

Water Resources Research®

RESEARCH ARTICLE

10.1029/2025WR040149

Key Points:

- An edge-based scheduling method is proposed to cut uploads, saving sensor energy for large-scale water distribution systems
- The spatiotemporal redundancy of sensors is leveraged to enhance prediction while preserving data accuracy
- This approach achieves up to 75% energy savings with a mere 0.7% relative error

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Chu,
chushipeng@zju.edu.cn

Citation:

Wei, S., Yu, T., Ostfeld, A., Liu, C., Chu, S., & Shen, H. (2026). Edge computing for energy-efficient sensor scheduling in water distribution systems. *Water Resources Research*, 62, e2025WR040149. <https://doi.org/10.1029/2025WR040149>

Received 20 FEB 2025

Accepted 9 JAN 2026

Author Contributions:

Conceptualization: Shaosong Wei, Shipeng Chu

Funding acquisition: Tingchao Yu, Avi Ostfeld

Investigation: Shipeng Chu

Methodology: Shaosong Wei

Resources: Tingchao Yu, Chengyin Liu, Hao Shen

Supervision: Shipeng Chu

Writing – review & editing:

Shaosong Wei, Tingchao Yu, Avi Ostfeld, Shipeng Chu

Edge Computing for Energy-Efficient Sensor Scheduling in Water Distribution Systems



Shaosong Wei¹, Tingchao Yu^{1,2} , Avi Ostfeld³, Chengyin Liu⁴, Shipeng Chu¹ , and Hao Shen⁴

¹College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China, ²Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing, China, ³Civil and Environmental Engineering, Technion - Israel Institute of Technology, Haifa, Israel, ⁴Ningbo Water Environment Group Co., Ltd, Ningbo, China

Abstract Water distribution systems (WDSs) utilize battery-powered sensors to monitor essential parameters like flow rate and pressure. Limited battery life requires reducing data upload frequencies to conserve energy, potentially compromising real-time monitoring vital for system reliability and performance. This challenge is addressed by leveraging temporal redundancies from daily cycles and spatial redundancies from sensor data correlations, enabling data extrapolation instead of continuous transmission. This study proposes an edge computing-based sensor scheduling method that optimizes data transmission frequency while maintaining high data accuracy, thereby extending sensor longevity without sacrificing monitoring capabilities. The proposed approach uses predictive models to forecast future sensor values over multiple time steps based on existing data redundancies. If the deviation between predicted and actual measurements is within a predefined threshold, data transmission is skipped, reducing sensor power consumption; otherwise, data is transmitted to ensure accuracy. Applied to a realistic WDS sensor network, the method achieved up to a 75% reduction in sensor energy consumption with 48 estimation steps and a 0.5 m error threshold, while maintaining a relative data error of only 0.7%. These results demonstrate the method's effectiveness in balancing energy savings with data reliability, suggesting a viable solution for enhancing WDS sustainability and efficiency.

Plain Language Summary Most sensors in water distribution systems are battery-powered, with data uploads accounting for 80% of their total energy consumption. To conserve energy, traditional approaches adopted by water utilities involve uploading system status data only once or twice per day. However, this results in a lack of real-time data availability, significantly limiting its practical applications. This study introduces an edge-based sensor scheduling method that leverages the spatiotemporal redundancy of sensors in water distribution systems. By establishing predictive models for each sensor, the method reduces the upload of redundant information. Importantly, this approach significantly conserves sensor energy while maintaining data accuracy, effectively resolving the conflict between real-time data uploads and energy consumption.

1. Introduction

Water distribution systems (WDSs) are essential components of urban infrastructure, delivering water to residents, businesses, and industries (Dave & Layton, 2020). Rapid urbanization and population growth have expanded and complicated modern WDSs (Tchórzewska-Cieślak et al., 2024). Ensuring their efficient, stable, and long-term operation presents significant challenges and requires advanced technologies to guarantee reliable supply, system resilience, and water quality.

To tackle issues like pipe bursts (Qi et al., 2018), water leakage (Zhang et al., 2016), high energy consumption (Salomons & Housh, 2020), and water contamination (Bazargan-Lari, 2014), water utilities are increasingly adopting smart technologies such as hydraulic modeling (Ormsbee et al., 2022), machine learning (Sousa et al., 2023), and deep learning (Wu et al., 2024). These technologies enhance system reliability, reduce operational costs, and enable proactive management. However, they inherently depend on real-time, high-quality sensor data (Geelen et al., 2019; Li et al., 2024), making data quality assurance and optimized data transmission frequency critical challenges.

Sensors distributed throughout WDSs collect critical data (e.g., flow rate, pressure, and chlorine levels) every 15 min, providing essential insights into system performance. Because WDSs are dynamic and their states change continuously over time (Ulanicki & Beaujean, 2021; Wang et al., 2023), real-time data transmission is essential for accurate monitoring. In China, advancements in the Internet of Things and smart sensor technology have led to

the installation of numerous pressure and flow sensors in WDSs. However, due to the large scale of these systems and the high costs associated with cabling, only a few sensors at key locations are connected to the power grid for high-frequency data uploads. Most sensors are battery-powered (Du et al., 2015). Data transmission consumes approximately 80% of a sensor's total energy (Anastasi et al., 2009), and frequent uploads increase energy consumption, necessitating regular battery replacements. This significantly raises the management costs of large sensor networks. To reduce the frequency of battery replacements, water companies often bundle and transmit data once or twice a day, which conflicts with the need for real-time information. Therefore, the primary engineering challenge is to minimize sensor energy consumption while ensuring real-time data transmission—a problem that has not yet been explored in WDS sensor networks.

Although real-time sensor data provide immediate insights into the states of WDSs, uploading every data point to the cloud is unnecessary. Many sensor readings exhibit low variability due to consistent water usage patterns and fixed network structures, resulting in regular state data at each node. Studies have shown that WDS state data are predictable rather than entirely random (Truong et al., 2024; Zhou et al., 2023). Recognizing that most sensor data points can be estimated, this paper proposes a sensor scheduling strategy that involves developing a predictive model for each sensor on the cloud platform and sharing it with the sensors. Sensors use this model to forecast multiple future data points and compare these estimates with actual measurements. If the error is within a pre-defined threshold, the sensor skips data transmission, and the cloud uses the estimated value as the actual measurement. If the error exceeds the threshold, the sensor uploads the actual value. This approach replaces highly regular and minimally fluctuating data points with predicted values, significantly reducing data transmissions and conserving energy while maintaining accurate WDS state information. Although the aforementioned sensor scheduling strategy holds significant energy-saving potential, its practical application effect is affected by the accuracy of data prediction. Improving prediction accuracy helps reduce the frequency of data uploads and conserve energy. Therefore, data prediction accuracy is another research focus that requires attention.

Water distribution systems possess unique characteristics that enhance the accuracy of sensor data prediction. Consistent daily water usage patterns cause periodic fluctuations and introduce temporal correlations in the state data (Huang et al., 2018). The interconnected nature of WDSs leads to strong spatial correlations among the states of different nodes (Li et al., 2021). Leveraging these spatiotemporal correlations involves integrating periodic historical data and exploiting redundancies among correlated sensors, which aids in accurate predictions. However, the dynamic nature of WDSs also introduces noise and outliers into the data, hindering prediction accuracy. By applying noise reduction techniques or time-series decomposition, these adverse effects can be mitigated, improving the reliability of data estimation (Morales et al., 2021; Shao et al., 2024). Enhancing prediction accuracy through these methods allows the proposed sensor scheduling strategy to achieve greater energy savings while maintaining precise system monitoring.

To implement this strategy, this paper addressed the challenges of high energy consumption and maintenance costs in wireless sensor networks within WDSs, along with the noise and spatiotemporal correlations inherent in WDS data. Spatially related sensors were grouped using k-medoid clustering. Sensor readings were denoised using low-pass (LP) filtering, and periodic components were removed through time-series decomposition to improve data quality. A multi-step estimation model for each sensor group was then developed using Long Short-Term Memory (LSTM) neural networks within a cloud-based platform. The trained LSTM models were deployed on both cloud and edge devices. Sensors compared the predicted values with actual measurements; if the difference was below a predefined threshold, data transmission was skipped, conserving energy. When the difference exceeded the threshold, actual measured data were sent to the cloud to maintain data accuracy.

