Optimization of Water Quality Sensor Placement in Sewer Networks

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

[1]Department of Civil, Materials, and Environmental Engineering, the University of Illinois Chicago, Chicago, IL 60607, USA. Assistant lecturer, Faculty of Engineering, Cairo University, Giza; E-mail: asalem22@uic.edu

[2]Department of Civil, Materials, and Environmental Engineering, the University of Illinois Chicago, Chicago, IL 60607, USA; Email: abokifa@uic.edu

Abstract

Wastewater surveillance has received significant attention in recent years due to the valuable environmental and epidemiological information it reveals. Yet, limited attempts have been made in the previous literature to determine the best locations for sample collection. In this study, we are introducing a model to solve the sensor placement (SP) problem that considers source identification (SI) as the main criterion for sensor placement. The objective of this study is to maximize the ability of the sensor design to identify the characteristics of the injection source (i.e., the injection location and concentration) of any given species of interest. To that end, we developed a python-based model that couples an SI model with a Greedy Algorithm to determine the best locations to place water quality sensors. The developed SP model was demonstrated on a benchmark, mid-sized sewer network. The results revealed that a clear tradeoff exists between the overall identifiability and the detection reliability of the sensor design. In general, placing the sensors in the downstream end of the network resulted in higher overall identifiability, while better detection reliability was achieved by the sensors placed in the upstream section of the network.

INTRODUCTION

Monitoring the water quality in sewer systems is crucial for informing the operations of wastewater treatment plants (Nourinejad et al., 2021). Water quality monitoring is particularly of significant importance in combined sewer systems where excess untreated effluents are discharged directly to water bodies (Banik et al., 2017a). Additionally, the identification of the sources of any unusual water constituents (e.g., contaminants) in sewer systems is of great importance (Sambito and Freni, 2021). This problem is formally known as the Source Identification (SI) problem, whereby the injection characteristics of a specific constituent/contaminant in the sewer network need to be identified. The SI problem is typically formulated as an optimization problem that aims to identify the locations and concentrations of constituent injection into the sewer system. For instance, Banik et al., (2014) proposed an optimization model for solving the SI problem by integrating the Storm Water Management Model (SWMM) with the Genetic Algorithm. Later, Banik et al., 2015a aimed to reduce the complexity of the SI problem through developing a pre-screening procedure to decrease the number of candidate junctions.

The solution to the SI problem is typically derived based on concentration data collected at one or more observation junctions (i.e., sensors). Accordingly, accurate concentration data is key to successful source identification. This can be achieved by optimizing the sensor placement (SP) in the network to maximize the quality of the collected monitoring data. Several attempts have been made in previous literature to solve the SP problem. Banik et al., 2015b proposed optimizing the sensor placement using the Non-dominated Sorting Genetic Algorithm (NSGA-II). Additionally, Banik et al., 2017a implemented the Greedy Algorithm (GR) to optimize sensor placement. Later, Banik et al., 2017b tested the adequacy of the GR in solving different objective function formulations (e.g., single and multi-objective functions). Finally, a probabilistic approach was introduced by Sambito et al., 2020 to optimize the SP.

Although the main motivation of the aforementioned studies was to develop sensor placement strategies that can better serve source identification in sewer systems, none of these studies employed model-based SI metrics within their SP optimization formulations. In this study, the SP problem is tackled from a different perspective, where an SI module was integrated within the sensor placement optimization process. The objectives of this study are to i) show how to employ an SI module in the SP optimization to enhance the sensors' identification performance, ii) introduce a ranked-based approach to place multiple sensors in the network, and iii) understand the relationship between the location of the sensor and the information it is providing.

METHODS

In this study, a model for determining the best locations to place water quality sensors in sewer networks is developed. The SP model aims to identify the sensor design alternative with the highest ability to identify the source characteristics of any given species of interest. To that end, the SP model integrates the results of an SI module (Salem and Abokifa, 2022) with a Greedy Algorithm. In this section, the SI module and its implementation within the SP model are explained. In addition, the case study used to show the capabilities of the proposed SP model is described.

