Journal of Cleaner Production 365 (2022) 132676

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Asynchronous sensor networks for Nodal water demand estimation in water distribution systems based on sensor grouping analysis

Tingchao Yu [a], Ben Lin [a], Zhihong Long [b], Yu Shao [a], Iran E. Lima Neto [c], Shipeng Chu [a,*]

[a] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, CHINA
[b] Guangzhou Water Supply Co., Ltd., Guangzhou, 510600, CHINA
[c] Department of Hydraulic and Environmental Engineering, Federal University of Ceará, Brazil

ARTICLE INFO

Handling Editor: Kathleen Aviso

Keywords:
State estimation
Water distribution
Asynchronous uploading sensors
Nodal water demand

ABSTRACT

Real-time hydraulic models are important tools for the management of water distribution systems (WDS). In such systems, high frequency nodal water demands are required for real-time modeling. Due to the limitations of the sensors' technology and power supply, the sensors distributed in the WDS typically upload data at a low frequency. However, replacing all existing low frequency sensors with new high frequency sensors is not cost effective. To solve this problem, an asynchronous data uploading strategy is proposed, which uses current low frequency sensors to estimate nodal water demand at high frequency while maintaining model accuracy. Based on an innovative clustering algorithm, the method splits the information redundancy sensors into multiple groups. Sensors in different groups asynchronously upload data at different time points to estimate nodal water demand. Applications to a simple hypothetical WDS and a realistic WDS demonstrate that the developed approach can efficiently improve the data upload frequency of the sensor network, thus boosting the demand estimation frequency. The developed method is expected to reduce the cost of upgrading sensor networks and increase the efficiency of WDS modeling, thus facilitating the cleaner production and sustainable management of WDS.

1. Introduction

The water distribution system (WDS) is an important component of a city's public facilities, delivering water to meet the needs of residents, industry, and commerce (Dave and Layton, 2020). However, the expansion of WDS has encountered many challenges, including excessive energy consumption (Salomons and Housh, 2020), water leakage (Zhang et al., 2016), pipe bursts (Qi et al., 2018), and water contamination (Bazargan-Lari, 2014). These challenges highlight the requirements to develop strategies for the cleaner production and sustainable management of WDS (Duan et al., 2020). In this context, hydraulic models are effective tools for water utility personnel to facilitate the control and management of WDS. WDS networks' working circumstances, such as nodal pressure, pipe flow, and water quality, can be dynamically simulated using real-time models. Thus, real-time WDS models have been used for system scheduling, leakage control, and burst pipe analysis. Among the relevant parameters in WDS models, nodal water demands vary with time and cannot be directly measured due to current technical and economic constraints. In this sense, the nodal water demand is a critical state variable, which can have a big impact on modeling accuracy (Kang and Lansey, 2009; Chu et al., 2020; Huang et al., 2017). Therefore, nodal water demands must be estimated with high frequency to meet the requirement of cleaner production and sustainable management of WDS.

A WDS model may contain thousands of nodes with unknown water demands that must be estimated (Savic et al., 2009). In contrast, only limited measurements are available as gauged measurements are taken only at a few selected locations (Kang and Lansey, 2009). To obtain an acceptable modeling result, efficient and precise algorithms are required to provide an acceptable modeling result. In recent decades, explicit and optimization methods have been widely used to estimate the nodal water demands in WDS. The explicit method has excellent computational performance (Davidson and Bouchart, 2006) and can give an explicit solution to hydraulic equations in large-scale WDS (Boulos et al., 1991; Rajakumar et al., 2019). However, it requires that the number of unknown variables (nodal water demand) is equal to or less than the

* Corresponding author.
E-mail addresses: yutingchao@zju.edu.cn (T. Yu), tslza@zju.edu.cn (B. Lin), lzh73@126.com (Z. Long), shaoyu1979@zju.edu.cn (Y. Shao), iran@deha.ufc.br (I.E. Lima Neto), chushipeng@zju.edu.cn (S. Chu).

https://doi.org/10.1016/j.jclepro.2022.132676
Received 31 October 2021; Received in revised form 9 June 2022; Accepted 11 June 2022
Available online 14 June 2022
0959-6526/

number of known variables (measurements) (Okeya et al., 2014; Liu et al., 2017). To make the equation solvable, the unknown variables are usually aggregated into some groups, and the group number should be equal to or fewer than the total number of sensors (Du et al., 2015). Furthermore, the method seldom gives a proper measurement of the modeling uncertainties (Muste et al., 2012). Regarding the optimization method, it considers the nodal water demands as decision variables and constructs an objective function to eliminate the residuals between the model output values and the real-time measurements. Evolutionary algorithm (EA)-based methods (Sabbaghpour et al., 2012; Dini and Tabesh, 2014; Do et al., 2016) and sensitivity matrix-based methods (Cheng and He, 2011; Sanz and Pérez, 2015) are two typical methods of this category. The EA-based method does not rely on complex mathematical concepts and computations, but it generally evaluates the fitness value of each individual, leading to high computational complexity (Zhang et al., 2018). With the increase of decision variables in large-scale networks, this becomes the biggest shortcoming of the EA-based method (Zheng et al., 2017). The sensitivity matrix-based method commonly has a less computational cost because it takes advantage of some features of the WDS (Piller et al., 2017). However, due to the insufficient number of sensors compared to the total number of state variables, even if the model outputs match the measurements, the estimated nodal water demand still has potential uncertainty.

As for requiring no inherent assumptions and being suitable for quantifying uncertainties, Bayesian approaches have become popular in real-time hydraulic modeling of WDS (Kapelan et al., 2007). Hutton et al. (2014) discussed formal and informal Bayesian methods in their performance in reducing uncertainty. Besides, some methods avoid making normal assumptions or likewise (Malve et al., 2007; Xie et al., 2017; Qin and Boccelli, 2019). In recent years, data assimilation methods have been frequently employed to estimate nodal water demand in real-time (Bragalli et al., 2016). Data assimilation can estimate nodal water demand by combining the prior information with real-time measurements (Chu et al., 2021; Do et al., 2017; Shang et al., 2008; Zhou et al., 2018). Compared to other methods, the data assimilation methods can estimate the nodal water demand in a probability form, typically a normal distribution characterized by mean and variance (Chu et al., 2020, 2021).

The aforementioned techniques try to enhance modeling accuracy by combining measurements and other information (e.g., customer billing), but they ignore the impact of sensor features such as data upload frequency. A lower sensor upload frequency indicates a large time delay between two consecutive estimates. The data assimilation method typically obtains a prior probability of the nodal water demand from the previous time step or the historical data (Law et al., 2015). Thus a lower sensor upload frequency will reduce the reliability of prior prediction and decrease the estimation accuracy. Moreover, this makes it difficult to simulate WDS work conditions in a timely manner, reducing the usefulness of real-time models (Chu et al., 2020).

In the last 30 years, China has placed a vast number of sensors in WDS networks. Due to the limitations of the sensors' technology and power supply, these sensors typically upload data at a low frequency. This creates a problem: the sensor upload frequency is insufficient to fulfill the relatively high frequency requirement of real-time modeling. For example, as a requirement for system operation, the real-time model should estimate the nodal water demand every 15 min, while existing sensors are limited to upload data every 60 min. Although new sensors with higher upload frequency are already available, installing a significant number of new sensors in a short period of time or replacing all existing low frequency sensors with new sensors is not cost effective. Additionally, given the integrity of the current sensor networks and the reliability of the available time series, one strategy to deal with the above dilemma is to change the sensor upload strategy from synchronous to asynchronous, in order to estimate higher frequency nodal water demands at a lower cost.