To address the challenge of high energy consumption associated with sensor data uploads in WDSs (an issue that prevents real-time monitoring), this study proposes a sensor scheduling methodology that reduces the frequency of data uploads, thereby conserving energy. The key innovations of this work are as follows: (a) An edge computing framework is proposed, which enables energy-efficient sensor scheduling by deploying sensor data prediction models on sensor devices to determine whether communication and uploads are necessary; (b) A sensor data prediction model that leverages the redundancy of WDS monitoring data to enhance data prediction accuracy is developed. This improved accuracy further reduces the number of required communication uploads, ultimately achieving sensor energy conservation. Testing on a real network data set demonstrated that this method can reduce sensor energy consumption by up to 75%, substantially lowering energy use while maintaining real-time data transmission.

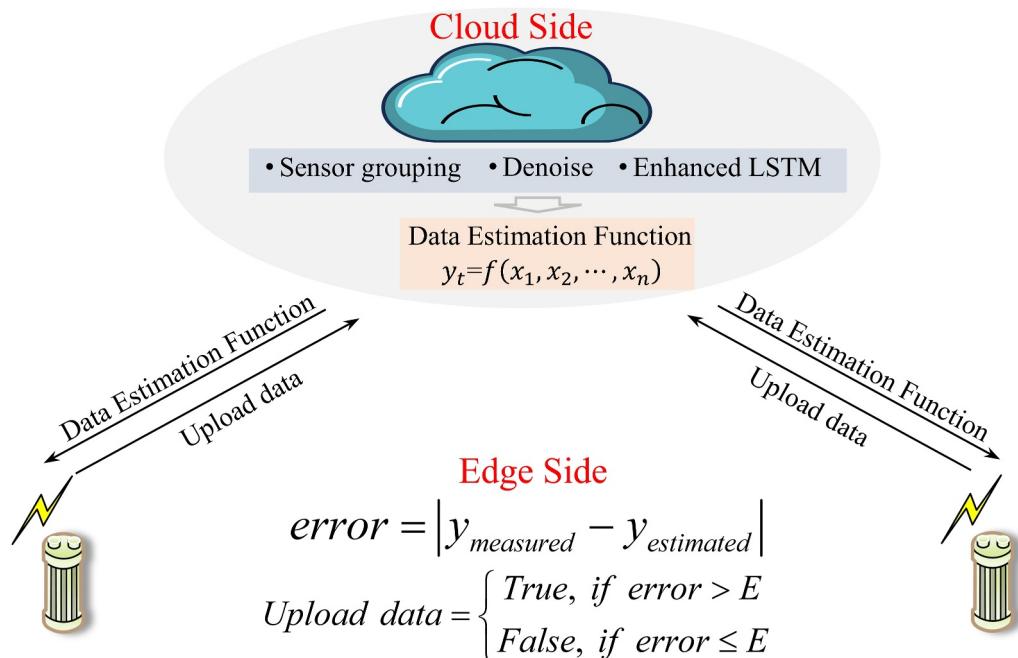


Figure 1. Edge computing-based data transmission mechanism.

2. Methodology

2.1. Design of Sensor Scheduling Strategy

As illustrated in Figure 1, the WDS sensor network consists of two main components: the cloud side and the edge side. The cloud side serves as the central data processing unit, equipped with robust computing, analytical, and storage capabilities. In contrast, the edge side comprises smart sensors with limited data analysis and storage capacity, capable of basic data processing and simple decision-making. Edge-side sensors collect data and transmit it to the cloud side via 4G/5G cellular networks. These sensors are typically battery-powered, and data uploads to the cloud account for 80% of their total energy consumption.

To conserve sensor energy while maintaining real-time data acquisition, this study proposes an edge computing-based sensor scheduling strategy. As shown in Figure 1, this strategy comprises two components: a cloud-side strategy and an edge-side strategy. The cloud side receives data from the edge side and uses it to train a predictive model, which is then transmitted back to the edge side via the wireless sensor network. The edge side collects data at fixed intervals in real time and stores it locally. Simultaneously, it uses the received predictive model to estimate data at each time step. The collected data are compared with the estimated values; if the deviation exceeds a predefined threshold, the locally stored data are packaged and uploaded to the cloud side, followed by clearing the local memory. If the deviation is within the threshold, the data are not uploaded. Additionally, when memory usage reaches a specified limit, the stored data are packaged and uploaded, and the local memory is cleared. Upon receiving new data from the edge side, the cloud side retrains and updates the predictive model and sends it back to the edge side.

2.2. Sensor Data Prediction

Accurately predicting sensor data is a crucial step in the proposed sensor scheduling strategy. Improved prediction increases the number of data points with low prediction errors, reducing the frequency of data uploads and conserving energy. Sensor data prediction relies on identifying redundancies in the data, which can be temporal or spatial. Temporal redundancy arises from regular, periodic fluctuations due to consistent water usage patterns. Spatial redundancy results from correlations between nearby sensors that tend to produce similar data. These redundancies are leveraged through clustering analysis to enhance prediction accuracy.

However, irregular noise and unstable fluctuations can negatively affect predictive model performance. To mitigate these issues, noise reduction and time-series decomposition techniques were applied to the sensor data after grouping the sensors, aiming to enhance data stability and improve predictability. Additionally, an attention mechanism was incorporated after the LSTM model to assign different weights to sensor inputs at various time steps, optimizing the use of temporal redundancy.

2.2.1. Sensor Grouping

Sensors closer in hydraulic distance typically exhibit stronger data correlations (Gomes et al., 2021). To exploit this spatial correlation for improved sensor data prediction, a clustering algorithm was employed to group sensors with strong internal correlations. Data from sensors within the same group were then jointly used as inputs for the prediction algorithm.

In this study, the k-medoid algorithm was utilized as the method for grouping sensors (Li et al., 2023). Unlike clustering techniques that use averaged values as centers, k-medoid selects actual data points, called medoids, to serve as cluster centers. The algorithm begins by randomly choosing K data points as initial medoids. Each sensor is then assigned to the nearest medoid, forming clusters. For each cluster, the medoid is updated by selecting the data point that minimizes the total distance to all other points within that cluster. This iterative process of reassigning sensors and updating medoids continues until the medoids stabilize. A key advantage of the k-medoid algorithm is its robustness against outliers and noise (Modak, 2024), making it particularly effective for clustering sensors in WDSs (Shao et al., 2024).

An essential aspect of the k-medoid algorithm is selecting an appropriate distance metric between data points. Sensor time series used for clustering are often long, resulting in high-dimensional data that make direct computation of Euclidean distances impractical due to the “curse of dimensionality” (Cai et al., 2005). To address this issue, a distance metric based on the Pearson correlation coefficient is employed. Specifically, the distance between two sensor time series is defined as one minus their Pearson correlation coefficient:

$$\text{corr}(x_1, x_2) = \frac{\text{cov}(x_1, x_2)}{\sqrt{\text{var}(x_1)} \sqrt{\text{var}(x_2)}} \quad (1)$$

$$\text{dist}_{i,j} = 1 - \text{corr}(x_i, x_j) \quad (2)$$

where, x_1, x_2 represent the sensor time series, $\text{cov}(x_1, x_2)$ denotes the covariance between them, and $\text{var}(x_1)$ and $\text{var}(x_2)$ are their variances.

Using this distance metric, the k-medoid algorithm is formulated as the following optimization problem:

$$\min : \sum_{i=1}^n \sum_{j=1}^n \text{dist}_{i,j} z_{i,j} \quad (3)$$

$$\sum_{i=1}^n z_{i,j} = 1, \quad j = 1, 2, \dots, n \quad (4)$$

$$z_{i,j} \leq y_i, \quad i = 1, 2, \dots, n \quad (5)$$

$$\sum_{i=1}^n y_i = k \quad (6)$$

$$y_i, z_{i,j} \in \{0, 1\}, \quad i, j = 1, 2, \dots, n \quad (7)$$

where, n denotes the number of sensors, k represents the total number of clusters, $z_{i,j}$ is a binary variable indicating whether sensor j is assigned to the cluster with medoid i , and y_i is a binary variable indicating whether sensor i is selected as a medoid (cluster center).