Source Identification (SI) module

The main criterion used for assessing the quality of a specific sensor placement design is its ability to identify the characteristics of an injection event (i.e., injection location and concentration). The identification process is conducted based on the observed concentration data by the sensors. In this study, we are considering multiple simultaneous continuous injections. The injection event can be identified by minimizing the normalized root mean square error ($nRMSE$) between the simulated concentrations and those observed by the sensors:

$$\text{Minimize:}\, nRMSE = \sum_{j=1}^{m} \frac{\left[\sum_{i=1}^{n} \left(C_{ji}^{obs} - C_{ji}^{sim}\right)^2 / n\right]^{1/2}}{\left(\sum_{i=to}^{n} C_{ji}^{obs}\right)/n} \qquad (1)$$

where C^{obs} and C^{sim} are the concentration observed and simulated at the sensor location; j is the index of the sensor, m is the total number of sensors, i is the index of the time step, and n is the total number of time steps. The SI module takes the observed concentrations as inputs and implements a simulation-optimization approach to inversely search for the injection event that reproduces the concentrations observed by the sensors. Detailed information on the SI module can be found in Salem and Abokifa, 2022.

Water quality simulation

In the SI module, water quality simulations are conducted by a machine learning-based surrogate model in the form of a multi-layer perceptron neural network (MLP-NN). Initially, a pre-processing step is performed by running a SWMM simulation to identify the connectivity status of the network junctions to each of the sensors within a specific sensor design. This step is important to eliminate the junctions that do not affect the concentrations observed by any of the sensors, which helps reduce the computational redundancy and computational complexity of the SI module.

After the determination of the junction's connectivity status, the input dataset is generated. This dataset consists of a pre-defined number of randomly generated injection events in the form of [(L_1, C_1),(L_1, C_1), ...,(L_{n_J}, C_{n_J})], where (L_i, C_i) represents the injection location and concentration of an injection source. Then, the generated injection events are then simulated by SWMM to form the output dataset. Later, the input and output datasets are split to train and test the MLP-NN surrogate model of each sensor. Through the training process, the MLP-NN can learn the input-output relationship between injection source characteristics and the signal generated at the sensors, and can hence be used to simulate other injection events. A python script was coded to perform the aforementioned tasks, with the use of the PySWMM package (McDonnell et al., 2020) to automate the SWMM simulations, and the Sklearn package (Pedregosa et al., 2011) to train and test the MLP-NN surrogate models.

Error minimization

To minimize the error between the simulated and the observed concentrations (Eq. (1)), the Genetic Algorithm (GA) is utilized by the SI module. First, the GA produces several injection events (i.e., individuals), and sends them to the MLP-NN to be simulated. Then, the $nRMSE$ (i.e., fitness) of each individual is calculated according to Eq. (1) and reported back to the GA. The next generation of individuals is then created by applying the selection, crossover, and mutation operators as explained in Salem et al., 2022. The same cycle is then repeated until the $nRMSE$ reaches zero, or 100 generations are completed. The GA implemented in the SI module was done by using the PyGAD package (Gad, 2021).

Sensor Placement (SP) model

The SP model employs the result of the SI model to evaluate the performance of different sensor design alternatives. This is done by computing a score (S_{score}) to measure the value of the information provided by each sensor design alternative in identifying the injection characteristics for several injection scenarios (i.e., events). To calculate the S_{score}, the identified injection characteristics are compared to the true ones, so that higher S_{score} means more accurate identification. The S_{score} can be mathematically represented as follows,

$$S_{score} = \frac{100}{N_{sc} \times N_{so}} \sum_{n=1}^{N_{sc}} \sum_{o=1}^{N_{so}} \frac{1}{\left\{\left[1 + \left(\frac{D_o}{D_o^{avg}}\right)\right] + \left[1 + C_o^{err}\right]\right\}} \tag{2}$$

where n is the scenario index out of N_{sc} scenarios; o is the source index out of N_{so} sources; D_o is the distance between the true and the identified injection source, D_o^{avg} is the average distance between the true injection source and the sensors; and C_o^{err} is the absolute percentage error between the true and the identified injection concentration.