In this study, an asynchronous data uploading approach is proposed to estimate nodal water demand at a high frequency using existing low frequency sensors. The method divides the low frequency sensors into several groups based on an innovative clustering algorithm. Sensors in different groups asynchronously upload data at different time points. The uploaded data is used by a data assimilation approach to estimate the nodal water demand. The proposed approach makes full use of the existing low frequency sensors in the WDS to estimate high frequency nodal water demands, which is expected to reduce the cost of upgrading the sensor network and increase the efficiency of WDS modeling.

2. Methodology

A schematic of the proposed methodology is given in Fig. 1, with details presented in the subsequent sections.

2.1. Nodal water demand estimation

The proposed asynchronous sensor networks are used for nodal water demand estimation in WDS steady-state flow hydraulic model. In this paper, a data assimilation method proposed by Chu et al. (2020) was adopted to estimate the nodal water demand. The data assimilation method consists of two stages. In the first stage, the nodal water demand at the current time step was estimated based on historical data or other information in a probability form, which is called the prior probability density function (PDF). In the second stage, the estimated prior PDF is further corrected by the real-time measurements.

In the first stage, historical demand can be used to predict nodal water demand in the current time step, as shown in Eq. (1).

$$x_{t|t} = f\left(x_{t-1|t-1}, x_{t-2|t-2}, \ldots\right) + \varepsilon_t \tag{1}$$

where $x_{t|t} \in \mathbb{R}^{n\times 1}$ is the nodal water demand predicted from historical data $(x_{t-1|t-1}, x_{t-2|t-2}, \ldots)$ and $f()$ is the prediction function; ε_t is the prediction error; n is the number of nodal water demand. Typically, the prediction error ε_t is a normally distributed variable, $\varepsilon_t \sim N(0, V_{t|t-1})$ and $V_{t|t-1}$ is the covariance matrix; $N()$ is the normal distributed PDF. Denoting $f(x_{t-1|t-1}, x_{t-2|t-2}, \ldots)$ as $x_{t|t-1}$, the nodal water demand ($x_{t|t} \in \mathbb{R}^{n\times 1}$) at time t also obey the normal distribution (prior PDF).

$$x_{t|t} \sim N\left(x_{t|t} \middle| x_{t|t-1}, V_{t|t-1}\right) \tag{2}$$

The growing use of automatic water meters enables for direct estimation of the nodal water demand. Furthermore, the nodal water demand can also be estimated based on the consumer billings or the served population. Chu et al. (2020) developed a method to use that information as well as the historical prediction data (Eq. (1)) to formulate the prior PDF (Eq. (2)).

In the second stage, the prior PDF (Eq. (2)) is combined with the real-time measured data to estimate nodal water demand. For a steady-state based WDS hydraulic model with m sensors, the measured data ($y_t \in \mathbb{R}^{m\times 1}$) differs from the model outputs due to the existence of noise η_t.

$$y_t = g\left(x_{t|t}\right) + \eta_t \tag{3}$$

where $g()$ is the steady-state based WDS hydraulic model and $g(x_{t|t}) \in \mathbb{R}^{m\times 1}$ is the model outputs corresponding to the measurements, given the nodal water demand $x_{t|t}$; $y_t \in \mathbb{R}^{m\times 1}$ is the measured data and $\eta_t \in \mathbb{R}^{m\times 1}$ is the measurement noise.

The steady-state model implies that the sensor distributed in the network can measure the steady-state pressure or pipe flow rate, despite being inevitably contaminated by background noises. Typically, the background noises are assumed normally distributed. However, transient waves are inevitably present in the WDS network. It usually occurs as a result of pump startup or abnormal shut-down, quick valve closing or opening, etc. (El-Ghandour et al., 2021; Zhang et al., 2022). In

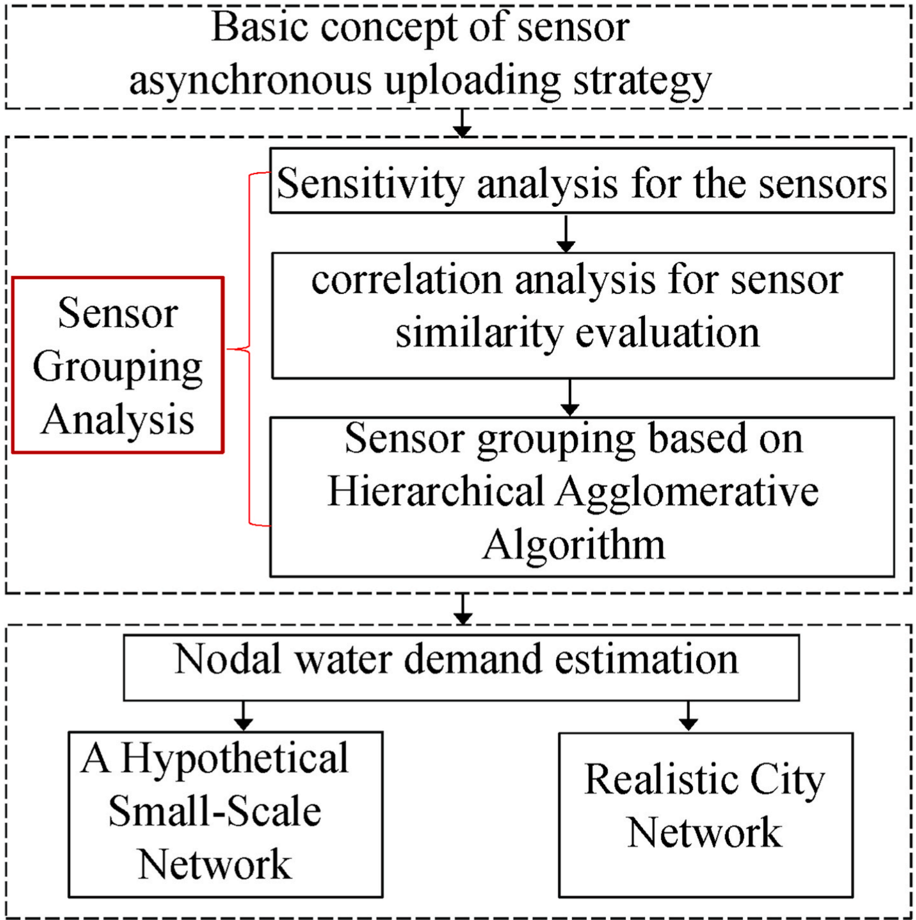

Fig. 1. Schematic of the proposed method.

addition to background noises, data anomalies due to transient events are also widely present in measurements. Therefore, noise (η_t) probably has a more complex and time-dependent structure. Yet the complexity can make it difficult to linearize the nonlinear systems in real-time network modeling.

Here, this difficulty is analyzed by assuming a normal distribution of noise $\eta_t \sim N(\eta_t|0, R)$ and $R \in \mathbb{R}^{m\times m}$ is the noise covariance matrix. Based on Eq. (3), the conditional PDF for measurement y_t can be written as Eq. (4).

$$P(y_t|x_{t|t}) = N[y_t|g(x_{t|t}), R] \tag{4}$$

By fusion of the prior PDF (Eq. (2)) and conditional PDF (Eq. (4)), the posterior PDF can be written as,

$$P(x_{t|t}|y_t) \propto N[y_t|g(x_{t|t}), R] \times N(x_{t|t}|x_{t|t-1}, V_{t|t-1}) \tag{5}$$

The nodal water demand $x_{t|t}$ is estimated by maximizing the logarithm of $P(x_{t|t}|y_t)$, which is equal to minimizing the objective function as shown in Eq. (6).

$$J(x_{t|t}) = [g(x_{t|t}) - y_t]^T R^{-1} [g(x_{t|t}) - y_t] + (x_{t|t} - x_{t|t-1})^T V_{t|t-1}^{-1} (x_{t|t} - x_{t|t-1}) \tag{6}$$

The right-hand side of Eq. (6) consists of two parts: the first tries to minimize the square deviation between model outputs ($g(x_{t|t})$) and measurements (y_t), while the second part aims to reduce the deviation between the nodal water demand ($x_{t|t}$) to its prior value ($x_{t|t-1}$). The solution of Eq. (6) can be found in Chu et al. (2020).