The objective function in Equation 3 minimizes the total distance between sensors and their assigned medoids, effectively clustering together sensors with strong correlations. Equation 4 ensures that each sensor is assigned to

exactly one cluster, while Equation 5 enforces that a sensor can only be assigned to a cluster if that cluster's medoid exists. Equation 6 specifies that exactly k medoids are selected, corresponding to the desired number of clusters. By clustering highly correlated sensors using the k-medoid algorithm with this distance metric, the predictive performance within each cluster is enhanced.

2.2.2. Noise Removal and Time-Series Decomposition

High-quality sensor data is crucial for accurate prediction, as significant noise can greatly diminish the predictive performance of models. Sensor data noise can generally be classified into two types: outliers and random noise. To address these issues, the Z-score method was first applied to eliminate outliers, followed by the use of LP filter to remove random noise. The Z-score method is defined as:

$$Z = \frac{x - \mu}{\sigma} \quad (8)$$

where x represents a data point, μ denotes the mean value of the sensor data, and σ is the standard deviation.

LP filtering, widely used in image and audio processing, has also proven effective in removing noise from WDS pressure sensor data (Shao et al., 2024). This method operates under the assumption that the true signal primarily contains low-frequency components, while noise is predominantly composed of high-frequency components. By allowing low-frequency signals to pass and attenuating high-frequency signals, noise can be reduced while retaining essential information in the signal (Bornou & Grigore, 2013; Lee et al., 2019). In this study, the sensor time series were first transformed into the frequency domain using the Fourier transform. Amplitudes of signals exceeding the cutoff frequency were set to zero, effectively removing high-frequency components. The filtered signal was then transformed back to the time domain using the inverse Fourier transform.

After filtering, the sensor data often remain non-stationary, which can negatively affect prediction accuracy. To address this issue, the filtered sensor data are decomposed into three components: a trend component, a seasonal component, and a residual component (Oliveira et al., 2017), as shown in Equation 9. The seasonal component represents periodic variations over fixed intervals, the trend component captures long-term patterns, and the residual component reflects noise and irregular fluctuations.

To enhance prediction accuracy, the trend and residual components are combined into a new time series that serves as input to the LSTM model, as shown in Equation 10. The LSTM model is trained to predict this combined trend-residual series. The trained model forecasts the trend and residual components, which are then added together to form the predicted trend-residual series. Finally, this predicted series is combined with the original seasonal component to generate the predicted sensor data.

$$Y_t = T_t + S_t + R_t \quad (9)$$

$$SR_t = S_t + R_t \quad (10)$$

where Y_t represents the original sensor data at time t ; T_t , S_t , and R_t denote the trend, seasonal, and residual components, respectively. SR_t is the combined trend-residual series.

2.2.3. LSTM Model With Attention Mechanism

The trend-residual series SR_t for each sensor in a group are combined to form a multidimensional time series, which is then used as input to train the LSTM model. The model predicts the trend-residual series for all sensors, and this predicted series is subsequently combined with the original seasonal component to generate the predicted sensor data.

The LSTM network, a specialized type of recurrent neural network (RNN), is designed to capture long-term dependencies in sequential data (Hochreiter & Schmidhuber, 1997). Compared to traditional RNNs, LSTM networks are better at preserving memory over time and mitigating the vanishing gradient problem. These advantages make LSTM particularly effective for time series forecasting and have led to its widespread use in WDSs in recent years (Fan et al., 2023; Kühnert et al., 2021; McMillan et al., 2023). Like traditional RNNs, LSTM

models are composed of a chain of “cells,” each containing three gates: an input gate, a forget gate, and an output gate. These gates regulate the flow of information through the network.

In time series forecasting, it is often assumed that each input data point at a given time step holds equal importance. However, this assumption may not be accurate, as the significance of data points can vary across time steps. To address this issue and improve prediction accuracy, an attention mechanism is incorporated that allows the LSTM model to assign different weights to data points at each time step. This enables the model to focus on more important information. The attention mechanism operates by calculating the correlation between the hidden states at each time step and adjusting the weights of each input accordingly. The process is outlined as follows:

Given a time series $x = (x_1, x_2, \dots, x_T)$, the attention mechanism computes the correlation between the hidden states of previous time steps and the current hidden state. For the current time step t , the correlation between each previous hidden state $h_j (j = 1, 2, \dots, t-1)$ and the current hidden state h_t is calculated using an attention score function α :

$$e_{ij} = \alpha(h_t, h_j) \quad (11)$$

where, e_{ij} represents the unnormalized attention score between h_t and h_j .

To determine the relative importance of each previous hidden state, the scores are normalized using the softmax function:

$$\beta_{ij} = \frac{\exp(e_{ij})}{\sum_{j'=1}^{t-1} \exp(e_{ij'})} \quad (12)$$

where β_{ij} is the attention weight assigned to each hidden state.

Subsequently, each previous hidden state h_j is multiplied by its corresponding attention weight β_{ij} and the weighted states are summed to obtain the context vector c_t for the current time step:

$$c_t = \sum_{j=1}^{t-1} \beta_{ij} h_j \quad (13)$$

Finally, the context vector is concatenated with the current hidden state h_t to produce the final output.

$$\hat{h}_t = [h_t; c_t] \quad (14)$$

In this study, a two-layer LSTM model is employed, followed by a two-layer attention mechanism. The model is trained using a backpropagation algorithm with a Mean Squared Error (MSE) loss function and the Adam optimizer to compute and update the model parameters.

3. Application

The proposed sensor scheduling strategy was applied to a realistic large-scale WDS using field-measured data. As depicted in Figure 2, this network consists of three reservoirs, 4,841 pipes, and 4,242 nodes. Forty-four pressure sensors were installed throughout the network, with their locations indicated in the figure. These sensors collected pressure readings at 15-min intervals from 1 February 2020, to 30 April 2020, resulting in 8,640 samples per sensor. The collected data served as the basis for implementing and evaluating the proposed method.

3.1. Proficiency Metric

To assess the performance of the scheduling algorithm, the historical data was backtracked to reconstruct the process of sensor data acquisition and upload. Since the actual monitoring data for all sensors are available, the estimated values from the scheduling algorithm were compared with the measured values to evaluate the accuracy of the estimated sensor data. Additionally, the number of data uploads for each sensor was recorded after the application of the scheduling algorithm and compared with the original upload frequency to assess the energy-saving performance of the algorithm.

