Foul sewer model development using geotagged information and smart water meter data

Jia, Yueyi; Zheng, Feifei; Zhang, Qingzhou; Duan, Huan Feng; Savic, Dragan; Kapelan, Zoran

DOI
10.1016/j.watres.2021.117594

Publication date
2021

Document Version
Final published version

Published in
Water Research

Citation (APA)
Jia, Y., Zheng, F., Zhang, Q., Duan, H. F., Savic, D., & Kapelan, Z. (2021). Foul sewer model development using geotagged information and smart water meter data. *Water Research*, *204*, Article 117594. https://doi.org/10.1016/j.watres.2021.117594

Water Research 204 (2021) 117594

Contents lists available at ScienceDirect

Water Research

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

Foul sewer model development using geotagged information and smart water meter data

Yueyi Jia [a], Feifei Zheng [b,*], Qingzhou Zhang [c], Huan-Feng Duan [d], Dragan Savic [e,f,h], Zoran Kapelan [g,h]

[a] College of Civil Engineering and Architecture, Zhejiang University, China
[b] College of Civil Engineering and Architecture, Zhejiang University, A501 Anzhong Building, Zijingang Campus, 866 Yuhangtang Rd, Hangzhou 310058, China
[c] School of Civil Engineering and Mechanics, Yanshan University, Qinhuangdao 066004, China
[d] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon 999077, Hong Kong
[e] Chief Executive Officer, KWR Water Research Institute, Netherlands
[f] Distinguished Professor, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Malaysia
[g] Department of Water Management, Delft University of Technology, the Netherland
[h] Centre for Water Systems, University of Exeter, North Park Road, Exeter, EX4 4QF, United Kingdom

ARTICLE INFO

Keywords:
Foul sewer system (FSS)
Hydraulic models
Geotagged data
Smart water meter
Uncertainty

ABSTRACT

Hydraulic modeling of a foul sewer system (FSS) enables a better understanding of the behavior of the system and its effective management. However, there is generally a lack of sufficient field measurement data for FSS model development due to the low number of *in-situ* sensors for data collection. To this end, this study proposes a new method to develop FSS models based on geotagged information and water consumption data from smart water meters that are readily available. Within the proposed method, each sewer manhole is firstly associated with a particular population whose size is estimated from geotagged data. Subsequently, a two-stage optimization framework is developed to identify daily time-series inflows for each manhole based on physical connections between manholes and population as well as sewer sensor observations. Finally, a new uncertainty analysis method is developed by mapping the probability distributions of water consumption captured by smart meters to the stochastic variations of wastewater discharges. Two real-world FSSs are used to demonstrate the effectiveness of the proposed method. Results show that the proposed method can significantly outperform the traditional FSS model development approach in accurately simulating the values and uncertainty ranges of FSS hydraulic variables (manhole water depths and sewer flows). The proposed method is promising due to the easy availability of geotagged information as well as water consumption data from smart water meters in near future.

1. Introduction

As a result of population growth and rapid urbanization, spatial scales and structural complexities (e.g., the number of pipes, pumps and weirs) of many foul sewer systems (FSSs) have substantially increased over the past few decades (Rokstad and Ugarelli, 2015). These physical changes combined with system ageing result in a number of challenges for FSS management or operation (Sweetapple et al., 2018). Typical issues include pipe blockages (Montes et al., 2020), manhole overflows (Liu et al., 2016), odor problems (Talaiekhozani et al., 2016), illicit inflows (e.g., toxic discharges from local factories, rainwater infiltration, groundwater intrusion (McCall et al. 2016), and sewer exfiltration (Lepot et al. 2016; Beheshti and Saegrov 2018)). These issues can either directly induce serious contamination to the surrounding water environments (Lepot et al., 2016; Beheshti and Saegrov, 2018), or cause functional failures of wastewater treatment plants and consequently result in significant contamination of the receiving water body (McCall et al., 2016). Therefore, an efficient and effective management strategy for the FSS is vital to the urban environment safety as well as sustainable development of the society (Bailey et al., 2019).

One promising approach to enable effective FSS management is through hydraulic modeling (See et al., 2009; Draude et al., 2019). Typically, simulations of the FSS hydraulic variables (water depth and flows) can be compared with the *in-situ* observations, thereby identifying anomalies when observed water depths differ significantly from the simulation results (Ahm et al., 2016; Bailey et al., 2019). However,

* Corresponding author.
E-mail address: feifeizheng@zju.edu.cn (F. Zheng).

https://doi.org/10.1016/j.watres.2021.117594
Received 28 May 2021; Received in revised form 18 August 2021; Accepted 19 August 2021
Available online 25 August 2021

ensuring the high performance of an FSS hydraulic model is not a trivial task. This is because manhole inflow data, i.e., dry weather flows (DWFs), is typically unavailable (Breinholt et al., 2013). In addition, the true manhole inflow is a result of an inherently stochastic process that can be affected by many external conditions (e.g., temperature, user behavior, Abdel-Aal et al. 2015) and hence it is difficult to simulate. To this end, this study aims to investigate the challenge of accurately simulating the FSS hydraulics including the underlying stochastic properties.

Regarding the manhole inflow data, a number of different methods have been developed to estimate dry weather flows (DWF) for FSS models. These include the domestic appliance usage survey methods (Butler et al., 1995; Almeida et al., 1999), various empirical prediction models (Carstensen et al., 1998; Bechmann et al., 1999; Langergraber et al., 2008; Rodríguez et al., 2013) and the time-series sewer generation approaches (Mannina et al., 2009; De Keyser et al., 2010). These studies have also recognized that there are sources of variability that cannot be represented entirely deterministically and that adding a stochastic component to the model is beneficial (Almeida et al., 1999; Rodríguez et al., 2013). While these DWF methods have made contributions in developing FSS hydraulic models, their practical applications are restricted due to large efforts and insufficient data accuracy associated with these approaches (Bailey et al., 2019).

In recent years, a widely used approach is to calibrate the FSS model to estimate manhole inflows (i.e., DWFs) based on limited in-sewer observations (Korving and Clemens, 2005). Currently, the majority of the calibration algorithms aim to identify the inflows for each manhole at each particular time of the day, which is kept the same across different days (Bailey et al., 2019). Such a calibration approach is referred to as *static or offline calibration*. The approach is based on an engineering assumption that inflows at each manhole at a particular time period (say 6:00 am–6:30 am) are similar across different days (Bailey et al., 2019). This, however, neglects the stochastic nature and variability associated with these inflows. More importantly, the static calibration results often exhibit the so-called “equifinality” problem (Khu et al., 2006). This refers to a situation where many manhole inflow combinations produce a similar agreement between simulated and observed water levels or sewer flows at monitoring locations. As a result, it is very difficult, if not impossible, to identify a unique parameter set (i.e., a manhole inflow combination) that represents the true underlying temporal and spatial distribution of manhole inflows. The “equifinality” issue can significantly hamper practical application of FSS models due to model performance suffering at locations without sensors and also under different sewer discharge scenarios (Zhang et al., 2021).

To address the “equifinality” problem, some domain knowledge can be incorporated into the calibration process. For example, the length of sewer pipes or the contributing area can be used as prior knowledge for manhole inflow calibration (Maurer et al., 2013). This is because, typically, a long pipe or a large contributing area often collects a relatively large amount of wastewater. While these heuristics can improve the quality of the static calibration and partially alleviate the “equifinality” problem, the resulting model may not match the real situation in a sewer system. For example, some long sewer pipes may be only used to transport wastewater collected in upstream regions. In that case, manhole inflows are rather low because the house/commercial building density around these pipes is rather low. Conversely, some short pipes may receive a large amount of wastewater discharged from surrounding regions with a high population density. Therefore, the use of pipe length or the contributing area as the domain knowledge for FSS calibration may not be able to identify the true inflows into the manholes. Another heuristic is the use of the pipe diameter size since an increase in pipe diameter at a given location may indicate larger local sewer flows. However, it is also not ideal as a pipe in the downstream not only collects the sewer discharges from its local resident buildings, but also delivers sewer flows that are from its upstream pipes. Therefore, there is no direct relationship between the pipe size and the amount of the local sewer inflows. More recently, Zhang et al. (2021) developed an FSS model using a high density of real-time water consumption data, but this approach is not ideal for practical application as many water utilities have a relatively low number of smart water meters (mainly for large water users, e.g., factories, hospitals or schools).

Relative to the studies focused on the static FSS modeling, investigations on the stochastic properties of the manhole inflow data (i.e., DWFs) are rare. Some previous studies have assumed a particular distribution function, e.g., Uniform distribution, Gaussian distribution or Poisson distribution (Jin and Mukherjee, 2010; Sun et al., 2014) to describe the stochastic process of water consumption. However, their effectiveness with applications to FSS models has not been demonstrated. More importantly, the parameters of the specified distributions (e.g., $\pm$ 15% around the expected value) are mainly assumed subjectively, and hence may not be realistic. Therefore, there is still a need of an effective uncertainty analysis method to describe the underlying variation of the expected manhole inflows.

The objective of this study is to propose a novel FSS modeling method that can accurately simulate manhole inflows and their underlying uncertainty ranges. This goal is achieved with the aid of geotagged information and smart water meter data. More specifically, in the proposed method, the population information is derived based on the geotagged data (e.g., building area and height) taken from public databases. This information is used as prior knowledge to facilitate the static calibration of inflows for each manhole. The rationale behind this is that the population density can better indicate the inflow magnitudes at manholes when compared to the pipe length previously considered. In addition, uncertainty ranges associated with manhole inflows are derived from the stochastic properties of water consumption data from smart water meters. The idea behind this uncertainty analysis approach is that: (i) a given number of smart water meters that record water consumption in a near real-time manner (say every 30 min, Creaco et al. 2018) can be used to derive stochastic properties of the water consumption, and (ii) stochastic characteristics of manhole inflows can be derived from water consumption properties due to the intrinsic relationship between the water consumption and wastewater discharge in the same area.

The main contributions and novelties of this study include (i) the use of geotagged information from public databases to estimate the FSS manhole inflows, which can greatly improve the simulation accuracy and address the problems of “equifinality”, and (ii) the use of water consumption data from smart water meters to accurately characterize uncertainty associated with manhole inflows. To our best knowledge, this is the first work where the geotagged information and water consumption data are used to improve the accuracy of FSS hydraulic modeling.

This paper is organized as follows. The proposed methodology is described in Section 2, followed by the descriptions of the case studies considered in Section 3. Results and discussions are given in Section 4. Finally, the conclusion section (Section 5) shows the main findings and implications of this paper.

2. Methodology

Fig. 1 illustrates the overall framework of the proposed methodology, which involves three phases of FSS model development as well as the demonstration of the method on real-world case studies. Phase 1 aims to estimate the population size associated with each sewer manhole based on geotagged data. In this phase, the geotagged data from public databases are used to build physical relationship between each FSS manhole and its surrounding buildings, with details given in Section 2.1. This is followed by the estimate of population size based on the established relationship between each manhole in the FSS and the associated buildings, as described in Section 2.1. In Phase 2, the daily pattern of the inflows (i.e., DWFs) for each manhole is identified using a two stage optimization approach applied to the FSS subsystems partitioned by the

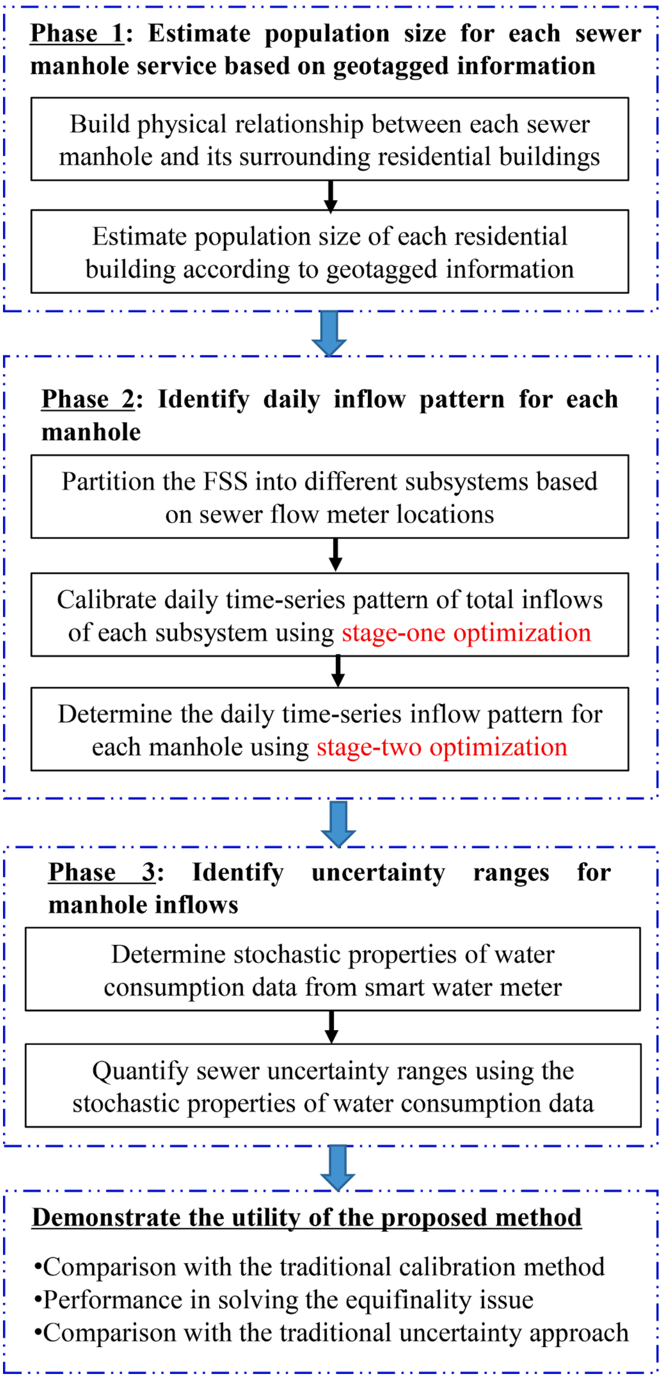


Fig. 1. The overall framework of the proposed method.

