Enhancing inflow and infiltration detection in urban sewer networks with a new deterministic sensor placement method

Wu, Yuling; Zheng, Feifei; Yang, Yongfei; Zhang, Kaiming; Du, Kun; Duan, Huanfeng; Savic, Dragan; Kapelan, Zoran

DOI
10.1016/j.watres.2024.122591

Publication date
2024

Document Version
Final published version

Published in
Water Research

Water Research 268 (2025) 122591

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Enhancing inflow and infiltration detection in urban sewer networks with a new deterministic sensor placement method

Yuling Wu [a], Feifei Zheng [a,*], Yongfei Yang [b], Kaiming Zhang [b], Kun Du [c], Huanfeng Duan [d], Dragan Savic [e,f], Zoran Kapelan [g]

[a] College of Civil Engineering and Architecture, Zhejiang University, China
[b] China Energy Construction Group Guangdong Electric Power Design and Research Institute, Guangzhou, China
[c] Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, China
[d] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong
[e] KWR Water Research Institute, Nieuwegein, the Netherlands
[f] Centre for Water Systems, University of Exeter, Exeter, UK
[g] Department of Water Management, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Urban sewer network
Sensor placement
Inflow and infiltration
Joint entropy

ABSTRACT

The inflow and infiltration (I&I) is an issue for many urban sewer networks (USNs), which can significantly affect system functioning. Placing sensors within the USNs is a typical approach to detect large I&I event, but deploying a limited number of sensors while achieving maximum detection reliability is challenging. While some methods are available for sensor placement, they are generally heuristic search-based methods (HSBMs) and hence the resultant sensor placement strategies (SPSs) are variable over different algorithm runs or parameterizations. This paper develops a new deterministic two-stage clustering method for SPS optimization based on information entropy. Within the first stage, the Spectral Clustering method is applied to assign USN nodes to different clusters according to their joint entropy. In the second stage, the topology structure property is considered to enable further clustering for improving detection reliability. Average I&I detection reliability is used to select clusters and the optimal SPS is identified by maximizing joint entropy of all possible solutions where a single sensor is assigned to each selected cluster. The proposed method and two existing HSBMs are applied to a real USN and their performance is compared. The results obtained show that: (i) a strong correlation coefficient R ($R > 0.95$) is observed between joint entropy and SPS's detection reliability, which has not been revealed before, (ii) the proposed method consistently outperforms the other two approaches in efficiently offering SPSs with about 7–15 % higher detection reliability, and (iii) the proposed method provides the optimal SPS in a deterministic manner, which makes it attractive for engineering applications.

1. Introduction

An urban sewer network (USN) plays a crucial role in the conveyance of wastewater in urban cities, which is essential for public health and environmental protection (Jia et al., 2021a). The USNs are extensively present in city areas and are typically buried underground. However, due to the aging and a lack of timely maintenance, many pipes in the USNs are subject to significant structural deterioration, such as pipe corrosion or damage (Rokstad and Ugarelli, 2015). Consequently, infiltration and inflow (I&I) has become a widespread issue within many USNs (Cahoon and Hanke, 2019), where I&I refers to water in the surrounding environment entering the system through manholes, imperfect pipe joints and pipe cracks.

The I&I may present significant challenges to the structural integrity of the USN and pose a considerable threat to urban water security. These include that (i) I&I can reduce the effective capacity of the sewer system, making it less efficient at transporting wastewater to treatment facilities (Diem et al., 2022), (ii) excessive I&I can cause combined sewer overflows (CSOs), leading to untreated wastewater being discharged into the environment (Flood and Cahoon, 2011), (iii) I&I can increase the risk of basement flooding and surface flooding during heavy rain events (Jia et al., 2021a), (iv) The additional volume of water from I&I increases the

* Corresponding author.
E-mail addresses: wyl00@zju.edu.cn (Y. Wu), feifeizheng@zju.edu.cn (F. Zheng), yangyongfei@gedi.com.cn (Y. Yang), gdutwater_zkm@163.com (K. Zhang), dukun_cq@foxmail.com (K. Du), hf.duan@polyu.edu.hk (H. Duan), Dragan.Savic@kwrwater.nl (D. Savic), z.kapelan@tudelft.nl (Z. Kapelan).

https://doi.org/10.1016/j.watres.2024.122591
Received 8 August 2024; Received in revised form 7 September 2024; Accepted 6 October 2024
Available online 9 October 2024

load on wastewater treatment plants, leading to higher pumping and treatment costs (Ge et al., 2024), and (v) Overflows and flooding caused by I&I can result in environmental contamination and pose health risks to the public (Jia et al., 2021b). It is noted that I&I can be also caused by the misconnection of the rainwater pipes and sewer pipes, which often occurs in many developing countries such as China (Panasiuk et al., 2016). Under such circumstances, the flows in the sewer pipes can significantly increase due to large rainwater inflows through the misconnection point between these two different pipe types.

Typically, detecting I&I within the USN, especially the large I&I events, is crucial to ensure the protection of the urban water environment. This motivates many efforts to develop various methods to enable efficient and effective I&I detection. For instance, the dye and smoke testing method can roughly determine whether there is I&I in the USN by injecting a specific substance into a designated location within the USN and then observing the substance's leakage (Beheshti et al., 2015). Alternatively, robots can be placed into the pipe of the USNs to identify damaged points that can lead to I&I (Harris & Dobson, 2006). While these physical methods are straightforward and can be very useful to localize the I&I location within a particular pipe, they are often inefficient and expensive, hence they cannot be used to check I&I for the entire USNs (Ge et al., 2024).

In recent years, with the quick development of the Internet of Things, online sensors have been increasingly employed to monitor and detect I&I in USNs. These sensors are typically used to monitor hydraulic variables such as water level or flow rate with a high time resolution (Oliker and Ostfeld, 2015). The data collected can be processed using advanced analytics techniques, such as machine learning algorithms, to identify occurrence and existence of I&I in the USNs (Tanda et al., 2023; Wang et al., 2024). Compared to traditional physical methods, the use of sensors has the merit of high efficiency as their I&I analysis is conditioned on the data collected. In addition, due to the great need for real-time monitoring and detection of I&I in recent years, the use of online sensors is promising for engineering practice (Edmondson et al., 2018).

However, due to the high cost of sensor purchase and maintenance, it is impracticable to place a sensor at each node of the USN. Therefore, optimally placing a limited number of sensors with a maximum I&I detection reliability is an important research question (Rieckermann et al., 2010). This motivates efforts to develop various methods to optimally deploying sensors, ranging from the engineering experience based approaches (Banik et al., 2017a; Yazdi, 2018) to the heuristic search-based methods (HSBMs). For example, Alfonso et al. (2010) proposed a multi-objective genetic algorithm (GA) framework for optimizing sensor placement, where the joint entropy of the sensors is considered as the objective. Subsequently, Banik et al. (2017b) proposed a few GA-based methods to optimally locate sensors in USNs with various objectives considered including the joint entropy, total correlation, and detection reliability of the sensors. Similarly, Yazdi (2018) applied the differential evolution algorithm to find the best monitoring sites for the USNs, where the joint entropy of the sensors is again considered as the objective. More recently, Tomperi et al. (2023) developed a low-cost distance sensor to detect I&I for USNs, and Ge et al. (2024) proposed a data-driven model to estimate I&I based on temperature and conductivity monitoring, with a comprehensive review regarding I&I given in Zeydalinejad et al. (2024). It is acknowledged that these heuristic optimization methods are likely to identify optimal sensor placement strategies (SPSs), but their solutions are often variable across different optimization runs and the solution quality can be heavily dependent on algorithm parameterizations (Yang and Shami, 2020). These are two main reasons why practitioners are reluctant to apply these methods to deal with real problems (Acar et al., 2021).

Interestingly, within the above optimization approaches, the information entropy has been widely used as a metric to assess the performance of the sensor systems, where a larger information entropy value represents a better SPS. This is because information entropy has been tailored to facilitate the design of the sensor system for various networks, such as streamflow, precipitation, groundwater level and sewer networks (Keum and Coulibaly, 2017). More specifically, the information entropy is able to quantify the information content of different SPSs, and the solution with the maximum information entropy indicates a minimum shared information among sensors, which represents the optimal spatial distribution of sensors for abnormal event detection (e.g., I&I event detection, Alameddine et al., 2013; Keum et al., 2017). However, while the information entropy has been frequently used to enable the SPS optimization, the extent to which this metric can represent the SPS's detection performance (if say the detection reliability) has not been explicitly explored.

This paper proposes a novel deterministic optimization method for SPS optimization, where the information entropy is used as the assessment metric and a new two-stage clustering approach is developed to identify the optimal SPS. The most important feature of the proposed method is that it discards the use of the heuristic search based methods (HSBMs) as these produce SPSs in a stochastic manner. Instead, conditioned on the information entropy, the two-stage clustering is proposed to identify the optimal SPS deterministically and with a great efficiency. In other words, the proposed method offers a single optimal solution for a given sensor placement problem in an efficient manner, which can facilitate the decision-making process compared to many different available solutions provided by the HSBMs.

The main contributions of this study include: (i) the demonstration of the relationship between information entropy and SPS's detection performance, which has not been explicitly investigated before; (ii) the proposal of a new deterministic two-stage clustering method to determine the optimal SPS for I&I monitoring and detection with a great efficiency; (iii) a comprehensive performance comparison between the proposed method and the two heuristic search based methods (HSBMs) based on a real USN. It is noted that relatively large I&I is considered in this study as small I&I may not significantly affect the operation of the USNs.

2. Methodology

The proposed method involves the information content assessment for the USN, followed by the joint entropy estimation, based on which the first stage clustering is performed. Conditioned on the result of the first stage, the second stage clustering is carried out based on the topology structure property (i.e., the property of the upstream nodes and pipes of the sensors) of the USN, aimed to further improve the effectiveness of the SPS. Finally, a single sensor is assigned to a cluster with the location determined based on the maximization of the joint entropy from all sensors. The details of the proposed method are given below.

2.1. Information content measurement

Suppose there are N nodes in a USN, and the total number of I&I events is M. Consider each node i of the USN as the potential location for sensor placement. For node i, its information content R_i^m based on the event m is computed using Eq. (1).

$$R_i^m = \begin{cases} 1, \text{event } m \text{ is detected} \\ 0, \text{event } m \text{ is not detected} \end{cases} \tag{1}$$

As shown in this equation, $R_i^m = 1$ if event m is detected by the sensor located at node i. Detection can be determined by the change of the analyzed variable (e.g., flow, water depth or water quality parameters) over a pre-specified threshold. For example, for the I&I event considered, the threshold can be considered as 15 % above the flow records at the same time of day, where this threshold is often used in many water utilities in China.

