



A multi-point leakage prediction statistical method using Bayesian inference in water distribution networks

Chenlei Xie ^{a,b}, Zheng Tian ^{a,*}, Jie Chen^a, Qiansheng Fang^a and Jie Wang^a

^a Anhui Province Key Laboratory of Intelligent Building and Building Energy-saving, Anhui Jianzhu University, Hefei, China

^b Anhui Institute of Strategic Study on Carbon Emissions Peak and Carbon Neutrality in Urban-rural Development, Anhui Jianzhu University, Hefei, China

*Corresponding author. E-mail: tz@stu.ahjzu.edu.cn

 CX, 0000-0002-4597-8585; ZT, 0009-0008-9752-8990

ABSTRACT

Leakage in water distribution networks (WDNs) not only leads to serious water loss and pipe contamination but also affects residents' daily water. Accurate localization of leaks in WDN is significant to conserve water resources and reduce economic losses. However, in traditional optimization and verification methods for leak detection, factors such as modeling and pressure monitoring point errors are not taken into account, resulting in the deviation from the actual location of leaks. To address the mentioned issues, this article presents a WDN leakage probability prediction method based on Bayesian inference. The method converts the prior probability of leakage events into posterior probability. Then, by utilizing the posterior probability density function, the uncertainty of modeling and measurement errors in the hydraulic simulation model are quantified. The probability distribution of leakage pipes and leakage quantity within the leakage area is calculated, allowing for the location prediction and corresponding magnitude of leakage. The experimental results indicate that the prediction model can detect unknown quantities of leakage events. By collecting multiple sets of leakage data, it is possible to accurately predict the location and quantity of leaks, enhancing the efficiency of leakage detection in large-scale water supply networks and providing decision-making assistance for water utilities.

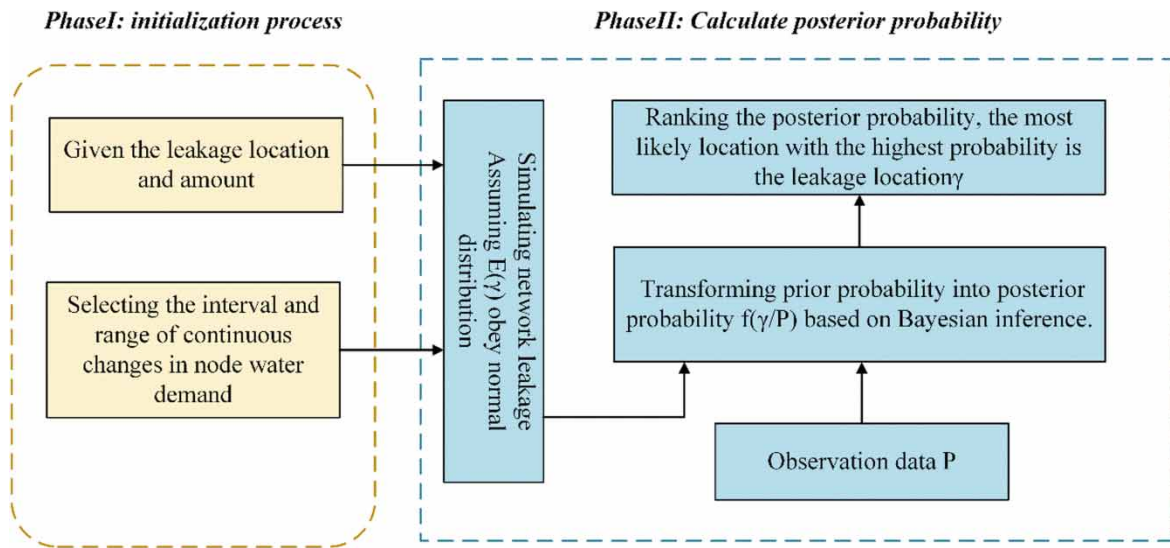
Key words: Bayesian inference, leakage amount, leakage location, prediction of leakage probability, water distribution networks

HIGHLIGHTS

- Presenting a method for predicting leakage probability based on probability statistics, which calculates the leakage probability of each leakage event and infers the most likely leakage event to occur.
- Reducing the adverse effects of observation and modeling errors through multiple data collection, achieving multi-point leakage detection in pipe networks.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



INTRODUCTION

The water supply network is an essential infrastructure in modern urban construction, playing a crucial role in urbanization and people's daily lives. With the accelerated process of urban modernization in our country, the urban population has been increasing dramatically. Urban water demand has been growing, leading to rapid development in the construction and management of large-scale water supply pipes. However, as urban water supply networks continue to expand in scale and the service time increases, the leakage detection and management level of China's water supply industry still lags behind, and the leakage rate remains high. Thus, addressing water leakage is essential in the field of leakage control for large-scale water supply networks. According to statistics, approximately 60 million cubic meters of water are lost daily due to leakage in the water supply networks in Asia. The leakage could meet the water demand of 230 million people, resulting in an economic loss of over 14 billion US dollars per year. According to the 'China Urban and Rural Construction Statistical Yearbook 2021' released by the Ministry of Housing and Urban-Rural Development of the People's Republic of China (Zhu 2022), the total municipal water supply in urban and county areas of China was 79.53 billion cubic meters, with a leakage volume of 9.408 billion cubic meters and a total leakage rate of 11.83%. Among the leakage data released from 687 cities nationwide, most cities have managed to control leakage rates between 10 and 20%. There are 109 cities that have achieved a leakage rate of below 10%, while 25 cities have a leakage rate above 30%, with some even exceeding 50%. Addressing the issue of water leakage in large-scale urban water supply networks has become an urgent challenge for the water supply industry both domestically and internationally.

In the past few decades, both domestically and internationally researchers have proposed a series of leak detection techniques, continuously exploring key technologies in addressing water pipe network leaks. Currently, the commonly used methods for leak localization can be summarized as acoustic signal-based and pressure signal-based. The point to the acoustic signal method lies in the cross-correlation technique, which involves placing specific sensors at both ends of a pipe for leak detection. Due to the complex environment of actual leakage situations and the influence of low signal-to-noise ratios, it is sometimes difficult to effectively collect leakage signals (Chen *et al.* 2011). To solve such problems, researchers often use wave decomposition and transformation techniques to reduce or eliminate the noise in leakage signals. Brennan *et al.* (2019) demonstrated through the use of polarity co-incidence correlation on simulated and measured data, accurate time delay estimate can be obtained by retaining only the zero crossings in the noisy data. The results indicated that the severe clipping has no significant effect on the leakage location, which is crucial to the development of leakage noise correlators. Cody *et al.* (2020) applied linear prediction (LP) techniques to address the detection and localization problem of water distribution pipeline leaks. This data-driven method extracts features from the linear programming coefficients that represent the underlying acoustic signals. The results showed that using shorter segments of LP reconstructed

signals can achieve similar levels of accuracy as those using longer segments of raw time series, which is a key advantage in long-term online implementation applications. [Gao *et al.* \(2006\)](#) compared the ability of various cross-correlation-based delay estimators to locate leakage in plastic pipes. The experimental results indicated that random noise has little impact on the measurement error of the signal relative to the time delay estimation resolution of the low-pass filtering characteristics of the pipe. Therefore, small-scale leak signals often get submerged in the background noise, making them difficult to detect.