2.2. Sensor upload strategy

The sensors distributed in the WDS network typically upload measurements at a regular time interval (e. g. 5 min, 15 min, 30 min). Then, these data (y_t) are used to estimate the nodal water demand (Eq. (6)). The frequency of measurement uploading is determined by the properties of the device (sampling frequency, battery life, etc). For the operation and management of WDS, a large number of sensors have been installed in the WDS network in the past few decades. Existing sensors typically upload the data with a relatively low frequency (e.g. 30 min, 60 min) due to the limitation of the sensors' hardware and power supply. As shown in Fig. 2(a), the network has a certain number of sensors installed and these sensors upload the measurements every 30 min synchronously (synchronous uploading systems). The upload frequency of the existing sensors allows us to estimate the nodal water demand every 30 min. This frequency typically cannot meet the requirement of real-time modeling and management of WDS (typically 5–10 min in China). There is a need to improve the sensor upload frequency. A simple strategy is to remove the existing low frequency sensors and install high frequency sensors. Obviously, this strategy will greatly increase the cost, especially for large-scale networks. This study proposes a more economical strategy by switching the sensor uploading strategy from synchronous to asynchronous uploading.

Fig. 2(b) gives the basic concept of sensor asynchronous uploading strategy. As shown in Fig. 2(b), the sensors are divided into several groups. The upload time for the sensors in each group is set at a specific interval. As shown in Fig. 2(b), the sensors in Group1 upload the data at the original time point (0:00, 0:30, 1:00, …), whereas the uploading time for the sensors in Group 2 and Group 3 is delayed by 10 min (uploading data at 0:10, 0:40,1:10, …) and 20 min (uploading data at

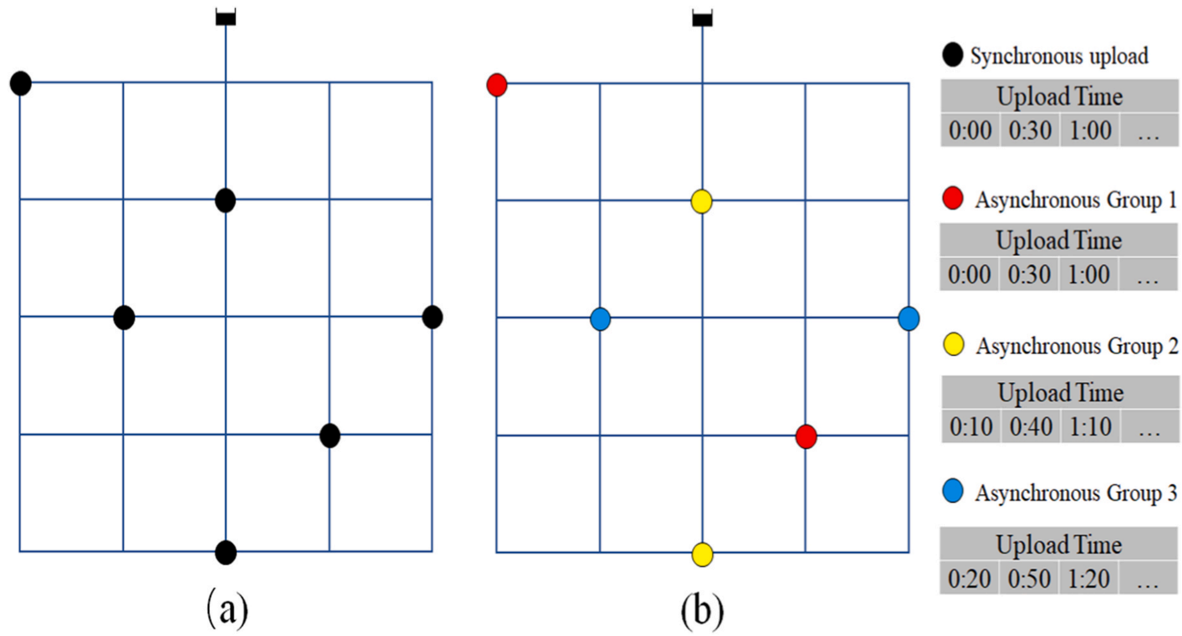

Fig. 2. Synchronous (a) and asynchronous (b) uploading systems.

0:20, 0:50,1:20, …), respectively. This asynchronous uploading strategy allows one to estimate the nodal water demand every 10 min using the data uploaded from different sensor groups.

The asynchronous system, in comparison to the synchronous system, increases the data upload frequency, while reducing the amount of available data (number of measurements) at each time step. For example, for the synchronous system, the estimator can receive 6 data every 30 min (Fig. 2a). However, for the asynchronous system, the estimator only receives 2 data every 10 min (Fig. 2(b)). The number of measurements has a significant impact on the accuracy of demand estimation. As a result, the main drawback of the asynchronous strategy is that it reduces the number of available measurements at each time step, which in turn has the potential to reduce the estimation accuracy. To address this issue, an efficient sensor grouping technique is proposed to make the data uploaded by each group of sensors more representative.

2.3. Sensor grouping strategy

Fig. 3 shows two alternative sensor grouping examples to help illustrate the sensor grouping technique. There are 12 sensors installed in the network and these sensors are divided into two groups. The sensors allow us to estimate the nodal water demand every 15 min. Intuitively, the sensor grouping in Fig. 3(a) is less reasonable than that in Fig. 3(b). This is because the neighboring sensors are aggregated into one group in Fig. 3(a), resulting in too much hydraulic correlation of sensors in the same group and serious homogenization of measured data. On the one hand, sensors in the same group upload homogenization measurements, leading to information redundancy; On the other hand, the measurements within the group only provide local hydraulic information about the WDS network, which makes it difficult to effectively estimate the nodal water demand in the entire WDS network. This indicates that sensors that collect quite similar information should be assigned to different groups to improve the representativeness of the data (Fig. 3(b)). Moreover, sensors in the same group should be distributed in the entire WDS network, which can ensure that each group has sufficient information about the entire network.

To find those sensors that upload similar information in the WDS, the sensitivity matrix analysis is applied to measure the redundancy of the measurements. Farley et al. (2010) presented a methodology to generate the sensitivity matrix by simulating a leakage at every node in the hydraulic model. A typical sensitivity matrix is shown in Eq. (7). Detailed information about the sensitivity matrix can be found in Farley et al.