For the total number of events M, an information content vector $\mathbf{NI}_i$ can be obtained using Eq. (2),

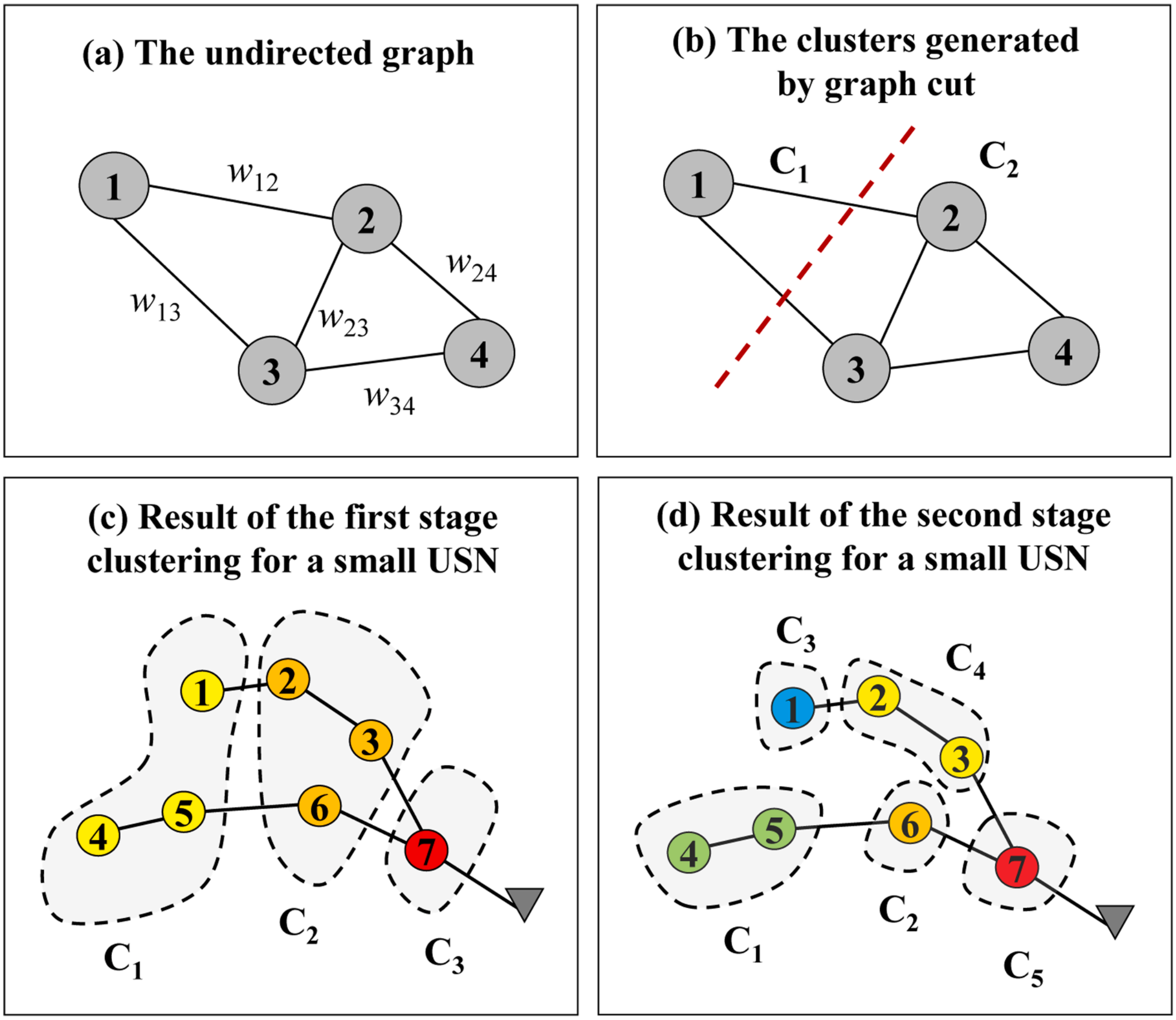

Fig. 1. The illustration of the of spectral clustering process.

$$\mathrm{NI}_i = \left[R_i^1, R_i^2, ..., R_i^M\right]^T \tag{2}$$

For the USN, the I&I events can occur at any pipe, but typically a pipe with a longer length has a higher likelihood of abnormal event occurrence, which should be considered in generating ***M*** events as did in this paper (discuss later).

2.2. Joint entropy computation

After obtaining the information content of each node i, an information matrix **IM** of the entire USN can be formed using Eq. (3), and this matrix is subsequently used to compute the joint entropy for each pair (i and j) of the nodes, as shown in Eq. (4).

$$\mathrm{IM} = \{\mathrm{NI}_1, \cdots, \mathrm{NI}_N\} = \begin{pmatrix} R_1^1 & \cdots & R_N^1 \\ \vdots & \ddots & \vdots \\ R_1^M & \cdots & R_N^M \end{pmatrix}_{M\times N} \tag{3}$$

$$H(\mathrm{NI}_i, \mathrm{NI}_j) = -\sum_{s=1}^{S}\sum_{k=1}^{K}\left\{p\left(R(\mathrm{NI}_i, s), R(\mathrm{NI}_j, k)\right)\log p\left(R(\mathrm{NI}_i, s), R(\mathrm{NI}_j, k)\right)\right\} \tag{4}$$

where $H(\mathrm{NI}_i, \mathrm{NI}_j)$ is joint entropy of the node i and j; $R(\mathrm{NI}_i, s)$ and $R(\mathrm{NI}_j, k)$ are the distinct events or records in NI_i and NI_j respectively; S and K are the number of distinct events or records in NI_i and NI_j respectively; $p\left(R(\mathrm{NI}_i, s), R(\mathrm{NI}_j, k)\right)$ is the probability that $R(\mathrm{NI}_i, s)$ and $R(\mathrm{NI}_j, k)$ occur simultaneously. For the given information content defined in Eqs. (1) and (2), S and K is 2 as R_i^m is a binary variable. In other words, $R(\mathrm{NI}_i, s)$=0 or 1.

If the joint entropy of two vectors $H(\mathrm{NI}_i, \mathrm{NI}_j)$ is high, it indicates that these two vectors offer distinct information with a high level, i.e., low redundant information. Conversely, a low value of $H(\mathrm{NI}_i, \mathrm{NI}_j)$ suggests that they contain redundant information, and hence they provide a relatively small amount of information. Therefore, the joint entropy between two vectors can offer crucial information about their correlation and hence it can be used to measure the similarity between nodal sensor locations in detecting abnormal events. Generally, a high joint entropy value of the sensors represents that these sensors are optimally distributed as they provide abundant information about event detection.

By calculating the joint entropy between every two nodes using Eq. (4), a joint entropy matrix **EM** can be constructed as shows in Eq. (5):

$$\mathrm{EM} = \begin{pmatrix} H(\mathrm{NI}_1, \mathrm{NI}_1) & \cdots & H(\mathrm{NI}_1, \mathrm{NI}_N) \\ \vdots & \ddots & \vdots \\ H(\mathrm{NI}_N, \mathrm{NI}_1) & \cdots & H(\mathrm{NI}_N, \mathrm{NI}_N) \end{pmatrix}_{N\times N} \tag{5}$$

EM in Eq. (5) is the joint entropy of each pair of nodes in the USN. If the number of sensors to be placed is two, the two nodes with the maximum $H(\mathrm{NI}_i, \mathrm{NI}_j)$ can be selected to place sensors. However, in engineering practice, the number of sensors (*NS*) is typically significantly larger than two, and hence **EM** cannot be directly used to identify the optimal SPS. It is noted that the joint entropy of the *NS* nodes can be theoretically estimated, but the computational overheads can be huge. For example, for a USN with $N = 100$ and the number of sensors to be placed NS=10, the total number of computing the joint entropy for each 10 nodes is about 1.7×10^{13}. This requires huge computational resources, making it impossible in engineering practices. While the heuristics methods can be available to enhance efficiency, their final results are uncertain over different algorithm runs and parameterization as previously stated. To address this issue, the two-stage clustering is proposed in this study.

2.3. The first stage clustering

The aim of the first stage clustering is to identify the clusters based on the given number of sensors to be placed, where all nodes within the same cluster have a low joint entropy. In other words, the nodes within the same cluster have overall similar information in detecting I&I events, and hence placing a single sensor at each cluster is sufficient to detect I&I events. The use of the clusters can significantly reduce the computational overhead in estimating the joint entropy for the given *NS* sensors compared to the direct joint entropy estimation for all possible *NS* combinations.

Within the first stage clustering, the Spectral Clustering method is used due to its excellent performance in dealing with high-dimensional data (Lucińska and Wierzchoń, 2012; Zhou et al., 2017). In Spectral Clustering, each sample (i.e. **NI** in Eq. (2)) is considered as a point in space. This is followed by the estimation of the similarity (i.e., weight) between each pair of nodes (w_{ij}). While many different methods are available to estimate similarity, Eq. (6) as shown below is a simple approach that has been widely used in literature (Chen Erjing, 2017). Based on the results of w_{ij} for all pairs of nodes, a similarity matrix **SM** can be constructed as shown in Eq. (7).

$$w_{ij} = \frac{1}{1 + H(\mathrm{NI}_i, \mathrm{NI}_j)} \tag{6}$$

$$\mathrm{SM} = \begin{pmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{pmatrix}_{N\times N} \tag{7}$$

For the Spectral Clustering method, the Graph Theory is used to enable the clustering (Di Nardo et al., 2017), where an undirected graph **G** is built based on w_{ij} for all pairs of nodes as illustrated in Fig. 1(a) with seven nodes. The Spectral Clustering method is typically used to construct the similarity between nodes according to physical properties of the pipe network (Di Nardo et al., 2017), which significantly differs to the present study that the joint entropy is used as a similarity measure to explore the I&I detection performance of nodes. As for the spectral clustering method, the principle is the same. This is followed by the definition of a loss function as a result of the graph cut into different clusters **C** as follows:

$$\min L(\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_g) = \min \sum\nolimits_{(i,j)\in \mathrm{CT}} w_{ij} \tag{8}$$

where $L(\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_g)$ is the loss function of the clusters (the total number of is *g*) that is formed from the original undirected graph **G; CT** is the set removed by the graph cut. For the example given in Fig. 1(a), when two subgraph is generated ($g = 2$) as illustrated in Fig. 1(b), $\mathbf{C_1}$={1}, $\mathbf{C_1}$={2,3,4}, **CT**={(1,2), (1,3)} and hence $L(\mathrm{C}_1, \mathrm{C}_2) = w_{1,2} + w_{1,3}$.

For the given number of sensors (*NS*) to be placed, the Spectral Clustering method aims to identify the equivalent number of clusters with a minimum loss function (Eq. (8)). More specifically, the N-cut graph cutting method can be used to minimize the loss function for a given *NS*, with details given in Ding et al. (2024). Fig. 1(c) illustrates the three clusters ($\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$) obtained using the Spectral Clustering method based on the information content vector **NI**, where the nodes at the same cluster have similar information content. When determining the locations of the three sensors, a single sensor is placed at each cluster. Consequently, based on the clustering results of the small USN in Fig. 1(c), a sensor should be placed at Node 7 as this single node is assigned to a cluster. Similarly, the second sensor can be deployed at Node 2, 3, or 6, and the third sensor can be placed at Node 1, 4, or 5.

2.4. The second stage clustering

The first stage clustering ensures the maximum joint entropy value between different clusters and the minimum joint entropy for the nodes within the same cluster. However, directly placing sensors at the clusters identified by the first stage clustering can be problematic due to the topology structure property (i.e., the property of the upstream nodes and pipes of the sensors) of the USN. For instance, in the example given above, the information content for Node 3 and 6 (Fig. 1c) are similar in numerical value and hence they are assigned to the same cluster, but their corresponding detection events are different. This is because Node 3 and 6 have different upstream nodes and these upstream nodes are not interconnected, indicating that these two nodes have no redundant information. This is not the case for Node 2 and 3 in the same cluster as they have identical upstream nodes. To solve this problem, the second stage clustering is proposed to further classify the nodes into different clusters based on their topology information.