The point of the pressure signal method lies in the construction of a hydraulic model to simulate the operation of the pipeline network, obtaining simulated monitoring data under different leakage conditions. By minimizing the difference between the actual values and corresponding pressure and flow monitoring points in the pipeline network, the inverse problem based on the steady-state model is formulated and resolved, thereby achieving the leakage localization ([Pérez *et al.* 2011](#)). This method was first proposed by [Pudar & Liggett \(1992\)](#), where the leakage detection problem was defined as a least squares parameter estimation problem ([Pudar & Liggett 1992](#)). [Zhang \(2013\)](#) simulated pipe network leakage by calculating the orifice flow rate. They optimized the leakage locations and node emitter coefficients using the Cuckoo algorithm, then validated the effectiveness of the method through two actual pipe network cases. [Du \(2014\)](#) utilized the first-order second moment method to quantify the normal fluctuation range of pressure values and transformed leakage localization into an optimization and verification problem similar to nodal flow. They constructed an objective function by minimizing the difference between observed values and simulated values to achieve leakage localization and leakage alarm in pipeline networks. Furthermore, [Wu *et al.* \(2019\)](#) introduced a pressure-related leakage detection method. This approach constructed the objective function related to the leakage node positions and their corresponding emitter coefficients, and solved using a genetic algorithm (GA) to achieve leak detection and localization in pipeline networks. [Berglund *et al.* \(2017\)](#) compared the pressure changes between simulated normal operating conditions and actual leaking pipeline networks. Then they determined the linear combination of individual simulated leaks and used linear constraint minimization to estimate the leakage amount. The deviation between the calibration requirements given by [Sanz *et al.* \(2016\)](#) and the expected requirements indicated the presence of leaks and their approximate locations. The demand nodes are grouped based on the sensitivity of the monitoring points to changes in node requirements. [Hajibandeh & Nazif \(2018\)](#) considered operating in areas with lower pressure, where the likelihood of low leakage is higher, and took into account the probability of corresponding leakage occurring during the detection process. [Sophocleous *et al.* \(2019\)](#) adopted a method with a two-phase approach: (1) search space reduction (SSR) and (2) leak detection and location (LDL). During the SSR phase, the number of decision variables is reduced with the range of possible values, while attempting to maintain an optimal solution. Subsequently, the size and area of the leak are detected in the LDL phase.

Except for the mentioned methods, some researchers have also employed probability statistical methods to locate leakage events in the research of pressure signals. [Poulakis *et al.* \(2003\)](#) proposed a Bayesian system identification method to estimate the most probable leakage events (leakage location and amount) and the uncertainties of these estimates based on flow test data. The experimental results showed that the model with an error of less than 5% is able to correctly identify the leakage locations in the pipeline network. [Lei *et al.* \(2020\)](#) investigated the impact of node flow uncertainty on the leakage localization ability of water supply networks. In different conditions, they built a hydraulic model for the demand distribution of node flow. The results indicated that under the same leakage scenario, the decomposed model is more efficient in leakage localization compared to the uniform model. [Díaz *et al.* \(2018\)](#) proposed a leakage detection method for state estimation that evaluates leakage events by analyzing a single hydraulic state and perceives significant leakage incidents. The experimental results indicated that the probability calculation method of the measurement-estimate joint binary distribution is more efficient than the sampling method of normalized residual calculation. [Jerez *et al.* \(2021\)](#) proposed a Bayesian model updating method for locating pollution sources in water distribution networks (WDNs). By combining Bayesian model updating techniques with water quality simulation models, the most reliable location of the pollution source can be determined, along with providing information about the pollution source. [Jensen & Jerez \(2020\)](#) investigated the hydraulic characterization and evaluation of large-scale water supply pipe networks in the presence of uncertainty. They defined the categories of hydraulic models using connectivity events, calculated the posterior probabilities of each hydraulic model category, and estimated the most probable connectivity scenarios of the water supply pipe network.

Traditional methods for locating leaks do not consider errors in WDN modeling and pressure monitoring points. If errors are considered, the observed values from simulation modeling will not match the actual values, making it impossible to obtain an optimal solution. This results in the leak location deviating from the true location (Zhang 2013). Modeling errors are the errors in the roughness coefficients of pipes and the node demand in the actual network compared to the hydraulic model. Measurement errors are the differences between the node pressures in the hydraulic model and the pressures at the corresponding monitoring points in the actual network. Therefore, the key issue in the leakage probability prediction model based on the hydraulic model lies in quantifying the hydraulic model error, and addressing the deviation in the positioning of leakage caused by modeling and measurement errors. In the preliminary work of this study, we investigated a method for identifying leakage areas based on virtual areas. A leakage area identification model was constructed to identify virtual leakage areas and estimate the magnitude of leakage in each area using the improved Gray Wolf Optimization algorithm (Fang *et al.* 2023). In the previously identified virtual leakage area, this paper proposes a probability prediction method for leakage in a WDN based on Bayesian inference. First, the prior probability of leakage events is transformed into a posterior probability, using the posterior probability density function (PDF) to quantify the uncertainty associated with modeling errors and measurement errors in hydraulic simulation models. Then, the probability distributions for leakage segments and associated leakage volumes within the leakage area are calculated. Finally, the method predicts the locations of leakage segments and corresponding leakage volumes, enhancing the efficiency of leakage detection in large-scale water supply networks. This provides valuable support for decision-making in water supply enterprises.

METHODS

When an actual water supply network leaks, how to simulate the actual operation of the network by leveraging the hydraulic model under normal operating conditions, is a significant problem in leak probability prediction models (Berglund *et al.* 2017). Hydraulic modeling is used to simulate the behavior of water flow in hydraulic systems. Although data does not change when modeling at a single point in time, as modeling is conducted at different time points, measurement errors and modeling errors will continuously arise. Therefore, by utilizing Bayesian inference based on probability statistics, we can reduce these errors through multiple data collection iterations. In previous research, it has been indicated that constructing a leakage area identification model based on optimization methods enables the recognition of virtual leakage areas and estimation of the magnitude of leakage for each virtual area (Fang *et al.* 2023). Based on that, the probability estimation of each leakage event is provided using Bayesian inference, and all simulated leakage events are prioritized. That means, the event with the highest probability is considered to be the leakage event most likely to occur in the WDN, thereby achieving accurate leakage prediction. The proposed leakage localization framework based on Bayesian inference consists of two stages: the initialization process and posterior probability calculation.

Phase I initialization process

Leakage occurring at any position in the WDN can lead to changes in network pressure, and these changes depend on the area and the corresponding amount of leakage. In this paper, it is assumed that the leak occurred at the node. Given the actual total leakage quantity Q in the pipe network, the leakage is simulated within the intervals and ranges where the water demand continuously varies, as shown in the following equation.

$$Q = q_1 + q_2 + \dots + q_s, s \in \{1, \dots, S\} \quad (1)$$

where Q represents the total leakage of the pipe network, q_s represents the leakage at leakage points, S represents the number of leaks in the pipe network, and S is less than or equal to the total number of network nodes N . If there is only one leak in the pipe network, S equals 1, as shown in the following equation.

$$0 \leq \Delta q_n \leq Q_L, n \in \{1, \dots, N\} \quad (2)$$

where Δq_n represents the change interval of leakage amount and leakage position. Q_L represents the maximum leakage amount at a node, while N represents the total number of leakage nodes in the pipe network. Q_L is less than or equal to Q .

