

Improving the Performance of Bayesian Decision Networks for Water Quality Sensor Deployment in UDNs through a Reduced Search Domain [†]

Mariacrosetta Sambito ¹  and Antonietta Simone ^{2,*} 

¹ Department of Engineering and Architecture, University of Enna “Kore”, 94100 Enna, Italy; mariacrosetta.sambito@unikore.it

² Department of Engineering and Geology, University “G. D’Annunzio” of Chieti Pescara, 66100 Pescara, Italy

* Correspondence: antonietta.simone@unich.it

[†] Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: The contamination of urban drainage systems (UDNs) represents a serious threat to the environment and public health. Treatment plants are often inefficient in their removal, making timely identification and isolation interventions necessary. In this regard, various monitoring strategies have been proposed, among which the Bayesian decision network (BDN) approach has proven to be very effective, although also very complex. To reduce their level of complexity, it is usual to optimize them using approaches based on preconditioning. The present work fits into this framework by proposing a two-phase strategy aimed at identifying an optimal monitoring system for UDNs. The first phase involves reducing the search domain of the system using a complex network theory (CNT) topological metric adapted to infrastructure systems; the second phase implements the Bayesian approach to the new search space to optimize the position of the sensors in the network. The results are promising and reveal that the strategy could be valuable to water utilities.

Keywords: Bayesian decision network; UDN monitoring; contaminant; complex network theory; reduced search domain



Citation: Sambito, M.; Simone, A.

Improving the Performance of Bayesian Decision Networks for Water Quality Sensor Deployment in UDNs through a Reduced Search Domain. *Eng. Proc.* **2024**, *69*, 58. <https://doi.org/10.3390/engproc2024069058>

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 4 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The discharge of polluting substances into urban drainage networks (UDNs) represents a serious threat to the receptor bodies, resulting in acute or cumulative impacts on the environment and public health also. Most strategies aimed at pollutant identification involve the optimal planning of a monitoring system [1,2], and among the various approaches, the Bayesian Decision Network (BDN) [3] is gaining momentum. It represents a useful decision support tool for the optimal design of monitoring systems, with the advantage of assimilating information from the system in an upgradable and updateable way (e.g., from numerical simulations and data from the real system), which acquires value especially for uncertain systems. In fact, a Bayesian decision network is an acyclic graphical structure that allows us to represent an uncertain domain and the conditional dependencies between independent and dependent variables in a probabilistic way. At the same time, however, Bayesian networks require a large number of computational resources to solve the optimization problem, which makes their implementation complex. For this reason, the use of pre-conditioning approaches is usually suggested to mine relevant probability information from available data, e.g., system topology, potential source positions, previous contamination events, etc.

In light of this consideration, the present paper proposes a two-step procedure for the optimal planning of a monitoring system for UDNs. The first step involves the use of a CNT topological metric [4] to perform a shrinking of the network domain by identifying a

range of the more relevant elements: this being the narrow domain results characterized by the presence of only the most suitable points to host sensors with respect to the specific problem to be analyzed. In the second step, the Bayesian approach [3] is applied to the domain obtained by the topological analysis to define the position of the sensors in the network. The integration of the two approaches allows us to obtain a better performing and effective strategy for the planning of advanced monitoring systems. In fact, the approach identifies the optimal sensor location gaining advantage from additional information but still reduces the computational effort needed to obtain the solution. The strategy represents a useful tool for both technicians and water utilities who must increasingly face problems relating to the contamination, often illicit, of urban systems.

2. Materials and Methods

The proposed study presents an innovative integration of two different methodologies to improve effectiveness in the design of advanced monitoring systems for UDNs. The strategy is composed of two well-defined phases. In the first phase, a CNT topological metric is used to reduce the search space of the system. Next, a Bayesian optimization procedure is applied to determine the optimal location of the sensors.