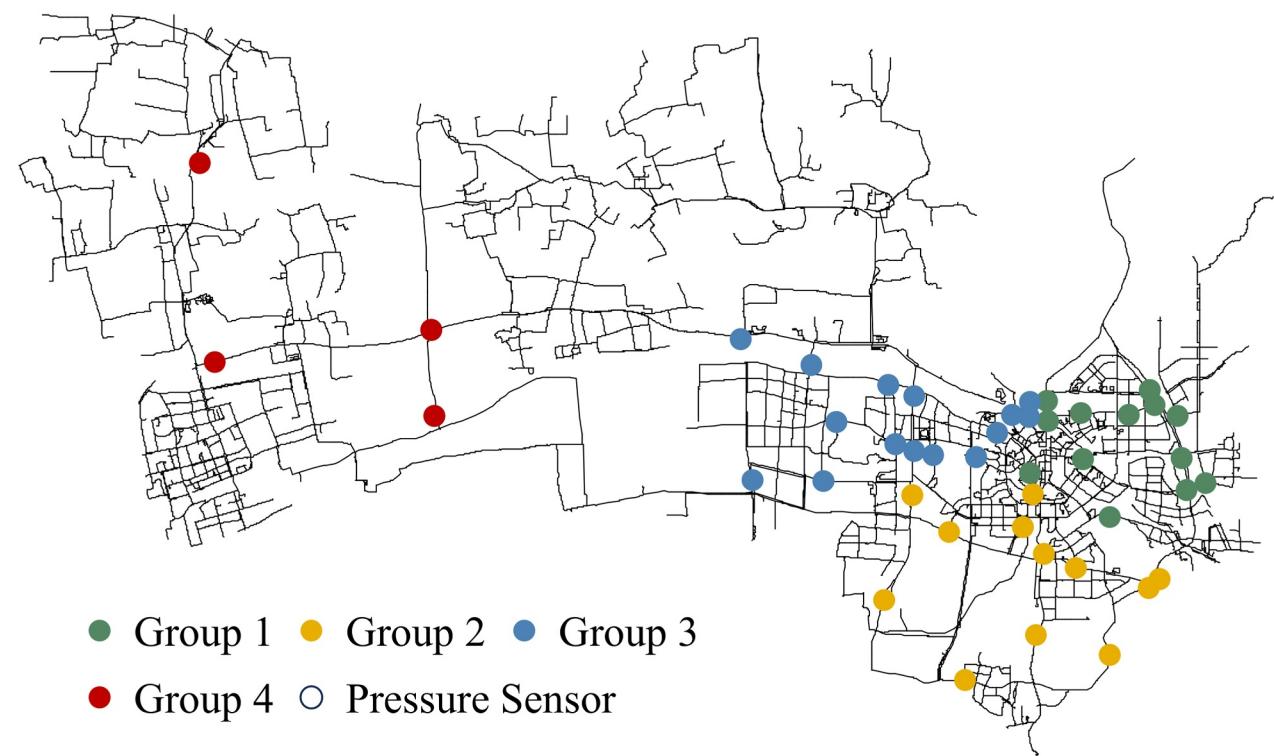


Figure 2. Layout of the case network.

The accuracy of the estimation is evaluated using the Root Mean Square Error (RMSE), calculated as follows:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

where y_i represents the measured value, \hat{y}_i denotes the predicted value, and n is the number of data points.

To evaluate the energy-saving effectiveness, the Energy Saving Rate (ESR) is calculated, which compares the number of uploads with and without the scheduling method. The ESR is given by:

$$ESR = \frac{UT_{\text{scheduling}}}{UT_{\text{original}}} \quad (16)$$

where $UT_{\text{scheduling}}$ refers to the number of uploads with the scheduling method, and UT_{original} represents the number of uploads without it.

3.2. Sensor Grouping and Data Processing

To group the sensor data, the k-medoids algorithm was applied, effectively clustering sensors based on their data correlations in the WDS. As shown in Figure 2, the sensors were divided into four distinct groups, with clearly defined boundaries and a compact distribution within each group. In this study, all sensor data were categorized into these four groups. By enhancing the predictive accuracy of the LSTM model, sensor grouping contributes to a lower ESR, as evidenced by the comparative data in Tables S3 and S4 of Supporting Information S1.

Following the clustering, the Z-score method and LP filtering were used to remove outliers and random noise from the data of the 44 sensors. The Z-score was set to 3 based on the standard three-sigma rule to identify outliers. For noise reduction, the LP filter cutoff frequency was optimized through trial and error and set to 0.6 times the Nyquist frequency (0.6×2 samples per hour). To simplify the analysis, data from one sensor,

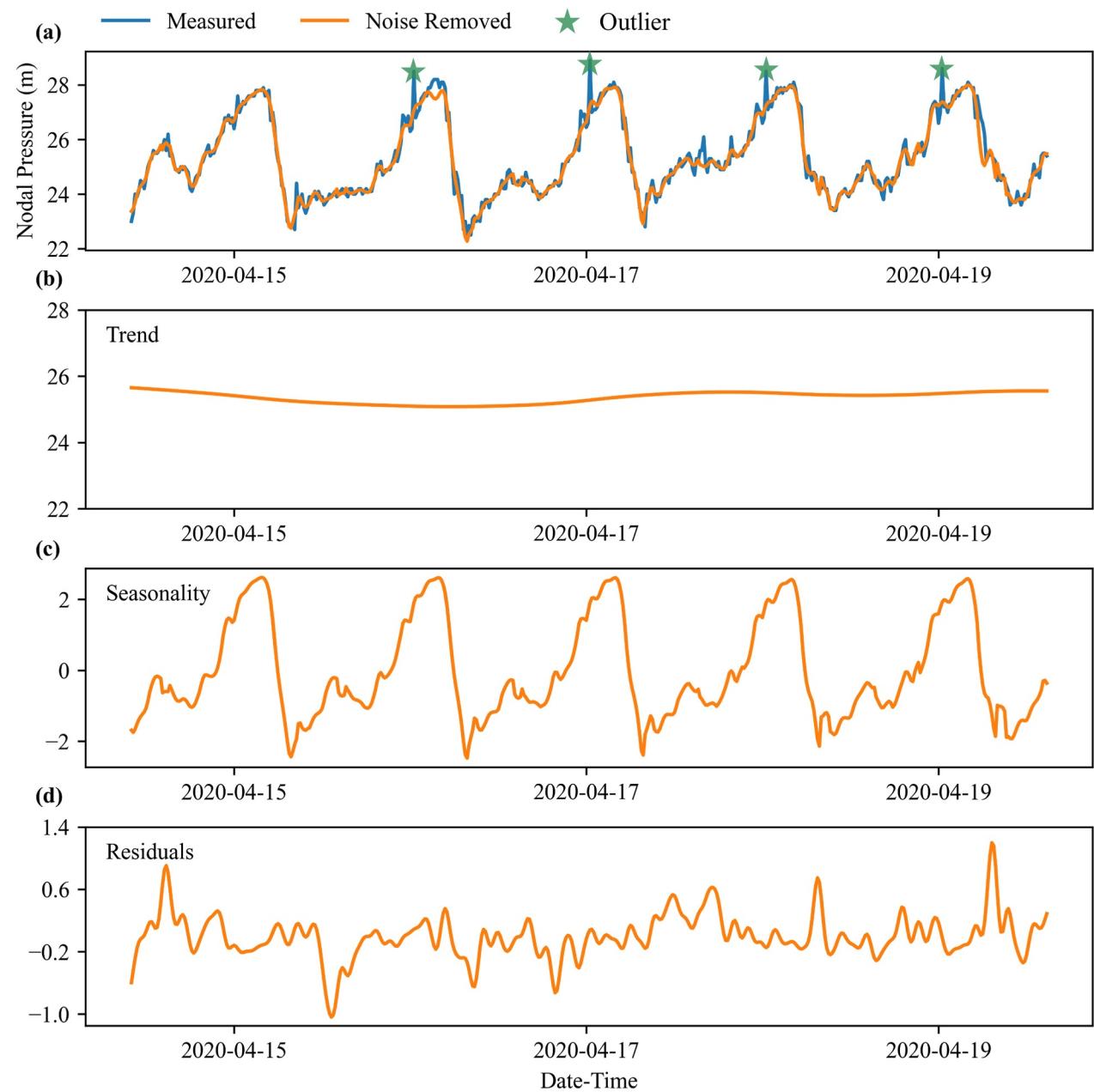


Figure 3. Denoising and time series decomposition of data for sensor #CLD0034: (a) Measured and denoised data; (b) Trend component from the decomposition of the denoised data; (c) Seasonal component from the decomposition of the denoised data; (d) Residual component from the decomposition of the denoised data.

#CLD0034, were selected for detailed review. Figure 3a shows the sensor data before (blue line) and after (orange line) noise removal. The denoised data closely match the field measurements while appearing smoother. The applied denoising technique effectively removed outliers (represented as green pentagrams) and random noise.

Next, Locally Weighted Regression (LWR) was applied to decompose the denoised sensor data. Figures 3b–3d present the results of the decomposition for sensor #CLD0034, breaking the data into three components: trend, seasonal, and residual. The seasonal component, which exhibited a consistent pattern, was removed from the data estimation process. The trend and residual components were then combined and used for estimating the sensor data. The denoising and time-series decomposition results for the representative sensors from the three additional groups are provided in Figures S1–S3 of Supporting Information S1.

