan, B., S. Carpitella, D. Barros, G. Meirelles, A. Certa, and J. Izquierdo. 2021. "Water quality sensor placement: A multiobjective and multi-criteria approach." *Water Resour. Manage.* 35 (Mar): 225–241. https://doi.org/10.1007/s11269-020-02720-3.

Butler, D., C. Digman, C. Makropoulos, and J. W. Davies. 2018. *Urban drainage*. 4th ed. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781351174305.

Calle, E., D. Martínez, R. Brugués-i-Pujolràs, M. Farreras, J. Saló-Grau, J. Pueyo-Ros, and L. Corominas. 2021. "Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks." *Environ. Int.* 157 (Dec): 106768. https://doi.org/10.1016/j.envint.2021.106768.

Diaz-Fierros, T. F., J. Puerta, J. Suarez, and V. F. Diaz-Fierros. 2002. "Contaminant loads of CSOs at the wastewater treatment plant of a city in NW Spain." *Urban Water* 4 (3): 291–299. https://doi.org/10.1016/S1462-0758(02)00020-1.

Edmondson, V., M. Cerny, M. Lim, B. Gledson, S. Lockley, and J. Woodward. 2018. "A smart sewer asset information model to enable an 'Internet of Things' for operational wastewater management." *Autom. Constr.* 91 (Dec): 193–205. https://doi.org/10.1016/j.autcon.2018.03.003.

Elbeltagi, E., T. Hegazy, and D. Grierson. 2005. "Comparison among five evolutionary-based optimization algorithms." *Adv. Eng. Inf.* 19 (1): 43–53. https://doi.org/10.1016/j.aei.2005.01.004.

Gad, A. F. 2024. "PyGAD: An intuitive genetic algorithm Python library." *Multimed. Tools Appl.* 83: 58029–58042. https://doi.org/10.1007/s11042-023-17167-y.

Guadagno, V., G. Del Giudice, C. Di Cristo, A. Leopardi, and A. Simone. 2023. "Impact coefficient evaluation for sensor location in sewer systems." *J. Water Resour. Plann. Manage.* 149 (11): 04023063. https://doi.org/10.1061/JWRMD5.WRENG-6093.

Hart, W. E., and R. Murray. 2010. "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems." *J. Water Resour. Plann. Manage.* 136 (6): 611–619. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000081.

Haupt, R. L., and S. E. Haupt. 2003. *Practical genetic algorithms*. New York: Wiley.

Huntington, D. E., and C. S. Lyrintzis. 1998. "Improvements to and limitations of Latin hypercube sampling." *Probab. Eng. Mech.* 13 (Mar): 245–253. https://doi.org/10.1016/S0266-8920(97)00013-1.

Larson, R. C., O. Berman, and M. Nourinejad. 2020. "Sampling manholes to home in on SARS-CoV-2 infections." *PLoS One* 15 (10): 1–20. https://doi.org/10.1371/journal.pone.0240007.

Lin, W., Z. Huang, S. Gao, Z. Luo, W. An, P. Li, S. Ping, and Y. Ren. 2021. "Evaluating the stability of prescription drugs in municipal wastewater and sewers based on wastewater-based epidemiology." *Sci. Total Environ.* 754 (Feb): 142414. https://doi.org/10.1016/j.scitotenv.2020.142414.

McDonnell, B., K. Ratliff, M. Tryby, J. Wu, and A. Mullapudi. 2020. "PySWMM: The Python interface to stormwater management model (SWMM)." *J. Open Source Software* 5 (52): 2292. https://doi.org/10.21105/joss.02292.

Mu, T., M. Huang, S. Tang, R. Zhang, G. Chen, and B. Jiang. 2022. "Sensor partitioning placements via random walk and water quality and leakage detection models within water distribution systems." *Water Resour. Manage.* 36 (May): 5297–5311. https://doi.org/10.1007/s11269-022-03312-z.

Nourinejad, M., O. Berman, and R. C. Larson. 2021. "Placing sensors in sewer networks: A system to pinpoint new cases of coronavirus." *PLoS One* 16 (Apr): e0248893. https://doi.org/10.1371/journal.pone.0248893.

Pan, T.-C., and J.-J. Kao. 2009. "GA-QP model to optimize sewer system design." *J. Environ. Eng.* 135 (Apr): 17–24. https://doi.org/10.1061/(ASCE)0733-9372(2009)135:1(17).

Preis, A., and A. Ostfeld. 2008a. "Genetic algorithm for contaminant source characterization using imperfect sensors." *Civ. Eng. Environ. Syst.* 25 (Dec): 29–39. https://doi.org/10.1080/10286600701695471.

Preis, A., and A. Ostfeld. 2008b. "Multiobjective contaminant response modeling for water distribution systems security." *J. Hydroinf.* 10 (Mar): 267–274. https://doi.org/10.2166/hydro.2008.061.

Rathi, S., and R. Gupta. 2014. "Sensor placement methods for contamination detection in water distribution networks: A review." *Procedia Eng.* 89 (Jan): 181–188. https://doi.org/10.1016/j.proeng.2014.11.175.

Rathi, S., and R. Gupta. 2016. "A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks." *KSCE J. Civ. Eng.* 20 (Mar): 597–608. https://doi.org/10.1007/s12205-015-0024-x.

Salem, A. K., and A. A. Abokifa. 2023. "Machine learning–based source identification in sewer networks." *J. Water Resour. Plann. Manage.* 149 (8): 04023034. https://doi.org/10.1061/JWRMD5.WRENG-6050.

Sambito, M., C. Di Cristo, G. Freni, and A. Leopardi. 2020. "Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification." *J. Hydroinf.* 22 (1): 46–60. https://doi.org/10.2166/hydro.2019.036.

Sambito, M., and G. Freni. 2021. "Strategies for improving optimal positioning of quality sensors in urban drainage systems for non-conservative contaminants." *Water* 13 (7): 934. https://doi.org/10.3390/w13070934.

Taha, A. F., S. Wang, Y. Guo, T. H. Summers, N. Gatsis, M. H. Giacomoni, and A. A. Abokifa. 2021. "Revisiting the water quality sensor placement problem: Optimizing network observability and state estimation metrics." *J. Water Resour. Plann. Manage.* 147 (7): 1–13. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001374.

Tatiparthi, S. R., Y. G. De Costa, C. N. Whittaker, S. Hu, Z. Yuan, R. Y. Zhong, and W. Q. Zhuang. 2021. "Development of radio-frequency identification (RFID) sensors suitable for smart-monitoring applications in sewer systems." *Water Res.* 198 (Jun): 117107. https://doi.org/10.1016/j.watres.2021.117107.

Wang, Y., C. L. Moe, S. Dutta, A. Wadhwa, S. Kanungo, W. Mairinger, Y. Zhao, Y. Jiang, and P. F. Teunis. 2020. "Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation." *Epidemics* 31 (Jun): 100391. https://doi.org/10.1016/j.epidem.2020.100391.

Xuesong, Y., S. Jie, and H. Chengyu. 2017. "Research on contaminant sources identification of uncertainty water demand using genetic algorithm." *Cluster Comput.* 20 (Jun): 1007–1016. https://doi.org/10.1007/s10586-017-0787-6.