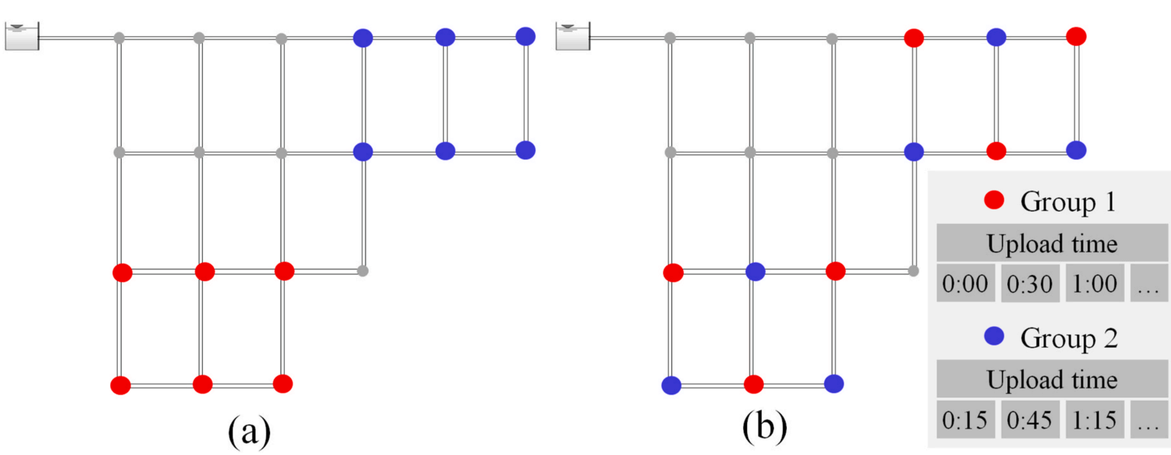

Fig. 3. Different sensor grouping scenarios for asynchronous upload system.

(2010).

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1m} \\ S_{21} & S_{22} & \dots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \dots & S_{nm} \end{bmatrix} \tag{7}$$

where S_{ij} is the sensitivity of the pressure at node i and the pressure at node j; n is the number of nodes; m is the number of sensors.

The column vector of the sensitivity matrix can be used to evaluate the sensitivity of the corresponding sensor to every node in the network. This means that the similarity of two column vectors can measure the similarity of the corresponding two sensors. Sensors with similar column vectors indicate that they provide redundant information. Therefore, these sensors should be assigned to different groups. In other words, sensors that are dissimilar to each other should be assigned to the same group.

In this study, the hierarchical agglomerative algorithm (Murtagh and Legendre, 2014) is used to group the sensors. The hierarchical agglomerative algorithm merges the most similar pair of groups gradually based on the distance between groups. The distance is calculated from features that reflect the monitoring capability of sensors. Clustering with too many features is often infeasible due to the curse of dimensionality (Cai et al., 2005). However, the number of features in this study is equal to the number of nodes in the WDS (typically more than 1000), far beyond the ability that the clustering algorithm can handle. To overcome this difficulty, the correlation coefficient of the sensor sensitivity vectors is used to measure the similarity between two sensors. The correlation coefficient is defined as the distance between two sensors. The distance matrix can be formulated as:

$$D = \begin{bmatrix} 1 & dist_{1,2} & \dots & dist_{1,m} \\ dist_{2,1} & 1 & \dots & dist_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ dist_{m,1} & dist_{m,2} & \dots & 1 \end{bmatrix} \tag{8}$$

$$dist_{i,j} = corr(S(i),\ S(j)) \tag{9}$$

where $dist_{i,j}$ is the distance between the ith sensor and the j th sensor; $S(i)$ is the i th column vector of the sensitivity matrix S; $corr(S(i),\ S(j))$ is the correlation coefficient of vector $S(i)$ and vector $S(j)$.

Equations (8) and (9) give the distance for the m sensors. As shown in Eq. (9), the distance between sensor i and sensor j is their correlation coefficient. Considering that correlation coefficients can measure the similarity of two sensors, Eq. (9) allows similar sensors to have greater distance. Based on Eqs. (8) and (9), the average distance between two clusters can be calculated by Eq. (10).

$$L_{i,j} = \frac{\sum_{k=1}^{n_i} \sum_{l=1}^{n_j} dist\left(x_{i,k}, x_{j,l}\right)}{n_i \cdot n_j} \tag{10}$$

where $L_{i,j}$ = the distance between group i and group j; n_i = the number of sensors in group i; $x_{i,k}$ = the kth sensor in the group i.

A pair of groups that have the shortest average distance will be merged into a new group. The clustering process will continue until the desired number of groups is reached. Ideally, all sensors can be clustered into several groups for asynchronous upload and nodal water demand estimation. Nevertheless, the clustering algorithm may get a seriously unbalanced number of sensors in each group. There will likely be that a group contains several sensors, while some other groups only include a small number of sensors. In this situation, the group with a large number of sensors should be re-clustered and divided into smaller groups. The newly generated groups need to merge with other existing groups to make the total number of groups equal to the desired number.

3. Case studies and discussion

3.1. Case study 1: A hypothetical small-scale network

This hypothetical network contains 381 nodes, 469 pipes, and 1 reservoir (Fig. 4), which has been investigated in previous studies (Zhang et al., 2016; Chu et al., 2020). In this case study, 16 pressure sensors are placed in the network. These pressure sensors are all assumed to have an average noise of 0.3 m and upload the nodal pressure every 30 min. The reservoir head is considered a known value. In addition, a flow meter with a noise of 1 L/s collects the outflow from the reservoir. The real nodal water demands at each time step are generated randomly. The initial value of prior nodal water demands is established based on the length of the connected pipes and the location of the nodes.

The sensor clustering strategy divides the 16 sensors into two groups, each of which has 8 sensors (Fig. 5(a)). The sensors in Group1 upload the data at the original time point (0:00, 0:30, 1:00, …), whereas the uploading time for the sensors in Group 2 is delayed for 15 min (uploading data at 0:15, 0:45,1:15, …). As a result, the grouped sensors upload data every 15 min asynchronously, and 8 measurements are used to estimate the nodal water demand at each time step. A random grouping of the 16 sensors is used for comparison to test the usability of the clustering strategy, as shown in Fig. 5. (b). The above two scenarios are denoted as Cluster-Grouping_ASY and Random-Grouping_ASY. Besides, the synchronous uploading scenario that uses all 16 pressure sensors and a 30 min estimation frequency is also provided. The above three different data uploading scenarios are adopted to estimate the nodal water demand by solving Eq. (6).

The modeling results of the 16 sensors are shown in Figs. S1–S16 (Supplementary 1). To facilitate the analysis, the modeling result of Sensor 6 is shown in Fig. 6. As shown in this figure, the deviations (errors) between estimation results (green, red and blue lines) and real values (black line) are less than those between the measured value (gray line). This is because the nodal water demand estimation algorithm can assimilate both the current and historical measurements to improve the modeling accuracy. Besides, the deviations (errors) for the asynchronous strategy (green line and red line) are slightly higher than those for the synchronous uploading strategy (blue line). The main reason is that the asynchronous strategy (green and red lines) uses only 8 measurements to estimate the nodal water demand, whereas the synchronous strategy (blue line) uses all 16 measurements in one time step. We can also find that the errors for the Cluster-Grouping_ASY are much less than that of Random-Grouping_ASY. As shown in Table 1, the average error for the Random-Grouping_ASY is 0.17 m, whereas the error for Cluster-Grouping_ASY is 0.14 m, just 0.01 m lower than that for the synchronous strategy (0.13m). This highlights the usability of the sensor grouping strategy to increase the nodal demand estimation frequency while maintaining accuracy compared to the synchronous uploading system.