sewer flow meter locations (Section 2.2). Phase 3 focuses on the uncertainty analysis of manhole inflows (Section 2.3). In this phase, stochastic properties of water consumption are derived using data from smart water meters deployed in the water distribution system (WDS) that is overlapping with the FSS. The stochastic properties of water consumption data are then used to quantify the uncertainty ranges for sewer manhole inflows (Section 2.3). The utility of the proposed method is demonstrated through two real case studies. The performance of the proposed method is compared with traditional calibration and uncertainty analysis methods in accurately estimating hydraulic variables.

2.1. Estimate population size for each sewer manhole based on geotagged data

For a manhole receiving residential wastewater, the population data associated with this manhole is an important indicator of inflows. However, it is usually difficult to obtain accurate population data for a particular area or an individual building level due to unknown occupancy rates and population mobility. In addition, privacy issues may also limit the availability of population mobility data in some areas. To this end, the proposed method uses maps taken from publicly available databases, such as Google Earth, OpenStreetMaps, Bing Maps (Zheng et al., 2018). These map databases often possess comprehensive geotagged data as illustrated in Fig. 2(a), which in this study are employed to estimate the population size associated with each manhole.

Typically, the density of residential buildings and the heights of these buildings can reflect the population size of an area, as illustrated in Fig. 2(a). Accordingly, the population size can represent an important indicator of the magnitude of dry-weather wastewater flows, thus providing a link between the building information and sewer manhole inflows (Sitzenfrei et al., 2010). The specific information of each building includes the building height and width, representing the number of floors and the number of households at each floor, respectively. This information can be obtained from geotagged data within the public databases. In addition, the occupancy of the building also needs to be accounted for in order to estimate the population size.

In addition to the residential buildings, the sewers from commercial buildings or public buildings (e.g., hospitals or schools) also need to be considered when developing the FSS hydraulic models. Typically, sensors (e.g., smart water sensors) are deployed to monitor the water consumption or discharges for these large water users in a near real-time (as illustrated in Fig. 2(a)). Therefore, the manhole inflows associated with these buildings can usually be directly acquired from local water utilities (Zhang et al., 2021). Prior to the population size estimate, it is necessary to build a physical connection between each manhole and the surrounding buildings. This physical connection represents that the discharges of these buildings are received by this manhole, with details given below.

2.1.1. Build the physical connection between each manhole and its surrounding buildings

In this study, the physical connection between a building (can be a residential, commercial or public building) and a manhole is determined based on their Euclidean distance. The rationale behind this is that the discharges of a building are most likely to flow to its nearest manhole. The Euclidean distance between the building and the manhole can be estimated using the following equation

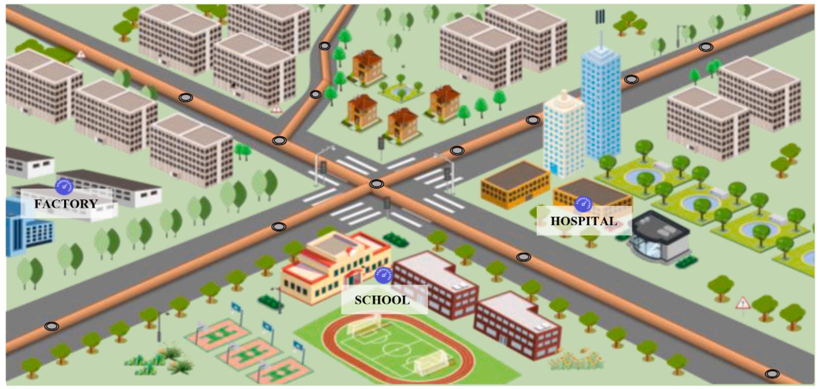
$$d(r,h) = \sqrt{(x_h - x_r)^2 + (y_h - y_r)^2 + (z_h - z_r)^2} \quad (1)$$

where (x_r,y_r,z_r) is the three-dimensional coordinate of the geometric center at the base of the building r and (x_h,y_h,z_h) are the coordinates of the manhole h. All these coordinates are available in the geotagged data of the public map databases. Consequently, for a given building r, its associated manhole can be identified by

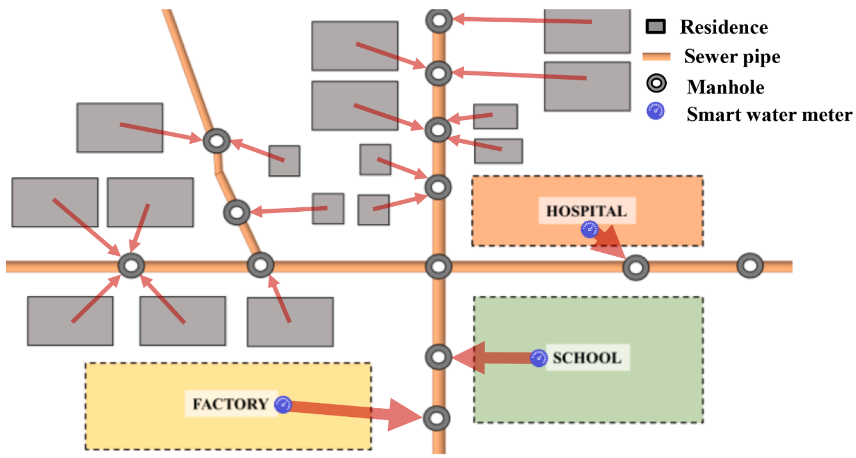
$$h(r) = \underset{h=1,2,\ldots,H}{\text{argmin}}\{d(r,h)\} \quad (2)$$

where $h(r)$ represents the hth manhole assigned to rth building; H is the total number of manholes in the FSS model.

Using Eqs. (1) and (2), the physical connections between the buildings and the manholes are established as shown in Fig. 2(b). For a real FSS, a single manhole is very likely to physically connect multiple buildings, especially when the buildings are small in size, as shown in Fig. 2(b). In a real FSS, there also might exist multiple manholes that potentially drain wastewater from a single building, which is often the case for large buildings. For this case, it is necessary to identify the



(a) Illustration of the building and sewer pipe information in a map



(b) Physical connections between buildings and manholes

Fig. 2. The conceptual figure of the proposed method to build physical connections between buildings and manholes.

proportion division of total discharges from a building across different surrounding sewer manholes, which is often difficult. For the sake of simplicity, only one manhole is assigned to a building in this study even though the fact is that multiple manholes are jointly used to deliver discharges of this building. It is acknowledged that such an assumption may lead to possible unrealistic hydraulic behavior in the local region of the FSS, but its influence on the hydraulic results of the entire FSS is negligible (Zhang et al. 2021).

2.1.1. Estimate population size of each residential building based on geotagged data

While it is ideal to have detailed population information for each building to enable FSS modeling, gaining such data is challenging and also time-consuming. Therefore, two assumptions are made in this study to estimate the population size of each residential building, as shown below.

(i) *Assumption 1: The population size is linearly correlated with the volume of the residential building.* This assumption is practically reasonable as a residential building with a relatively large area and height is often associated with a large population size.
(ii) *Assumption 2: All the residential buildings are fully occupied.* It is believed that such an assumption is again practically reasonable as the manhole inflows are estimated based on the fraction of the population associated with each manhole, rather than the exact population number. Given that the occupation rate of each residential building should be statistically similar in a local region, this assumption should not significantly affect the final results.

Conditioned on the two assumptions stated above, the following equation can be used to estimate the population size associated with each manhole,

$$P(h) = A_r \sum_{r=1}^{R_h} \eta \times V_r(h) \tag{3}$$

where $P(h(r))$ is the estimated population size associated with manhole h; $V_r(h)$ (m^3) is the building volume associated with manhole h, which can be computed based on geotagged data from public map databases as shown in Fig. 3; R_h is the total number of buildings that has physically connected to manhole h; η is the average number of living persons (np) per building volume (np/m^3); A_r is the occupation rate of each residential building, which is 100% in this study as stated in Assumption 2. Fig. 3 illustrates the proposed method for estimating the population size for each manhole associated with the residential buildings.

To enable the computation of Eq. (3), it is necessary to estimate the value of η, which can be different at different cities. In this study, a simple survey can be conducted to enable the determination of η. More specifically, within the area of the FSS, the model practitioners can investigate a few housing estates in the city to acquire the total number of population of a particular set of residential buildings, thereby estimating the value of η. In many countries, such as China, the average number of persons per building volume can be easily acquired from the local government. In this study, a constant value of η is determined and used in the entire FSS model based on the total building capacity and total population data from the local government.

Note that Eq. (3) is only used for residential buildings. For the

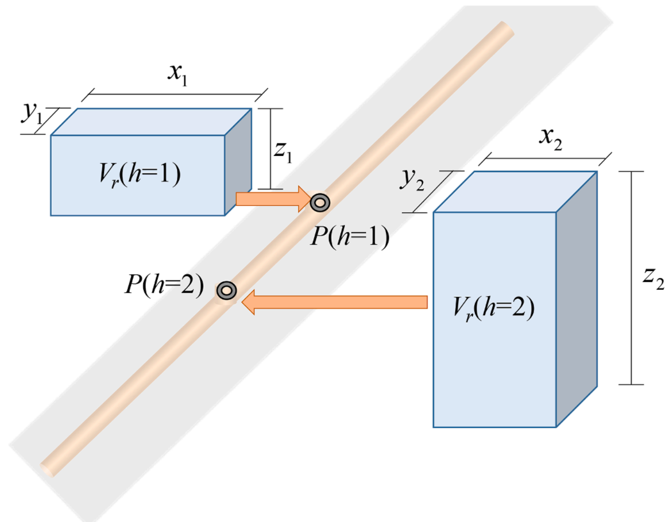


Fig. 3. Illustration of the population size estimate for each manhole.

commercial/public buildings, their corresponding manhole inflows are estimated from water consumption data recorded by the smart water meters (Bailey et al., 2019) as shown below,

$$DS_j(t) = TF_j(t) \times WS_j(t) \tag{4}$$

where $DS_j(t)$ is the discharges of the jth commercial/public buildings at time t; $WS_j(t)$ is the water consumption data of the jth commercial/public buildings provided by smart water meters at time t; $TF_j(t)$ is the transfer factor between water consumption and discharges at time t, which is caused by the inevitable loss during the transporting process within the facilities of the users (Behzadian and Kapelan, 2015).

2.2. Identify daily inflow pattern for each manhole

2.2.1. Partitioning the FSS into different subsystems based on sewer flow meters

This study aims to develop an accurate offline model (i.e., static model), where each manhole has only one inflow value at each time across different days. This is because, despite their variations at a certain degree caused by many external factors such as temporary population mobility, the total discharges from each building with many users are statistically similar at each time over different days (Bailey et al., 2019).