First, a pre-defined number of injection scenarios are randomly generated, each with a pre-defined number of injection sources. Then, these scenarios are simulated so that the concentration results at the sensors are identified, which are then sent to the SI module to solve the inverse SI problem and to identify the injection characteristics of the injection scenarios. Once the SI is completed, the identification results are linked to their corresponding sensor design alternative. Finally, the S_{score} of each alternative is calculated according to Eq. (2).

Greedy Algorithm

In this study, a Greedy Algorithm (GR) is used as a rank-based approach to find the best placement of multiple sensors in the network. In this algorithm, the placement of N sensors is achieved by performing N ranking rounds. In each round, the highest-ranking location is selected as a sensor and then fixed while performing the following round to select the location of the next sensor. Even though the GR doesn't guarantee finding the optimal solution, it has proved to be an efficient and simple approach for finding near-optimal solutions for large search domain problems with low computational cost and was hence commonly used in various SP optimization studies (Banik et al., 2017b; Cheifetz et al., 2015; Sela and Amin, 2018).

Case study

To demonstrate the performance of the proposed SP model, we applied it to the SWMM example 8 network, which has been repeatedly used in previous literature. The case study network data was collected from the SWMM applications manual (Gironás et al., 2009). The SWMM example 8 network drains the dry weather flow (DWF) of 5 sub-sewer sheds to 1 outfall through 16 conduits and 16 junctions (Figure 1). The DWF assigned to the exterior 5 junctions comes directly from the sub-sewer sheds, whereas a direct DWF was assigned to the other intermediate junctions, the DWF assignments are summarized in Table 1. The pattern of the DWF was assumed to follow the wastewater pattern developed by Butler et al., 2018. In this study, two sensor placement cases were examined, one sensor was considered in the first case (C1), whereas two sensors were considered in the second case (C2). In both C1 and C2, it was considered that a non-conservative constituent (i.e., contaminant) with a decay rate (k) of $17\ d^{-1}$ is injected in two junctions simultaneously.

Table 1: DWF assignments of SWMM example 8 network

Junction	J1	J2a	J3	J13	J12	All others
DWF (liter/min.)	13.6	17.0	6.8	20.9	21.2	8.5

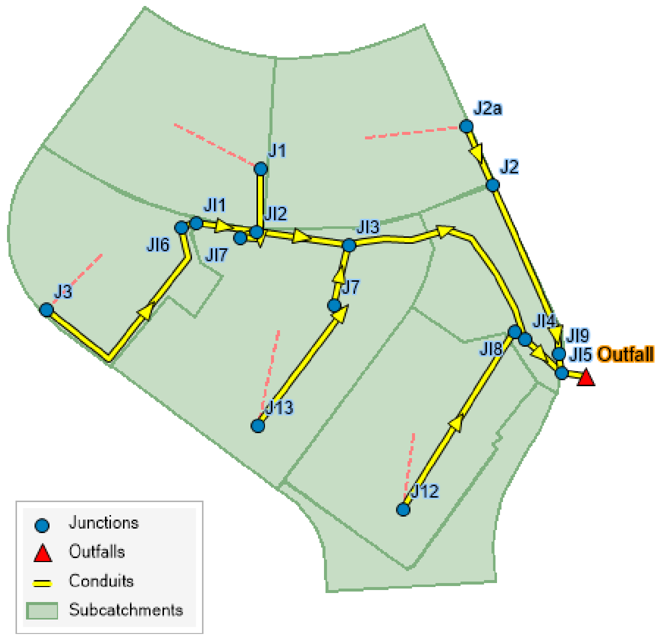

Figure 1: The layout of SWMM Example 8 network

Injection scenarios

To evaluate the performance of different sensor design alternatives, 100 injection scenarios were generated randomly by using the Latin Hypercube Sampling method (Huntington and Lyrintzis, 1998). The injection scenarios were generated considering a uniform distribution over the injection location to ensure that all junctions have an equal opportunity to be selected as an injection source (Figure 2). Similarly, a uniform distribution over the injection concentration is considered between 1 and 100 mg/L.