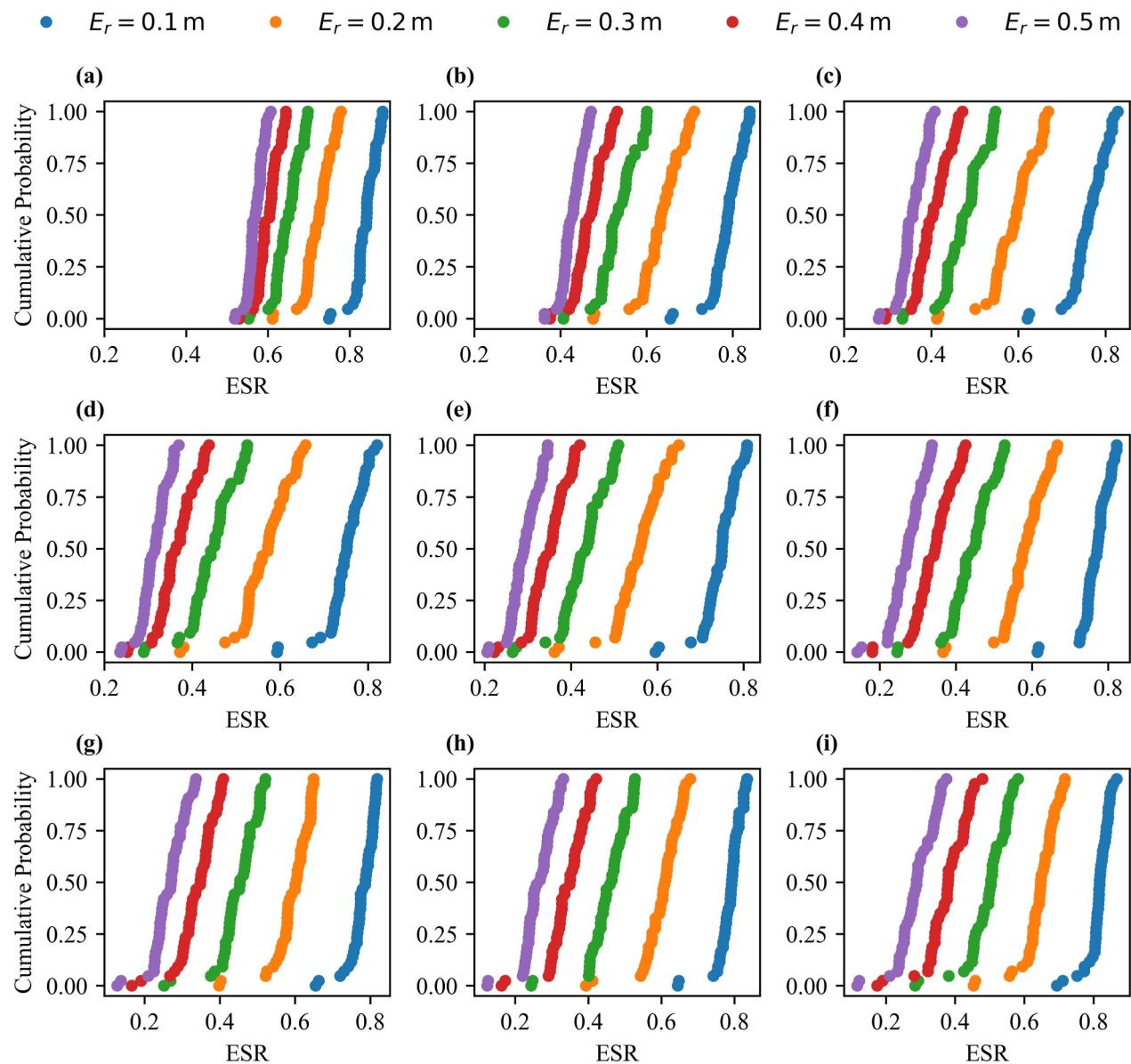


Figure 4. ESR statistics via ES and E_r : (a) ES = 2; (b) ES = 3; (c) ES = 4; (d) ES = 5; (e) ES = 6; (f) ES = 12; (g) ES = 24; (h) ES = 48; (i) ES = 96.

3.3. Sensor Energy Saving

The proposed scheduling method is implemented on sensor data that has undergone clustering, denoising, and decomposition. To evaluate its energy-saving performance, 45 different scenarios were designed based on varying estimation steps (ES) and error thresholds (E_r). The ES values were set to 2, 3, 4, 5, 6, 12, 24, 48, and 96 steps, and the E_r values ranged from 0.1 to 0.5 m. The optimal model parameters and hyperparameter settings for training are detailed in Tables S1 and S2 of Supporting Information S1.

Figure 4 shows the cumulative probability distribution of the ESR for all sensors after applying the scheduling method, with the average ESR provided in Table S4 of Supporting Information S1. The results indicate a significant reduction in energy consumption. For example, with an ES of 48 steps and an E_r of 0.5 m, the average sensor energy consumption is reduced to 26.3% of the original method. The ESR varies based on different combinations of ES and E_r . Specifically, at a fixed ES, the ESR decreases as the E_r increases. This is because more data points fall below the E_r threshold as the E_r rises, leading to fewer uploads. Thus, increasing the E_r can reduce data transmission frequency and, consequently, save more energy when the prediction accuracy is not

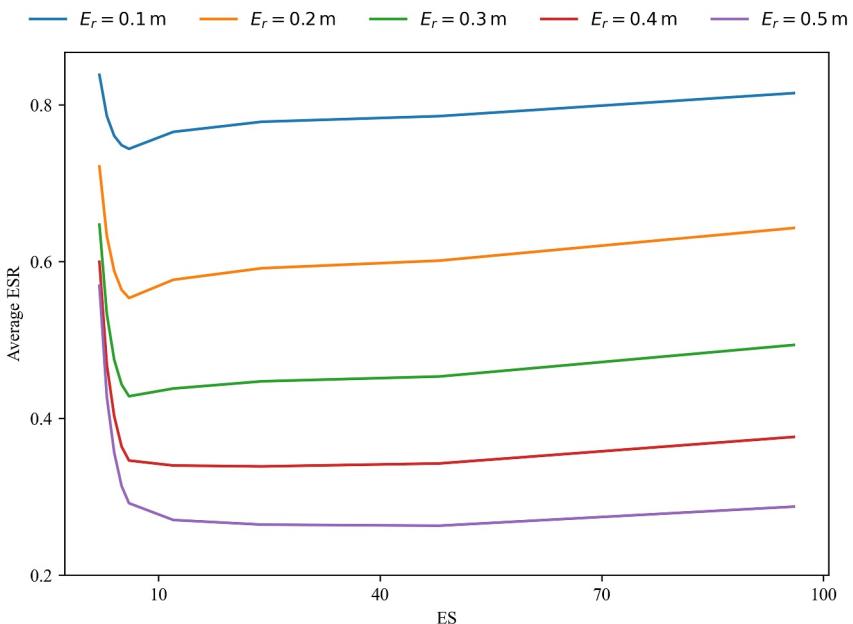


Figure 5. Average ESR of sensors at varying E_r and ESs.

significantly impacted. For a detailed performance analysis, the spatial distribution of the ESR and the pressure data from representative sensors (before and after scheduling) are provided in Figures S4–S5 of Supporting Information S1.

Figure 5 illustrates the relationship between ESR and ES for different E_r values (0.1, 0.2, 0.3, 0.4, and 0.5 m). The graphs consistently show that ESR decreases initially and then increases as ES increases, suggesting an optimal ESR for maximizing energy savings. For instance, when E_r is set to 0.5 m, the optimal point occurs at ES = 48, where the ESR reaches 0.264. In the proposed sensor scheduling approach, the deviation between predicted and measured data is always smaller than the E_r , which represents an ideal scenario. In this case, the sensor upload frequency is inversely proportional to the ES. For example, at ES = 48, data is uploaded every 12 hr (48 × 15 min). This results in lower upload frequencies and greater energy savings as the ES increases. However, in real-world applications, longer prediction steps lead to larger deviations between predicted and actual data, which triggers more uploads and increases energy consumption. Therefore, a practical scheduling system must strike a balance between prediction accuracy and upload frequency to avoid excessive uploads caused by large prediction errors, ultimately minimizing energy consumption.