Typically, a FSS is often large in spatial scale, resulting in challenges for the calibration of model parameters, such as manhole inflows. To this end, this study proposes a two-sage optimization approach, aimed to reducing the calibration complexity. As part of the proposed two-sage optimization approach, the entire FSS is partitioned into different subsystems based on the available sewer flow meter sensors. The rationale behind such a partitioning approach is that a FSS often possesses a tree-like structure and hence observations of each sewer sensor primarily represent the hydraulic properties of the upstream part of the sensor location. In this study, each subsystem is assumed to have an identical time-series pattern of manhole inflows as the properties of the water users (user types and habits of water usages) in a local region is likely to be the same. Such an engineering heuristic has been widely used in urban water supply and drainage research area (Zhang et al., 2018; Bailey et al., 2019). It is noted that only flow meter sensors are considered for FSS partitioning in this study. This is done because (i) the residential users within each local region/subsystem (the outlet is typically monitored by a sewer flow meter) are highly likely to have a similar discharge pattern, (ii) the water depth data is overall less sensitive compared to the flow data as a result of inflow changes due to that the diameter size of a sewer is often relatively large, and (ii) the consideration of the water depth sensor may result in a significantly increased number of decision variables. For instance, if a 30 min time resolution is used (Zhang et al., 2021), 48 decision variables have to be optimized in order to identify the flow patterns in each subsystem. For a realistic FSS, if the number of water depth sensors is 30 (this number is often significantly larger than that of the sewer flow meters), the total number of decision variables can be up to 1440. This can bring large challenges for model calibration."

By using this partitioning method, the entire system can be divided into N subsystems, where N is the total number of sewer flow meters in the FSS. Fig. 4 illustrates the results of the proposed partitioning method. As shown in this figure, a total of three sewer flow meter sensors are available and hence three subsystems are identified (shaded regions in Fig. 4). Flow observations in Sensor A represent the manhole inflow properties at its upstream FSS. Similarly flow data in Sensor B and C can be used to calibrate the manhole inflows within its corresponding subsystem. In this study, the hydraulic interactions between different subsystems are handled by a hydraulic software called Storm Water Management Model (SWMM, Gironas et al., 2010).

2.2.2. Calibrate time-series pattern of total inflows for each subsystem (stage-one optimization)

As previously stated, the time-series pattern of flows associated all residential manholes in each subsystem is considered to be identical in this study. This is done to reduce the number of variables to be calibrated. Note that such an assumption has been widely used for nodal demand calibration in water distribution systems, which has achieved great success within practical applications (Zhang et al., 2018).

The formulation of the stage-one optimization problem is as follows,

$$\begin{aligned} \text{Min} : F(\mathbf{Q}) \\ = \sum_{t=T_w}^{T_e} \left(\sum_{i=1}^{M} \left[g\left(w_i^o(t)\right) - g\left(w_i^s(t)\right)\right]^2 + \sum_{j=1}^{N} \left[g\left(f_j^o(t)\right) - g\left(f_j^s(t)\right)\right]^2 \right) \end{aligned} \tag{5}$$

$$\mathbf{Q} = \begin{bmatrix} q_1(\Delta t), q_1(2\Delta t), \ldots, q_1(T) \\ q_2(\Delta t), q_2(2\Delta t), \ldots, q_2(T) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ q_N(\Delta t), q_N(2\Delta t), \ldots, q_N(T) \end{bmatrix} \tag{6}$$

$$MI_h(t_a) = \begin{cases} \dfrac{q_j(t_a) \times P(h)}{\sum_{h=1}^{H_j} P(h)}, h \in \mathbf{H}_j, \text{ if } h \text{ is associated with residential buildings} \\ \sum_{j=1}^{h(r)} DS_j(t_a), \text{ if } h \text{ is associated with commercial} \Big/ \text{public buildings} \end{cases} \tag{7}$$

$$\begin{aligned} F_m(\mathbf{MI}(t_a)) &= [\mathbf{W}^s(t_a); \mathbf{f}^s(t_a)] \\ &= \left[w_1^s(t_a), w_2^s(t_a), \ldots, w_M^s(t_a); f_1^s(t_a), f_2^s(t_a), \ldots, f_N^s(t_a)\right] \end{aligned} \tag{8}$$

where $\mathbf{Q}$ is the decision variable matrix, representing the total inflow of each subsystem at each time step, which is defined as $q_j(t_a)$ in Eq. (6); $j = 1,2,\ldots,N$ is the jth subsystem, where N is the total number of subsystems (i.e., total number of sewer flow meters); t_a=$\Delta t, 2\Delta t, \ldots, T$ with T representing 24 h as the daily time-series inflow pattern is considered in this study; T_e is the time period with observations used for FSS calibration with a model time resolution of Δt; T_w is the warming-up time period for model setting up (Guo et al., 2020); M is the total number of water depth sensors at the manholes; $w_i^o(t)$ and $f_j^o(t)$ are observed water depth at manhole i and observed flow rate at sewer pipe j at time t, respectively; $w_i^s(t)$ and $f_j^s(t)$ are simulated water depth at manhole i and simulated flow rate at sewer pipe j at time t, respectively; $g()$ is a linear function used to convert water depths and pipe flow rates into the same

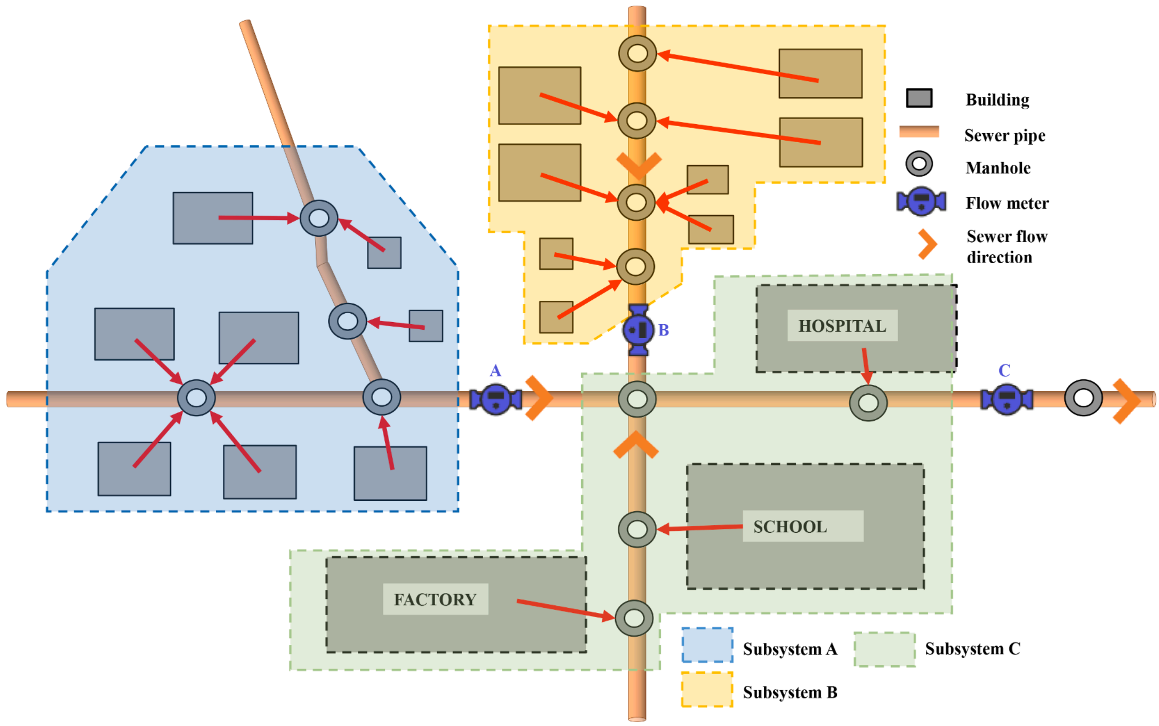


Fig. 4. The identified subsystems for a FSS with three sewer flow meters.

scale, thereby enabling both terms in the right side of Eq. (5) are approximately equivalent in terms of the objective function value. In this study, $g(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ is used, where $x_{\min}$ and $x_{\max}$ is the minimum and maximum value of the variable x being considered, respectively. $\mathbf{MI}(t_a)$ in Eq. (8) is the manhole inflow vector at time t_a and the $\mathbf{MI}(t_a)$ value is determined by Eq. (7); $\mathbf{W}^s(t_a) == [w_1^s(t_a), w_2^s(t_a), ..., w_M^s(t_a)]$ and $\mathbf{f}^s(t_a) == [f_1^s(t_a), f_2^s(t_a), ..., f_N^s(t_a)]$ are the vector of the water depth and flow predictions at all sensor locations at time t_a, respectively.

The aim of the stage-one optimization is to identify $\mathbf{Q}$ through minimizing $F(\mathbf{Q})$ (Eq. (5)). As shown in Eq. (6), for a FSS with N subsystems and with Δt time resolution (Δt can be half of an hour), the total number of the decision variables (daily dry-weather inflows at manholes) in the matrix of $\mathbf{Q}$ is $N\frac{T}{\Delta t}$ ($T = 24$ h), which is calibrated using the stage-one optimization in this study. As shown in Eq. (7), for the manhole h that is physically connected to residential buildings, if it belongs to the subsystem j ($h \in \mathbf{H}_j$), its manhole inflows at time t_a are estimated by the total inflow $q_j(t_a)$ times by the fraction of the population size of manhole h ($P(h)$) relative to the all manholes (H_j) in this subsystem (n), i.e., $\sum_{h=1}^{H_j} P(h)$. If the manhole h is physically connected to commercial or public buildings, its manhole inflows at time t_a are estimated by the total discharges of these buildings, with $h(r)$ representing the total number of commercial or public buildings associated with manhole h (Eq. (7)). $DS_j(t_a)$ is defined in Eq. (4). For the case that a manhole receives discharges from both residential and commercial/public buildings, its inflows are the sum of the two terms in the right side of Eq. (7).

After each manhole has been assigned an inflow estimate at time t_a using Eq. (7), a hydraulic simulation model (SWMM is used in this study) is used to solve Eq. (8), thereby generating predictions at all sensor locations. These predictions are then compared with the observations as shown in Eq. (5). In this study, an evolutionary algorithm (EA) combined with the FSS hydraulic software SWMM Zhang et al., 2021) is used to solve Eqs. (5)–((8).

2.2.3. Determine the daily time-series inflow pattern for each manhole (stage-two optimization)

The stage-one optimization has identified the total inflow time-series pattern for each subsystem, where daily time-series inflows of each manhole within the subsystem are proportionally assigned based on its estimated population size. Given that the population size estimate at each manhole may deviate from the true value to a certain extent due to the two assumptions stated in Section 2.1.2, the stage-two optimization is conducted to further improve manhole inflow estimates based on the results of the stage-one optimization. The formation of the stage-two optimization problem is as follow,

$$\text{Min}: F(\mathbf{K}) = \sum_{t=T_w}^{T_e}\left(\sum_{i=1}^{M}\left[g\left(w_i^o(t)\right) - g\left(w_i^s(t)\right)\right]^2 + \sum_{j=1}^{N}\left[g\left(f_j^o(t)\right) - g\left(f_j^s(t)\right)\right]^2\right) \quad (9)$$

$$MI_h(t_a) = k_h \times \frac{q_n(t_a) \times P(h)}{\sum_{h=1}^{H_n} P(h)}, h \text{ is associated with residential buildings} \quad (10)$$

$$F_m(\mathbf{MI}(t_a)) = [\mathbf{W}^s(t_a); \mathbf{f}^s(t_a)] \quad (11)$$

$$k_h \in [k_{\min}, k_{\max}] \quad (12)$$

where $\mathbf{K} = [k_1, k_2, ...k_H]^T$ with k_h representing the inflow adjusting coefficient for manhole h (only for the residential users). This indicates that Stage-two optimization aims to identify k_h for each manhole based on the given time-series inflow $q_n(t_a)$ determined by Stage-one optimization as shown in Eq. (10). Therefore, the total number of decision variables in Stage-two optimization is the number of manholes that are physically connected to residential buildings. Eq. (11) is used to simulate values of the hydraulic variables to enable the objective function computation Eq. (9)) based on the $\mathbf{MI}_h(t_a)$ that is defined in Eq. (8). $k_{\min}$ and k_{max} are the minimum and maximum adjustment coefficients, respectively. As the same for the stage-one optimization, an EA with the SWMM software are jointly used to minimize the objective function defined in the stage-two optimization stage (Eqs. (9)–((12)).

The main merit of the proposed two-stage optimization is that the optimization complexity is significantly reduced. This is because the number of decision variables considered at each stage is substantially lower than the traditional approach where all manhole inflows are directly considered. For example, for a FSS with four flow meters (i.e. four subsystems) and 100 manholes with a time step of 30 min, the number of decision variables considered at the stage-one and stage-two optimizations are $4 \times 48 = 192$ and 100 (100 different k values), respectively in the proposed method. The total number of decision variables considered in the traditional optimization approach for this case is also 292. Using the proposed two-stage optimization method, the number of decision variables at stage-one and stage-two are 192 and 100, respectively. Consequently, the complexity of the proposed optimization method can be significantly lower than the traditional optimization approach with 292 variables simultaneously considered.