Fig. 7 gives the cumulative probability for the absolute errors of 381 estimated nodal water demands. It is expected that the errors for the asynchronous strategy (green and red line) are greater than those for the synchronous uploading strategy (black line). The proposed asynchronous uploading strategy has the advantage of increasing upload frequency, but it has the disadvantage of reducing the number of available measurements at each time step, and thus in turn reduces the estimation accuracy. Such shortcoming can be mitigated by the proposed grouping strategy. As shown in Fig. 7, the estimation errors of Cluster-Grouping_ASY (red line) are smaller than those of the Random-Grouping_ASY (green line). For the Cluster-Grouping_ASY, the estimated nodal demand error for 95% of the nodes is within 0.12 L/s and the errors for about 2% of the nodes exceed 0.2 L/s. However, for the Random-Grouping_ASY, the errors for 5% of the nodes exceed 0.2 L/s and the errors for some nodes reach 0.5 L/s, which is a relatively large error for this network. This improvement can be attributed to the sensor grouping strategy, which makes the data uploaded by each group of

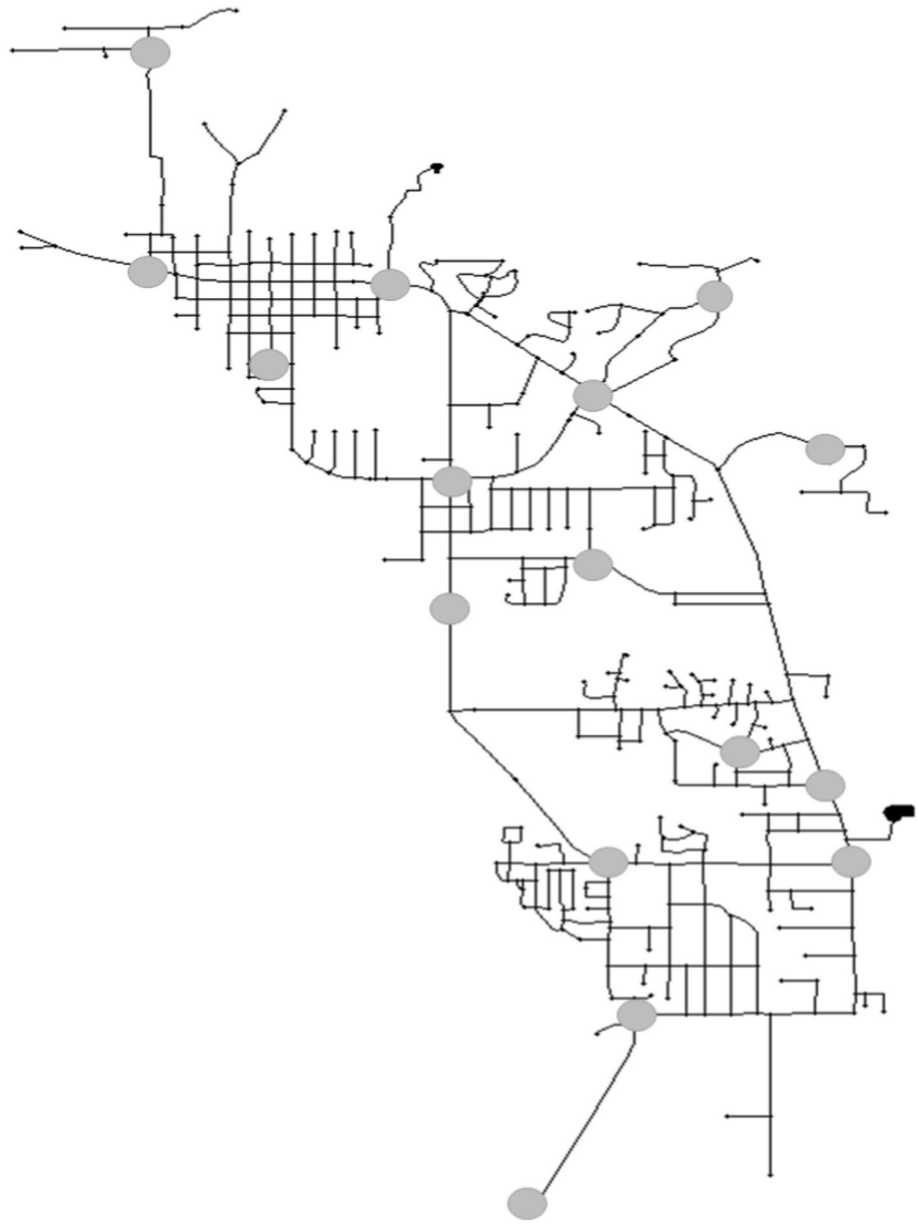

Fig. 4. Schematic of the hypothetical small-scale network and location of sensors.

sensors more representative and thus improves the accuracy of the estimation.

3.2. Case 2: realistic city network

Case 2 is used to demonstrate the performance of the developed method when applied to a realistic large-scale WDS network with field data. As shown in Fig. 8, the network consists of three reservoirs, 4242 nodes, 4841 pipes, and 63 pressure sensors installed in the network. Among the 63 pressure sensors, 48 sensors are used to estimate the nodal water demand, which are set to upload data every 30 min. The other 15 sensors are set to upload data every 15 min. These 15 sensors, which are not included in the demand estimation process, are used to validate the modeling accuracy by computing the deviation between the model output and measured data. The 48 sensors are assigned to two groups by the proposed sensor grouping algorithm. Each group contains 24 sensors, and the data is uploaded asynchronously. The sensors in Group 1 upload data at 0:00, 0:30, 1:00, …, while the sensors in Group 2 upload data at 0:15, 0:45, 1:15, …. Therefore, the estimator receives data every 15 min to estimate the nodal water demand. The standard deviation of the noise of each pressure measurement is 1 m because of outdated sensors. The outflow from three reservoirs is monitored by flow sensors. One day's data are used to verify the usability of the proposed approach.

The comparison of the model output and measured data from the 48 sensors can be found in Figs. S17–S64 (Supplementary 2). Specifically, the modeling results of Sensor *31 are shown in Fig. 9. The model output value always approximates the measured value (Fig. 9). This implies that the proposed approach can efficiently solve the estimation problem by minimizing the square deviation between model output and measurements (Eq. (6)). Table S1 (Supplementary 3) shows the average absolute residuals (AAR) for the 48 sensors. The AARs of the 48 sensors are within 0.90 m, indicating a very good matching with the measurements.

Figure S65–S79 (Supplementary 4) gives the model outputs and measured data of the 15 validation sensors. Furthermore, the AARs for the 15 validation sensors are shown in Table S2 (Supplementary 5). From Figs. S65–S79, we can find that the model outputs can match the measured data for most of the validation sensors. The AARs of the 15

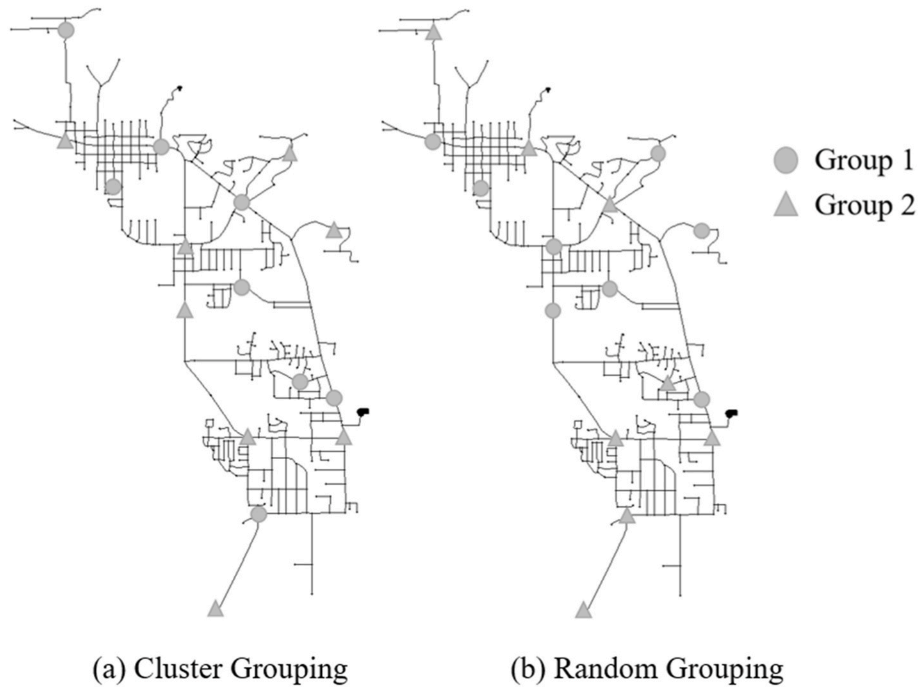

Fig. 5. Result of grouping applied to all pressure sensors.