3.4. Sensor Data Quality

The previous section demonstrated that the proposed scheduling method effectively reduces sensor data upload frequency, conserving energy. However, this approach may lead to a decrease in data accuracy at the cloud. To assess the impact of this method on data accuracy, a statistical analysis was performed comparing the sensor data obtained through the scheduling method with the actual monitoring data. As shown in Figure 6, the proposed scheduling method delivers high data accuracy across different ES and E_r , with the average RMSE provided in Table S5 of Supporting Information S1. In the worst-case scenario, with an ES of 96 steps and an E_r of 0.5 m, the mean RMSE between the sensor data and actual monitoring data is 0.21, corresponding to a relative error of only 0.7% (relative to the average pressure). This error is well within an acceptable range. How data accuracy varies with different ESs and E_r s was also examined. When ES is fixed, data accuracy decreases as E_r increases. This is because a higher E_r excludes more data points from transmission, leading to a higher number of errors in the scheduled data. On the other hand, when E_r is fixed, data accuracy does not change significantly as ES increases. This is due to the similar number of data points exceeding the E_r at different ESs, as shown in Figure 4, resulting in negligible differences in data accuracy.

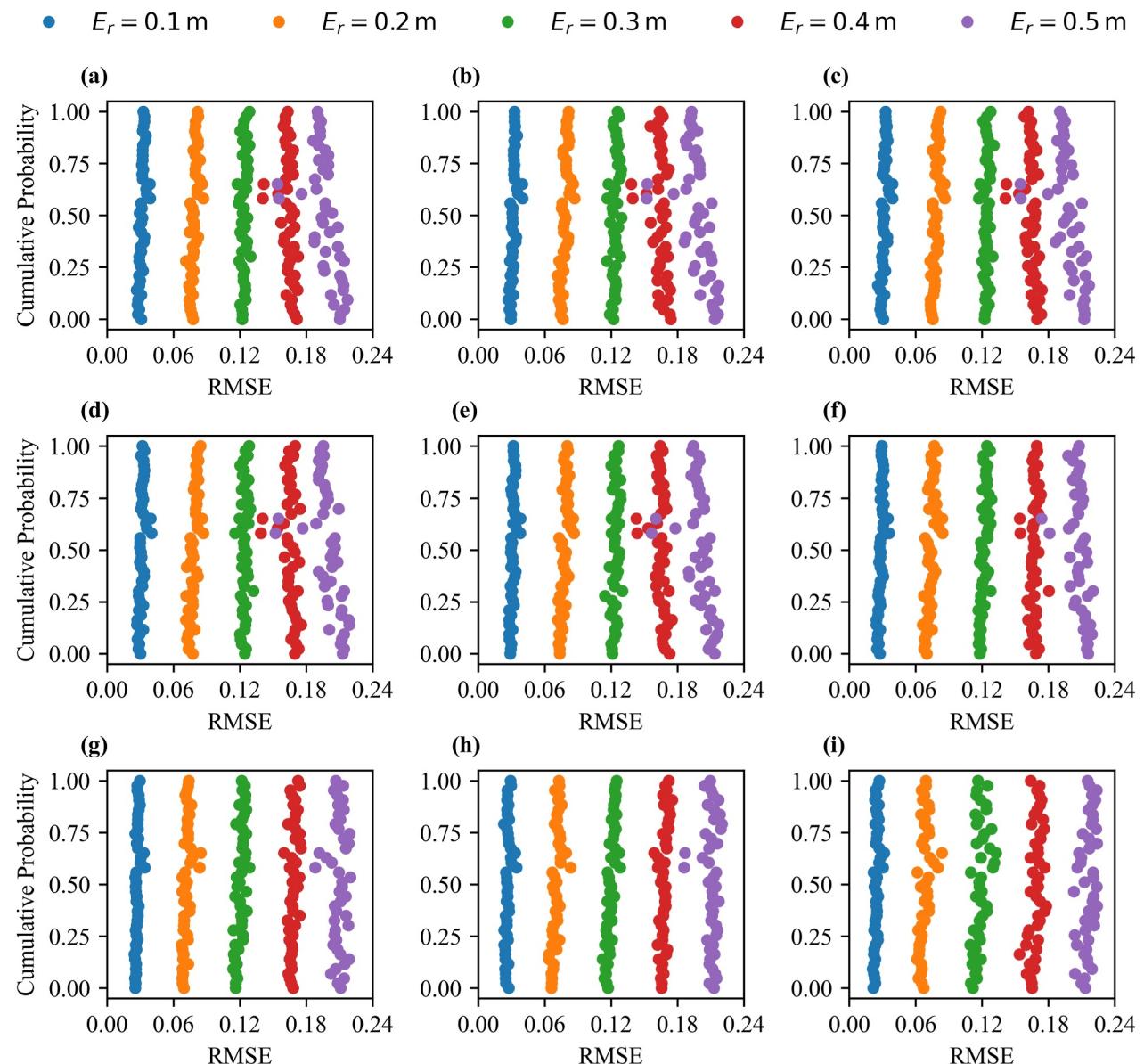


Figure 6. Data accuracy of the sensors after applying the scheduling method: (a) ES of 2; (b) ES of 3; (c) ES of 4; (d) ES of 5; (e) ES of 6; (f) ES of 12; (g) ES of 24; (h) ES of 48; (i) ES of 96.

3.5. Performance Under Optimal Conditions

Sections 3.3 and 3.4 have systematically evaluated the ESR and data accuracy of the proposed method under 45 operational scenarios. By optimizing the average ESR, the optimal ES (48) and E_r (0.5 m) were determined. This section further analyzes the performance under these optimal conditions. As shown in Figure 7, under the optimal parameter combination, the ESR of individual sensors ranges from 0.12 to 0.33, with a mean value of 0.26. This indicates that the proposed scheduling algorithm achieves a 74% reduction in the number of data transmissions on average, demonstrating significant energy conservation. In terms of data accuracy, when E_r is set to 0.5 m, the RMSE between the scheduled data and the original data varies between 0.18 and 0.22, with an average of 0.21, corresponding to a mean relative error of 0.7%. Therefore, the proposed sensor scheduling method effectively reduces energy consumption while maintaining data accuracy.

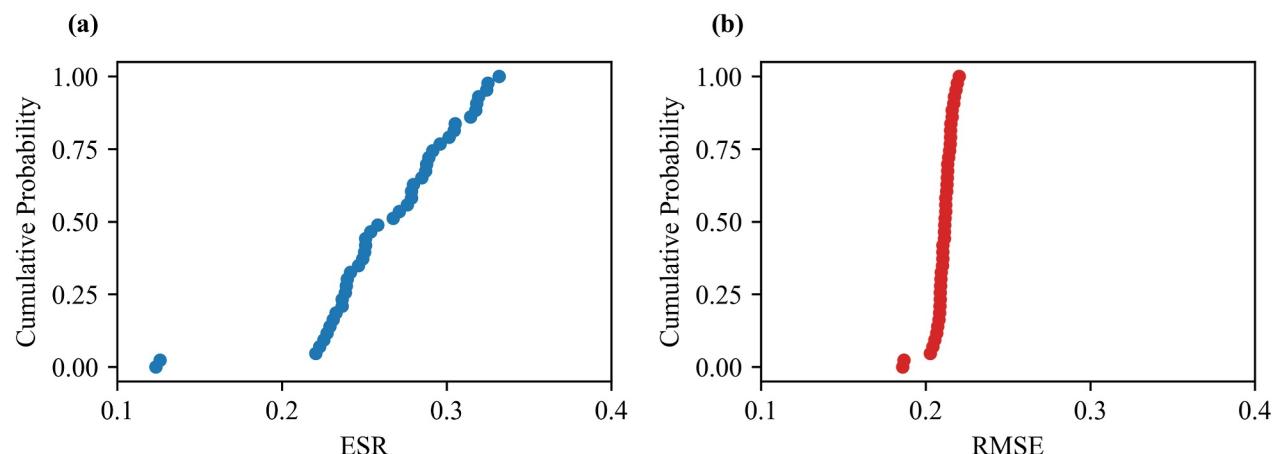


Figure 7. Performance evaluation under an ES of 48 and an E_r of 0.5 m: (a) ESR; (b) RMSE.