2.3. *Identify uncertainty ranges for manhole inflows*

The proposed two-stage optimization method provides the averaged or expected daily time-series dry-weather inflow pattern for each manhole. These simulations may neglect the potential variability associated with these inflows. To address this issue, an uncertainty analysis approach is proposed in this study. The proposed uncertainty analysis method for manhole inflows is based on the stochastic properties of water consumption data that are taken from smart water meters. The rationale for this analysis is based on the existing physical connection between water supply and the wastewater discharges for each residential building (Bailey et al., 2019).

Fig. 5 illustrates the physical relationship between water consumption and wastewater discharge within a specific building. Generally, a large proportion of clean water (delivered by the water distribution system) at time t ($WS(t)$ in Eq. (4)) is discharged into the sewer system ($DS(t)$) after a short time delay Δt (water travelling time period within the building). The transfer factor between water supply and discharges is *TF* (Eq. (4)) as shown in Fig. 5(a), which is caused by various losses during the consumption process. Despite the deviation between water supply and wastewater discharge at time t, it is reasonable to map the demand time series and discharge pattern using similar trends (Fig. 5b). In other words, the expected manhole inflows are expected to have a similar time pattern as water consumption data, with the former slightly decreased by a factor of *TF* compared to latter after Δt, as illustrated in Fig. 5b. Consequently, both the water supply and its corresponding discharges should have a similar stochastic distribution (Fig. 5b), and thus the uncertainty ranges of the manhole inflows can be mapped from the water consumption data analysis based on records from smart water meters. It is noted that this study does not consider the infiltration/exfiltration within the sewer pipes, in order to focus the main methodology of this proposed method. However, it is straightforward to add an infiltration/exfiltration estimate within the calibration process of the proposed method.

2.3.1. *Determine stochastic properties of water consumption data*

In this study, the stochastic properties of water supply flows are determined based on real-time data collected by available smart water meters installed for residential buildings. More specifically, the following steps are used to quantify the stochastic properties of water consumption data.

Step 1: Determine the daily average time-series water consumption data. For each building or water user with a smart water meter, their real-time water consumption data are collected often with an half an hour time resolution. This is followed by the computation of the averaged water consumption at each time of the day based on records over many different days. Consequently, the daily average/expected time-series water supply data with a particular time-resolution can be determined for each smart water meter.
Step 2: Compute the coefficient of variation for each time a day. For each time a day, all the records from smart water meter divides their corresponding average values, thereby producing the coefficient of variation (*CV*, Zhang et al., 2018). Using this approach, a large number of *CV* values (some are greater than 1 and some are smaller than 1) is generated for each time of the day based on each smart water meter.
Step 3: Establish a sampling pool for each time t at the day. For each time t of the day, all *CV* values over different smart water meters are collected to form a sampling pool ($\Psi(t)$). In other words, if the time resolution is 30 min, a total of 48 sampling pools are generated using the proposed method. The *CV* values in different $\Psi(t)$ can be significantly different, representing various stochastic properties at different time periods at a day. This is a novel aspect of the proposed uncertainty analysis method as it can capture the underlying variation of the manhole inflows at different time periods.

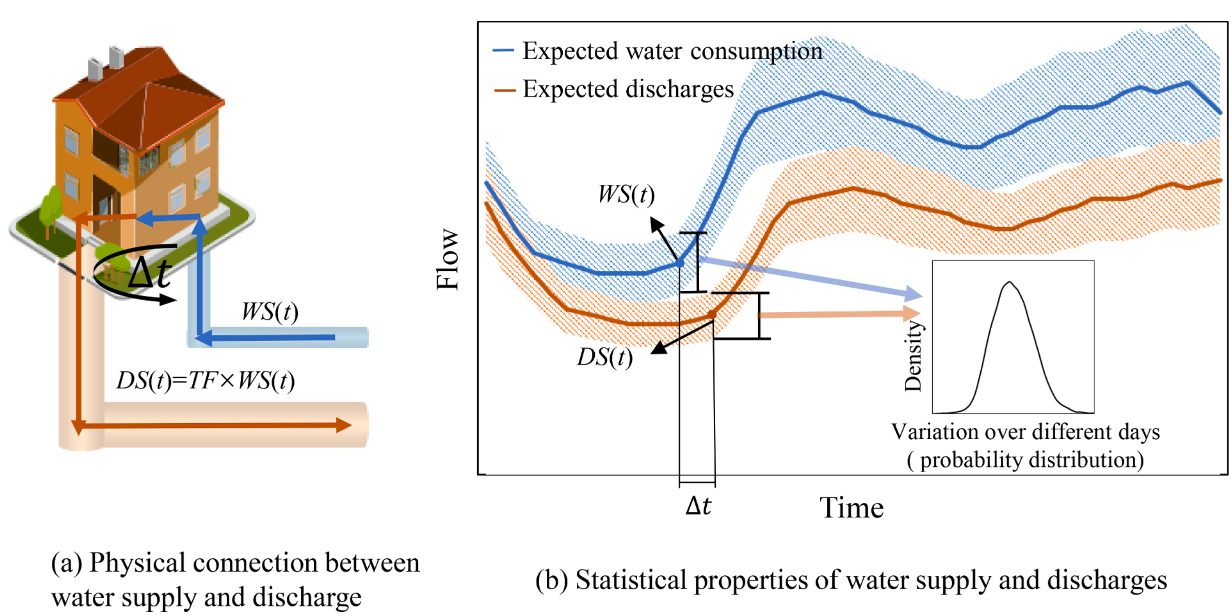


Fig. 5. Uncertainty mapping between water consumption and wastewater discharge of a single residential building.

These established sampling pools based on water consumption data ($\Psi(t)$) represent the stochastic properties of the water supply data at each time of the day, which will be used to for uncertainty analysis for the manhole inflows.

2.3.2. Quantify sewer uncertainty range based on stochastic properties of water consumption data

Typically, the causes of hydraulic variability within sewer systems can be divided into two types: random and systematic factors. The random factor mainly includes the temporal population mobility as well as the natural variability of water used by persons (e.g., different shower time over different days). The systematic factor mainly includes the sudden temperature changes that can affect the water use habits (e.g., shower time or frequency) of many persons in the residential buildings, as well as the holiday time-period where many people leave the city. It is noted that many countries such as China, the population density of some cities can be significantly varied during the holiday time-period due to the economic structure properties (i.e., many people work in a city but may live in another city). Therefore, the number of people is consistently reduced or increased for each building during the holiday time-period (this is a systematic factor), but the population mobility in working time-period is a random factor as it can increase for some residential buildings but decrease for some others.

In recognizing the two different types of causes that affect the sewer variability, this study proposes a new uncertainty analysis method to account for both types of causes, as shown in the following,

$$\begin{aligned} CV_h(t) &= Rand(\Psi(t)) \\ MI_h^u(t) &= CV_h(t) \times MI_h(t) \end{aligned} \tag{13}$$

where $CV_h(t)$ is the coefficient of the variation for manhole h at time t, which is randomly selected from the established sampling pools ($\Psi(t)$) based on water consumption data ($\Psi(t)$); $Rand()$ is a function for random sampling. $MI_h^u(t)$ is the updated inflows for the manhole h ($h = 1,2,\ldots,H$) that is physically connected to residential buildings at time t; $MI_h(t)$ is the manhole inflows at time t determined by the proposed two-stage optimization method (See Section 2.2).

In addition to Eq. (13) that considers the random factor of the manhole inflows, Eqs. (14) and (15) are used to account for the systematical factor,

$$MI_h^u(t) = CV_h^L(t) \times MI_h(t), CV_h^L \in \Psi(t) \tag{14}$$

$$MI_h^u(t) = CV_h^S(t) \times MI_h(t), CV_h^S \in \Psi(t) \tag{15}$$

where $CV_h^L(t)$ and $CV_h^S(t)$ are the coefficients of the variation for manhole h at time t. More specifically, $CV_h^L(t)$ is greater than 1, and hence it is randomly selected from the values that are greater than 1 in $\Psi(t)$. Conversely, $CV_h^S(t)$ is smaller than 1, and hence it is randomly selected from the values that are smaller than 1 in $\Psi(t)$.

Fig. 6 illustrates the proposed uncertainty analysis method for a FSS with seven manholes Fig. 6b) at a particular t, where Fig. 6a and c represent the sampling results using Eqs. (13) and ((15). As shown in Fig. 6(a), for the seven CV values generated using Eq. (13), some values are greater than 1 and the others are smaller than 1; but all CV values are smaller than 1 for those produced by Eq. (15).

2.4. Demonstrate the utility of the proposed method

2.4.1. Traditional calibration and uncertainty analysis methods

To demonstrate the effectiveness of the proposed method in this study, its performance is compared to the traditional calibration

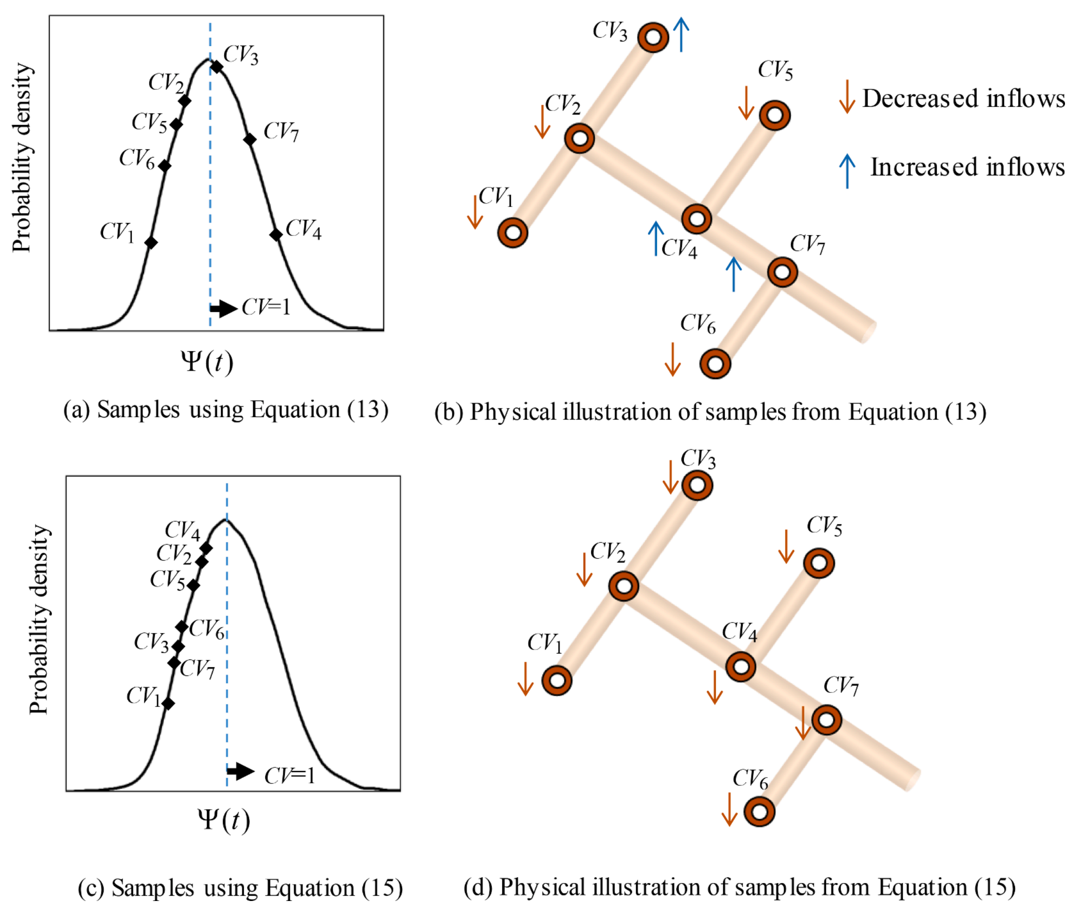


Fig. 6. Variability of sewer inflows due to random (Eq. (13)) and systematic (Eq. (15)) factors.

methodology on real-world case studies. The traditional calibration method often takes runoff contributing area or/and sewer pipe lengths as prior information to enable the manhole inflow allocation (Chu et al., 2021). While various heuristics can be used as prior knowledge for FSS hydraulic modeling, i.e., based on pipe length or on contributing areas, they have similar implications for simulation results. In this particular case (the two case studies considered), the pipe-length heuristics procedure is considered as the traditional approach due to it simple implementation (Zhang et al., 2018). It is highlighted that the only difference between the proposed method and the traditional approach in this study is that the former considers the population sizes associated with each manhole as the prior information, but the latter considers the pipe length as the initial knowledge. In other words, the proposed two-stage optimization is also used in the traditional approach. The proposed uncertainty analysis method is also compared to the traditional uncertainty analysis approach that uses assumed specified distributions overall all manholes across different time periods at the day (Jin and Mukherjee, 2010; Sun et al., 2014).