Phase II calculates the posterior probability

Bayesian inference describes the situation of unknown variables through probability distributions. This probability distribution is an estimating process about the unknown variables based on existing knowledge. The flexibility and generality of Bayesian inference enable it to handle quite complex problems, and it does not hinder the fitting of models with many parameters and complex multilevel probability specifications.

In the leakage hydraulic model (denoted as Ω), we use leakage amount and leakage location to describe the leakage events. Therefore, different leakage positions and leakage amounts constitute a series of parameter sets γ , $\gamma = (j, Q_j^l)$. For multi-point leakage, subset j and Q_j^l are column vectors, where the number of leakage points corresponds to the dimension of this vector. For pressure monitoring points, assuming there are S monitoring points, and each monitoring point has undergone T measurements when leakage occurs, we can use p_{ij}^m to represent the pressure measurement value of the i th monitoring point at the j th time. Therefore, P^m is a matrix of size $S \times T$ with elements p_{ij}^m .

Assuming $P^c(\gamma) = [p_{ij}^c(\gamma)]$ is the simulated data of pressure monitoring points obtained through the leakage hydraulic model, the error between the observed data of pressure monitoring points and the simulated data is denoted as $E(\gamma)$, as shown in the following equation.

$$E(\gamma) = P^m - P^c(\gamma) \quad (3)$$

where $E(\gamma)$ is the $S \times T$ matrix composed of the error $e(\gamma)$ between observed data and simulated data from pressure monitoring points under each operating condition.

In Bayesian inference, the mentioned error is considered as a random number from the probability error model Π , where the corresponding parameter is denoted as σ . Therefore, hydraulic model and model Π parameters can be represented by a parameter set $[\gamma, \sigma]$. On the basis of continuously updating the parameter set $[\gamma, \sigma]$, we first calculate the uncertainty of each leakage event using Bayesian inference, and quantify the uncertainty of the leakage event through PDFs. Here, the uncertainty of the parameter set $[\gamma, \sigma]$ is quantified based on engineering experience and set as $\Pi[\gamma, \sigma]$, which represents a non-informative prior distribution, i.e., a constant. This means that all possible parameter values are assigned the same initial probability density. This choice ensures unbiased treatment of all potential parameter values in the absence of additional information. Then, according to Bayesian theory, the prior probability is transformed into the posterior probability, as shown in the following equation.

$$f(\gamma, \sigma | P^m) = c_1 f(P^m | \gamma, \sigma) \Pi(\gamma, \sigma) \quad (4)$$

where $f(\gamma, \sigma | P^m)$ represents the posterior PDF of a leakage event, $f(P^m | \gamma, \sigma)$ represents the prior PDF of the leakage event in the measurement data, c_1 is a constant that ensures the integral of the posterior PDF equals 1.

Assuming that each element $e(\gamma)$ in $E(\gamma)$ follows a normal distribution $N \sim (0, \gamma)$, the expression of $f(\gamma, \sigma | P^m)$ is shown in the following equation.

$$f(P^m | \gamma, \sigma) = \prod_{i=1}^S \prod_{j=1}^T \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{(p_{ij}^c(\gamma) - p_{ij}^m)^2}{2\sigma^2} \right] \quad (5)$$

As mentioned, it is known that $\Pi[\gamma, \sigma]$ is a constant. And $(p_{ij}^c(\gamma) - p_{ij}^m)$ is $e_{ij}(\gamma)$. Therefore, substituting Equation (4) into Equation (5) yields the resulting equation, as shown in the following equation.

$$f(\gamma, \sigma | P^m) = c_2 \frac{1}{(\sqrt{2\pi\sigma})^{ST}} \exp \left[-\frac{\sum_{i=1}^T \| p_j^i(\gamma) - p_j^m \|^2}{2\sigma^2} \right] \quad (6)$$

where $\| \cdot \|$ represents the Euclidean norm, The normalization constant c_2 ensures that the posterior PDF integrates to 1 over the entire parameter space. Since the prior distribution is a constant, c_2 includes this constant and is determined during the normalization process, ultimately not affecting the posterior probability.

When $f(\gamma, \sigma|P^m)$ reaches its maximum value, the model parameter $[\hat{\gamma}, \hat{\sigma}]$ is the most likely leakage event to occur. Therefore, the optimal $[\hat{\gamma}, \hat{\sigma}]$ can be found by finding the maximum value of $f(\gamma, \sigma|P^m)$. Take the negative natural logarithm of $f(\gamma, \sigma|P^m)$, as shown in the following equation.

$$g(\gamma, \sigma) = -\ln[f(\gamma, \sigma|P^m)] \\ = \frac{1}{2\sigma^2} \sum_{j=1}^T \|p_j^c(\gamma) - p_j^m\|^2 + \frac{ST}{2} \ln \sigma^2 + c \quad (7)$$

Converting the task of seeking the maximum value of $f(\gamma, \sigma|P^m)$ into the task of finding the minimum value of $g(\gamma, \sigma)$. According to Equation (7), we can obtain the optimal value of $\hat{\gamma}$ by minimizing the function $h(\gamma)$, as shown in the following equation.

$$h(\gamma) = \frac{1}{ST} \sum_{j=1}^T \|p_j^c(\gamma) - p_j^m\|^2 \quad (8)$$

Equation (8) represents the error between the simulated values and observed values of the model. By calculations, the optimal value for the predicted error $\hat{\sigma}^2$, can be obtained, as shown in Equation (9). The calculation of the marginal probability density is shown in the following equation.

$$\hat{\sigma}^2 = h(\gamma) \quad (9)$$

$$f(\gamma|P^m) = \int f(\gamma, \sigma|P^m) d\sigma \quad (10)$$

By combining the given Equations (8)–(10), we can obtain a posterior probability, as shown in the following equation.

$$f(\gamma|P^m) = c[h(\gamma)]^{-(ST-1)/2} K(\gamma) \quad (11)$$

where $K(\gamma) = \Pi(\gamma, \sqrt{h(\gamma)})$ represents a non-informative prior distribution, which is a constant. As shown in the following equation.

$$f(\gamma|P^m) = c[h(\gamma)]^{-(ST-1)/2} \quad (12)$$

Thus, we obtain the posterior probability of the optimal estimation of the model parameter γ . By calculating formula (12) to prioritize the model parameter set γ , values with higher probabilities are considered as the optimal values (\hat{j}, \hat{Q}_j^l) , representing the most probable locations and corresponding amounts of leakage.

EXPERIMENT AND RESULT ANALYSIS

To assess the reliability of the proposed model, we conducted experiments on the NET3 benchmark model network, which is publicly available from the Rossman team. In previous work, a method for identifying leakage areas based on virtual partitioning was studied. The hydraulic model was used to simulate leakage in the pipe network, establish rules for virtual partitioning, reduce the search space of the model, and identify abnormal areas, providing predictive targets for leakage probability prediction (Fang *et al.* 2023).