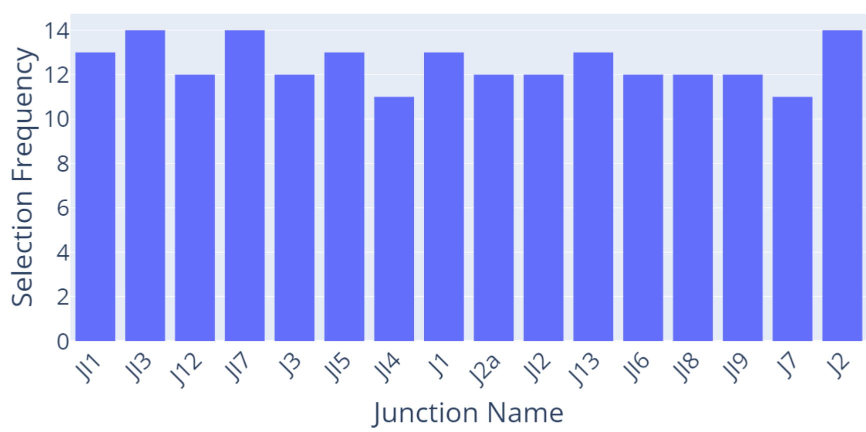

Figure 2: The selection frequency of the junctions as an injection source

RESULTS AND DISCUSSION

Case 1 (C1): one sensor placement

To find the best location for placing one sensor in the network, the SP model was run to rank the junctions based on their S_{score}. In addition to the S_{score}, two other metrics were computed to help assess the performance of different sensor design alternatives, namely the sensor coverage (S_C) and the detection reliability (D_R). The S_C is the percentage of the network's junctions covered by each sensor. The higher the value of S_C, the greater the ability of the sensor to observe more of the network junctions. The D_R is the percentage of the correctly detected injection scenarios relative to the injection scenarios covered by a sensor. The higher the value of D_R, the more reliable the sensor is in detecting the injection characteristics of the scenarios it is observing.

In Figure 3, the x-axis represents the junction name, and the y-axis represents the i) S_{score}, ii) S_C, and iii) D_R. Figure 3 shows that there is a strong correlation between the sensor coverage (S_C) and the sensor score (S_{score}). This indicates that, as expected, sensors observing more of the network junctions tend to achieve higher S_{score}. Accordingly, the network outfall was found to be the best location to place the first sensor as it observes all the network junctions.

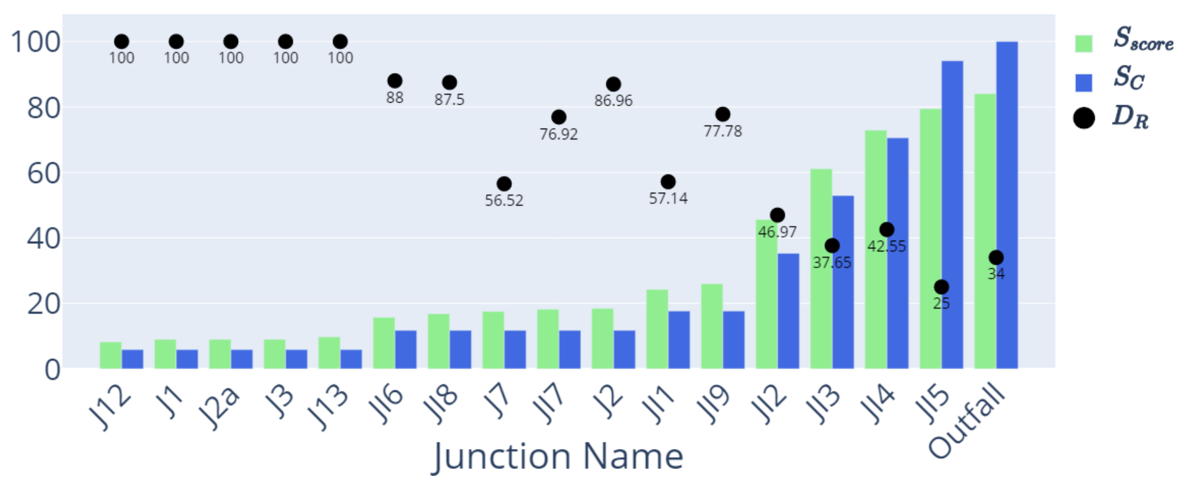

Figure 3: Results of placing one sensor (C1). The bar plots represent the S_{score} and the S_C, and the scatter plot represents the D_R