2.4.2. Comparison with the traditional calibration method

In this study, four statistical metrics are used to evaluate the performance of the proposed method for calibrating FSS hydraulic models, including the relative error (*RE*) or absolute percentage error (*APE*), the coefficient of determination (R^2), the Nash-Sutcliffe model efficiency (*NSE*), and the Kling-Gupta Efficiency (*KGE*). Note that these assessment matrices have been widely used for hydraulic model evaluation in the field of water system analysis (Guo et al., 2020). These equations are defined as follows.

(1) Relative error (*RE*) and absolute percentage error (*APE*):

$$RE = \frac{\widehat{Y}_i - Y_i}{Y_i} \times 100\%, APE = \left|\frac{\widehat{Y}_i - Y_i}{Y_i}\right| \times 100\% \tag{16}$$

where Y_i is the ith observation and $\widehat{Y}_i$ is its corresponding simulated value. *APE* is the absolute value of *RE*.

(1) Coefficient of determination (R^2):

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(Y_i - \widetilde{Y}\right)\left(Y_i - \overline{Y}\right)\right)^2}{\sum_{i=1}^{n}\left(Y_i - \widetilde{Y}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \tag{17}$$

where $\overline{Y}$ and $\widetilde{Y}$ are the mean values of observed and simulated data, and n is the total number of data points.

(1) Nash-Sutcliffe model efficiency (*NSE*) (Nash and Sutcliffe, 1970):

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \tag{18}$$

(2) Kling-Gupta efficiency (*KGE*) (Knoben et al., 2019):

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{19}$$

where r is the Pearson product-moment correlation coefficient; σ_{sim} and σ_{obs} are the standard deviation of simulations and observations; μ_{sim} and μ_{obs} are the mean values of simulations and observations. A lower value of *RE* or *APE* represents a better model performance. In contrast, a large value of R^2, *NSE* or *KGE*) indicates that the simulations can match observations better, with the value of 1 representing the best model performance.

2.4.3. Performance in addressing the "equifinality" issue and comparison with the traditional uncertainty analysis approach

In this study, the proposed method is compared to the traditional method in addressing the "equifinality" issue, i.e., the simulation performance of hydraulic variables at locations without sensors. Specifically, for the FSS locations without sensors but with available water smart meters, the water consumption data are used to indirectly assess the accuracy of the simulated sewer discharges. To assess the performance of the proposed uncertainty analysis approach, its results as well as the uncertainty ranges determined by the traditional uncertainty analysis method are compared with observations collected by the installed water depth sensors and sensor flow meters in the FSS.

3. Case studies

3.1. Case study description

The proposed method is demonstrated on two real-world FSSs in China, namely the Benk network (BKN) and the Xiuzhou network (XZN). These two FSS with significantly different scales can also be used to explore how the proposed method performs when dealing with the increased system complexity. The BKN case study has 64 manholes, 64 sewer pipes (9.4 km length) and one outlet, and the XZN case study has 1,214 manholes, 1,214 sewer pipes (86 km pipe length) and one outlet as shown in Fig. 7. The average pipe slopes of the BKN and XZN case studies are 0.65% and 0.27%, respectively. As shown in Fig. 7, one sewer flow meter and three water level sensors have been installed in the BKN. For the XZN case study, three flow meters and eight water level sensors have been deployed in the system. All sensors in these two systems collect real-time data with a 30 min time resolution. While two FSS case studies are designed to solely deliver wastewater discharges, runoff in the rainy days may inevitably affect the hydraulics of the sewer pipes through infiltration. Therefore, observations for a period of consecutive 31 days without rainfall events are used for FSS model development and uncertainty analysis, in order to minimize the impacts of the infiltration.

For the BKN and XZN case studies, 16 and 152 residential users have smart water meters, respectively (red circles in Fig. 7), where these water consumption data with an 30-min time resolution are used for uncertainty analysis and model performance demonstration. In addition to these residential users with water smart meters, all commercial/public buildings also have water smart meters (red squares in Fig. 7) and these data facilitate the model development and calibration. The records of the water smart meters at the same time period with the sewer sensors (a period of consecutive 31 days) are considered in this study.

3.2. Parameterization of the proposed method

In this study, SWMM5.1 Gironas et al., 2010) has been used to simulate the hydraulic behavior of these two FSSs. The model simulations are implemented with a time resolution of 30 min, matching the time resolution of the measurement data. For the entire simulation period of 31 days (i.e., the data collection period), the first three days ($T_w = 3$ days in Eqs. (5) and ((9)) are regarded as the warming-up time for model set up, to ensure appropriate initial conditions for FSS simulation. The observations between the 4 and 17th day are used for model calibration, and the remaining observed data are utilized to validate the model simulation performance on unseen data.

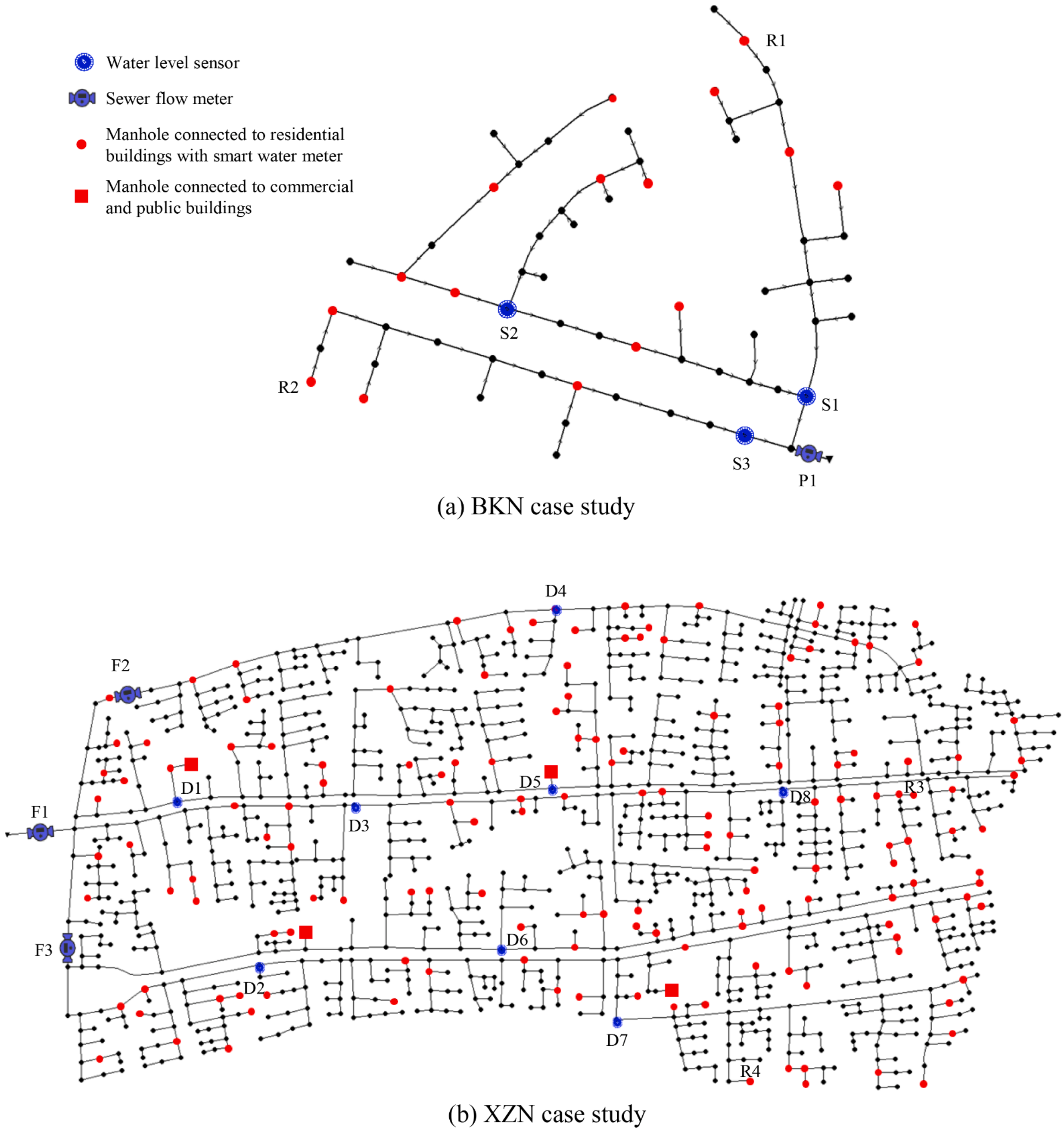


Fig. 7. The layouts of two FSS case studies and the information of the smart water meters (P1 and F1-F3 represents sewer flow meters in the two case studies, respectively, S1-S3 and D1-D8 represents manhole water level sensors in the two case studies, respectively, R1-R4 represent four typical manholes without sensors which will be used in Fig. 11).

For each case study, the water consumption data from smart meters are used to derive the stochastic properties of the water use with method described in Section 2.3.1. This leads to an establishment of the total sampling pool $\Psi(t)$ for each time t a day, with various CV values included for inflow uncertainty analysis for residential users. Each stage of the proposed two-stage optimization Eqs. (5)–((12)) is optimized using the Borg evolutionary algorithm (Hadka and Reed, 2013). This optimization algorithm is chosen as it has been demonstrated to efficient in addressing complex problems in the area of urban water resources and engineering optimization (Zheng et al., 2016). For both case studies, the initial population size is set as 500, and the maximum number of allowable solution evaluations is 1,00,000 based on a preliminary algorithm parameter calibration. The other Borg parameters use the default values as presented in Hadka and Reed (2013).

For the BKN and XZN case studies, the population size per building volume η defined in Eq. (3) is 0.96 and 0.97 $np/(100\text{m}^3)$, respectively, as provided by the local government. For each commercial/public building, the transfer factor $TF_j(t)$ between water consumption and discharges (see Eq. (4)) is assumed constant over different time at a day, where $TF_j(t) = 0.8$ is used in this study Zhang et al., 2021). $k_{\min} = 0.85$ and $k_{max} = 1.15$ are used in Eq. (12) (Zhang et al., 2018), representing the inflow updating range in the stage-two optimization. To enable the uncertainty analysis for the manhole inflows (only for residential users), Eq. (13) is used to generate the random samples from the $\Psi(t)$. This is followed by the use of Eqs. (14) and ((15) to produce samples with CV values greater than 1 and smaller than 1, respectively. More specifically, for each time t of the day, 20,000 samples are randomly taken from the $\Psi(t)$ using Eqs. (13), (14) and (15), respectively for the BKN case study.

For the XZN case study, 50,000 samples are randomly taken from the $\Psi(t)$ using the same approach. For the traditional uncertainty analysis approach, a constant of value with $vc_h(t) = 0.85$ or 1.15 is randomly selected for each manhole (Zhang et al., 2018) based on the expected inflow values identified by the proposed two-stage optimization method.

4. Results and discussion

The proposed method is applied to the two FSS case studies, with identified physical connections between sewer manholes and residential buildings illustrated in Fig. 8(a), which is a small region of the XZN case study. The density distributions of the estimated population sizes for the two case studies are shown in Fig. 8(b) based on the geotagged data from public databases using the proposed method in Section 2.1. Given that one and three flow meters are installed in the BKN and XZN, respectively, one and three corresponding subsystems are identified for these two case studies based on the approach described in Section 2.1.1. This is followed by the application of the proposed two-stage optimization method, with results presented below.

4.1. Performance comparison of the hydraulic simulations at FSS locations with sensors

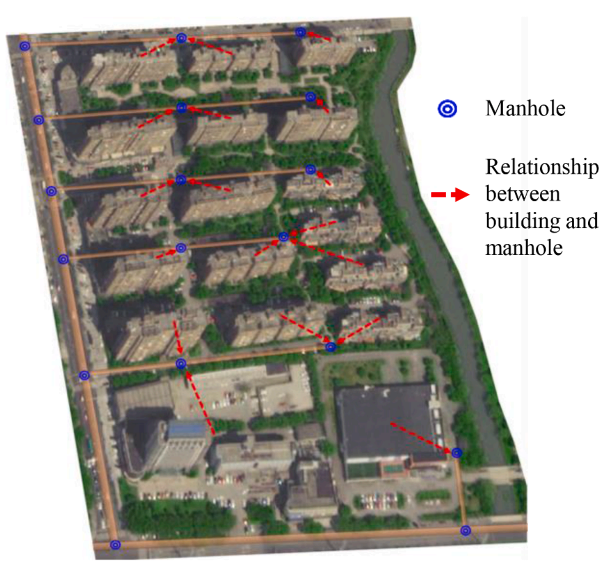
Fig. 9 compares the performance of the proposed method and traditional model in simulating hydraulic variables at FSS locations with sensors for both case studies. It is noted that simulation results at typical FSS sensor locations with seven days within the validation time period (from 18th day and 24th day) are presented in Fig. 9 to enable the clear presentation. Fig. 10 is the results of one day (18th day) taken from Fig. 9, in order to further clearly show the differences between the proposed and traditional methods.