As shown in Figure 1, the case network NET3 consists of 92 water consumption nodes, 117 pipes, 2 water sources, 2 pumps, 3 water tanks, and other physical components. The total inflow, total node demand, and total pipe length of the network are 828.95 L/s, 192.28 L/s, and 65.7 km, respectively. Furthermore, after constructing the topology structure of the water supply network, the node water demand is allocated based on the actual water usage data at each node of the water supply network. Considering the delayed operational status of the actual water supply network, it is necessary to set a water demand pattern for each water-consuming node in the network. This water demand pattern is a set of data that changes over time, simulating the variations in water demand in the supply network between 0 and 24 h, as shown in Figure 2.

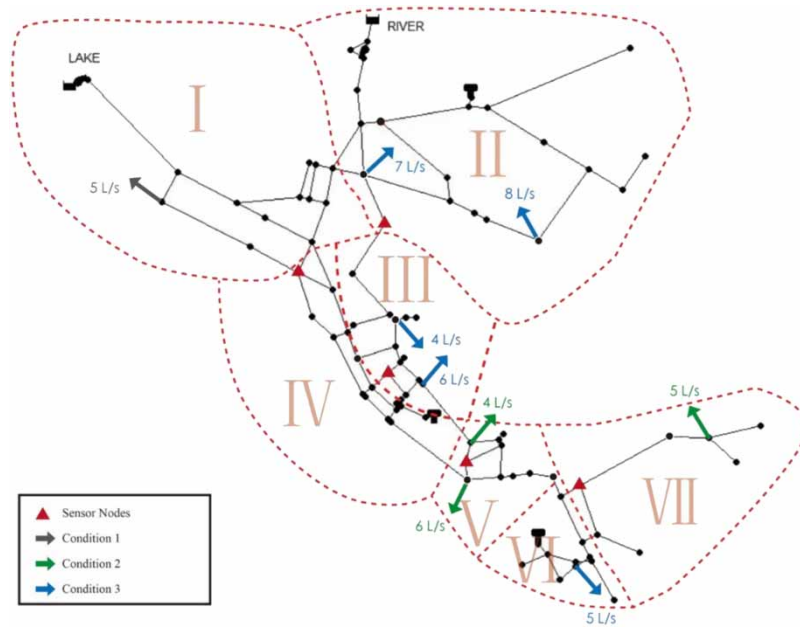


Figure 1 | Partitioning of the NET3 network and simulation design diagram for leakage conditions.

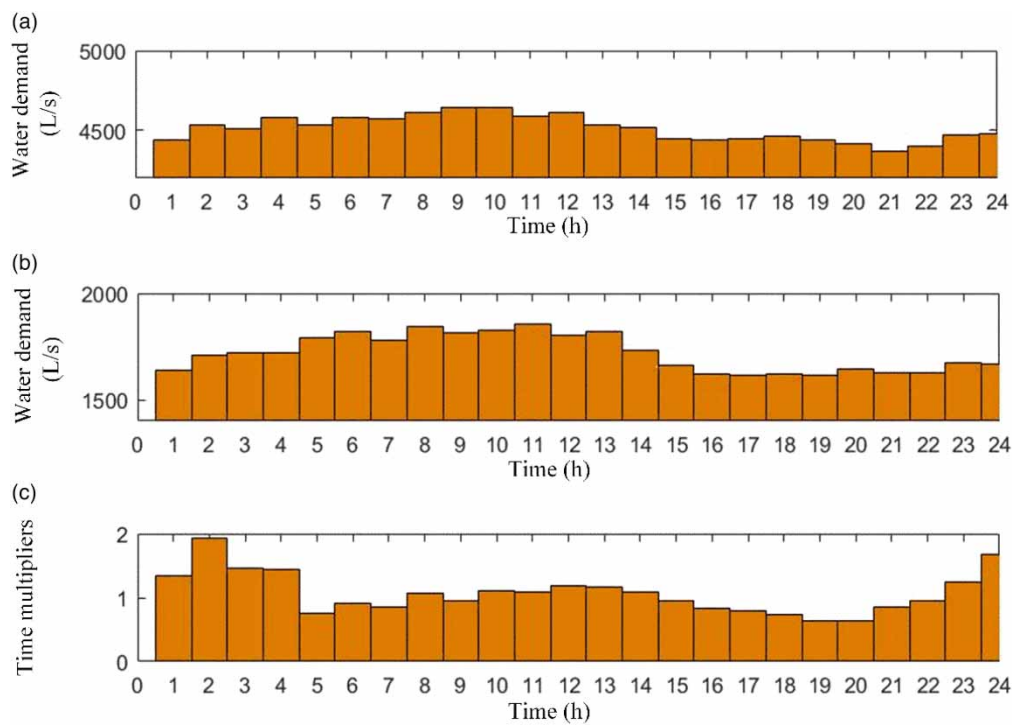


Figure 2 | 24-h variations: (a) first water demand mode; (b) second water demand mode; (c) time multipliers.

We focus on the issue of leakage probability prediction within virtual areas I and II. Single-point leakage and multi-point leakage scenarios are simulated separately, along with variation and pattern of node water demand. As shown in Figure 2(a) and 2(b) represents variations in node water demand (L/s), while the remaining set represents time multipliers (dimensionless). The time interval for all sets is 1 hour, during which the values remain constant indicating that the water demand remains unchanged and is equal to the product of the base water demand and the time multiplier for that time period. It is worth noting that nodes 15, 123, 203, and 35 in the network have different demand patterns, which are not detailed here.

During our experiment, we simulate two types of leakage scenarios: single-point leakage and multi-point leakage. Simulated single-point leakage occurs in virtual area I, which includes 14 water consumption nodes, 17 pipes, 1 water source, and 1 water pump, among other physical components. Simulated multi-point leakage occurs in virtual area II, which includes 22 water nodes, 25 pipes, 1 water source, 1 water tank, 1 water pump, and other physical components.

As shown in Figure 3, Figure 3(a) represents a simulated single-point leakage experiment. The simulation involves a leakage of 5 L/s at node 3. The number of monitoring points $S = 2$. They are located at nodes 7 and 14 in Figure 3(a). In the WDN, multi-point leakage events can be simulated by adding multiple single-point leakage events. As shown in Figure 3(b), it simulates two leakages occurring at nodes 6 and 17, with leakage magnitudes of 10 and 5 L/s, respectively. The number of monitoring points $S = 2$. They are located at nodes 8 and 14 in Figure 3(b) (Hu *et al.* 2021).