Specifically, the nodes at the same cluster but located on a different tree structure after the use of the first stage clustering are separated into different clusters. Compared to the first stage clustering (Section 2.3) based on node information, the second stage clustering is carried out according to the topology property of the USN. Consequently, as illustrated in Fig. 1(d), Node 6 is assigned to another cluster and Nodes 2 and 3 are in the same cluster. This is the same for Node 1 as it is located on a different tree structure from Node 4 and 5. Note that the second stage clustering can increase the number of clusters (*NC*) compared to that from the first stage. For the small USN in Fig. 3(c,d), there are three clusters (the same as *NS*) identified in the first stage clustering, but it increases to 5 after the second stage clustering.

It can be deduced that the second stage clustering is able to ensure the low joint entropy of the nodes within each cluster in detecting I&I events. This is in contrast to the first stage clustering which aims to identify the nodes with a small joint entropy value based on nodal information content.

2.5. The determination of the optimal sensor placement strategy (SPS)

After the two-stage clustering, the number of clusters *NC* can be larger than the number of sensors *NS*. However, the fundamental principle of the proposed method is that only one single sensor is placed at a cluster, and hence it leads to a challenge in how to determine clusters for sensor placement. In this study, the average detection reliability (*ADR*) across all nodes at each cluster is defined to rank the clusters, enabling the selection of the *NS* clusters from the total *NC* clusters. For each node *i*, its DR_i represents the probability of the total number of events that can be detected if the sensor is placed at node *i*, i.e. as follows:

$$DR_i = \frac{1}{M}\sum_{m=1}^{M} R_i^m \tag{9}$$

The ADR_j for each cluster *j* is then computed as follows:

$$ADR_j = \frac{1}{S(\Omega_j)}\sum_{i\in\Omega_j}^{S(\Omega_j)} DR_i \tag{10}$$

where Ω_j is the set of nodes in the j^{th} cluster, and $S(\Omega_j)$ is the total number of nodes in the j^{th} cluster. The clusters can be ranked based on

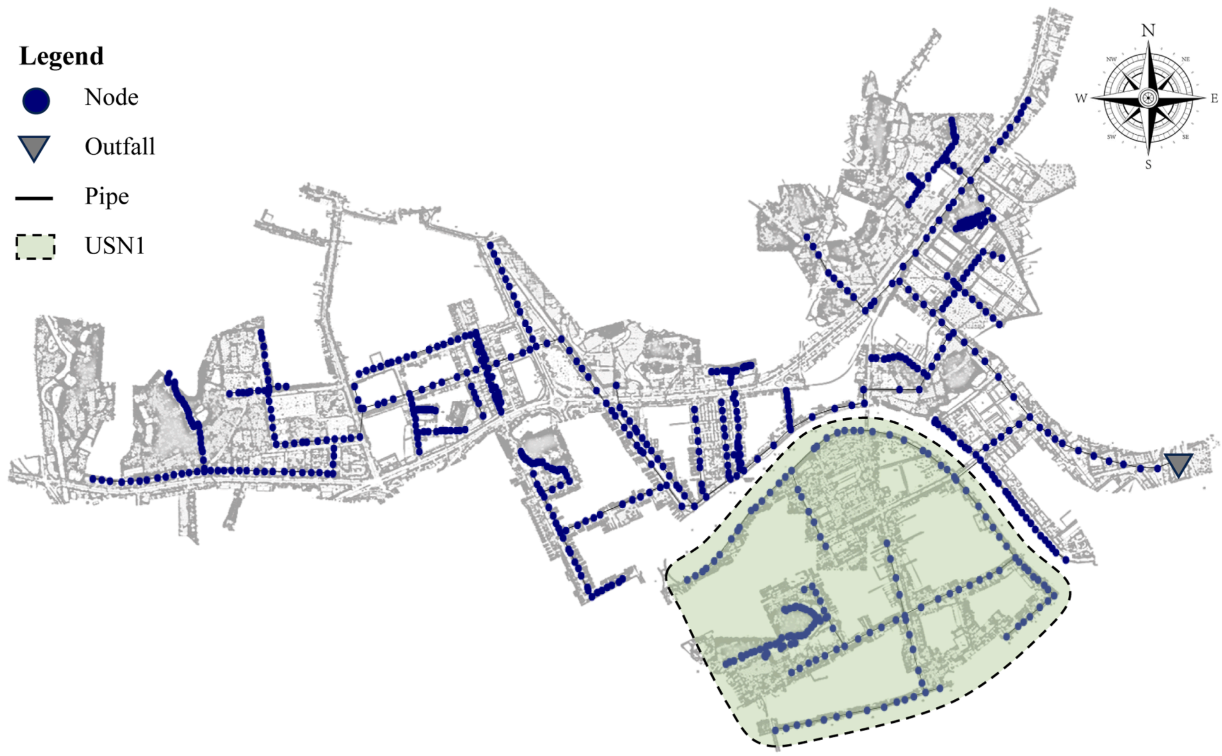

Fig. 2. The topology of the real USN case study.

Table 1
Information of the two case studies.

	The number of nodes	Daily flow DF (m^3/d)	Number. of events (M)	Q of the large I&I event
USN1	156	12,800	1000	3 %, 5 % and 10 % of the *DF*
FUSN	732	43,400	3000	1 %, 3 % and 5 % of the *DF*

the descending order of the *ADR* value, and the top *NS* clusters can be selected for sensor placement. For the small USN example in Fig. 1(d), the three clusters with the largest *ADR* values are $\mathbf{C}_5$, $\mathbf{C}_2$ and $\mathbf{C}_4$, and hence these are selected for placing three sensors.

Conditioned on the determined *NS* clusters, the joint entropy of all possible solutions are computed, where a single sensor is placed at each cluster. The final optimal SPS is taken as the solution with the largest joint entropy. While Eq. (4) shows the joint entropy for two nodes within the USN, it can be easily extended to compute the joint entropy for *NS* nodes, with details given in Li et al. (2012). The identified SPS with the largest joint entropy represents the comprehensive information of the sensors that can provide for detecting I&I events of the entire USN.

It is highlighted that the two-stage clustering method proposed in this study can significantly reduce the number of joint entropy evaluations. Taking the small USN in Fig. 1(d) as an example, it has 7 nodes and 3 sensors are to be placed. If one is to identify the optimal SPS by enumerating the joint entropy for all possible solutions, there would be 35 entropy evaluations required. However, the proposed method only needs to evaluate the joint entropy for two potential solutions in order to identify the optimal SPS as shown in Fig. 1(d). This indicates that the proposed method can be significantly more efficient in identifying the optimal SPS than the traditional approach without the clustering. More importantly, such computational benefit can be more prominent for real USNs with a large number of sensors to be placed, which makes the proposed method very attractive for engineering implementation. One would argue that the use of heuristic search based methods can also improve efficiency, but its underlying uncertainty in identifying optimal solutions is always a challenge for engineering practice. In contrast, the proposed method uses a deterministic approach, resulting in a single solution with no variability for a given sensor placement problem.

3. Case studies

3.1. Case study description

The real USN used in this study is located in southeast China and is shown in Fig. 2. The service area of this USN is about 15 km^2, and the total pipe length is about 60 km, with pipe diameters ranging from 0.3 m to 1.5 m. The daily total flow of this system is about 43,400 m^3/d. The USN is located in a city with many wet days and many rivers, and hence the I&I is an important issue. The presence of significant I&I is reflected in the low BOD concentrations of the flow entering the downstream wastewater treatment plant. The local water authority plans to deploy 20 flow meters for this USN, in order to monitor and detect the locations with large I&I. The purpose of this study is to optimally locate these sensors in the analyzed USN.

A hydraulic model is built for this USN with the aid of the SWMM software (Rossman, 2010). The model consists of 819 pipes, 820 nodes and one outlet located in the eastern part of the study area (see Fig. 2). The model has been calibrated by the local water authority and hence its calibration is not the focus of this study. To better explain the application of the proposed method, a small independent region is selected (the shaded region in Fig. 1) for illustration purposes. This small USN is referred to as USN1 with a daily flow (*DF*) of about 12,800 m^3/d, and the full USN is referred to as FUSN.

3.2. Simulation experiment design

A total of *M* I&I events are generated to enable the computation of the nodal information content (Eq. (1)), the joint entropy (Eq. (4)) and the detection reliability (Eq. (9)). For the USN1 and FUSN, *M* is 1000 and 3000 respectively. As previously mentioned, this study considers large

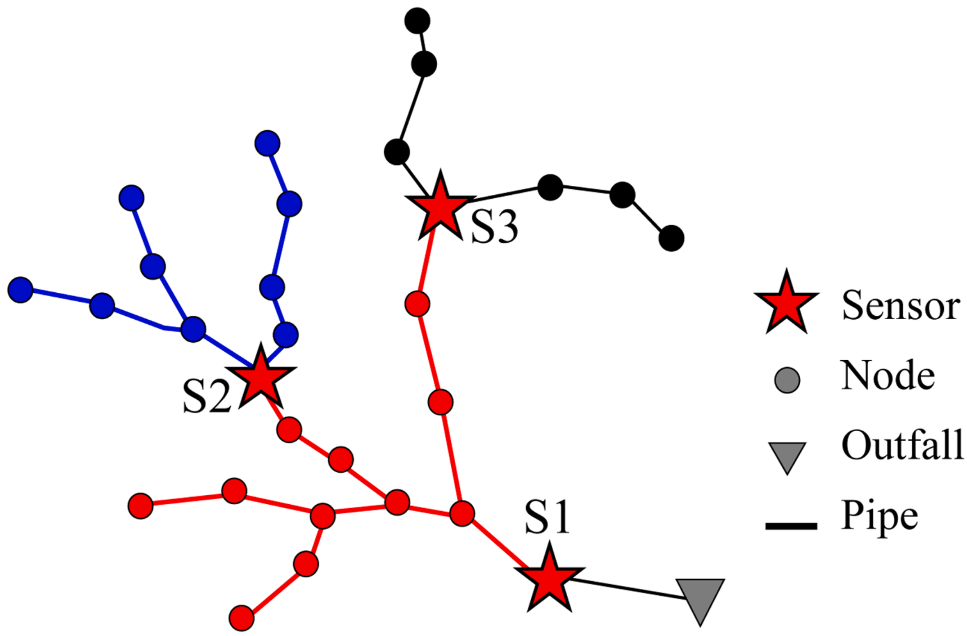

Fig. 3. The illustration of the pipes monitored by each sensor.