As shown in Figs. 9 and 10, both the proposed and traditional methods are able to capture the overall trends of the manhole water depth and pipe flow observations at P1 and S1 of the BKN case study (see Fig. 7(a)), as well as F1 and D1 in the XZN case study (see Fig. 7(b)). For the BKN case study, the average *APE* values for the simulated flows of the proposed and traditional methods are 8.78% and 9.67%, respectively (Fig. 9(b)), and these two values are 3.57% and 3.63%, respectively for the water depth simulations at S1 (Fig. 9(d)). For the XZN case study, the average *APE* value is 6.29% for the flow simulations at F1 from the proposed method, and this value is 6.46% from the traditional approach. In terms of the water depth simulations at D1, the mean *APE* values of the proposed and traditional methods are 4.50% and 7.60%, respectively. This implies that both the proposed and traditional approaches can overall accurately simulate hydraulic variables at P1, S1, F1 and D1 sensor locations (Fig. 7), but the former performs consistently slightly better than the latter.

It can be seen from Fig. 9 that while the mean *APE* value is consistently below 10% for the manhole water depth and pipe flow variables, its maximum value can be up to about 30% for the both the proposed and traditional methods. We also observe that the majority of the large *APE* values occur at the time periods with relatively low manhole water depths or pipe flows. Therefore, it can be deduced that the large *APEs* can be related to the low values of the denominator in Eqs. (16).

Tables 1 and 2 present the values of performance metrics for simulations at FSS locations with sensors for both case studies. It can be seen from these two tables that the proposed method shows an overall similar performance for the small BKN case study, but a slightly better performance for the large XZN case study relative to the traditional method. This can be proven by that the mean *NSE* and *KGE* values across all FSS sensor locations of the proposed method are 0.90 and 0.93, which are all larger than those from the traditional approach (0.81 and 0.88). More specifically, the *NSE* values of the traditional approach at D1-D5 in the XZN are consistently lower than 0.75, which are significantly lower than those from the proposed method (consistently larger than 0.85). Results in Tables 1 and 2 can demonstrate that the proposed method is able to exhibit a better performance than the traditional approach in accurately simulating hydraulic variables for relatively large FSSs. This is because the manhole inflow combinations for a larger FSS can be larger relative to a small FSS, resulting in a more complex calibration process. For such cases, the use of the population size as the domain knowledge as did in the paper exhibits a more prominent performance compared to the traditional approach.

As previously stated, given that the static simulation is considered in this study (i.e., the water depth or flow time-series pattern is identical over different days), the simulations (expected simulations of hydraulic variables) are unable to capture the variations of the hydraulic variables over different days as shown in Fig. 9. To mitigate this, an uncertainty range is often combined with the static simulation results, in order to



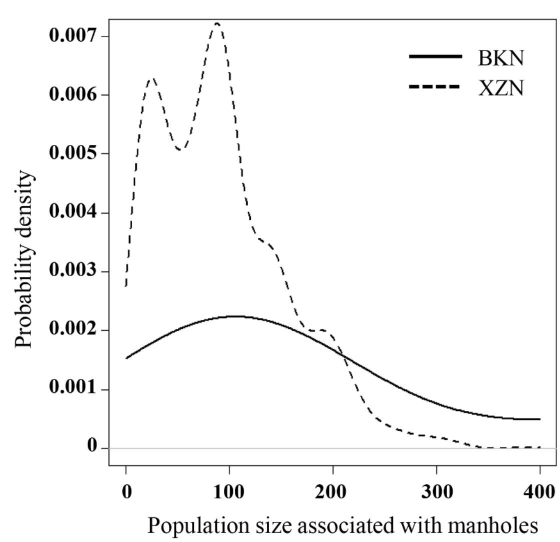


Fig. 8. Results of the physical connections and estimate population sizes of the manholes for the two case studies.

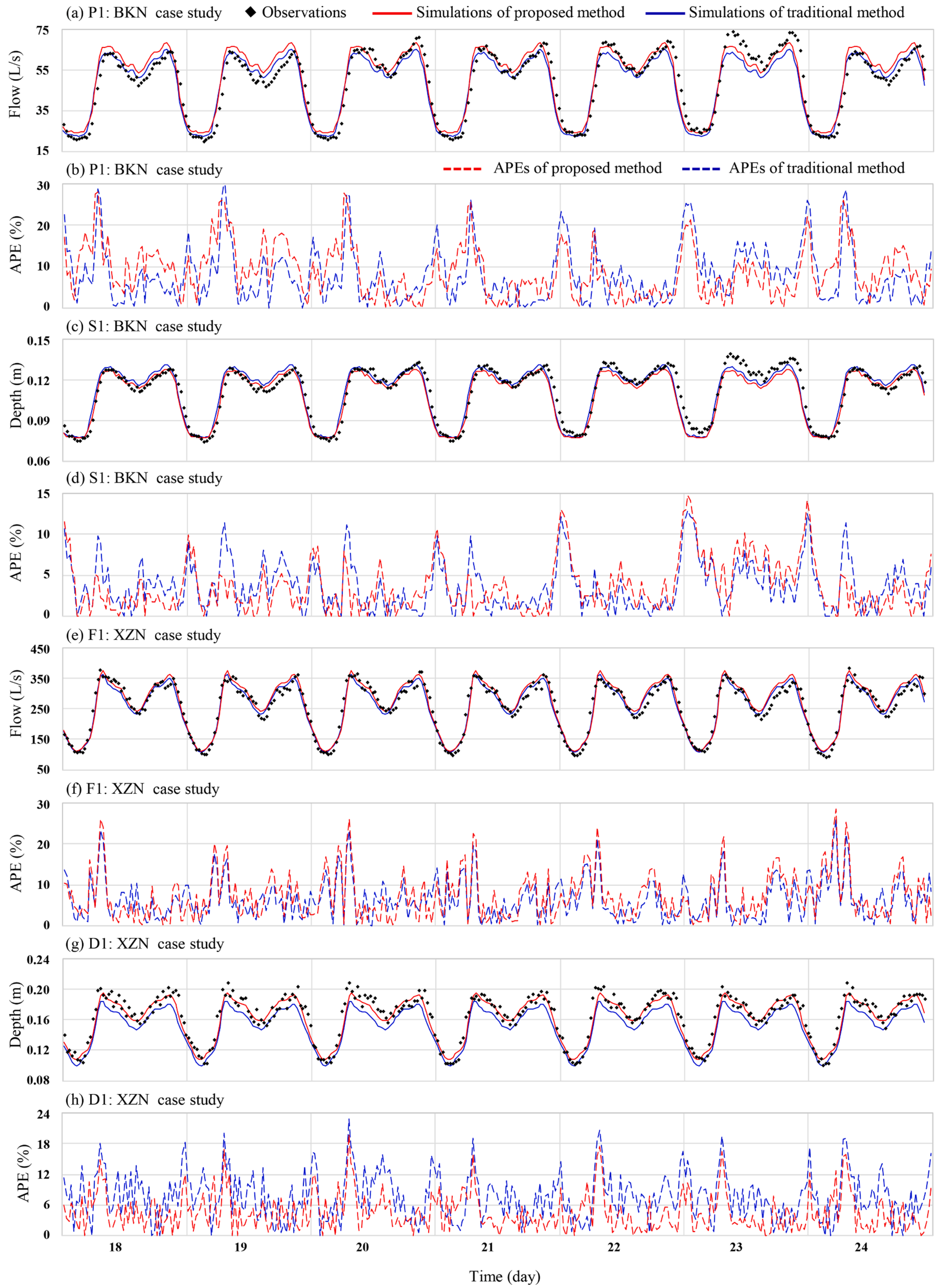


Fig. 9. Results of observations versus simulations and the absolute percentage error (*APE*, %) values at the typical FSS sensor locations (P1, S1, F1 and D1 are shown in Fig. 7).

provide abnormal warning, with results presented in Section 4.3.

4.2. Performance of the proposed method in addressing the "equifinality" issue

It is noted that Section 4.1 focuses on the performance analysis at the

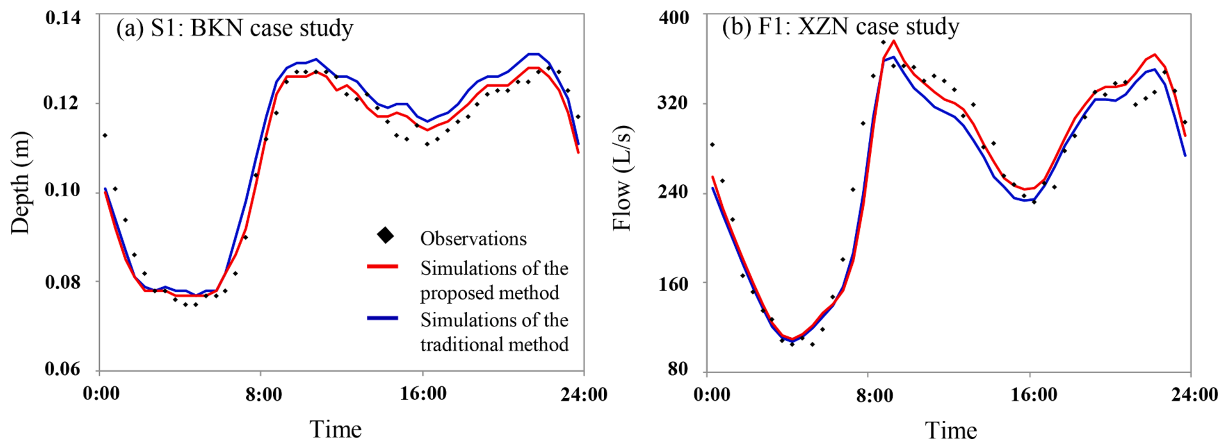


Fig. 10. Observations versus simulations at a typical day (18th day) of two sensor locations (S1 and F1 are shown in Fig. 7).

Table 1
Metric values of simulations at validation time period for the BKN case study.

Monitoring locations	The traditional method			The proposed method		
	R^2	*NSE*	*KGE*	R^2	*NSE*	*KGE*
S1	0.92	0.92	0.96	0.93	0.92	0.95
S2	0.91	0.89	0.90	0.92	0.90	0.91
S3	0.88	0.87	0.80	0.90	0.87	0.78
P1	0.91	0.91	0.92	0.92	0.91	0.94
Mean	***0.91***	***0.89***	***0.89***	***0.92***	***0.90***	***0.89***

Table 2
Metric values of simulations at validation time period for the XZN case study.

Monitoring locations	The traditional method			The proposed method		
	R^2	*NSE*	*KGE*	R^2	*NSE*	*KGE*
D1	0.91	0.73	0.85	0.90	0.90	0.93
D2	0.92	0.70	0.82	0.92	0.89	0.89
D3	0.90	0.74	0.88	0.89	0.88	0.94
D4	0.93	0.73	0.82	0.93	0.92	0.91
D5	0.90	0.68	0.81	0.89	0.89	0.91
D6	0.91	0.82	0.88	0.90	0.89	0.92
D7	0.90	0.86	0.86	0.90	0.90	0.90
D8	0.88	0.86	0.93	0.86	0.85	0.92
F1	0.94	0.94	0.96	0.93	0.92	0.95
F2	0.96	0.96	0.95	0.96	0.96	0.96
F3	0.94	0.94	0.95	0.93	0.93	0.96
Mean	***0.92***	***0.81***	***0.88***	***0.91***	***0.90***	***0.93***

FSS locations with sensors where observations are available. This section aims to compare the performance of the proposed and traditional methods in accurately simulating the sewer variables at FSS locations without sensor observations, i.e., the ability in addressing the "equifinality" issue. To attain this goal, water consumption data are compared with the inflow simulations of the manholes (without sewer observations) that are physically connected the residential buildings with installed water smart meters.