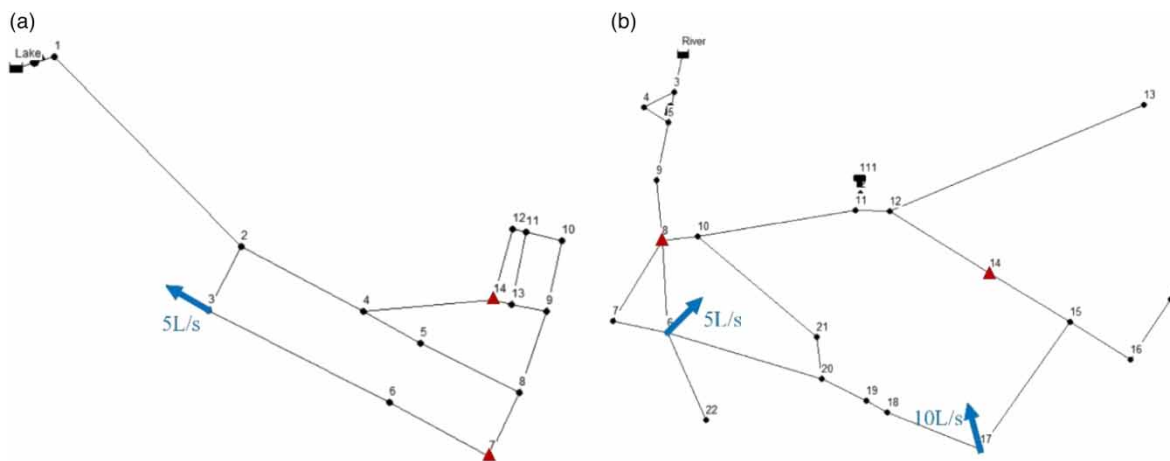


Figure 3 | Leakage simulation design diagram within the pipe network area: (a) single-point leakage; (b) multi-point leakage.

Using simulated pressure values under leakage conditions as field actual pressure observations. Quantifying the uncertainties of modeling errors and measurement errors in hydraulic simulation models through a leak probability prediction model. Calculate the probability distribution of leakage nodes and leak quantities within the leakage area, and then search potential leakage events. Finally, by invoking the EPANET dynamic link library, select any node in the pipe network as the leakage node to simulate leakage conditions occurring in an actual pipe network.

Prediction of single-point leakage probability

Due to the differences between the actual WDN and the hydraulic model, random errors are added to the model's predictions to simulate the differences between the actual network predictions and the modeling process, and the measurements from sensors. Furthermore, it introduces the random errors α , β , and γ , which respectively represent the roughness coefficient error of the pipe network model, the error in node water demand, and the error in pressure monitoring points. Assuming that the variations of each error term follow a zero-mean uniform distribution, with boundaries of $(-\alpha, \alpha)$, $(-\beta, \beta)$, and $(-\gamma, \gamma)$, these values represent the magnitude of model errors, expressed as a percentage of the actual WDN.

As mentioned earlier, the error is simulated by adding zero-mean uniform noise to the actual parameters, and the error is equal to 5% of the actual system parameters (Poulakis *et al.* 2003). First, consider collecting data once to calculate the posterior probability. The impact of roughness coefficient error and node water demand error on the prediction results are shown in Figure 4.

As shown in Figure 4(a) and 4(b), respectively, represent the probability distribution of leakage events for each type after adding roughness coefficients and errors in water demand for nodes. The leakage probability is expressed as the posterior probability based on Bayesian inference. The experimental results indicate that node 3 has the highest probability of leakage, with the probability of leakage in other nodes being significantly lower than that of node 3. Near the topology structure of node 3, nodes 6–8 have a higher probability of leakage

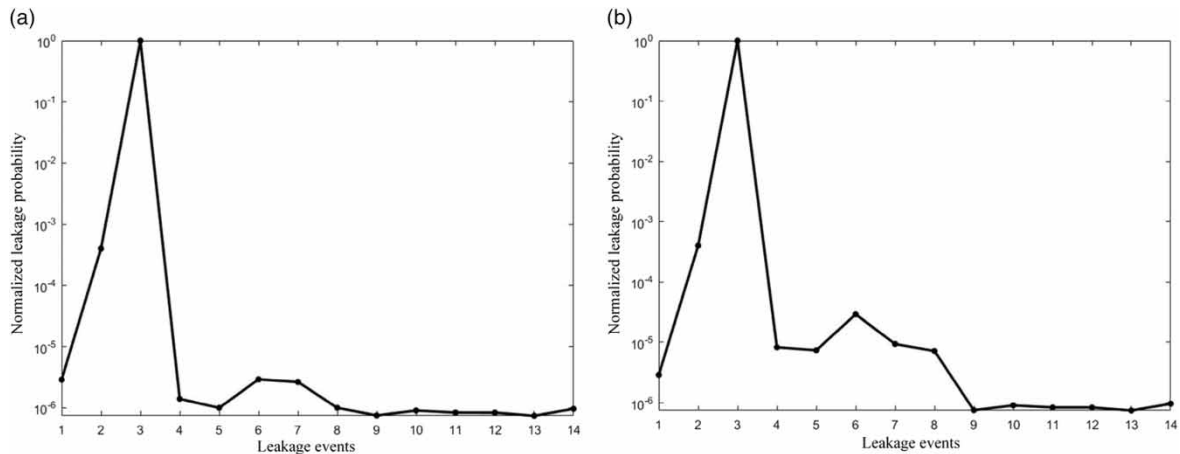


Figure 4 | Prediction results after adding 5% roughness coefficient error or node water demand error.

compared to other nodes. The closer the nodes are to the location of the leakage point, the higher the probability of leakage compared to other positions. This is because there was a leakage at node 3, which resulted in a decrease in pressure in the neighboring nodes. Therefore, even with the addition of 5% roughness coefficient error and error in node water demand, the model is still able to accurately predict the location of leakage.

As shown in Figure 5, after adding a measurement error of 5%, the model failed to accurately predict the location of the leakage (at node 3) under the single data collection scenario but instead detected a leakage occurring near node 6. This indicates that the accuracy of pressure monitoring points in the pipe network plays a crucial role in predicting the probability of leaks, and the detection results from a single set of data collection are not satisfactory. Therefore, consider using statistical methods, with intervals of one-time units, to reduce modeling and measurement errors through multiple data collection. As shown in the red curve in Figure 5, considering multi-group data, the model is able to correctly identify the location of the leakage. The position with the highest probability of leakage is located at node 3, while the leakage probabilities of other nodes are much lower than that of node 3. Therefore, compared to single data collection, the results of collecting data from 15 groups can accurately predict the location of leaks. Multiple data collections based on statistical methods contribute to enhancing leak prediction effectiveness.

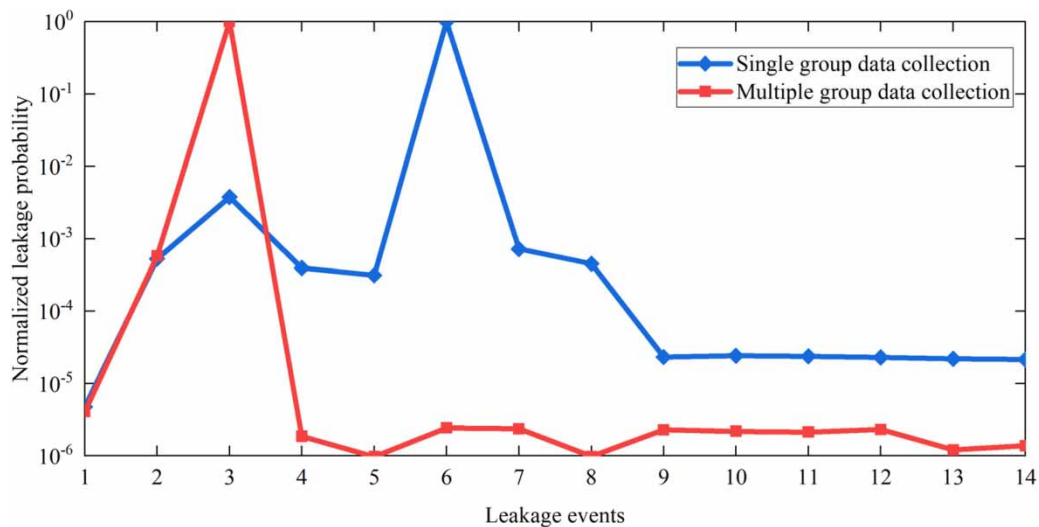


Figure 5 | With the addition of a 5% pressure monitoring point error, the predicted results of single and multiple data collections.