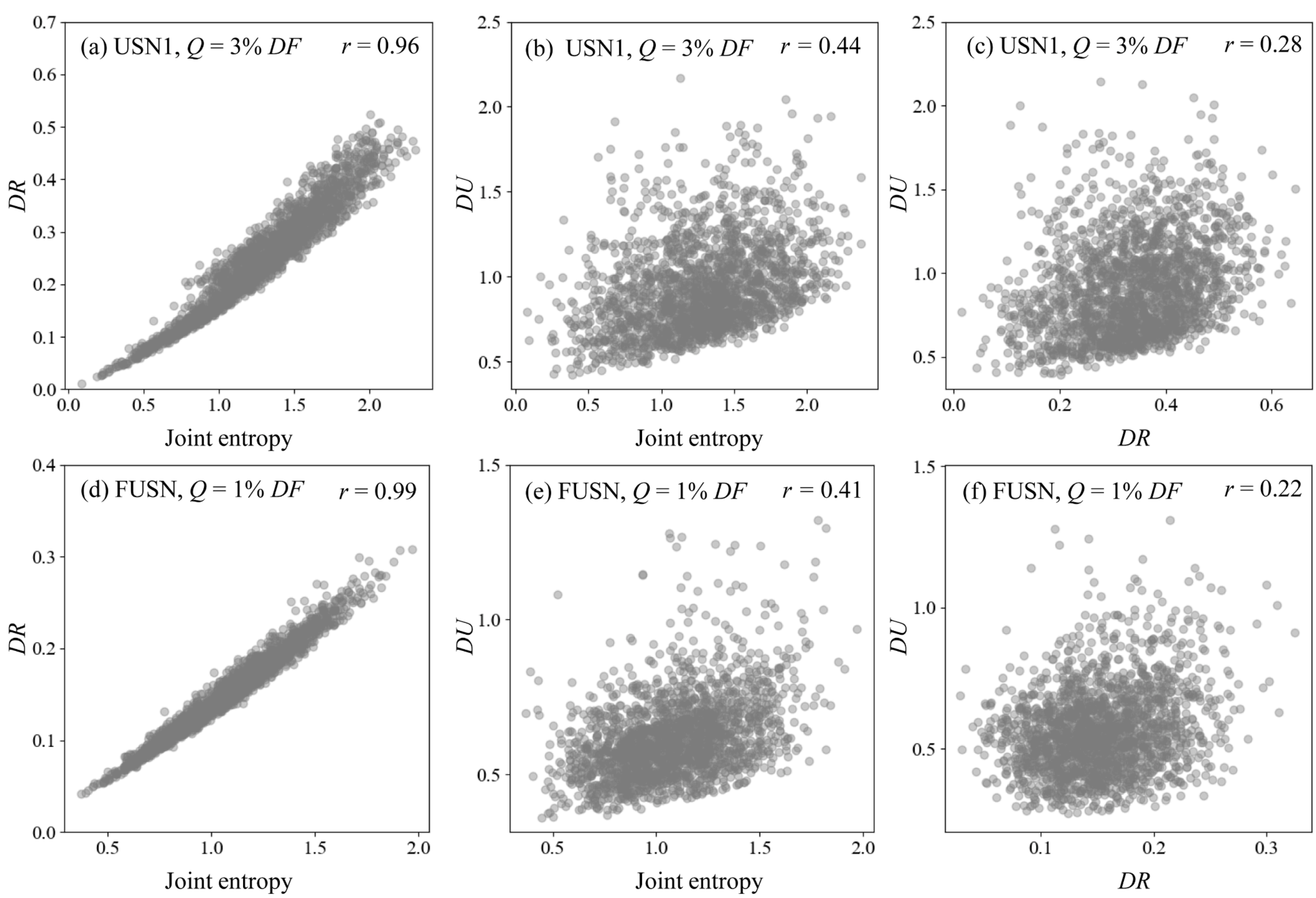

Fig. 4. The correlation between joint entropy, *DR* and *DU*, where *Q* is the I&I flow, *DF* is the total daily flow of the USN and *NS* is the number of sensors to be placed.

I&I events that can significantly affect the system safety (needs to take urgent actions), which differs to the small I&I that can occur simultaneously at multiple locations in the USN. Therefore, for each event, only a single pipe of the USN is injected with a flow *Q*, where the probability of each pipe being selected for I&I injection is based on its length relative to the total length of the USN.

The value of *Q* can be critical as it can affect the computation of the information content R_i^m in Eq. (1). In this study, a few different *Q* values

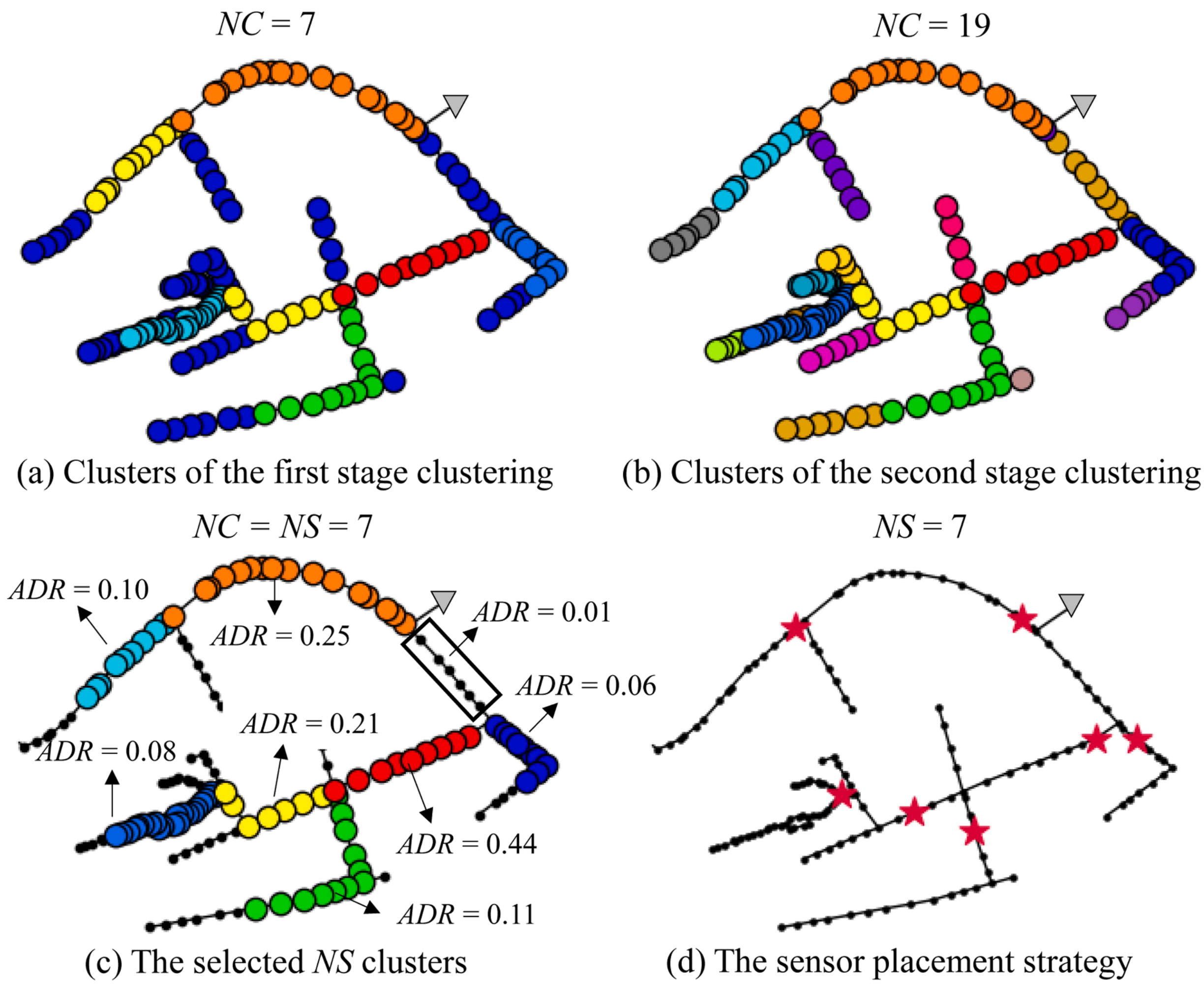

Fig. 5. The clusters and the final sensor placement strategy identified by the proposed method where the nodes with different colors are assigned to different clusters.

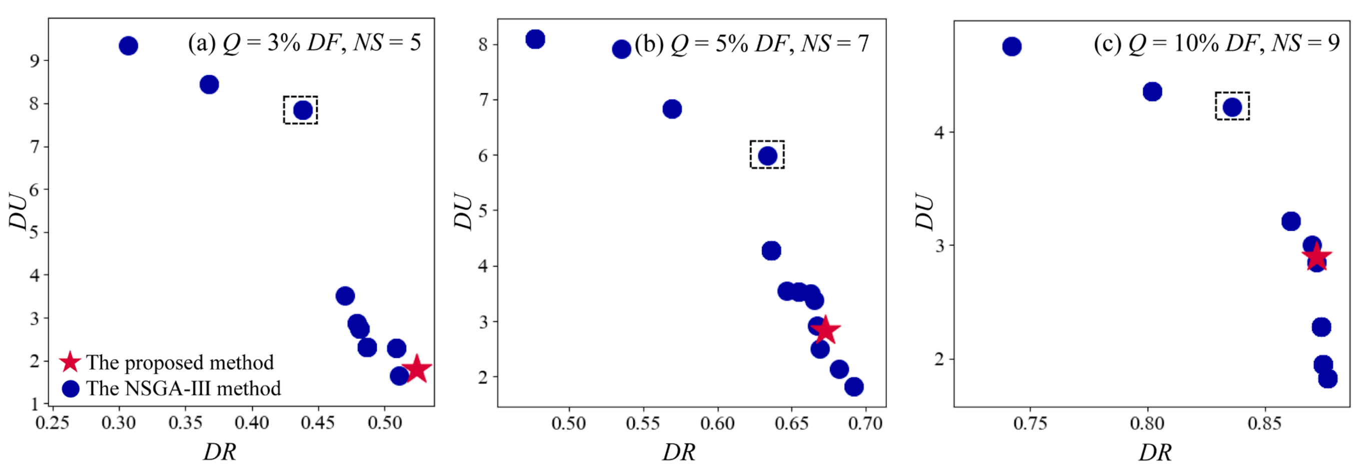

Fig. 6. Pareto fronts of the NSGA-III results applied to the USN1, where the *DR* and *DU* values of the optimal solution provided by the proposed method are also presented. The solution with dashed square will be further analyzed later.

are considered as given in Table 1, where this Q value is constant through the entire simulation time (24 h in this study). In this study, the timing of the Q is at the 8:00am in the morning, but the different timing of Q would not affect the results as Q is assumed to be constant within the entire event duration (24 h). In addition, the dry days are considered in this study. The threshold for I&I warning is set as the value above the daily flow record with a 15 % following Zhang et al. (2018), and this threshold is actually used for the local water utilities for I&I event warning.