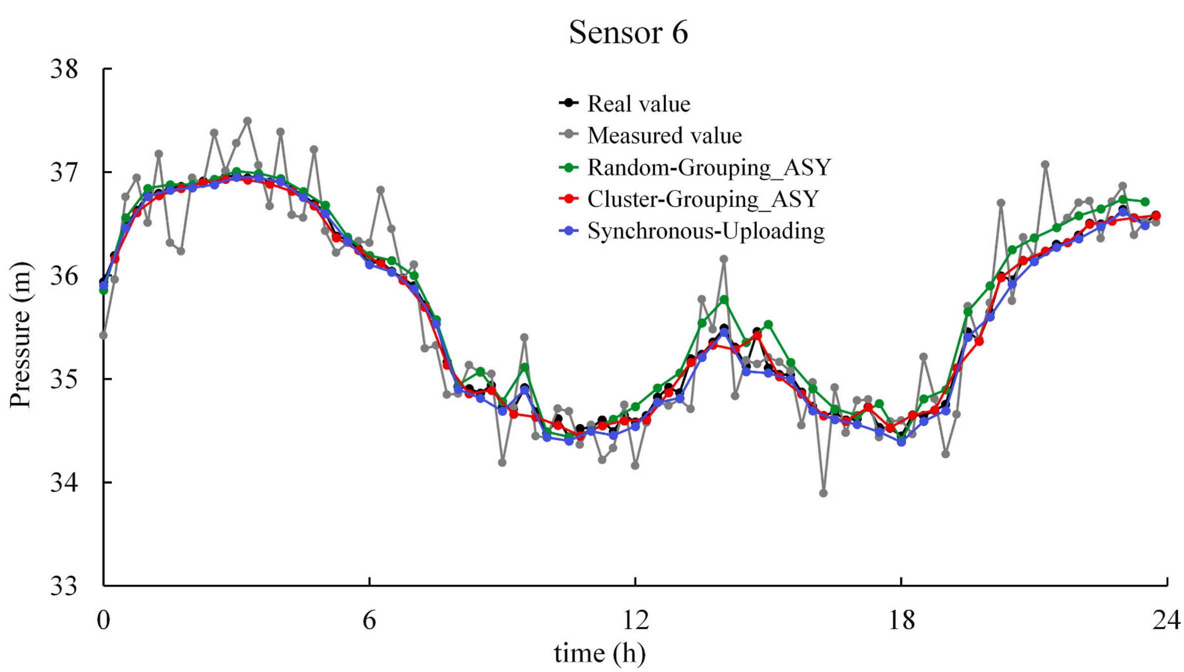

Fig. 6. Comparison of estimation results at Sensor 6.

validation sensors are within 0.77 m, indicating that the estimation is successful as the pressure residuals are lower than the criteria (2 m for all sensors) specified by the Water Research Centre (1989).

4. Discussion

The asynchronous data uploading approach proposed in this paper divides the sensors into several groups and uploads data at different time points to estimate the nodal water demand using a data assimilation approach. Although the increase in data upload frequency satisfies the needs of real-time modeling, the available data (the number of measurements) is reduced at each time step. This problem is addressed by employing a sensor grouping algorithm to improve the representativeness of the data and avoid information redundancy for the sensor data in each group. However, we should acknowledge that this approach is more suitable for WDS networks with a large number of sensors. For sensor dense networks, grouping optimization can still ensure data representativeness and thus ensure estimation accuracy. However, for WDS networks with sparse sensors, the modeling accuracy will inevitably be reduced due to the limited number of measurements at each time step. In this case, it is recommended to first use the proposed asynchronous upload strategy and sensor grouping method to obtain the

Table 1
Average absolute estimation errors for the 16 sensors (m).

Sensor ID	Random-Grouping_ASY	Cluster-Grouping_ASY	Synchronous-Uploading
Sensor 1	0.32	0.36	0.40
Sensor 2	0.32	0.27	0.39
Sensor 3	0.09	0.18	0.10
Sensor 4	0.09	0.14	0.08
Sensor 5	0.15	0.13	0.16
Sensor 6	0.12	0.12	0.03
Sensor 7	0.17	0.01	0.03
Sensor 8	0.17	0.13	0.02
Sensor 9	0.14	0.13	0.09
Sensor 10	0.16	0.01	0.02
Sensor 11	0.14	0.20	0.11
Sensor 12	0.14	0.08	0.07
Sensor 13	0.24	0.23	0.35
Sensor 14	0.12	0.12	0.04
Sensor 15	0.20	0.02	0.02
Sensor 16	0.14	0.12	0.09
average	0.17	0.14	0.13

asynchronous sensor network. Then, new high frequency sensors should be installed in some critical locations to compensate for the lack of sensors. Finally, the old low frequency sensors are combined with the high frequency sensors to estimate the nodal water demands. Therefore, the developed approach can be applied to both sensor dense and sensor sparse networks.

The demand estimation methods adopted in this paper assume that the measurement noise is normally distributed and ignore the impact of anomalous events caused by transient events, such as turning pumps on and off, opening and closing valves, having reservoir levels rise and fall, and water demand shifts caused by industrial users. This can significantly reduce the reliability and accuracy of the estimated nodal water demand. A possible solution to this problem is to develop an efficient algorithm to detect the transient event, then isolate and recover the measurements before the measurements are used to estimate the nodal water demand.

5. Conclusions

This study developed an asynchronous data uploading approach for the nodal water demand estimation in the WDS network. This approach uses existing low frequency sensors to estimate the nodal water demand at a high frequency. The asynchronous data uploading approach proposed in this paper divides the sensors into several groups and uploads data at different time points to estimate the nodal water demand using a data assimilation approach. A sensor grouping strategy is adopted to improve the representativeness of the data and avoid information redundancy for the sensor data in each group. Applications of the proposed approach to a simple hypothetical WDS and a realistic WDS demonstrate that the asynchronous uploading strategy ensures the accuracy of state estimation and improves the state estimation frequency for the sensor network.

The developed approach is more suitable for the sensor dense network. For the sensor sparse network, it is recommended to first use the proposed asynchronous upload strategy and sensor grouping method to generate an asynchronous sensor network. Then, new high frequency sensors should be installed in some critical locations. Therefore, the optimization method for the placement of high frequency sensors taking into account existing low frequency sensors will be explored in future research. This method can effectively improve the data upload frequency of the existing low frequency sensors, thus avoiding the disposal of these sensors and reducing the cost of sensor network upgrade. Besides, the increase in the data upload frequency allows for modeling the WDS network at a higher frequency. This would help for cleaner production and sustainable management of the WDS network, such as leakage control, pipe burst detection, and pump scheduling(for energy saving).

We would like to highlight that the this method is based on the assumption of clean signals, prior demand probability distributions, and steady-state models. These assumptions need further confirmation in a range of WDSs and at a variety of times. In fact, the measured data are inevitably contaminated by transient events. Using data contaminated by the transient event to calibrate steady-state models is a very difficult problem, which should be explored in future research.