Fig. 11 shows water consumption data versus sewer inflow simulations at four FSS manholes (shown in Fig. 7) without sensors. It can be seen from this figure that the simulation results of the traditional model at R1, R3 and R4 (blue lines in Fig. 11) are consistently substantially larger than the water consumption data. For the results at R2, the manhole inflows are always significantly lower than their corresponding water consumption data (Fig. 11(b)), implying that a rather low proportion of water consumption is discharged. Both cases above do not actually conform to the real engineering practice where the wastewater discharges of the residential buildings are often slightly lower than their corresponding water supply amount (*TF* in Eq. (4) is between 0.80 and 1.0 as stated in Zhang et al. 2021). Conversely, the manhole inflow simulations of the proposed method in this study (red lines in Fig. 11) are overall slightly lower than their corresponding water consumption data. This indicates a good performance in accurately simulating the sewer hydraulic variables at FSS locations without sensors (R1, R2, R3 and R4).

To further evaluate the overall performance of the proposed model in addressing the "equifinality" issue, the values of *TF* for all manholes (only for residential users) with available water consumption data are presented in Fig. 12. More specifically, for each of the two methods (the proposed and traditional methods), a *TF* value is computed for each manhole with available water consumption data at each time step (30 min) at the validation time period. The probability density distributions of these *TF* values from the proposed and traditional methods are plotted in Fig. 12 to enable the comparison. It is seen from this figure the majority of the *TF* values of the proposed method are around the value of 1.0, which is practically reasonable. However, many *TF* values from the traditional method are either significantly lower than 1 or substantially larger than 1. This implies that the proposed method can match better the real conditions than the traditional method at manholes without sensors. This means that the proposed method can better address the "equifinality" issue.

4.3. Performance with respect to uncertainty analysis

As previously stated, uncertainty analysis is essential to the static FSS model as it can assist modellers in identifying the potential impact of the stochastic nature of sewer formation and flow processes. The density distributions of the *CV* values over different smart water meters in the sampling pool ($\Psi(t)$) (see Section 2.3.1 for details) are presented in Fig. 13, where each line represents the density distribution of a particular time t at a day with 30 min resolution. As shown in this figure, while the stochastic property of the water consumption data is overall similar over different time at a day, small to moderate variations are still observed. Therefore, it can be derived that the use of the constant a *CV* value over different time periods at a day as did in the traditional method is not reasonable. This also highlights the novel aspect of the proposed uncertainty analysis method as it can capture the underlying variation of the manhole inflows at different time periods at a day.

As stated in Section 2.3.2, the sampling methods described in Eqs. (13)–(15) are used to estimate the uncertainty range of the sewer simulations based on the $\Psi(t)$, where the hydraulic simulations based on these samples are used to determine the uncertainty ranges (i.e., the maximum and minimum values) as well as the expected values (the mean value). Fig. 14 shows the uncertainty ranges and expected values based on the samples taken from the $\Psi(t)$ for the FSS sensor locations with observations within the validation time period. The red and blue dotted lines represent the results from the proposed and traditional

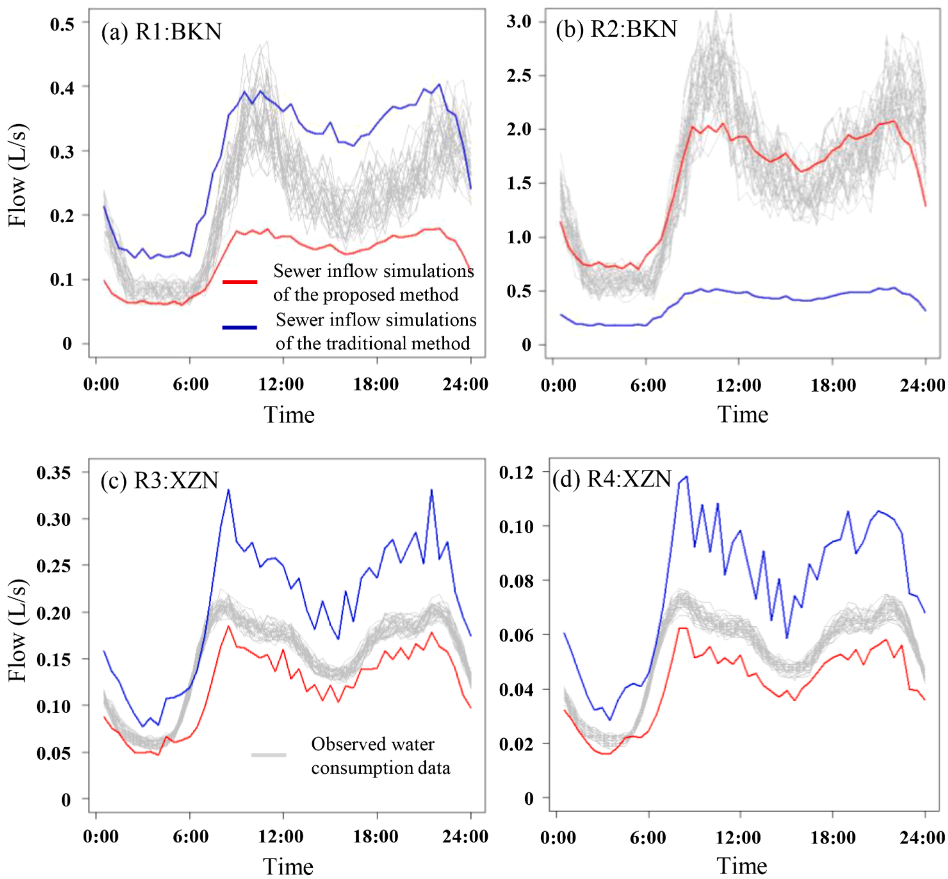


Fig. 11. Water consumption data versus sewer inflow predictions at four FSS manholes (R1-R4 are shown in Fig. 7) without sensors.

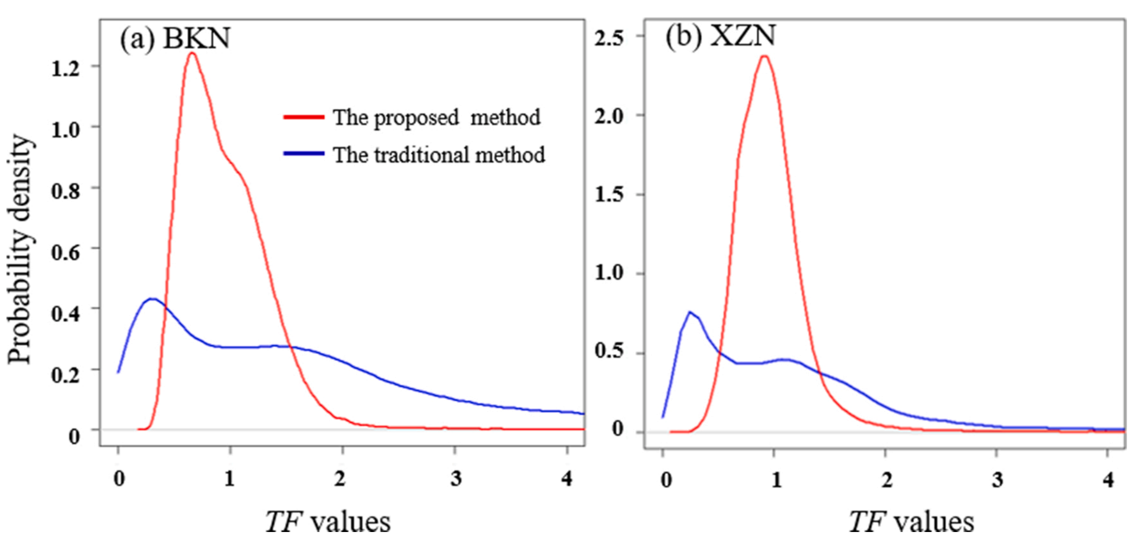


Fig. 12. Probability density distributions of the transfer factor (*TF*) values between the water consumption data and the corresponding wastewater discharges for residential users.

($CV_h(t) = 0.85 or 1.15$) uncertainty analysis method, respectively. As shown in this figure, the observations of the sewer hydraulic variables can be significantly varied at the same time periods but different days (grey lines in Fig. 14).

It can be observed from Fig. 14 that the proposed uncertainty analysis method is able to capture well the underlying variations of the observations at different FSS sensor locations. However, this is not the case for the traditional uncertainty analysis approach, as many of the observations are outside of the predicted ranges. To further visualize the performance of these two methods, Fig. 15 shows the uncertainty analysis results on the 24th day within the validation time period. As shown in this figure, the performance of the proposed uncertainty analysis method is appreciably better than the traditional approach in simulating the variations of the water depths or pipe flows. However, it is observed that few observations are still beyond the ranges identified by the proposed uncertainty analysis method (Fig. 15). This can be

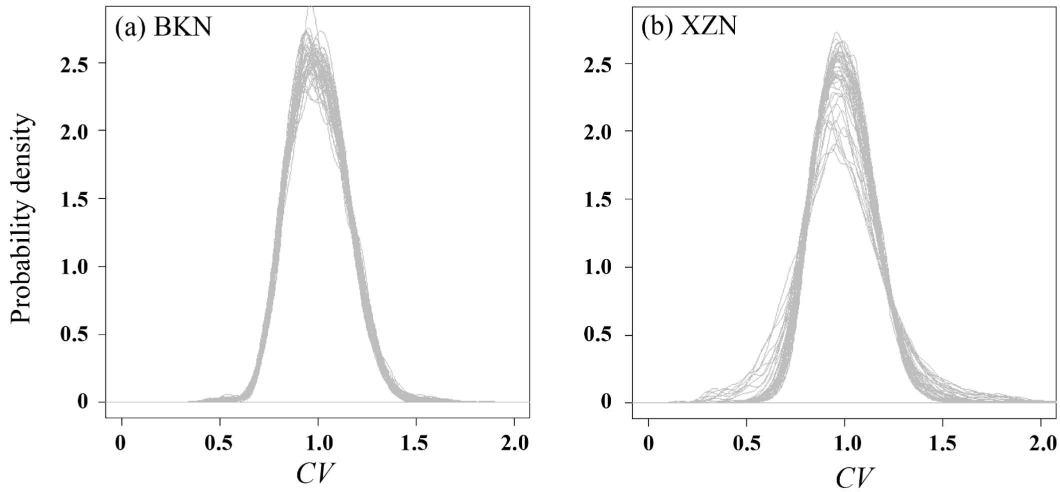


Fig. 13. The density distribution of *CV* values in each sampling pool ($\Psi(t)$), with 48 lines included for each case study.

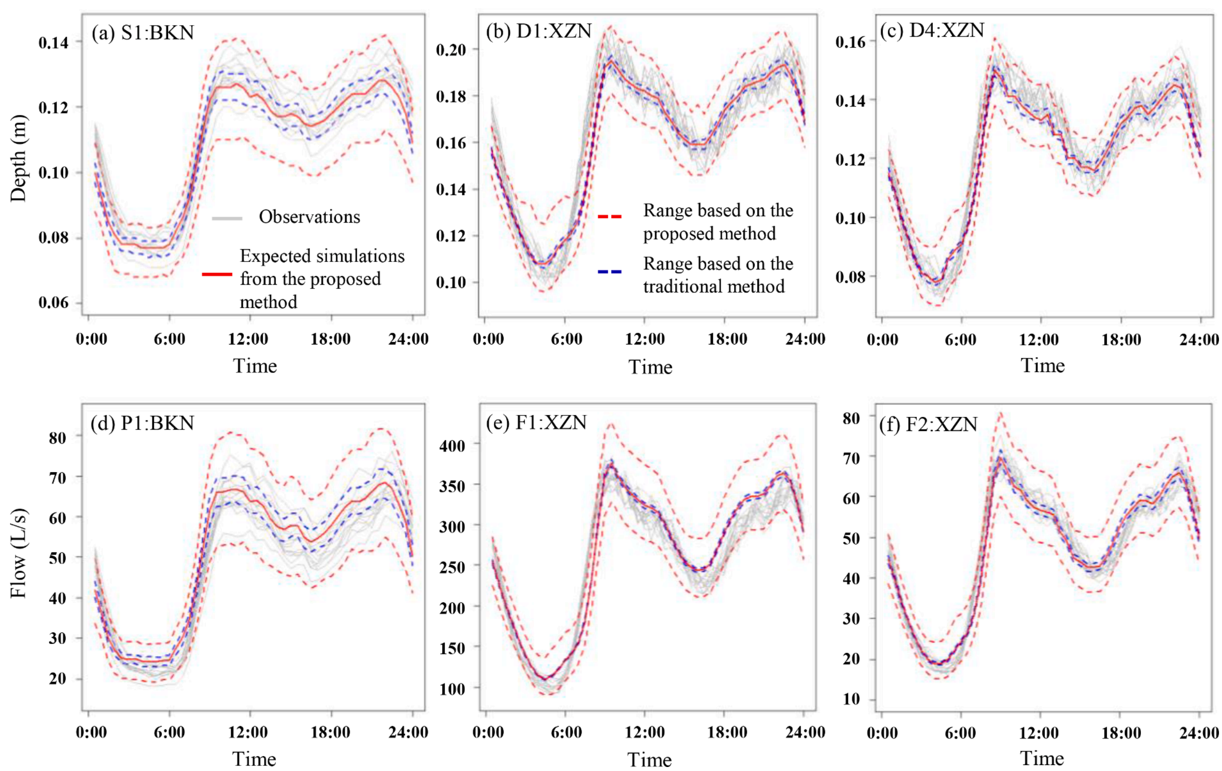


Fig. 14. Uncertainty ranges for the FSS sensor locations within the validation time period (S1, P1, D1, D4, F1 and F2 are shown in Fig. 7).

caused by a lack of the consideration of infiltration in this study, which should be accounted for in a future study. Similar observations can be made for other FSS sensor locations. This implies that the proposed uncertainty analysis method (based on the water consumption data) is significantly better than the traditional approach in representing the stochastic properties of the sewer hydraulic variables.