It is noted that the parameter values mentioned above are generally selected based on the discussion with the practitioners in the local water authority. While the results of the case study used in this paper are conditioned on the parameters considered, the use of different parameters (e.g., I&I flow or threshold value) would not affect the application

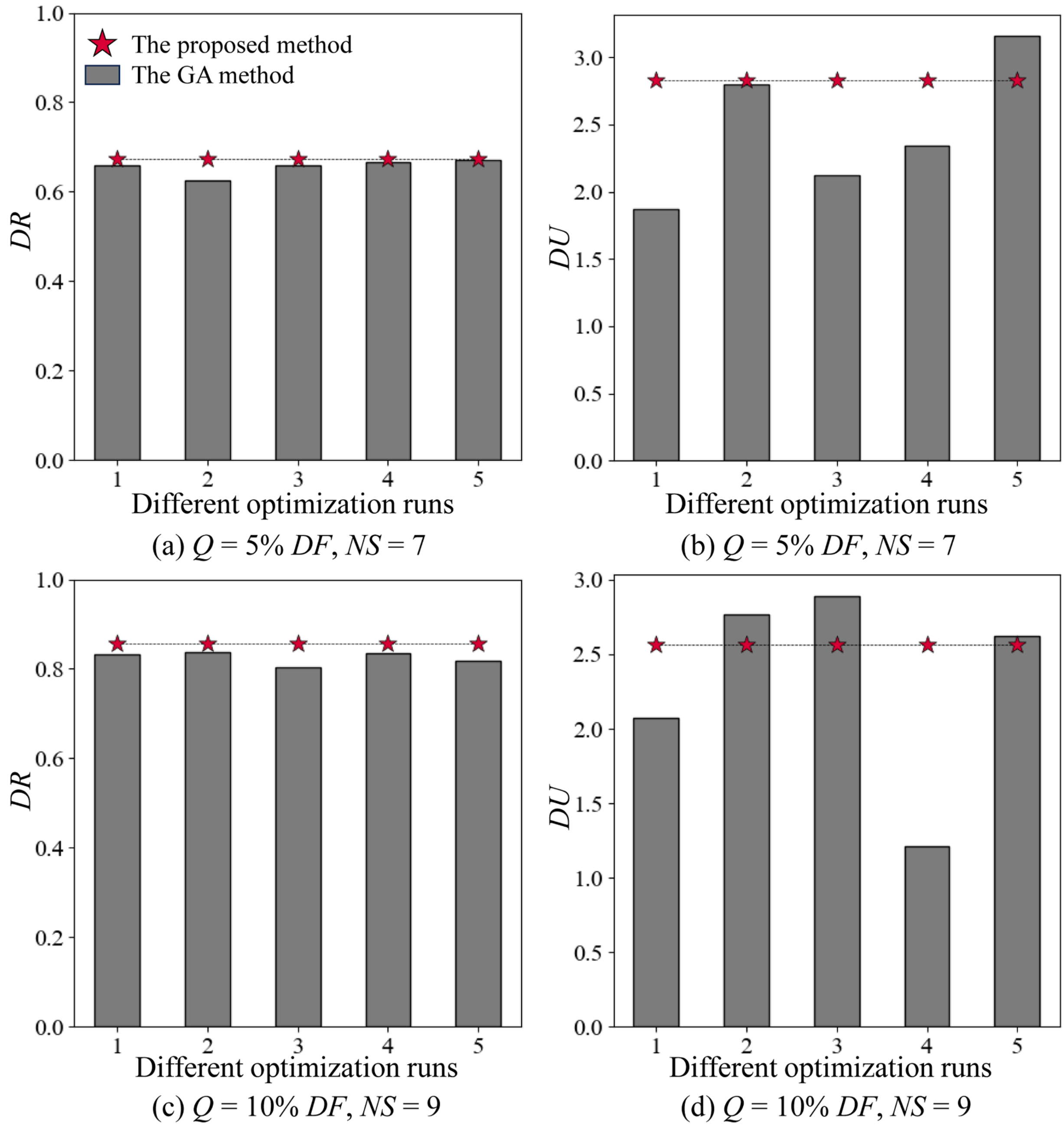

Fig. 7. Results of the GA method applied to the USN1, where the *DR* and *DU* values of the optimal solution provided by the proposed method are also presented.

of the proposed method. To enable performance comparison, two heuristic search based methods (HSBM) are also implemented in this study. These include a multi-objective optimization method using the non-dominated sorting genetic algorithm III (NSGA-III, Hu et al., 2020), and a typical single-objective GA method (Banik et al., 2016). The NSGA-III is an improved multi-objective optimization algorithm, which has been widely used in many research domains due to its great performance and robustness (Yadav et al., 2023).

3.3. Assessment metrics

To assess the effectiveness of the SPS, two quantitative metrics are utilized to enable performance evaluation. These are the I&I event detection reliability (*DR*) that has been widely used in literature (Banik et al., 2017b) and the distribution uniformity (*DU*) of the sensor location as proposed in this study. *DR* defined in Eq. (1) is slightly changed here when applied to the SPS with many sensors. More specifically, for a given SPS, *DR* indicates the proportion of *M* total events detected by any of the sensors in the SPS (rather than probability mentioned in Eq. (9)).

Typically, the *DR* is the primary objective for an SPS, but its spatial distribution property as measured by *DU* can be also important. This is because an SPS with a high *DU* value indicates that the length of the pipes associated with each sensor is overall similar. As illustrated in Fig. 3, the red, blue and black pipes are monitored by the sensors S1, S2 and S3 respectively. It is highlighted that while the I&I events that occurred at blue pipes may also be monitored by S1, these blue pipes are assigned to S2 in this study as these events are first monitored by S2. If the total pipe length of each sensor is overall similar, the effort required to pinpoint the exact location of the I&I event detected by each sensor is overall similar, representing a reasonable expected effort in localizing I&I.

In this study, *DU* can be approximated by the coefficient of variation

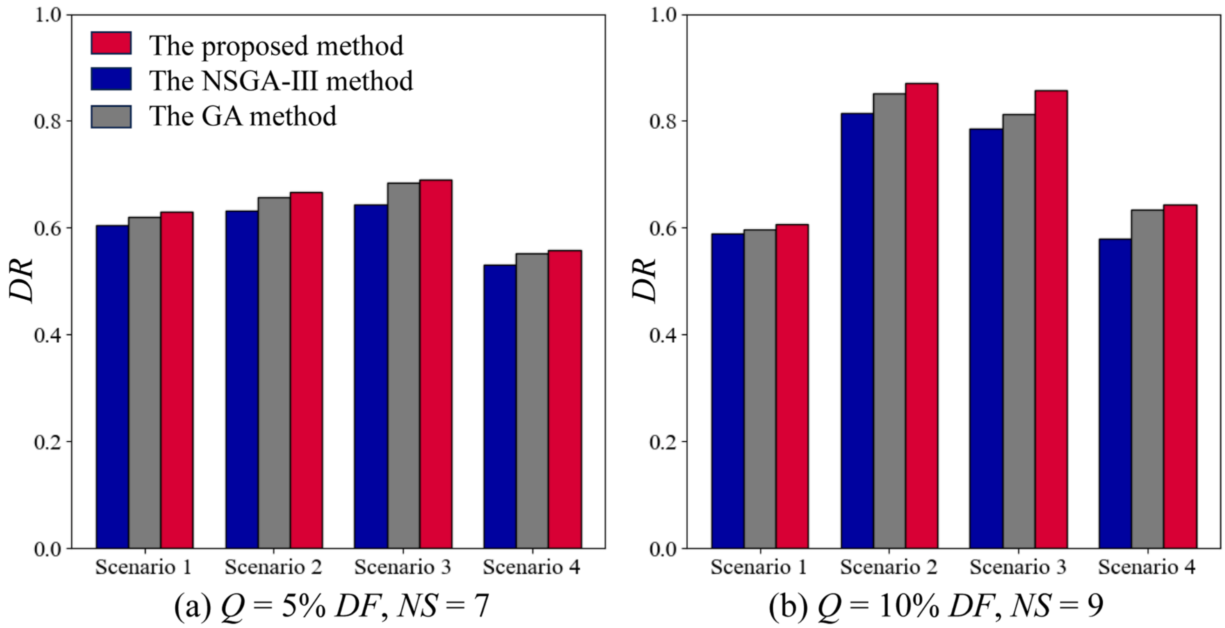

Fig. 8. *DR* results of the optimal sensor placement strategies identified by the three methods applied to the USN1 under different I&I flow (*Q*) scenarios.

(*CV*) of the total pipe length monitored by all different sensors. More specifically, the *CV* is computed as the variance of different total pipe lengths associated with different sensors divided by the mean of these different total pipe lengths. To better present the results, the *DU* is defined as the inversed value of the coefficient of variation (*CV*), and consequently, a larger *DU* indicates an overall greater spatial uniformity in sensor placement. Consequently, the *DU* is defined as the following,

$$DU = \frac{1}{CV} = \frac{\overline{L_S}}{\sqrt{\frac{1}{NS}\sum_{S=1}^{NS}\left(L_S - \overline{L_S}\right)^2}} \tag{11}$$

where L_S is the total length of pipes monitored by sensor *S*, and $\overline{L_S}$ is the average of L_S of all sensors.

4. Results and discussion

4.1. Analysis of correlations between the joint entropy, DR and DU

As previously stated, while joint entropy has been widely used as an indicator to assess the detection performance of the SPSs, its correlation with the detection reliability (*DR*) and distribution uniformity (*DU*) has not been investigated before. To this end, this paper attempts to address this gap. For the two case studies considered, a total of 2000 random solutions are generated for a given number of sensors *NS* ranging from 3 to 20. The joint entropy, the *DR* and the *DU* are computed for the corresponding solutions, with results given in Fig. 4. It is noted that other results with different I&I *Q* values and *NS* values are similar to those in Fig. 4 and hence they are not shown.

As shown in Fig. 4, a strong correlation is consistently observed between the joint entropy and the *DR*, with the correlation coefficient *R* value greater than 0.95. This suggests that the SPS with a high joint entropy value possesses greater detection reliability. In terms of the *DU* metric, a moderate correlation (*R* around 0.4) is observed between the joint entropy and the distribution uniformity. This indicates that while the joint entropy has accounted well for the *DR*, it cannot fully represent the distribution uniformity of the SPS. However, a weak correlation is noted for the *DR* and *DU* as shown in Fig. 4, suggesting that these two metrics need to be simultaneously considered when optimizing the SPSs. The main implication of these results is that the joint entropy is effective in representing the *DR* for the SPS, but it is insufficient to indicate the *DU* of the sensors. This implies that the direction optimization of joint entropy is unable to produce the optimal SPS with the great *DR* and *DU* values, and hence the two-stage clustering is proposed in this study to improve the solution quality in a deterministic manner.

4.2. Effectiveness of the proposed two-stage clustering method

Fig. 5 shows the clusters and the final SPS identified by the proposed method applied to the USN1 when seven sensors are to be placed (i.e., *NS*=7). As shown in this figure, a total of seven clusters (NS=7) are identified based on the joint entropy value with the aid of the spectral clustering method as described in Section 2.3 (the first stage clustering). However, the nodes at different tree structures are assigned to the same cluster based on the joint entropy value as shown in Fig. 5(a), which is not reasonable as these nodes respond to different I&I events. Therefore, the proposed second stage clustering is applied to further assign the nodes at the same cluster but with different tree structures to different clusters, leading to a total of 19 clusters (Fig. 5(b)).

Given that *NS*=7, these 19 clusters are ranked based on their *ADR* value, and the top seven clusters are subsequently identified, where the *ADR* values are given in Fig. 5(c). Finally, the optimal SPS is determined by selecting the combination with the largest joint entropy with a single sensor placed at each selected cluster (see Section 2.5). It is noted that some downstream nodes of the USN have a low *ADR* value compared to the upstream nodes (*ADR*=0.01 for the nodes surrounded by the black line in Fig. 5c). This is because the pipe flows of these nodes are rather large (not shown) and hence the sensors at these nodes cannot detect I&I events.