Funding

This work was supported by the National Natural Science Foundation

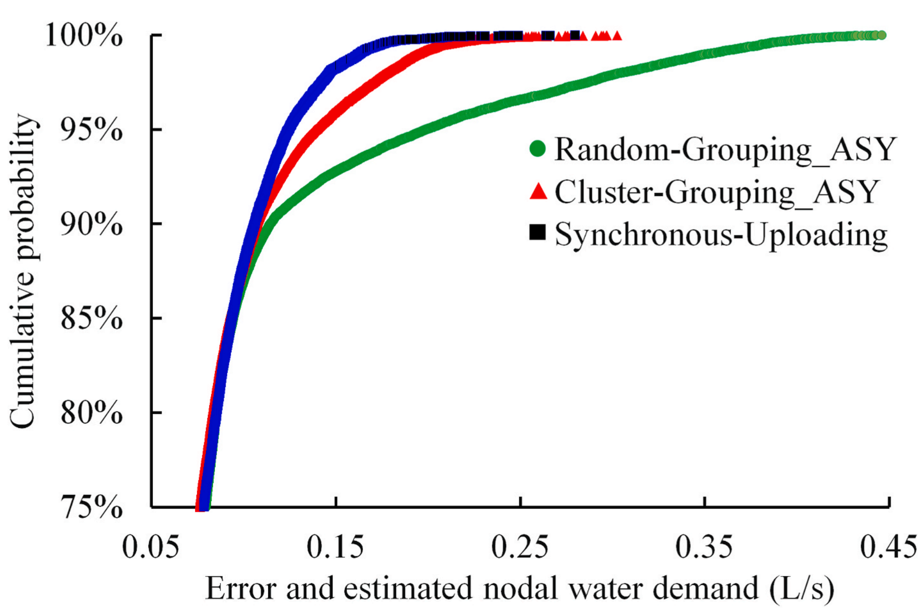

Fig. 7. Cumulative probability for the absolute errors of estimated nodal water demand.

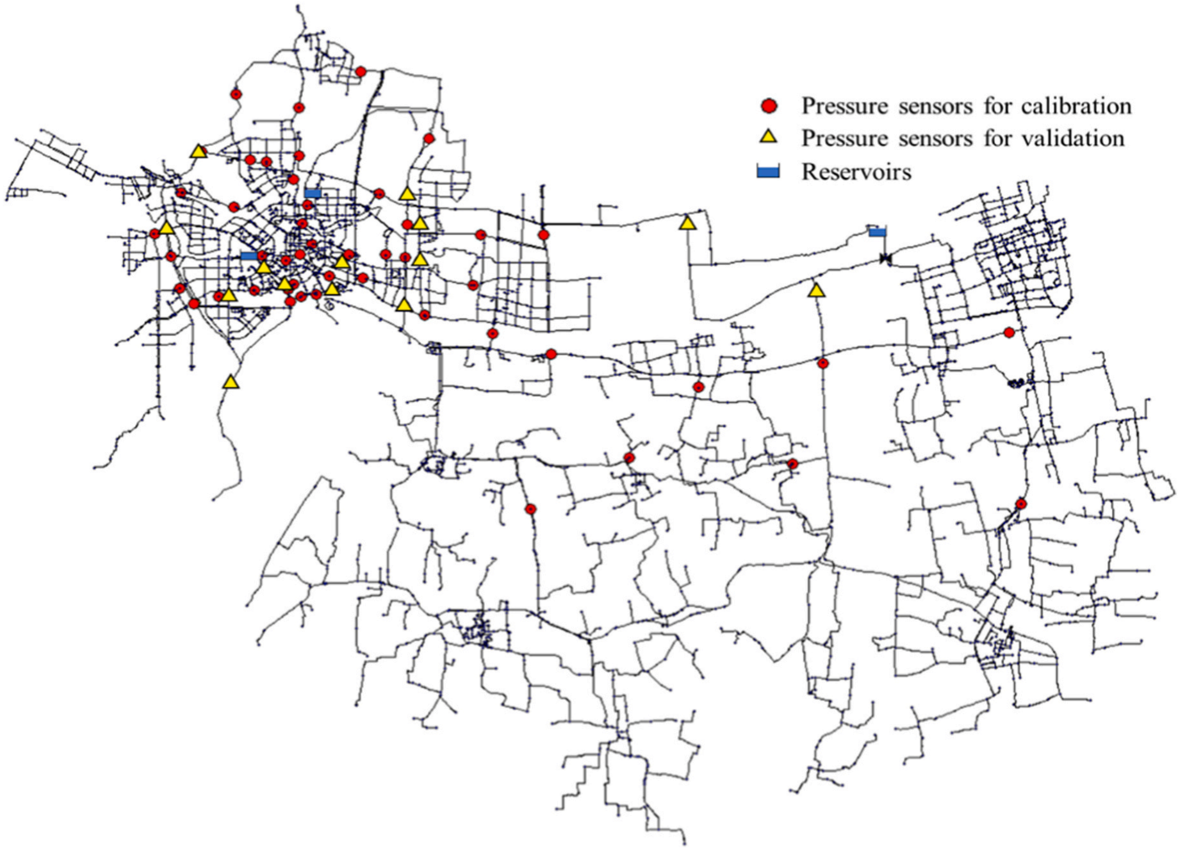

Fig. 8. Schematic of the realistic network and location of sensors and reservoirs.

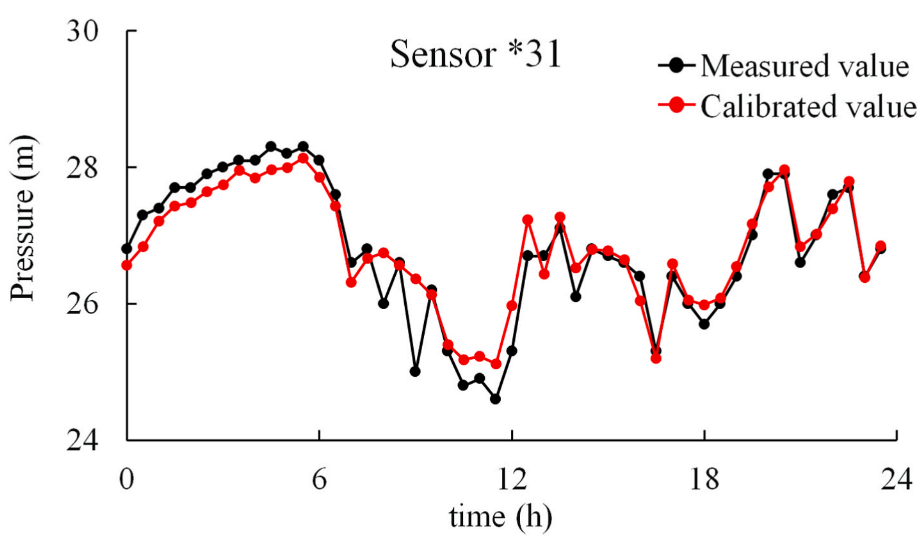

Fig. 9. Comparison of model output and measured data at Sensor *31.

of China [grant numbers 52070167]; and the National Natural Science Foundation of China [grant numbers 52070165].

CRediT authorship contribution statement

Tingchao Yu: Conceptualization, Methodology, Funding acquisition, Supervision. **Ben Lin:** Software, Validation, Writing – original draft. **Zhihong Long:** Funding acquisition, Acquisition of data. **Yu Shao:** Writing – review & editing, Funding acquisition. **Iran E. Lima Neto:** Writing – review & editing. **Shipeng Chu:** Writing – review & editing, Supervision, Writing – original draft.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2022.132676.

References

Bazargan-Lari, M.R., 2014. An evidential reasoning approach to optimal monitoring of drinking water distribution systems for detecting deliberate contamination events. J. Clean. Prod. 78, 1–14.

Boulos, P.F., Ormsbee, L.E., 1991. Explicit network calibration for multiple loading conditions. Civ. Eng. Syst. 8 (3), 153–160.