4. Discussion

The energy consumption of sensors in WDSs presents a significant challenge to achieving real-time data uploads. To address this challenge, an edge computing-based sensor scheduling method was developed, aimed at reducing energy usage without compromising data accuracy. The proposed approach leverages predictive models to decrease the frequency of data uploads. By setting a predefined E_r , the system determines whether the measured data deviates sufficiently from the predicted values to warrant an upload. This strategy effectively balances energy conservation with the need for accurate and timely data, ensuring reliable monitoring of the WDS.

The effectiveness of the sensor scheduling method in saving energy is influenced by both the E_r and the accuracy of the predictions. A higher E_r allows for more tolerance in prediction errors, resulting in fewer data uploads and greater energy savings. However, maintaining high prediction accuracy is crucial for minimizing erroneous data points. The proposed method enhances prediction accuracy through several mechanisms: clustering sensors based on spatial correlations, applying noise reduction techniques, utilizing robust prediction algorithms, and optimizing the prediction step size. By grouping sensors with high spatial correlation, spatiotemporal redundancies are exploited to improve the reliability of predictions. Additionally, the redundancy in historical data further supports accurate future predictions. These combined efforts significantly reduce the number of uploads triggered by prediction errors, thereby conserving sensor energy.

The ES plays a critical role in the sensor scheduling process. Ideally, increasing the ES reduces the frequency of data uploads, leading to lower energy consumption. However, in practical applications, a larger ES can decrease prediction accuracy, resulting in more frequent uploads when deviations exceed the E_r . Therefore, selecting an appropriate ES is essential to balance energy savings and data accuracy. Moreover, a higher ES means more predicted steps must be stored and processed by the edge device, which typically has limited storage and computational resources. This constraint necessitates careful consideration of the ES during the design of the sensor scheduling system to ensure both efficiency and practicality.

This study utilizes E_r to decide when data should be uploaded. The trigger threshold represents the allowable deviation between predicted and actual sensor data. Significant deviations, such as those caused by data failures or emergencies like pipe bursts, will exceed the E_r and prompt data uploads. An important area for future research is the ability to distinguish between random data faults and critical pipe network events at the edge. By effectively filtering out random noise and only uploading data related to significant events, the system can further optimize energy savings while ensuring that critical incidents are promptly detected and addressed. This capability would enhance the robustness and reliability of the sensor scheduling method in real-world WDS applications.

In the proposed method, employing a model at the sensor side for data prediction inevitably introduces additional energy consumption. However, during the data acquisition process, data transmission accounts for nearly 80% of the total energy consumption (Anastasi et al., 2009). In contrast, the computational cost of running the predictive model at the edge is negligible. Previous studies have shown that the use of various models at the edge, including LSTM (Mohanty et al., 2020), CNN (Cheng et al., 2019), Seq2Seq (Morales et al., 2021), combined models (Jain

et al., 2022), and even complex architectures (Njoya et al., 2022), results in energy consumption that is substantially lower than communication. The developed method is therefore considered highly effective for practical applications.

The spatiotemporal redundancy in WDSs provides the basis for achieving sensor energy savings with the proposed method. Leveraging this redundancy, an alternative approach could involve reducing sensor density while increasing the battery capacity of each sensor. However, this approach has certain limitations. First, reducing sensor density may compromise system resilience, as redundancy ensures that sufficient state information can still be obtained even if some sensors fail. Second, increasing battery capacity faces practical constraints. The specifications of installed sensors are often fixed and cannot be modified. In addition, sensors are typically installed in confined spaces such as valve chambers, where limited room for maintenance activities makes the installation of larger-capacity batteries challenging.

In the process of developing the sensor scheduling method, this study leveraged the spatiotemporal redundancy of sensor data within the WDS, establishing a unique connection between the proposed method and the system. However, it is recognized that the primary focus of this study is the data itself, rather than the WDS, which implies that the method may also be applicable to wireless sensor networks in other systems. Nevertheless, this study provides a sufficient case to demonstrate how energy savings in sensors can be achieved while maintaining data accuracy under conditions of data redundancy. Future work will further explore how the proposed sensor scheduling method can continue to function effectively under specific WDS conditions, such as pipe bursts, valve switching, and significant fluctuations in water consumption, thereby strengthening the connection between the method and the system.

In summary, the edge computing-based sensor scheduling method offers a promising solution to reduce sensor energy consumption while maintaining high data accuracy in WDSs. By carefully balancing the E_r and ES, and by leveraging spatiotemporal redundancies, the method achieves significant energy savings without sacrificing the reliability of the monitoring system. Future enhancements, such as advanced fault detection mechanisms, will further strengthen the effectiveness and applicability of this approach in diverse and dynamic water distribution environments.

5. Conclusions

Sensors in WDSs often rely on battery power, which limits their operational lifespan due to energy constraints. To prolong sensor service life without reducing data upload frequency, this research addresses the critical challenge of minimizing sensor energy consumption while maintaining high data accuracy. By developing advanced sensor scheduling methods, this study aims to support efficient WDS operations that require real-time data monitoring in increasingly complex urban infrastructures.

This study introduced an edge computing-based sensor scheduling method that optimizes data upload frequency by leveraging predictive models and predefined E_r . This strategy reduces the need for frequent data transmissions, thereby conserving sensor energy without compromising data quality. By clustering sensors based on spatial correlations, inherent spatial redundancy is exploited, enhancing prediction accuracy as collective data from grouped sensors provide a more reliable basis for forecasting. Additionally, integrating time redundancy through historical data allows the proposed method to capture consistent temporal patterns, further improving prediction accuracy. These combined strategies enable accurate estimation of future sensor readings, significantly reducing the need for frequent data uploads. For data prediction, noise reduction and time-series decomposition techniques were applied, and LSTM neural networks with attention mechanisms were utilized for reliable data estimation. In the case study, the proposed scheduling method achieved up to 75% reduction in sensor energy consumption while maintaining a mean RMSE of only 0.21, corresponding to a relative error of 0.7%. These results demonstrate the effectiveness of the proposed approach in balancing energy savings with data reliability, making it a viable solution for WDSs.

However, the proposed method relies heavily on accurate clustering of sensors and precise prediction models. Inaccurate clustering or suboptimal predictions can lead to increased data uploads or loss of critical information, potentially diminishing both energy savings and data reliability. Furthermore, the identification of random faults and pipe network events such as pipe bursts at the edge devices is also an important research direction. Future research should focus on enhancing the robustness of clustering algorithms and improving prediction accuracy,

possibly by integrating advanced validation techniques and adaptive mechanisms to ensure the scheduling method remains effective under diverse conditions.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The dataset used in the case study are available for free on Zenodo via DOI (ShaosongWei, 2025b). The software associated with the construction and validation of sensor scheduling is published on GitHub (ShaosongWei, 2025a).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.52570112), Ningbo science and technology plan project (2023Z057), the National Key Research and Development Program of China (2023YFC3208102, 2022YFF0606905), and by the Israeli Water Authority under project number 2033800.