For this small USN1 with seven sensors to be placed, 156 nodes are available as potential sensor locations. If the joint entropy is computed by enumeration for all these nodes in order to find the seven sensor locations with the maximum joint entropy, the total number of evaluations is 3.9×10^{11}. This requires massive computational resources that are typically unavailable in practice. However, the total evaluation of the joint entropy for the sensors located in the final seven clusters is about 1.5×10^{7}. This implies that the proposed two stage clustering can substantially reduce the computational time.

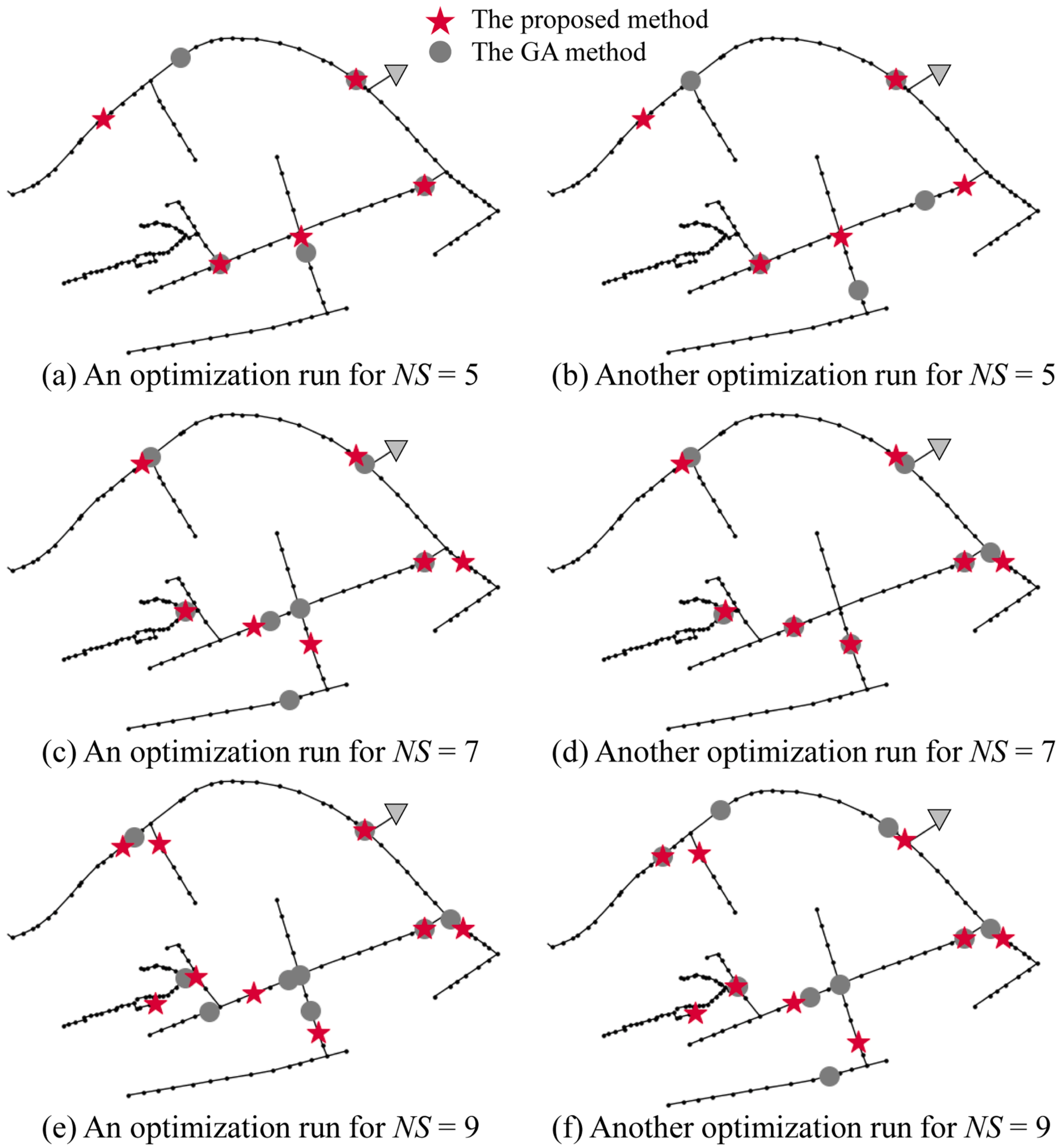

Fig. 9. Sensor locations of the optimal placement strategies identified by the proposed and GA methods applied to the USN1.

4.3. Comparison analysis between different methods applied to the USN1

4.3.1. Comparison between the proposed method and the NSGA-III method

Since the *DR* and *DU* are weakly correlated, it is natural to optimize these two metrics simultaneously. Hence, the NSGA-III is used to optimize the sensor placement strategy for the two case studies with the maximization of *DR* and *DU* considered as the two optimization objectives (denoted as the NSGA-III method). Similarly, since the joint entropy is effective in representing the *DR*, one may apply the single objective optimization method to the problem where the joint entropy is considered as the sole objective. Therefore, a typical GA is used to find the SPS with the maximization of the joint entropy value as the objective function for the two case studies (denoted as the GA method). It is noted the parameters of the NSGA-III and GA are all taken from Banik et al. (2016). More specifically, the population sizes of the NSGA-III and the GA are all 200 and the maximum allowable number of generations are 200. The single-point crossover with a rate of 0.9 and the single-point mutation with a rate of 0.1 are used for both algorithms. While many other advanced heuristic search based methods (HSBMs) are available, the GAs has gained the great popularity in engineering practice (Banik et al., 2016) and hence they are selected to enable the performance comparison.

Fig. 6 shows the Pareto fronts of the NSGA-III method applied to the USN1, and the *DR* and *DU* values of the solution provided by the proposed method are also presented to enable comparison. It is noted that many solutions are identical when a large number of sensors (if say *NS*=9) are located and hence the number of solutions in the Pareto front is reduced as shown in Fig. 6. As shown in this figure, the proposed method is consistently able to produce optimal SPSs that are not dominated by those from the NSGA-III method across different I&I *Q* and

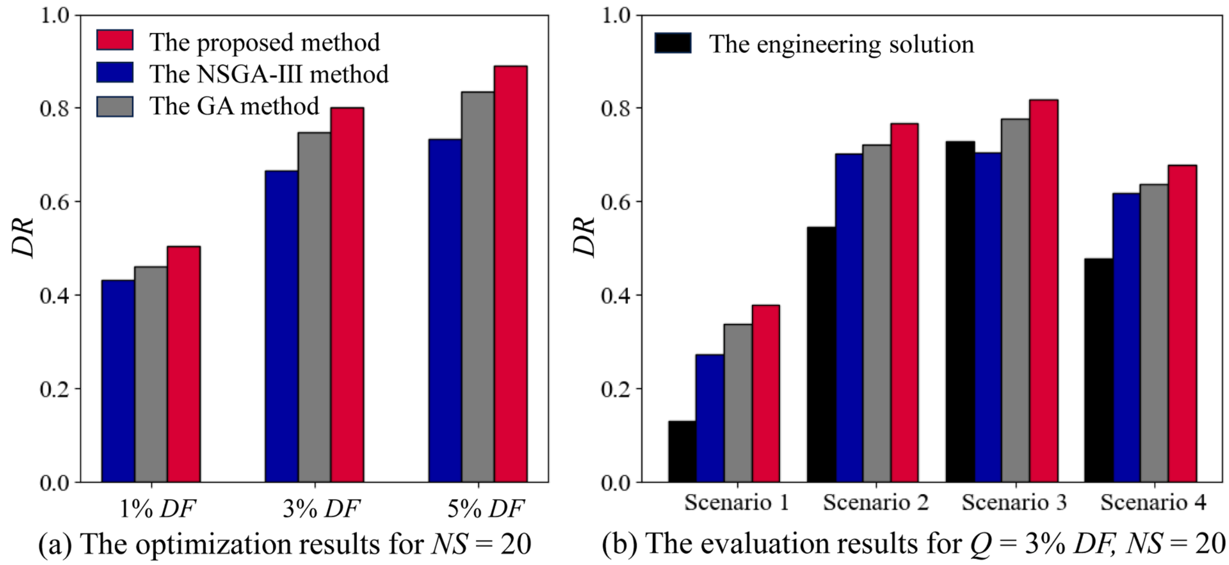

Fig. 10. *DR* results of the optimal sensor placement strategies identified by the four methods applied to the FUSN under different I&I flow (*Q*) scenarios.

NS values. This is because, within the proposed method, the entropy information considered is able to maximize the joint entropy of the sensors, and the two-stage clustering is capable of ensuring the distribution uniformity of these sensors.

More specifically, for $Q = 3$ %DF and *NS*=5, the *DR* and *DU* values of optimal SPS identified by the proposed method are 0.53 and 1.92 respectively. These two values are rather similar to its neighboring solution (*DR*=0.53, *DU*=1.89) in the Pareto front generated by the NSGA-III method. In addition, the solutions produced by the proposed method can be better than the NSGA-III when a large *NS* is considered. This is because the search space can be significantly enlarged for the NSGA-III when the *NS* is large, leading to challenge to identify optimal solutions. It should be highlighted that the proposed method identifies only a single solution for the given problem, whereas the Pareto front generated by the NSGA-III provides a trade-off between the *DR* and *DU*. Theoretically, the latter possesses the benefit as it provides many optimal solutions for analysis, but in engineering practice a single and deterministic solution can enable a straightforward and easy decision-making process.

4.3.2. Comparison between the proposed method and the GA method

Fig. 7 presents the optimization results of the GA method applied to the USN1, where results of five different runs with different random number seeds are presented. It is seen from this figure that the proposed method is overall able to offer SPSs with a greater *DR* value than those from the GA method which takes the joint entropy as the objective. However, in terms of the *DU*, while the GA method may produce solutions with a relatively better distribution uniformity than the proposed method, it has a large variability as some GA solutions can have rather low *DU* values. This highlights the benefit of the deterministic property of the proposed method in contrast to the high variability associated with the GA results.

4.3.3. Robustness comparison between different methods

Since the I&I flow *Q* is constant to enable the SPS optimization, it is necessary to assess the robustness of the final solutions for different *Q* values of I&I events. To this end, four different scenarios are considered in this study, where 500 random I&I events are generated with $Q = 3$ % of *DF*, $Q = 5$ % of *DF*, $Q = 10$ % of *DF*, and $Q = 0 \sim 15$ % *DF* for Scenario 1, 2, 3 and 4 respectively. The method for generating the random I&I events is described in Section 3.2. For Scenario 4, the value of *Q* is randomly selected between 0 and 15 % *DF* for each event. It is highlighted again that large I&I events are considered in this study and hence the flow changes caused by the many small I&I can be added to *DF* as the base flow for analysis.