2.1. First Phase: Search Space Reduction through a CNT Topological Metric

The behavior of UDNs is strongly influenced by their topology, and, for this reason, the use of CNT tools proves effective and useful in supporting and simplifying many analyses with respect to different problems. The proposed approach involves the use of the Harmonic centrality [4,5], a metric that indicates the ability of each node to collect and spread information in the system. The metric is used in its directed and relevance-based version [5] in order to consider both the information on the topological component (connective structure) and on the hydraulic one (intrinsic relevance of the nodes). The metric allows us to define a ranking of importance of the nodes, which is useful in identifying a reduced search space. Considering a network with N nodes and L links, the mathematical formulation of the relevance-based In-Harmonic centrality [6] is:

$$H_i^C = \sum_{j=1}^N \frac{1}{d_{ij}} f(R_i, R_j) \quad (1)$$

where R_i and R_j represent the intrinsic relevance R_n ($n = 1, \dots, N$) of the final nodes i and j of the generic link l ($l = 1, \dots, L$). Intrinsic relevance is integrated into the analysis through the function $f(R_i, R_j)$, which depends on the specific characteristics of the problem to be analyzed. Nodes with higher values of the metric detect the greater quantity of information on the network and are, therefore, included in the reduced search domain.

2.2. Bayesian Approach

In order to determine the position of the sensors through the Bayesian optimization procedure, the mathematical model described in [3] was used. The Bayesian decision support process can ideally be divided into three main phases. First, it is important to establish a probability model for all observable and unobservable quantities in the problem. Next, it is possible to proceed to calculate and interpret the appropriate posterior distribution and conditional probability distribution. In its original version, the Bayesian numerical simulation model includes a first step with a pre-screening procedure that allows us to remove all nodes that are not strategic for the rapid detection of contamination, such as head nodes, since placing the sensors in those nodes would give 1:1 information. The present work suggests replacing the classic pre-screening procedure with those proposed by the relevance-based In-Harmonic previously defined.

The strategy involves the combined use of the “MatLab” programming language and the “EPA-SWMM 5” software for the hydraulic simulation by recalling the functions via the “MatSWMM” toolbox [7]. It is important to evaluate how well the model fits the data and

whether the conclusions obtained are reasonable; then, it is possible to decide whether to modify the model and repeat all of the steps. The Bayesian probability calculation, which is based on the calculation of the percentage of contamination interception that a node has if a sensor was positioned there, can be applied to further reduce the search domain. All of the nodes are considered a possible source of contamination.

3. Results and Discussion

The proposed strategy is applied to a benchmark UDN, whose network model is composed of 77 manholes, 79 sewer pipes, and 1 outfall. In order to validate the approach, the results obtained were compared with the Bayesian numerical model as proposed in [3]. For both cases, the hypothesized contaminant is of a conservative type (i.e., it is not subject to degradation phenomena).

The classic Bayesian approach identified the set of candidate nodes for sensor positioning shown in Figure 1 (left) and reported in Table 1 (left). The highest values of the contamination interception percentage are at nodes 78, corresponding with the outfall, 74 (94.87%) and 60 (82.05%), while the lowest one is at node 13 (8.97%).

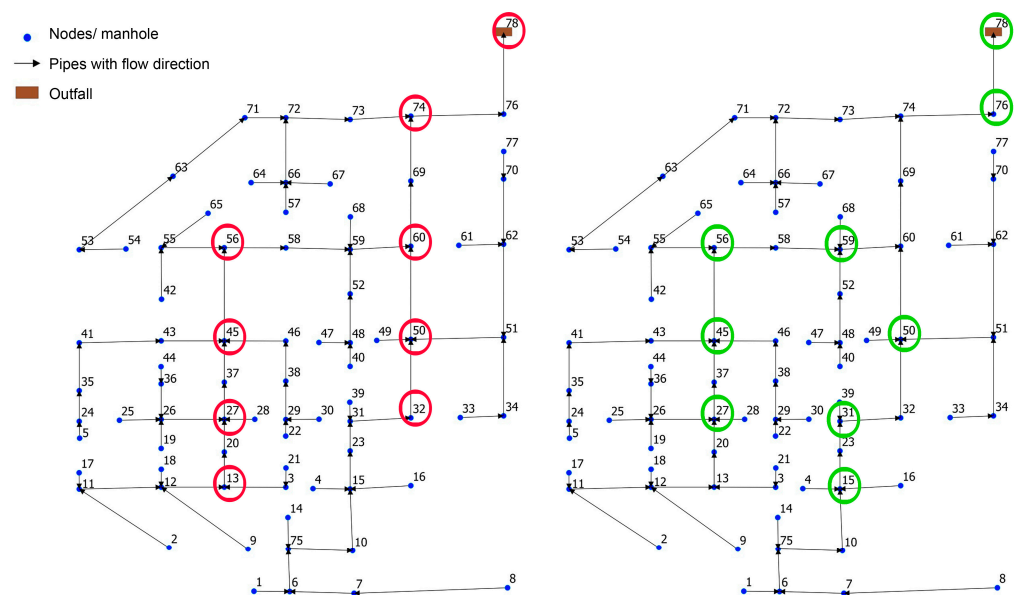


Figure 1. Results for the classic Bayesian (left) approach and the proposed two-phase strategy (right).