In general, junctions with low S_{score} seemed to perform better in terms of D_R. These are generally junctions located near the upstream sections of the network (e.g., J2, J7, and JI8). Even though these junctions observe a smaller portion of the network compared to downstream junctions (i.e., lower S_C), they receive a clearer signal from the junctions and can hence better detect their characteristics (i.e., higher D_R). On the other hand, junctions with high S_{score} tend to be the ones at/near the downstream sections of the network (e.g., the outfall, JI5, and JI4). These junctions observe more of the sewer network (i.e., higher S_C) but at the expense of their ability to accurately detect the true injection characteristics of the distant junctions (i.e., lower D_R). Taken together, these results highlight the trade-off between the overall identifiability represented by the S_{score} and the detection reliability represented by the D_R.

Case 2 (C2): two sensors placement

According to the Greedy Algorithm, a second ranking round is required to place the second sensor. Hence, the SP model was run to find the location of the second sensor considering that a sensor already exists at the highest-ranking location determined by the first round (i.e., the outfall). Figure 4 shows the results of two sensors placement (C2). The x-axis represents the junction name where the second sensor is placed. The left y-axis represents the S_{score} of the sensor design alternative (considering another sensor at the outfall), and the right y-axis represents the junction rank in C1 (considering only a sensor at that junction). It can be seen from Figure 4 that JI2 was the best location to place the second sensor, although it was ranked 5th in C1 (i.e., as the only sensor). More importantly, JI2 was selected as the second sensor in favor of JI3 and JI8, which according to Figure 1, are also considered central junctions. Figure 4 also shows that downstream junctions (e.g., JI5 and JI4) are not good locations for the second sensor, even though they are considered valuable in C1. The reason is that they provide similar (i.e., redundant) information to that provided by the outfall.

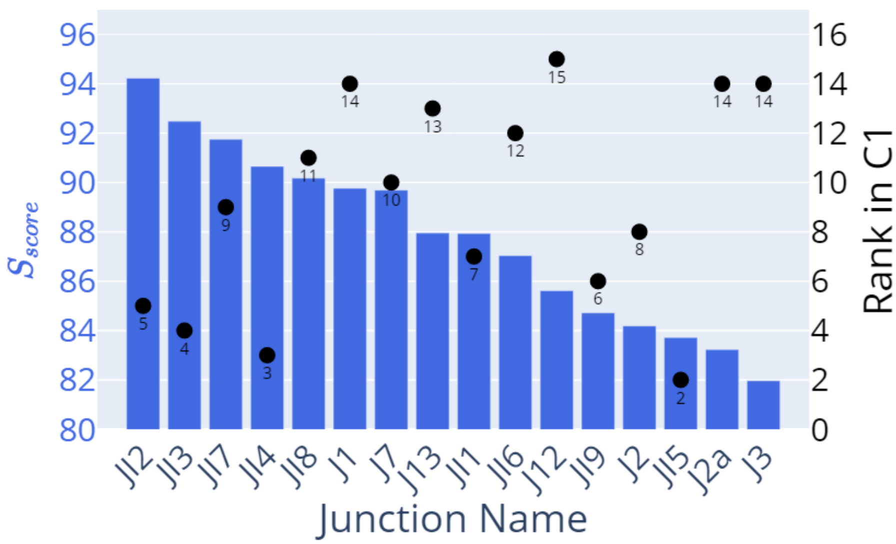

Figure 4: Results of placing two sensors (C2). The bar plot represents the S_{score} and corresponds to the left y-axis, and the scatter plot represents the junction rank in C1 and corresponds to the right y-axis

CONCLUSIONS

In this study, we proposed an approach for integrating model-based source identification within the optimization of water quality sensor placement in sewer systems. The developed SP model utilizes the Greedy Algorithm to optimize the placement of multiple sensors with the aim of maximizing the ability of sensors to identify the injection characteristics of water constituents of interest. The performance of the SP model was examined by applying it on a benchmark sewer network under different injection scenarios. The results revealed that a trade-off exists between the overall identifiability represented by the S_{score} and the detection reliability (D_R) of a sensor design alternative. Future studies are recommended to follow a similar approach while tackling sensor placement optimization to ensure that the information gained from the sensors achieves accurate source identification.