Fig. 8 shows the robustness testing results (*DR* values) for the SPSs determined by $Q = 5$ % *DF* (*NS*=7) and $Q = 10$ % (*NS*=9) *DF* using the three methods. We have also conducted the robustness analysis for other *Q* values and *NS* values, but the results are not presented as they are similar to Fig. 8. Since multiple solutions are available for the NSGA-III method, the solution representing a great trade-off between *DR* and *DU* is selected for robustness analysis, as shown in Fig. 6 (the solutions are surrounded by the black line).

As can be seen from Fig. 8, the proposed method consistently exhibits the best performance in detecting the I&I event with different *Q* magnitudes. This implies that the two-stage clustering conditioned on the joint entropy is an effective manner to enable the SPS optimization for USNs. It is noted that the NSGA-III method shows the worst performance, which can be caused by its relatively low convergence quality as the searching space associated with a multi-objective optimization framework can be significantly larger than the single GA method.

4.3.4. Sensor location comparison between different methods

As previously stated, the main benefit of the proposed method is that it can produce a single constant SPS for a given problem as it is a deterministic method. Fig. 9 shows the sensor locations of the optimal solutions identified by the proposed method and the GA method for various *NS* values. Clearly, the sensor locations specified by the GA method can be significantly varied across different optimization runs, which is especially the case when a large *NS* is considered. Similar observations can be made for the NSGA-III results and hence they are not represented. The uncertainty in sensor locations associated with the heuristics methods (the NSGA-III or GA method) can seriously hinder their adoption in engineering practice, further highlighting the advantage of the proposed method.

Within the engineering application, some nodes may not be accessible in installing sensors. This can be easily considered in the proposed method as these sensors can be excluded within the clustering process or the final process in determining the optimal sensor locations.

4.4. Applications to the large FUSN

The proposed method, the NSGA-III method and the GA method are all applied to the large FUSN with 20 sensors (*NS*=20) considered. In addition, the engineering experience-based solution provided by the local water authority is also included for comparison. Fig. 10(a) shows

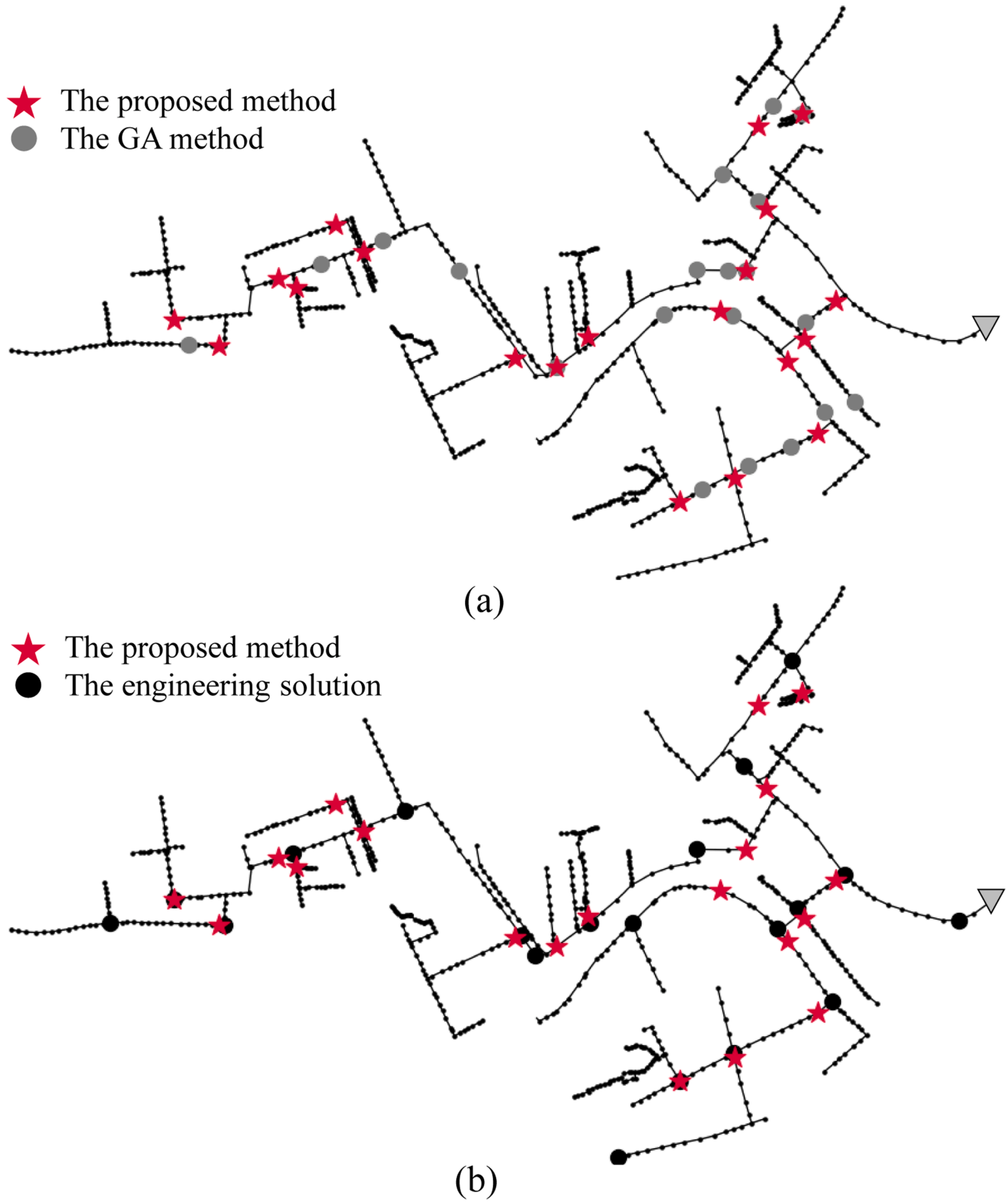

Fig. 11. Sensor locations of the optimal placement strategies identified by the proposed method, the GA method and the engineering solution for the FUSN.

the *DR* results of the three optimization methods with different *Q* values. As shown in this figure, the proposed method is able to find solutions with larger *DR* values compared to the NSGA-III and GA approaches across different I&I flow values (*Q*).

Fig. 10(b) shows the robustness tests for the solutions with $Q = 3\ \%\ DF$, where 1000 random I&I events are generated with $Q = 1\ \%\ DF$, $Q = 3\ \%\ DF$, $Q = 5\ \%\ DF$ and $Q = 0{\sim}10\ \%\ DF$ for Scenario 1, 2, 3 and 4 respectively. As can be seen from this figure, the proposed method significantly outperforms the engineering solution in *DR* under various *Q* values. This implies that the use of the optimization method is highly important for determining the optimal SPS that can maximize the detection ability of these sensors. Within the optimization approaches, the proposed method also shows the best performance, which is the same as for the USN1.

The locations of the 20 sensors of the optimal solutions identified by the proposed method and the GA method are given in Fig. 11(a). As seen from this figure, some sensors are located very close to each other for the GA method, but this is not the case for the proposed method. This is because the two-stage clustering is able to ensure the overall distribution uniformity of the sensors while maintaining the high detection

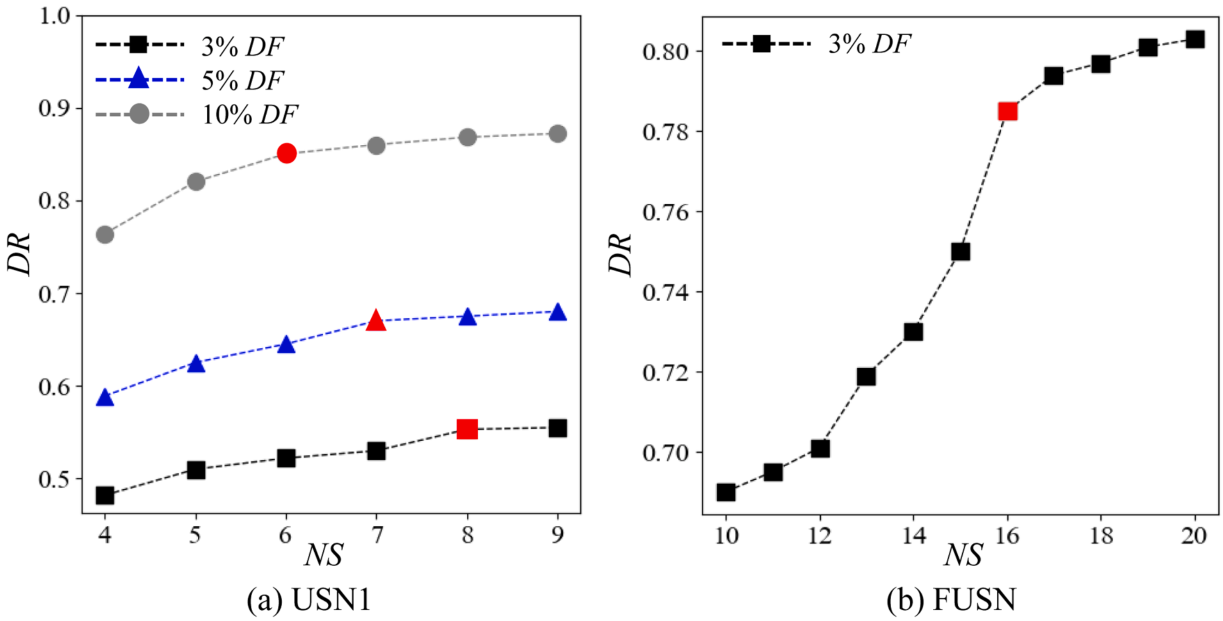

Fig. 12. *DR* versus *NS* for the USN1 and FUSN, where the red points represent the good trade-off between *DR* and *NS*.

Table 2
Comparison of computational efficiency between the proposed method and the HSBMs.

	M	NS	Number of model simulations (The proposed method)	Number of model simulations (the HSBMs)
USN1	1000	9	1.8×10^3	4×10^7
FUSN	3000	20	6.5×10^6	1.2×10^8

reliability of these sensors. It is seen from Fig. 11(b) that some sensors of the proposed method are located close to those provided by the engineering solution provided by the water utility engineers. It is noted that engineering solutions may be able to consider many additional criteria such as the accessibility of locations and location of critical customer, but these can be easily accounted for in the proposed method. This implies that the proposed method is capable of recognizing the critical locations of the USN such as the nodes with relatively high overflow likelihood. In addition, the proposed method can also mitigate the drawback of the engineering solution as such deploying sensors at the very upstream or placing sensors too close.

4.5. Impacts of sensor number on detection performance

While the results presented above are conditioned on the given number of sensors, the proposed method can be easily extended to identify the optimal number of sensors for a given USN. This can be achieved by running the proposed method multiple times with different number of sensors, followed by the identification of the optimal number based on trade-off analysis between the detection reliability (*DR*) and the number of sensors (*NS*). Fig. 12 shows the trade-offs between *DR* and *NS* for the two case studies considered. While the *DR* improves as the *NS* increases, a different *DR* enhancement degree is observed as shown in Fig. 12, representing different marginal improvement. For the USN1, the optimal number of sensors can be 8, 7 and 6 (red color in Fig. 12a) when the I&I inflow is 3 % *DF*, 5 % *DF* and 10 % *DF* respectively. For the large FUSN, *NS*=16 represents a good trade-off.