Table 1. Results in terms of sensor positioning for Bayesian approach and two-phase approach.

Sensor ID Bayesian Approach	Contamination Interception Bayesian Approach [%]	Sensor ID Two-Phase Approach	Contamination Interception Two-Phase Approach [%]
78	100.00	78	100.00
74	94.87	76	96.68
60	82.05	59	87.42
56	42.31	56	44.34
45	37.18	45	38.05
50	29.49	50	32.22
27	21.79	27	26.12
32	17.95	31	20.27
13	8.97	15	12.97

The two-phase approach starts with the evaluation of the relevance-based In-Harmonic for all of the nodes of the network. The lateral inflow is assumed as intrinsic relevance for each node, since it is proportional to the quantity of contaminant introduced into each node (expressed as mg/l). The topological analysis provided a ranking of the importance

of the nodes, useful for defining the reduced search domain, which is here composed of all of the nodes with topological metric values in the range [30–100]. The second phase of the strategy involved applying the Bayesian approach to this reduced search domain. The set of candidates for sensor positioning is shown in Figure 1 (right) and reported in Table 1 (right).

The results show how the final set of sensor positioning is very similar between the two approaches, with the advantage of presenting higher contamination interception percentages for the nodes identified with the two-phase strategy proposed here. The important role of topology in UDN contamination is widely recognized. Therefore, carrying out effective pre-screening before applying Bayesian logic allows for a significant improvement in identifying the best sensor placement strategy.

4. Conclusions

The problem of pollutants in UDNs is increasingly frequent and causes several negative cascading effects, both on the system itself and on the environment. The present work proposes the use of a topological approach as a pre-screening phase to be used in a Bayesian probabilistic numerical simulation model. The comparison between the results with and without the topological pre-screening procedure demonstrated how the additional information positively influences the model results, amplifying the number of contamination interceptions detected at the nodes. These results validate the proposed strategy and demonstrate how using topology-based approaches improves analyses, which is relevant also for contaminant detection.

Author Contributions: Conceptualization, M.S. and A.S.; methodology, M.S. and A.S.; writing—original draft preparation, M.S. and A.S.; writing—review and editing, M.S. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-Generation EU (National Recovery and Resilience Plan—NRRP, Mission 4, Component 2, Investment 1.3—D.D. 1243 2/8/2022, PE0000005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Simone, A.; Di Cristo, C.; Guadagno, V.; Del Giudice, G. Sewer networks monitoring through a topological backtracking. *J. Environ. Manag.* **2023**, *15*, 346. [[CrossRef](#)] [[PubMed](#)]
2. Sambito, M.; Piazza, S.; Freni, G. Optimal Deployment of the Water Quality Sensors in Urban Drainage Systems. *Environ. Sci. Proc.* **2022**, *21*, 42. [[CrossRef](#)]
3. Sambito, M.; Freni, G. Strategies for improving optimal positioning of quality sensors in urban drainage systems for non-conservative contaminants. *Water* **2021**, *13*, 934. [[CrossRef](#)]
4. Roachat, Y. Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index. In *Applied Social Network Analysis (ASNA)*; Institute of Applied Mathematics, University of Lausanne: Lausanne, Switzerland, 2009.
5. Simone, A.; Cesaro, A.; Del Giudice, G.; Di Cristo, C.; Fecarotta, O. Potentialities of complex network theory tools for urban drainage networks analysis. *Water Resour. Res.* **2022**, *58*, e2022WR032277. [[CrossRef](#)]
6. Simone, A.; Cesaro, A.; Di Cristo, C.; Fecarotta, O.; Morani, M.C. Two different approaches for monitoring planning in sewer networks: Topological vs. deterministic optimization. *J. Inform.* **2023**, *25*, 2323–2337. [[CrossRef](#)]
7. Riaño-Briceño, G.; Barreiro-Gomez, J.; Ramirez-Jaime, A.; Quijano, N.; Ocampo-Martinez, C. MatSWMM—An open-source toolbox for designing real-time control of urban drainage systems. *Environ. Model. Softw.* **2016**, *83*, 143–154. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.