Prediction of multi-point leakage probability

When leakages occur in the water supply network, the actual number of leaks often cannot be determined in advance. Even when the number of leaks is known, there may be thousands of possible combinations of leaks, resulting in a large computational burden and low accuracy. Therefore, based on the single-point leakage probability prediction model, we investigate a multi-point leakage prediction method. Assuming the total leakage of the pipe network is known, the total leakage can be calculated based on the pressure of the network (Maghrebi *et al.* 2014). With the increase in network leakage, the pressure decreases nonlinearly, but within a smaller range, it shows a linear correlation (Sophocleous *et al.* 2019). In the WDN, multi-point leakage events can be simulated by adding multiple single-point leakage events, and the location of the leakage is more likely to cause pressure fluctuations than the amount of leakage.

The multi-point leakage prediction framework also includes two phases: the initialization process and the calculation of posterior probability, as shown in Figure 6. The initialization process proposes that, given the known leakage areas and their corresponding leakage quantities, the leakage prediction problem is essentially a problem of how to allocate the total leakage to the nodes. As the number of leakage points in the pipe network cannot be obtained in advance, we adopt a K -value decrement method to achieve leakage prediction, where K represents the number of leakage points in the pipe network. If the value of K decreases to 0, it indicates that all leakage incidents in the pipe network have been thoroughly inspected, and K leak locations are outputted.

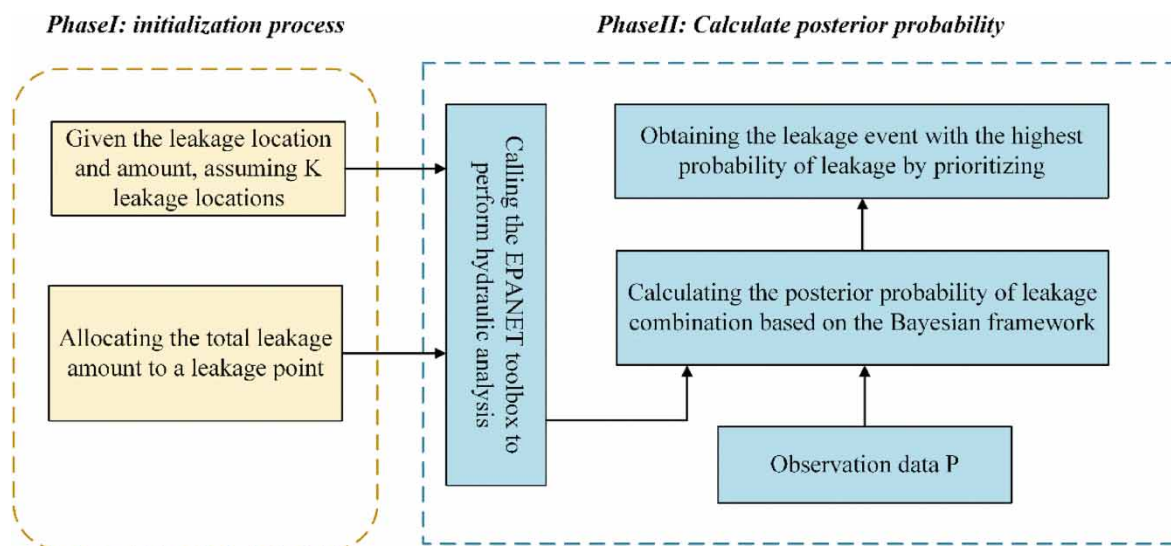


Figure 6 | Multi-point leakage probability prediction model framework.

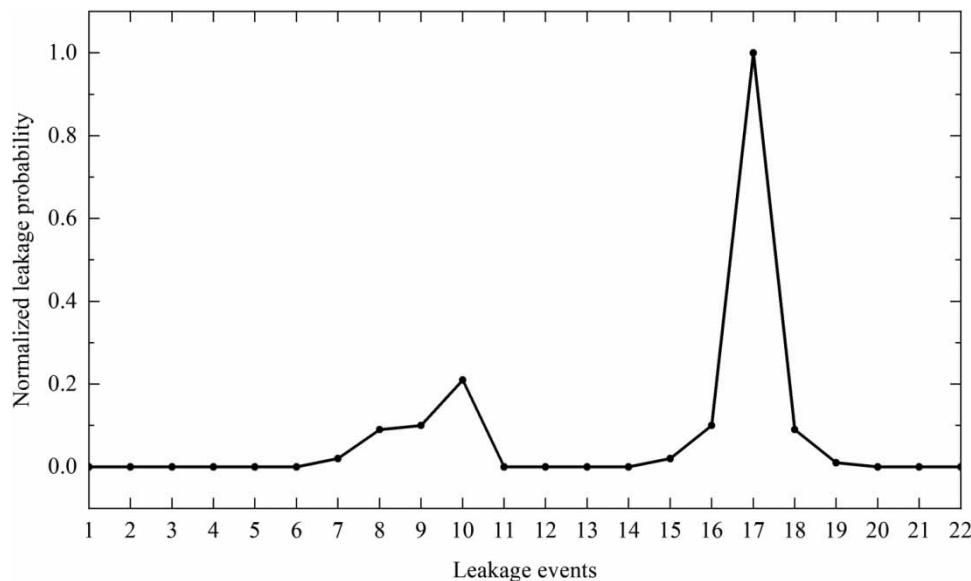
In phase II, the total network leakage is added to an arbitrary node. The EPANET toolkit is then called to perform hydraulic analysis and obtain simulated data for the leakage model. Calculating the posterior probability of leakage combinations based on the Bayesian framework and prioritizing them, thereby obtaining the leakage events with the highest probability of leakage.

As shown in Figure 3(b), the linear combination of individual leaks can simulate multiple leakage scenarios in the pipe network. In this simulation, two leaks occur at nodes 6 and 17, with leakage rates of 5 and 10 L/s, respectively. The number of monitoring points $S = 2$, with pressure monitoring points located at nodes 8 and 14. Under the condition of adding 5% measurement error, consider using statistical methods, with intervals of one-time units, to quantify the measurement errors through multiple data collection. According to Table 1, a multiple-point leakage event occurred in the simulated WDN. Collecting experimental results of 10 independent experiments under this leakage event were collected. The three possible leakage events with the highest probabilities of leakage were selected based on the ranking of leakage probabilities in each independent experiment. The experimental results indicate that, among ten independent experiments, the leakage event with the highest predicted probability of leakage occurred at node 17 in eight experiments. The predicted results of the remaining two independent experiments are nodes 18 and 20 (both nodes are located near node 17).