4.6. Computational analysis

All the codes of this study are run on a PC with AMD Ryzen 7 5800H CPU @ 3.20 GHz. To enable a fair comparison, the total time required for each method has been converted to the equivalent number of model evaluations, with details given in Table 2. For USN1 and FUSN, the total running time of the proposed method to identify the optimal SPS is equivalent to 800 simulations and 6.5×10^6 model evaluations respectively. For the NSGA-III and GA methods, the total number of model evaluations are 4×10^7 and 1.2×10^8 for the USN1 and FUSN respectively. This indicates that the computational overheads used by the proposed method are only 0.0045 % and 5.4 % of those used by the HSBMs for the USN1 and FUSN respectively. In addition, the final solutions provided by the proposed method are overall better than those offered by the HSBMs.

5. Summary and conclusions

The inflow and infiltration (I&I) in urban sewer networks (USN) can cause negative impacts on the normal operation of the system. How to effectively place sensors to monitor and detect I&I has been crucial and essential for effective USN management. This motivates the development of many sensor placement strategy (SPS) optimization methods, which are mainly based on stochastic heuristics algorithms. In contrast, this paper proposes a deterministic two-clustering method for SPS optimization based on the information entropy. Within the proposed method, the nodal information content in detecting I&I and the topology property are simultaneously considered to improve the effectiveness of the SPS. The proposed method as well as two heuristics algorithms (a multi-objective NSGA-III and a single objective GA) are applied to a real USN and their performance is compared under a wide range of I&I scenarios.

The findings and implications of this study are outlined below:

(i) The joint entropy exhibits a strong correlation with the SPS's I&I event detection reliability. This has to our best knowledge been explicitly established for the first time. This implies that the use of the joint entropy is effective for SPS optimization in USNs.

(ii) The proposed method works well and it consistently shows better performance and robustness in identifying optimal SPSs than heuristics-based search algorithms as measured by the I&I event detection reliability and distribution uniformity of sensor location. The main feature of the proposed method is that it can provide a single optimal SPS for a given problem as it is a deterministic method. This makes the proposed method promising for real-life engineering applications.

(iii) In addition to the deterministic nature of the proposed method, its great efficiency in identifying the optimal SPS is another merit for practical application, especially when large USNs are considered.

(iv) While I&I is considered in this study, the proposed method can be easily extended to optimally deploy sensors for other types of issues in the USN, such as pollution detection.

A few assumptions are made in this study to enable the demonstration of the proposed method. These include (i) a constant I&I flow Q is considered in this study, which may differ to the real situations that the Q can be variable over time, (ii) the impact of the rainwater is not considered as the proposed method is demonstrated in the dry days, and (iii) the small and widely distributed I&I are considered as the base flow to enable the large I&I detection. Future work should address these assumptions.

CRediT authorship contribution statement

Yuling Wu: Investigation, Formal analysis, Data curation. **Feifei Zheng:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Yongfei Yang:** Funding acquisition. **Kaiming Zhang:** Funding acquisition. **Kun Du:** Writing – review & editing. **Huanfeng Duan:** Writing – review & editing. **Dragan Savic:** Writing – review & editing. **Zoran Kapelan:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No. 2023YFC3208905, 2023YFC3208900), National Natural Science Foundation of China (NSFC, 52179080, 52261160379) and Yunnan Fundamental Research Projects (grant No. 202401AT070392). All data and codes used in this paper will be available on request to the corresponding author.

Data availability

Data will be made available on request.

References

Acar, E., Bayrak, G., Jung, Y., Lee, I., Ramu, P., Ravichandran, S.S., 2021. Modeling, analysis, and optimization under uncertainties: a review. Struct. Multidiscipl. Optimiz. 64 (5), 2909–2945. https://doi.org/10.1007/s00158-021-03026-7.

Alameddine, I., Karmakar, S., Qian, S.S., Paerl, H.W., Reckhow, K.H., 2013. Optimizing an estuarine water quality monitoring program through an entropy-based hierarchical spatiotemporal Bayesian framework. Water Resour. Res. 49 (10), 6933–6945. https://doi.org/10.1002/wrcr.20509.

Alfonso, L., Lobbrecht, A., Price, R., 2010. Optimization of water level monitoring network in polder systems using information theory. Water Resour. Res. 46 (12). https://doi.org/10.1029/2009WR008953.

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

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

Banik, B.K., Di Cristo, C., Leopardi, A., de Marinis, G., 2016. Illicit intrusion characterization in sewer systems. Urban Water J. 14 (4), 416–426. https://doi.org/10.1080/1573062X.2016.1176220.

Beheshti, M., Sægrov, S., Ugarelli, R. (2015). Infiltration/inflow assessment and detection in urban sewer system. 24–34. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2598406.

Cahoon, L.B., Hanke, M.H., 2019. Inflow and infiltration in coastal wastewater collection systems: effects of rainfall, temperature, and sea level. Water Environ. Res. 91 (4), 322–331. https://doi.org/10.1002/wer.1036.

Chen Erjing, J.E., 2017. Review of studies on text similarity measures. Data Anal. Knowl. Discov. 1 (6), 1–11. https://doi.org/10.11925/infotech.2096-3467.2017.06.01.

Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R., Santonastaso, G.F., 2017. Weighted spectral clustering for water distribution network partitioning. Applied Network Science 2 (1). https://doi.org/10.1007/s41109-017-0033-4.

Diem, J.E., Pangle, L.A., Milligan, R.A., Adams, E.A., 2022. How much water is stolen by sewers? Estimating watershed-level inflow and infiltration throughout a metropolitan area. J. Hydrol. 614, 128629. https://doi.org/10.1016/j.jhydrol.2022.128629.

Ding, L., Li, C., Jin, D., 2024. Survey of spectral clustering based on graph theory. Pattern Recognition 151, 110366. https://doi.org/10.1016/j.patcog.2024.110366.

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

Flood, J.F., Cahoon, L.B., 2011. Risks to coastal wastewater collection systems from sea-level rise and climate change. J. Coast. Res. 274, 652–660. https://doi.org/10.2112/JCOASTRES-D-10-00129.1.

Ge, J., Li, J., Qiu, R., 2024. A data-driven method for estimating sewer inflow and infiltration based on temperature and conductivity monitoring. Water Res. 261, 122002. https://doi.org/10.1016/j.watres.2024.122002.

Harris, R.J., Dobson, C., 2006. Sewer pipe infiltration assessment: comparison of electro-scan, joint pressure testing, and CCTV inspection. In: Pipelines 2006: Service to the Owner, pp. 1–10. https://doi.org/10.1061/40854(211)61.

Hu, C., Dai, L., Yan, X., 2020. Modified NSGA-III for sensor placement in water distribution system. Inf. Sci. (Ny) 509, 488–500. https://doi.org/10.1016/j.ins.2018.06.055.

Jia, Y., Zheng, F., Maier, H.R., Ostfeld, A., Creaco, E., Savic, D., Langeveld, J., Kapelan, Z., 2021a. Water quality modeling in sewer networks: review and future research directions. Water Res. 202, 117419. https://doi.org/10.1016/j.watres.2021.117419.

Jia, Y., Zheng, F., Zhang, Q., Duan, H.-F., Savic, D., Kapelan, Z., 2021b. Foul sewer model development using geotagged information and smart water meter data. Water Res. 204, 117594. https://doi.org/10.1016/j.watres.2021.117594.

Keum, J., Coulibaly, P., 2017. Information theory-based decision support system for integrated design of multivariable hydrometric networks. Water Resour. Res. 53 (7), 6239–6259. https://doi.org/10.1002/2016WR019981.

Keum, J., Kornelsen, K.C., Leach, J.M., Coulibaly, P., 2017. Entropy applications to water monitoring network design: a review. Entropy 19 (11). https://doi.org/10.3390/e19110613. Article 11.

Li, C., Singh, V.P., Mishra, A.K., 2012. Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy. Water Resour. Res. 48 (5). https://doi.org/10.1029/2011WR011251.

Lucińska, M., & Wierzchoń, S.T. (2012). Spectral Clustering Based On K-Nearest Neighbor Graph (A. Cortesi, N. Chaki, K. Saeed, & S. Wierzchoń, Eds.; Vol. 7564, pp. 254–265). Springer Berlin Heidelberg. 10.1007/978-3-642-33260-9_22.

Oliker, N., Ostfeld, A., 2015. Network hydraulics inclusion in water quality event detection using multiple sensor stations data. Water Res. 80, 47–58. https://doi.org/10.1016/j.watres.2015.04.036.

Panasiuk, O., Hedström, A., Ashley, R.M., Viklander, M., 2016. Detection of wastewater discharges into stormwater sewers: effects of travel distance on parameters. J. Environ. Eng. 142 (5), 04016016. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001086.

Rieckermann, J., Borsuk, M.E., Sydler, D., Gujer, W., Reichert, P., 2010. Bayesian experimental design of tracer studies to monitor wastewater leakage from sewer networks. Water Resour. Res. 46 (8). https://doi.org/10.1029/2009WR008630.

Rokstad, M.M., Ugarelli, R.M., 2015. Evaluating the role of deterioration models for condition assessment of sewers. J. Hydroinform. 17 (5), 789–804. https://doi.org/10.2166/hydro.2015.122.

Rossman, L.A., 2010. Storm Water Management Model User's Manual. US Environmental Protection Agency, Version 5.0.

Tanda, M.G., D'Oria, M., Secci, D., & Todaro, V. (2023). Identification of the inflow source in a foul sewer system through techniques of inverse modelling. 2444(1). 10.1088/1742-6596/2444/1/012011.

Tomperi, J., Rossi, P.M., Ruusunen, M., 2023. Estimation of wastewater flowrate in a gravitational sewer line based on a low-cost distance sensor. Water Pract. Technol. 18 (1), 40–52. https://doi.org/10.2166/wpt.2022.171.

Wang, Y., Huang, B., Zhu, D.Z., 2024. Assessment of rainfall-derived inflow and infiltration in sewer systems with machine learning approaches. Water Sci. Technol. https://doi.org/10.2166/wst.2024.115 wst2024115.

Yadav, D., Ramu, P., Deb, K., 2023. Finding robust solutions for many-objective optimization using NSGA-III. In: 2023 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. https://doi.org/10.1109/CEC53210.2023.10254160.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Yazdi, J., 2018. Water quality monitoring network design for urban drainage systems, an entropy method. Urban Water J. 15 (3), 227–233. https://doi.org/10.1080/1573062X.2018.1424215.

Zeydalinejad, N., Javadi, A.A., Webber, J.L., 2024. Global perspectives on groundwater infiltration to sewer networks: a threat to urban sustainability. Water Res. 262, 122098. https://doi.org/10.1016/j.watres.2024.122098.

Zhang, Q., Zheng, F., Duan, H.-F., Jia, Y., Zhang, T., Guo, X., 2018. Efficient numerical approach for simultaneous calibration of pipe roughness coefficients and nodal demands for water distribution systems. Journal of Water Resources Planning and Management 144 (10), 04018063. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000986.

Zhou, Z., Jia, X., Zhao, X., 2017. A new constraint spectral clustering algorithm. In: 2017 29th Chinese Control And Decision Conference (CCDC), pp. 6664–6668. https://doi.org/10.1109/CCDC.2017.7978376, 2017 29th Chinese Control And Decision Conference (CCDC).