ACKNOWLEDGEMENTS

This work was supported by funding from the University of Illinois Chicago, and the National Science Foundation under grant number 2015603.

REFERENCES

- Banik, B.K., Alfonso, L., Di Cristo, C., Leopardi, A., Mynett, A., 2017a. Evaluation of Different Formulations to Optimally Locate Sensors in Sewer Systems. J. Water Resour. Plan. Manag. 143, 04017026. https://doi.org/10.1061/(asce)wr.1943-5452.0000778
- Banik, B.K., Alfonso, L., Di Cristo, C., Leopardi, A., 2017b. Greedy algorithms for sensor location in sewer systems. Water (Switzerland) 9, 1–14. https://doi.org/10.3390/w9110856
- Banik, B.K., Di Cristo, C., Leopardi, A., 2015a. A pre-screening procedure for pollution source identification in sewer systems. Procedia Eng. 119, 360–369. https://doi.org/10.1016/j.proeng.2015.08.896
- Banik, B.K., Alfonso, L., Torres, A.S., Mynett, A., Di Cristo, C., Leopardi, A., 2015b. Optimal placement of water quality monitoring stations in sewer systems: An information theory approach. Procedia Eng. 119, 1308–1317. https://doi.org/10.1016/j.proeng.2015.08.956
- Banik, B.K., Di Cristo, C., Leopardi, A., 2014. SWMM5 toolkit development for pollution source identification in sewer systems. Procedia Eng. 89, 750–757. https://doi.org/10.1016/j.proeng.2014.11.503
- Butler, D., Digman, C.J., Makropoulos, C., Davies, J.W., 2018. Urban Drainage, Fourth Edi. ed. CRC Press, Taylor & Francis Group. https://doi.org/10.4324/9780203351673
- Cheifetz, N., Sandraz, A.C., Féliers, C., Gilbert, D., Piller, O., Lang, A., 2015. An incremental sensor placement optimization in a large real-world water system. Procedia Eng. 119, 947–952. https://doi.org/10.1016/j.proeng.2015.08.977
- Gad, A.F., 2021. PyGAD: An Intuitive Genetic Algorithm Python Library.
- Gironás, J., Roesner, L.A., Davis, J., Rossman, L.A., 2009. Storm Water Management Model Applications Manual. Cincinnati, OH. National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency.
- Huntington, D.E., Lyrintzis, C.S., 1998. Improvements to and limitations of Latin hypercube sampling. Probabilistic Eng. Mech. 13, 245–253. https://doi.org/10.1016/s0266-8920(97)00013-1
- McDonnell, B., Ratliff, K., Tryby, M., Wu, J., Mullapudi, A., 2020. PySWMM: The Python Interface to Stormwater Management Model (SWMM). J. Open Source Softw. 5, 2292. https://doi.org/10.21105/joss.02292

- Nourinejad, M., Berman, O., Larson, R.C., 2021. Placing sensors in sewer networks: A system to pinpoint new cases of coronavirus. PLoS One 16, 1–25. https://doi.org/10.1371/journal.pone.0248893
- Salem, A.K., Imam, Y.E., Ghanem, A.H., Bazaraa, A.S., 2022. Genetic Algorithm Based Model for Optimal Selection of Open Channel Design Parameters. Water Resour. Manag. https://doi.org/10.1007/s11269-022-03323-w
- Salem, A.K., Abokifa, A.A., 2022. Machine Learning-Based Source Identification in Sewer Networks. Under review
- Sambito, M., Di Cristo, C., Freni, G., Leopardi, A., 2020. Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification. J. Hydroinformatics 22, 46–60. https://doi.org/10.2166/hydro.2019.036
- Sambito, M., Freni, G., 2021. Strategies for improving optimal positioning of quality sensors in urban drainage systems for non-conservative contaminants. Water (Switzerland) 13. https://doi.org/10.3390/w13070934
- Sela, L., Amin, S., 2018. Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization. Adv. Eng. Informatics 36, 55–63. https://doi.org/10.1016/j.aei.2018.02.004
- Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: machine learning in python. Journal of machine learning research, 12(Oct), 2825–2830.