Bragalli, C., Fortini, M., Todini, E., 2016. Enhancing knowledge in water distribution networks via data assimilation. Water Resour. Manag. 30 (11), 3689–3706.

Cai, D., He, X.F., Han, J.W., 2005. Document clustering using locality preserving indexing. IEEE Trans. Knowl. Data Eng. 17 (12), 1624–1637.

Cheng, W., He, Z., 2011. Calibration of nodal demand in water distribution systems. J. Water Resour. Plann. Manag. 137 (1), 31–40.

Chu, S., Zhang, T., Shao, Y., Yu, T., Yao, H., 2020. Numerical approach for water distribution system model calibration through incorporation of multiple stochastic prior distributions. Sci. Total Environ. 708, 134565.

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, 116837.

Dave, T., Layton, A., 2020. Designing ecologically-inspired robustness into a water distribution network. J. Clean. Prod. 254.

Davidson, J.W., Bouchart, F.J.C., 2006. Adjusting nodal demands in SCADA constrained real-time water distribution network models. J. Hydraul. Eng. 132 (1), 102–110.

Dini, M., Tabesh, M., 2014. A new method for simultaneous calibration of demand pattern and Hazen-Williams coefficients in water distribution systems. Water Resour. Manag. 28 (7), 2021–2034.

Do, N.C., Simpson, A.R., Deuerlein, J.W., Piller, O., 2016. Calibration of water demand multipliers in water distribution systems using genetic algorithms. J. Water Resour. Plann. Manag. 142 (11), 04016044.

Do, N.C., Simpson, A.R., Deuerlein, J.W., Piller, O., 2017. Particle filter–based model for online estimation of demand multipliers in water distribution systems under uncertainty. J. Water Resour. Plann. Manag. 143 (11), 04017065.

Du, K., Long, T.-Y., Wang, J.-H., Guo, J.-S., 2015. Inversion model of water distribution systems for nodal demand calibration. J. Water Resour. Plann. Manag. 141 (9), 04015002.

Duan, H.-F., Pan, B., Wang, M., Chen, L., Zheng, F., Zhang, Y., 2020. State-of-the-art review on the transient flow modeling and utilization for urban water supply system (UWSS) management. J. Water Supply Res. Technol. - Aqua.

El-Ghandour, H.A., Elabd, S.M., Elbeltagi, E., 2021. Assessment of optimal water distribution systems design under steady-state and transient conditions due to pipe roughness uncertainty. Ain Shams Eng. J. 12 (1), 465–473.

Farley, B., Mounce, S.R., Boxall, J.B., 2010. Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. Urban Water J. 7 (6), 345–356.

Huang, Y., Duan, H.-F., Zhao, M., Zhang, Q., Zhao, H., Zhang, K., 2017. Probabilistic analysis and evaluation of nodal demand effect on transient analysis in urban water distribution systems. J. Water Resour. Plann. Manag. 143 (8), 04017041.

Hutton, C.J., Kapelan, Z., Vamvakeridou-Lyroudia, L., Savić, D., 2014. Application of formal and informal Bayesian methods for water distribution hydraulic model calibration. J. Water Resour. Plann. Manag. 140 (11), 04014030.

Kang, D., Lansey, K., 2009. Real-time demand estimation and confidence limit analysis for water distribution systems. J. Hydraul. Eng. 135 (10), 825–837.

Kapelan, Z.S., Savic, D.A., Walters, G.A., 2007. Calibration of water distribution hydraulic models using a Bayesian-Type procedure. J. Hydraul. Eng. 133 (8), 927–936.

Law, K.J.H., Stuart, A.M., Zygalakis, K.C., 2015. Data assimilation: a mathematical introduction. Revista Brasileira De Meteorologia 26 (3), 433–442.

Liu, N.D., Du, K., Tu, J.P., Dong, W.X., 2017. Analytical solution of Jacobian matrices of WDS models. Procedia Eng. 186, 388–396.

Malve, O., Laine, M., Haario, H., Kirkkala, T., Sarvala, J., 2007. Bayesian modelling of algal mass occurrences - using adaptive MCMC methods with a lake water quality model. Environ. Model. Software 22 (7), 966–977.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J. Classif. 31 (3), 274–295.

Muste, M., Lee, K., Bertrand-Krajewski, J.-L., 2012. Standardized uncertainty analysis for hydrometry: a review of relevant approaches and implementation examples. Hydrological Sciences Journal-Journal Des Sciences Hydrol Sci J-J Des Sci Hydrol. 57 (4), 643–667.

Okeya, I., Kapelan, Z., Hutton, C., Naga, D., 2014. Online modelling of water distribution system using data assimilation. Procedia Eng. 70, 1261–1270.

Piller, O., Elhay, S., Deuerlein, J., Simpson, A.R., 2017. Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters. J. Water Resour. Plann. Manag. 143 (2), 04016074.

Qi, Z., Zheng, F., Guo, D., Zhang, T., Shao, Y., Yu, T., Zhang, K., Maier, H.R., 2018. A comprehensive framework to evaluate hydraulic and water quality impacts of pipe breaks on water distribution systems. Water Resour. Res. 54 (10), 8174–8195.

Qin, T., Boccelli, D.L., 2019. Estimating distribution system water demands using Markov chain Monte Carlo. J. Water Resour. Plann. Manag. 145 (7), 04019023.

Rajakumar, A.G., Kumar, M.S.M., Amrutur, B., Kapelan, Z., 2019. Real-time water quality modeling with ensemble kalman filter for state and parameter estimation in water distribution networks. J. Water Resour. Plann. Manag. 145 (11), 04019049.

Sabbaghpour, S., Naghashzadehgan, M., Javaherdeh, K., Haddad, O.B., 2012. HBMO algorithm for calibrating water distribution network of Langarud city. Water Sci. Technol. 65 (9), 1564–1569.

Salomons, E., Housh, M., 2020. Practical real-time optimization for energy efficient water distribution systems operation. J. Clean. Prod. 275.

Sanz, G., Pérez, R., 2015. Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition. J. Water Resour. Plann. Manag. 141 (10), 04015020.

Savic, D.A., Kapelan, Z.S., Jonkergouw, P.M.R., 2009. Quo vadis water distribution model calibration? Urban Water J. 6 (1), 3–22.

Shang, F., Uber, J.G., Waanders, B.G.v.B., Boccelli, D., Janke, R., 2008. Real time water demand estimation in water distribution system. In: Water Distribution Systems Analysis Symposium 2006, pp. 1–14.

WRC (Water Research Centre), 1989. Network Analysis—A Code of Practice. Water Research Centre, Swindon, UK.

Xie, X., Zhang, H., Hou, D., 2017. Bayesian approach for joint estimation of demand and roughness in water distribution systems. J. Water Resour. Plann. Manag. 143 (8), 04017034.

Zhang, Q., Wu, Z.Y., Zhao, M., Qi, J., Huang, Y., Zhao, H., 2016. Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. J. Water Resour. Plann. Manag. 142 (11), 04016042.

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. Plann. Manag. 144 (10), 04018063.

Zhang, Y., Duan, H.-F., Keramat, A., Che, T.-C., 2022. On the leak-induced transient wave reflection and dominance analysis in water pipelines. Mech. Syst. Signal Process. 167, 108512.

Zheng, F., Zecchin, A.C., Newman, J.P., Maier, H.R., Dandy, G.C., 2017. An adaptive convergence-trajectory controlled ant colony optimization algorithm with application to water distribution system design problems. IEEE Trans. Evol. Comput. 21 (5), 773–791.

Zhou, X., Xu, W.R., Xin, K.L., Yan, H.X., Tao, T., 2018. Self-Adaptive calibration of real-time demand and roughness of water distribution systems. Water Resour. Res. 54 (8), 5536–5550.