5. Conclusions

The present study proposes a new method for effectively calibrating the foul sewer system (FSS) model by using geotagged data and water consumption data from smart water metering. Based on the results obtained from two real case studies, the following conclusions are made:

(1) The proposed method provides similar or slightly better FSS hydraulic prediction accuracy at the locations with sensors when compared to the traditional approach. However, the proposed method produces significantly better prediction results at the FSS locations without sensors. This indicates that the proposed method can significantly improve the model performance by addressing the “equifinality” problem.

(2) The proposed uncertainty analysis method provides means to accurately estimate the variation bounds for water depths and flows influenced by different uncertainty factors. Therefore, it has the potential to improve the performance of certain practical applications (e.g. detection of blockages) when compared to traditional uncertainty estimation methods currently used.

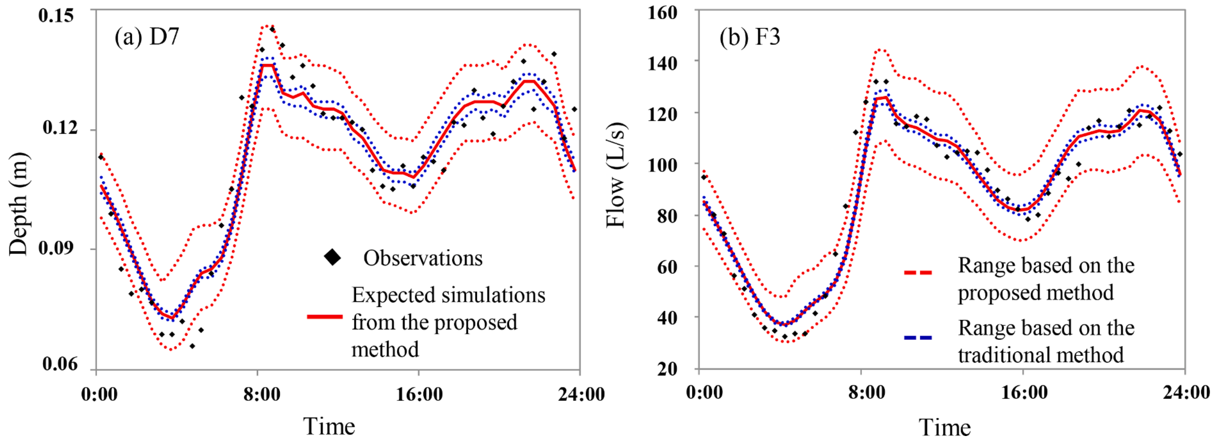


Fig. 15. Uncertainty ranges for two XZN sensor locations on the 24th day (validation period, D7 and F3 are shown in Fig. 7).

Having said above, some potential limitations remain to be addressed as part of future work of the proposed method, which are given as follows: (i) the inability to account for the impacts of the infiltration/exfiltration process, which may affect the model accuracy especially in an aged FSS or FSS in an area with groundwater; (ii) the incapability to deal with combined sewer systems where catchment runoff is present too; (iii) reliance on smart water metering data or geotagged data which may not be available and (iv) dealing with more complex FSSs that contain pumps, weirs and other control structures.

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.

Acknowledgments

This work is funded by the National Natural Science Foundation of China (Grant No. 51922096), and Excellent Youth Natural Science Foundation of Zhejiang Province, China (LR19E080003). The author Dr. HF Duan would like to appreciate the support from the Hong Kong Research Grants Council (RGC) (15200719).

References

Abdel-Aal, M., Mohamed, M., Smits, R., Abdel-Aal, R.E., De Gussem, K., Schellart, A., Tait, S., 2015. Predicting wastewater temperatures in sewer pipes using abductive network models. Water Sci. Technol. 71 (1), 137–144.
Ahm, M., Thorndahl, S., Nielsen, J.E., Rasmussen, M.R., 2016. Estimation of combined sewer overflow discharge: a software sensor approach based on local water level measurements. Water Sci. Technol. 74 (11), 2683–2696.
Almeida, M.C., Butler, D., Friedler, E., 1999. At-source domestic wastewater quality. Urban Water 1 (1), 49–55.
Bailey, O., Arnot, T.C., Blokker, E.J.M., Kapelan, Z., Vreeburg, J., Hofman, J.A.M.H., 2019. Developing a stochastic sewer model to support sewer design under water conservation measures. J. Hydrol. 573, 908–917.
Bechmann, H., Nielsen, M.K., Madsen, H., Poulsen, N.K., 1999. Grey-box modeling of pollutant loads from a sewer system. Urban Water 71–78.
Beheshti, M., Saegrov, S., 2018. Quantification assessment of extraneous water infiltration and inflow by analysis of the thermal behavior of the sewer network. Water 10 (8), 17.
Behzadian, K., Kapelan, Z., 2015. Modeling metabolism based performance of an urban water system using WaterMet(2). Resour. Conserv. Recycl. 99, 84–99.
Breinholt, A., Grum, M., Madsen, H., Thordarson, F.O., Mikkelsen, P.S., 2013. Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration inflow – consistency of containment ratios in calibration and validation? Hydrol. Earth Syst. Sci. 17 (10), 4159–4176.
Butler, D., Graham, N.J.D., 1995. Modeling dry weather waste-water flow in sewer networks. J. Environ. Eng. 121 (2), 161–173. Asce.
Carstensen, J., Nielsen, M.K., Strandbaek, H., 1998. Prediction of hydraulic load for urban storm control of a municipal WWT plant. Water Sci. Technol. 37 (12), 363–370.
Chu, S., Zhang, T., Yu, T., Wang, Q.J., Shao, Y., 2021. A noise adaptive approach for nodal water demand estimation in water distribution systems. Water Res. 192.
Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P.R., Walski, T., 2018. Real time control of water distribution networks: a state-of-the-art review. Water Res. 161, 517–530.
De Keyser, W., Gevaert, V., Verdonck, F., De Baets, B., Benedetti, L., 2010. An emission time series generator for pollutant release modeling in urban areas. Environ. Model. Softw. 25 (4), 554–561.
Draude, S., Keedwell, E., Hiscock, R., Kapelan, Z., 2019. A statistical analysis on the effect of preceding dry weather on sewer blockages in South Wales. Water Sci. Technol. 80 (12), 2381–2391.
Gironas, J., Roesner, L.A., Rossman, L.A., Davis, J., 2010. A new applications manual for the storm water management model (SWMM). Environ. Model. Softw. 25 (6), 813–814.
Guo, D., Zheng, F., Gupta, H., Maier, H.R., 2020. On the robustness of conceptual rainfall-runoff models to calibration and evaluation data set splits selection: a large sample investigation. Water Resour. Res. 56 (3).
Hadka, D., Reed, P., 2013. Borg: an auto-adaptive many-objective evolutionary computing framework. Evolut. Comput. 21 (2), 231–259.
Jin, Y., Mukherjee, A., 2010. Modeling blockage failures in sewer systems to support maintenance decision making. J. Perform. Constr. Facil. 24 (6), 622–633.
Khu, S.T., di Pierro, F., Savic, D., Djordjevic, S., Walters, G.A., 2006. Incorporating spatial and temporal information for urban drainage model calibration: An approach using preference ordering genetic algorithm. Adv. Water Res. 29 (8), 116–1181.
Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. 23 (10), 4323–4331.
Korving, H., Clemens, F., 2005. Impact of dimension uncertainty and model calibration on sewer system assessment. Water Sci. Technol. 52 (5), 35–42.
Langergraber, G., Alex, J., Weissenbacher, N., Woerner, D., Ahnert, M., Frehmann, T., Halft, N., Hobus, I., Plattes, M., Spering, V., Winkler, S., 2008. Generation of diurnal variation for influent data for dynamic simulation. Water Sci. Technol. 57 (9), 1483–1486.
Lepot, M., Torres, A., Hofer, T., Caradot, N., Gruber, G., Aubin, J.B., Bertrand-Krajewski, J.L., 2016. Calibration of UV/Vis spectrophotometers: A review and comparison of different methods to estimate TSS and total and dissolved COD concentrations in sewers, WWTPs and rivers. Water Res. 101, 519–534.
Liu, Y., Tugtas, A.E., Sharma, K.R., Ni, B.-J., Yuan, Z., 2016. Sulfide and methane production in sewer sediments: field survey and model evaluation. Water Res. 89, 142–150.
Mannina, G., Viviani, G., 2009. Separate and combined sewer systems: a long-term modeling approach. Water Sci. Technol. 60 (3), 555–565.
Maurer, M., Scheidegger, A., Herlyn, A., 2013. Quantifying costs and lengths of urban drainage systems with a simple static sewer infrastructure model. Urban Water J. 10 (4), 268–280.
McCall, A.K., Bade, R., Kinyua, J., Lai, F.Y., Thai, P.K., Covaci, A., Bijlsma, L., van Nuijs, A.L.N., Ort, C., 2016. Critical review on the stability of illicit drugs in sewers and wastewater samples. Water Res. 88, 933–947.
Montes, C., Kapelan, Z., Saldarriaga, J., 2020. Predicting non-deposition sediment transport in sewer pipes using random forest. Water Res. 189.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles - Science direct. J. Hydrol. 10 (3), 282–290.
Rodriguez, J.P., McIntyre, N., Diaz-Granados, M., Achleitner, S., Hochedlinger, M., Maksimovic, C, 2013. Generating time-series of dry weather loads to sewers. Environ. Model. Softw. 43, 133–143.
Rokstad, M.M., Ugarelli, R.M., 2015. Evaluating the role of deterioration models for condition assessment of sewers. J. Hydroinf. 17 (5), 789–804.
See, C.H., Horoshenkov, K.V., Tait, S.J., Abd-Alhameed, R.A., Hu, Y.F., Elkhazmi, E.A., Gardiner, J.G., Ieee, 2009. A Zigbee Based Wireless Sensor Network for Sewerage Monitoring. Asia-Pacific Microw. Conf., pp. 731–734

Sitzenfrei, R., Fach, S., Kinzel, H., Rauch, W., 2010. A multi-layer cellular automata approach for algorithmic generation of virtual case studies: VIBe. Water Sci. Technol. 61 (1), 37–45.
Sun, N., Hong, B.G., Hall, M., 2014. Assessment of the SWMM model uncertainties within the generalized likelihood uncertainty estimation (GLUE) framework for a high-resolution urban sewershed. Hydrol. Process. 28 (6), 3018–3034.
Sweetapple, C., Astaraie-Imani, M., Butler, D, 2018. Design and operation of urban wastewater systems considering reliability, risk and resilience. Water Res. 147, 1–12.
Talaiekhozani, A., Bagheri, M., Goli, A., Khoozani, M.R.T., 2016. An overview of principles of odor production, emission, and control methods in wastewater collection and treatment systems. J. Environ. Manag. 170, 186–206.
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. J. Water Resour. Plan. Manag. 144 (10).
Zhang, Q., Zheng, F., Jia, Y., Savic, D., Kapelan, Z., 2021. Real-time foul sewer hydraulic modeling driven by water consumption data from water distribution systems. Water Res. 188, 116544.
Zheng, F., Tao, R., Maier, H.R., See, L., Savic, D., Zhang, T., Chen, Q., Assumpcao, T.H., Yang, P., Heidari, B., Rieckermann, J., Minsker, B., Bi, W., Cai, X., Solomatine, D., Popescu, I., 2018. Crowdsourcing methods for data collection in geophysics: state of the Art, issues, and future directions. Rev. Geophys. 56 (4), 698–740.
Zheng, F., Zecchin, A.C., Maier, H.R., Simpson, A.R., 2016. Comparison of the searching behavior of NSGA-II, SAMODE, and Borg MOEAs applied to water distribution system design problems. J. Water Resour. Plan. Manag. 142 (7).