References

Anastasi, G., Conti, M., Di Francesco, M., & Passarella, A. (2009). Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks*, 7(3), 537–568. <https://doi.org/10.1016/j.adhoc.2008.06.003>

Bazargan-Lari, M. R. (2014). An evidential reasoning approach to optimal monitoring of drinking water distribution systems for detecting deliberate contamination events. *Journal of Cleaner Production*, 78, 1–14. <https://doi.org/10.1016/j.jclepro.2014.04.061>

Bornoiu, I.-V., & Grigore, O. (2013). A study about feature extraction for stress detection. In *2013 8th international symposium on advanced topics in electrical engineering, ATEE 2013* (pp. 1–4). IEEE. <https://doi.org/10.1109/ATEE.2013.6563421>

Cai, D., He, X., & Han, J. (2005). Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering*, 17(12), 1624–1637. <https://doi.org/10.1109/TKDE.2005.198>

Cheng, H., Xie, Z., Shi, Y., & Xiong, N. (2019). Multi-step data prediction in wireless sensor networks based on one-dimensional CNN and bidirectional LSTM. *IEEE Access*, 7, 117883–117896. <https://doi.org/10.1109/ACCESS.2019.2937098>

Dave, T., & Layton, A. (2020). Designing ecologically-inspired robustness into a water distribution network. *Journal of Cleaner Production*, 254, 120057. <https://doi.org/10.1016/j.jclepro.2020.120057>

Du, R., Gkatzikis, L., Fischione, C., & Xiao, M. (2015). Energy efficient sensor activation for water distribution networks based on compressive sensing. *IEEE Journal on Selected Areas in Communications*, 33(12), 2997–3010. <https://doi.org/10.1109/JSAC.2015.2481199>

Fan, X., Zhang, X., & Yu, X. B. (2023). Uncertainty quantification of a deep learning model for failure rate prediction of water distribution networks. *Reliability Engineering and System Safety*, 236, 109088. <https://doi.org/10.1016/j.ress.2023.109088>

Geelen, C. V. C., Yntema, D. R., Molenaar, J., & Keesman, K. J. (2019). Monitoring support for water distribution systems based on pressure sensor data. *Water Resources Management*, 33(10), 3339–3353. <https://doi.org/10.1007/s11269-019-02245-4>

Gomes, S. C., Vinga, S., & Henriques, R. (2021). Spatiotemporal correlation feature spaces to support anomaly detection in water distribution networks. *Water*, 13(18), 2551. <https://doi.org/10.3390/w13182551>

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Huang, P., Zhu, N., Hou, D., Chen, J., Xiao, Y., Yu, J., et al. (2018). Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning. *Water*, 10(12), 1765. <https://doi.org/10.3390/w10121765>

Jain, K., Agarwal, A., & Abraham, A. (2022). A combinational data prediction model for data transmission reduction in wireless sensor networks. *IEEE Access*, 10, 53468–53480. <https://doi.org/10.1109/ACCESS.2022.3175522>

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of LSTM networks for water demand prediction in optimal pump control. *Water (Basel)*, 13(5), 644. <https://doi.org/10.3390/w13050644>

Lee, G., Choi, B., Jebelli, H., Ahn, C. R., & Lee, S. (2019). Reference signal-based method to remove respiration noise in electrodermal activity (EDA) collected from the field. *Computing in Civil Engineering* 2019, 17–25. <https://doi.org/10.1061/9780784482438.003>

Li, L., Yan, J., Wen, Q., Jin, Y., & Yang, X. (2023). Learning Robust deep State space for unsupervised Anomaly detection in contaminated time-series. *IEEE Transactions on Knowledge and Data Engineering*, 35(6), 6058–6072. <https://doi.org/10.1109/TKDE.2022.3171562>

Li, X., Chu, S., Zhang, T., Yu, T., & Shao, Y. (2021). Leakage localization using pressure sensors and spatial clustering in water distribution systems. *Water Supply*, 22(1), 1020–1034. <https://doi.org/10.2166/ws.2021.219>

Li, Z., Liu, H., Zhang, C., & Fu, G. (2024). Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. *Water Research*, 250, 121018. <https://doi.org/10.1016/j.watres.2023.121018>

McMillan, L., Fayaz, J., & Varga, L. (2023). Flow forecasting for leakage burst prediction in water distribution systems using long short-term memory neural networks and Kalman filtering. *Sustainable Cities and Society*, 99, 104934. <https://doi.org/10.1016/j.scs.2023.104934>

Modak, S. (2024). Finding groups in data: An introduction to cluster analysis: Authored by Leonard Kaufman and Peter J. Rousseeuw, John Wiley and Sons. *Journal of Applied Statistics*, 51(8), 1618–1620. <https://doi.org/10.1080/02664763.2023.2220087>

Mohanty, S. N., Lydia, E. L., Elhoseny, M., Al Otaibi, M. M. G., & Shankar, K. (2020). Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks. *Physical Communication*, 40, 101097. <https://doi.org/10.1016/j.phycom.2020.101097>

Morales, C. R., Rangel de Sousa, F., Brusamarello, V., & Fernandes, N. C. (2021). Evaluation of deep learning methods in a dual prediction scheme to reduce transmission data in a WSN. *Sensors*, 21(21), 7375. <https://doi.org/10.3390/s21217375>

Njoya, A. N., Tchangmena, A. A. N., Ari, A. A. A., Gueroui, A., Thron, C., Mpinda, B. N., et al. (2022). Data prediction based encoder-decoder learning in wireless sensor networks. *IEEE Access*, 10, 109340–109356. <https://doi.org/10.1109/ACCESS.2022.3213671>

Oliveira, P. J., Steffen, J. L., & Cheung, P. (2017). Parameter estimation of seasonal arima models for water demand forecasting using the harmony search Algorithm. *Procedia Engineering*, 186, 177–185. <https://doi.org/10.1016/j.proeng.2017.03.225>

Ormsbee, L., Hoagland, S., Hernandez, E., Hall, A., & Ostfeld, A. (2022). Hydraulic model database for applied water distribution systems research. *Journal of Water Resources Planning and Management*, 148(8), 04022037. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001559](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001559)

Qi, Z., Zheng, F., Guo, D., Zhang, T., Shao, Y., Yu, T., et al. (2018). A comprehensive framework to evaluate hydraulic and water quality impacts of pipe breaks on water distribution systems. *Water Resources Research*, 54(10), 8174–8195. <https://doi.org/10.1029/2018WR022736>

Salomons, E., & Housh, M. (2020). Practical real-time optimization for energy efficient water distribution systems operation. *Journal of Cleaner Production*, 275, 124148. <https://doi.org/10.1016/j.jclepro.2020.124148>

Shao, Y., Xu, C., Zhang, T., Shentu, H., & Chu, S. (2024). Noise removal for the steady-state pressure measurements based on domain knowledge of water distribution systems. *Journal of Water Resources Planning and Management*, 150(3), 04023082. <https://doi.org/10.1061/JWRMD5.WRENG-6240>

ShaosongWei. (2025a). ShaosongWei/WRR2025-Papers-with-Code: WRR2025-Papers-with-Code (version v1.0) [Software]. *Zenodo*. <https://doi.org/10.5281/zenodo.14898362>

ShaosongWei. (2025b). ShaosongWei/WRR2025-Papers-with-Data-Files: WRR2025-Papers-with-Data-Files (Version v1.0) [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.14898372>

Sousa, D. P., Du, R., Mairton Barros da Silva Jr, J., Cavalcante, C. C., & Fischione, C. (2023). Leakage detection in water distribution networks using machine-learning strategies. *Water Supply*, 23(3), 1115–1126. <https://doi.org/10.2166/ws.2023.054>

Tchórzewska-Cieślak, B., Szpak, D., Żywiec, J., & Rożnowski, M. (2024). The concept of estimating the risk of water losses in the water supply network. *Journal of Environmental Management*, 359, 120965. <https://doi.org/10.1016/j.jenvman.2024.120965>

Truong, H., Tello, A., Lazovik, A., & Degeler, V. (2024). Graph neural networks for pressure estimation in water distribution systems. *Water Resources Research*, 60(7), e2023WR036741. <https://doi.org/10.1029/2023WR036741>

Ulanicki, B., & Beaujean, P. (2021). Modeling dynamic behavior of water distribution systems for control purposes. *Journal of Water Resources Planning and Management*, 147(8), 04021043. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001403](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001403)

Wang, S., Chakrabarty, A., & Taha, A. F. (2023). Data-Driven identification of dynamic quality models in drinking water networks. *Journal of Water Resources Planning and Management*, 149(4), 04023008. <https://doi.org/10.1061/JWRMD5.WRENG-5431>

Wu, Y., Ma, X., Guo, G., Jia, T., Huang, Y., Liu, S., et al. (2024). Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective. *Water Research*, 261, 121999. <https://doi.org/10.1016/j.watres.2024.121999>

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. *Journal of Water Resources Planning and Management*, 142(11), 04016042. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000661](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000661)

Zhou, X., Zhang, J., Guo, S., Liu, S., & Xin, K. (2023). A convenient and stable graph-based pressure estimation methodology for water distribution networks: Development and field validation. *Water Research*, 233, 119747. <https://doi.org/10.1016/j.watres.2023.119747>