Table 1 | The predicted results of the probability of multiple-point leakage under 10 independent tests

	ID	P	ID	P	ID	P	ID	P	ID	P
	One		Two		Three		Four		Five	
1	17	1	17	1	17	1	17	1	18	1
2	10	0.21	7	0.38	7	0.38	10	0.21	7	0.38
3	9	0.1	8	0.09	8	0.09	9	0.1	16	0.1
	Six		Seven		Eight		Nine		Ten	
1	17	1	17	1	17	1	17	1	20	1
2	6	0.69	10	0.21	10	0.21	6	0.69	10	0.21
3	8	0.09	9	0.1	9	0.1	8	0.09	16	0.1

As shown in Figure 7, the multi-point leakage probability prediction results exhibit two peaks, with a vertical axis interval of 0.1. The first peak points to node 17, and the second peak points to node 10. It can be observed from Figure 8 that the first peak refers to the location where leakage occurs in the actual pipe network, while the other peak refers to a nearby location of the actual leakage position. Therefore, in practical work, it is possible to eliminate the most likely leakage locations one by one, transforming the complex problem of multi-point leakage detection into multiple single-point leakage detection problems. Water utility workers can identify the actual location of leaks by conducting tests and inspecting nearby pipes.

**Figure 7** | Multi-point leakage probability prediction results.

As shown in Figure 9, after excluding the leakage event at node 17, the hydraulic information of the WDN is modified. Thus, the problem of predicting the probability of multiple-point leakage is transformed into the problem of predicting the probability of a single-point leakage. By collecting more sets of data, the model can accurately identify the locations of leakage, with the highest probability of leakage occurring at node 6. The probability of leakage at other nodes is significantly lower than that at node 6. Therefore, compared to single-group data collection, collecting multiple groups of data samples predicts the leakage location very stably. Multiple collections based on statistical methods help improve the effectiveness of leakage prediction.

In this study, we also compare our method with the leak localization approaches proposed by Lijuan *et al.* (2012) and Sophocleous *et al.* (2019), which utilize optimization algorithms. Each method underwent 10 independent experiments, with the runtime representing the duration required for a single experiment (conducted on an Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz, 16 GB RAM). The results are presented in Table 2.

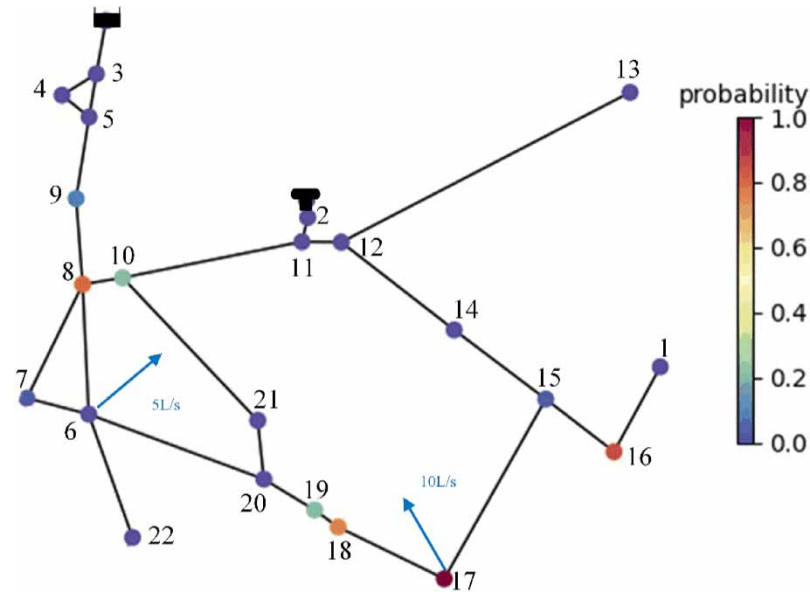


Figure 8 | Multi-point leakage probability prediction map.

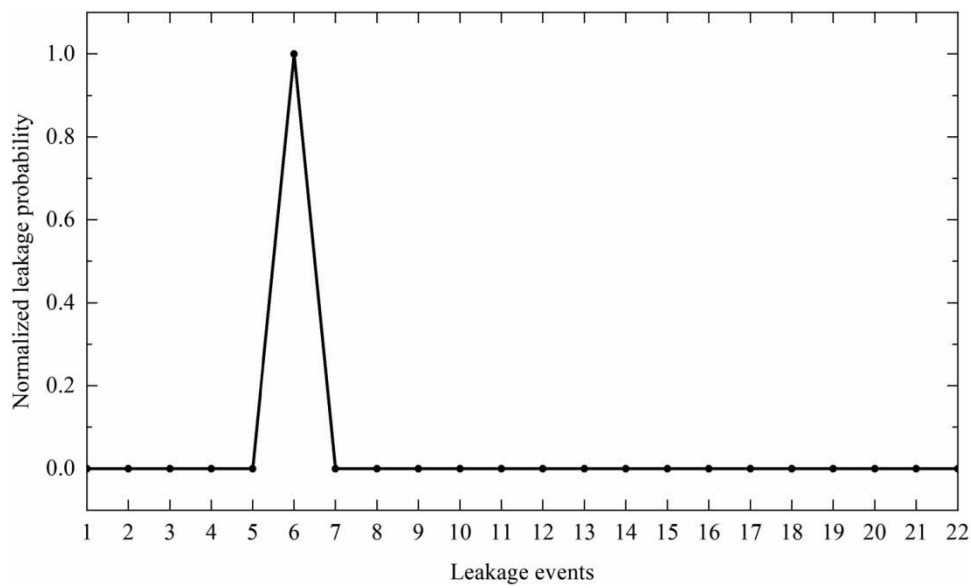


Figure 9 | Prediction results of single-point leakage probability after excluding node 17 leakage event.

Table 2 | Comparison results between Bayesian inference and GA

	Single-point leakage		Multi-point leakage	
No error added				
	Correct count	Run time	Correct count	Run time
Bayesian	10	3 s	10	9 s
GA	10	705 s	-	-
Add error				
	correct count	Run time	correct count	Run time
Bayesian	10	4 s	8	10 s
GA	7	550 s	-	-

As shown in Table 2, for the single-point leakage scenario, both Bayesian inference and the GA can effectively localize the leakage when no error is added. However, the GA's runtime is significantly longer than that of Bayesian inference. When errors are introduced, Bayesian inference still accurately localizes the leakage in a short amount of time, whereas the GA not only takes much longer but also fails to correctly localize the leakage in three out of ten trials. For a multi-point leakage scenario, Bayesian inference accurately localizes the leakages when no error is added. When errors are introduced, Bayesian inference fails to correctly localize the leakages twice, whereas the GA is unable to localize multi-point leakages altogether. Therefore, Bayesian inference is not only leakage localization but also effectively handles the impact of errors.

CONCLUSIONS

We build a leak prediction model based on Bayesian inference, achieving single-point and multi-point leak prediction within the leakage region. Initially, a leakage prediction model is constructed based on Bayesian inference to calculate the posterior probability of each leakage event. Subsequently, considering the uncertainties in the roughness coefficient and the water demand, modeling errors, and measurement errors should be added. Finally, considering the multi-point leakage in the pipe network, multiple sets of data are collected to solve the predictive model. The experimental results indicate that multiple collections based on statistical methods contribute to the accuracy and reliability of leakage prediction. Collecting multiple sets of pressure data can predict the leakage location very stably, and investigate leakage events one by one to achieve multi-point leakage probability prediction. This can provide water workers with more precise guidance.

This paper explores the feasibility of Bayesian methods in multi-point leakage detection but does not delve into scenarios with numerous leakage points, such as five or ten leakage points. Future research should further investigate the accuracy of this method in cases with numerous leakage points and its accuracy in larger WDNs.

ACKNOWLEDGEMENTS

This work is supported by the funds: National Key Research and Development Project No.2023YFC3807704; Scientific Research Program of Anhui Provincial Department of Education No.2022AH010018.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Berglund, A., Areti, V. S., Brill, D. & Mahinthakumar, G. (. (2017) [Successive linear approximation methods for leak detection in water distribution systems](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000784), *Journal of Water Resources Planning and Management*, **143** (8), 04017042. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000784](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000784).
- Brennan, M. J., Gao, Y., Ayala, P. C., Almeida, F. C. L., Joseph, P. F. & Paschoalini, A. T. (2019) [Amplitude distortion of measured leak noise signals caused by instrumentation: Effects on leak detection in water pipes using the cross-correlation method](https://doi.org/10.1016/j.jsv.2019.114905), *Journal of Sound and Vibration*, **461**, 114905. <https://doi.org/10.1016/j.jsv.2019.114905>.
- Chen, L., Liu, Y., Kong, F. & He, N. (2011) [Acoustic source localization based on generalized cross-correlation time-delay estimation](https://doi.org/10.1016/j.proeng.2011.08.915), *Procedia Engineering*, **15**, 4912–4919. <https://doi.org/10.1016/j.proeng.2011.08.915>.
- Cody, R. A., Dey, P. & Narasimhan, S. (2020) [Linear prediction for leak detection in water distribution networks](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000415), *Journal of Pipeline Systems Engineering and Practice*, **11** (1), 04019043. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000415](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000415).
- Díaz, S., Mínguez, R. & González, J. (2018) [Probabilistic leak detectability assessment via state estimation in water transport networks](https://doi.org/10.1007/s00477-018-1515-3), *Stochastic Environmental Research and Risk Assessment*, **32** (7), 2111–2128. <https://doi.org/10.1007/s00477-018-1515-3>.
- Du, K. (2014) *Study of Hydraulic Model Calibration and Leakage Location of Water Distribution System*. MS thesis, Chongqing University.
- Fang, Q., Chen, J. & Xie, C. (2023) [Research on leakage area detection method in water distribution network based on gray wolf optimization](https://doi.org/10.2166/ws.2023.014), *Water Supply*, **23** (2), 645–656. <https://doi.org/10.2166/ws.2023.014>.
- Gao, Y., Brennan, M. J. & Joseph, P. F. (2006) [A comparison of time delay estimators for the detection of leak noise signals in plastic water distribution pipes](https://doi.org/10.1016/j.jsv.2005.08.014), *Journal of Sound and Vibration*, **292** (3–5), 552–570. <https://doi.org/10.1016/j.jsv.2005.08.014>.

- Hajibandeh, E. & Nazif, S. (2018) Pressure zoning approach for leak detection in water distribution systems based on a multi objective ant colony optimization, *Water Resources Management*, **32** (7), 2287–2300. <https://doi.org/10.1007/s11269-018-1929-1>.
- Hu, X., Han, Y., Yu, B., Geng, Z. & Fan, J. (2021) Novel leakage detection and water loss management of urban water supply network using multiscale neural networks, *Journal of Cleaner Production*, **278**, 123611. <https://doi.org/10.1016/j.jclepro.2020.123611>.
- Jensen, H. A. & Jerez, D. J. (2020) A stochastic framework for hydraulic performance assessment of complex water distribution networks: Application to connectivity detection problems, *Probabilistic Engineering Mechanics*, **60**, 103029. <https://doi.org/10.1016/j.probenmech.2020.103029>.
- Jerez, D. J., Jensen, H. A., Beer, M. & Broggi, M. (2021) Contaminant source identification in water distribution networks: A Bayesian framework, *Mechanical Systems and Signal Processing*, **159**, 107834. <https://doi.org/10.1016/j.ymssp.2021.107834>.
- Lei, Y., Du, K., Ding, R., Mao, R. & Chen, P. (2020) Influence of uncertainty of flow rate in water supply network node on leakage location, *Journal of Water Resources and Water Engineering*, **19** (01), 144–147.
- Lijuan, W., Hongwei, Z. & Hui, J., (2012) A leak detection method based on epanet and genetic algorithm in water distribution systems. In: Wu, Y. (ed.) *Software Engineering and Knowledge Engineering: Theory and Practice*, Vol. 114. Berlin, Heidelberg: Springer, pp. 459–465. https://doi.org/10.1007/978-3-642-03718-4_57.
- Maghrebi, M. F., Aghaebrahimi, M. R., Taherian, H. & Attari, M. (2014). Determining the amount and location of leakage in water supply networks using a neural network improved by the Bat optimization algorithm. *Journal of Civil Engineering and Urbanism*, **4** (3), 322–327.
- Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. & Peralta, A. (2011) Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks, *Control Engineering Practice*, **19** (10), 1157–1167. <https://doi.org/10.1016/j.conengprac.2011.06.004>.
- Poulakis, Z., Valougeorgis, D. & Papadimitriou, C. (2003) Leakage detection in water pipe networks using a Bayesian probabilistic framework, *Probabilistic Engineering Mechanics*, **18** (4), 315–327. [https://doi.org/10.1016/S0266-8920\(03\)00045-6](https://doi.org/10.1016/S0266-8920(03)00045-6).
- Pudar, R. S. & Liggett, J. A. (1992) Leaks in pipe networks, *Journal of Hydraulic Engineering*, **118** (7), 1031–1046. [https://doi.org/10.1061/\(ASCE\)0773-9429\(1992\)118:7\(1031\)](https://doi.org/10.1061/(ASCE)0773-9429(1992)118:7(1031)).
- Sanz, G., Pérez, R., Kapelan, Z. & Savic, D. (2016) Leak detection and localization through demand components calibration, *Journal of Water Resources Planning and Management*, **142** (2), 04015057. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000592](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000592).
- Sophocleous, S., Savić, D. & Kapelan, Z. (2019) Leak localization in a real water distribution network based on search-space reduction, *Journal of Water Resources Planning and Management*, **145** (7), 04019024. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001079](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001079).
- Wu, Z. Y., Sage, P. & Turtle, D. (2010) Pressure-dependent leak detection model and Its application to a district water system, *Journal of Water Resources Planning and Management*, **136** (1), 116–128. [https://doi.org/10.1061/\(ASCE\)0773-9496\(2010\)136:1\(116\)](https://doi.org/10.1061/(ASCE)0773-9496(2010)136:1(116)).
- Zhang, Q. (2013) *Research on Leakage Detection Via Hydraulic Model Calibration in Water Distribution Systems*. MS thesis, Harbin Institute of Technology.
- Zhu, D. (2022) *Study on Leak Detection and Localization of Urban Water Distribution Networks Based on Data-Driven Approaches*. MS thesis, Chang'an University.

First received 4 March 2024; accepted in revised form 28 August 2024. Available online 12 September 2024