Research Article

A Multivariate Statistical Model of Water Leakage in Urban Water Supply Networks Based on Random Matrix Theory

Yumo Chen¹ and Ying Yang²

¹School of Xi'an Vocational and Technical College, Xi'an, Shaanxi 710077, China
²School of Xi'an Physical Education University, Xi'an, Shaanxi 710068, China

Correspondence should be addressed to Ying Yang; 501029@tea.xaipe.edu.cn

Received 21 April 2022; Revised 25 May 2022; Accepted 1 June 2022; Published 20 June 2022

1. Introduction

The urban water supply network is the city’s infrastructure and an essential facility to maintain the regular operation of the town. With the rapid development of the urban economy, the urban water supply network has taken on more and more heavy water supply tasks. The water supply network has become more and more complex, and the management methods of water supply enterprises have also undergone significant changes [1]. As a result, the previous management style of empirical management is not fully capable of operating the network under high pressure. Various problems in the operation of the network have caused a lot of contradictions in the direction of the water supply industry. The frequent occurrence of contradictions in the water supply industry also means that the water supply industry must make revolutionary innovations in the management style [2]. However, although the number of urban residents has been increasing in recent years and the city has been expanding to meet the development needs, the pipeline network infrastructure has also been growing, resulting in an urban water supply pipeline network with multiple water sources, multiple water plants, multiple transmission and distribution ring networks, and frequent uncertainty accidents, but the way urban infrastructure is managed has not changed much [3]. The current experience management
mode of pipeline network time-scheduled plan can meet the water supply requirements. Still, it is not easy to be scientific and intelligent, which consumes human and material resources and makes it challenging to meet the higher standards of pipeline network users. It is difficult to fulfill the requirements through traditional backward management and empirical scheduling methods to achieve smooth and safe operation of the water supply network. Underwater quality, water quantity, and water pressure. In recent years, water supply accidents caused by the lack of unscientific management of urban water supply networks, such as local flooding, water quality pollution, electricity consumption waste, and other problems, have brought severe economic losses.

With the national emphasis on developing smart cities, the concept of intelligent water services came into being. As an essential part of the construction of intelligent cities, scientific and clever use of water resources, the construction of intelligent, refined water systems is the fundamental guarantee of the structure of smart cities [4]. Through the study of advanced water supply technology and concepts, and combined with their situation, the water division around the town is in full swing to carry out the construction of urban intelligent water services and independent metering area DMA (district metering area) zoning construction and gradually establish and improve the SCADA system for pipe network monitoring and data acquisition (supervisory control and data acquisition), Pipeline GIS (geographic information system), and permalog leakage monitoring system [5]. Using advanced technical means and intelligent platform equipment to provide reliable data support for water supply network status monitoring, optimal scheduling, leakage monitoring, and other research, thus improving water supply efficiency and pipeline safety monitoring capabilities, have become the primary development direction of the water supply industry. This method is a classification-based leakage detection method. Still, it has some practical applications because it simulates leakage by adding leakage quantities to the no-leakage flow data set as a training set without requiring a large amount of historical monitoring data with labels.

With the increasing awareness of water resources protection, the water supply department has increased the detection of pipelines into passive detection for active detection, without waiting for the leakage water to spill out of the ground before judging its occurrence of leakage, which has achieved specific results in reducing the amount of leakage and improving the efficiency of water supply [6]. There are many reasons for the leakage of pipe network, but the amount of leakage of the pipe network is only related to three factors: first, the appearance of the pipe breakage point is too long; second, the shape, size, and number of breakage points and other parameters related to pipe rupture; third, the pressure of water supply in the pipe. Experimental studies also show a close link between water supply pressure and leakage. The scientific management of the pressure in the network pipe can reduce the chances of leakage and burst [7]. Pressure management is a very economical and effective loss reduction method, which plays a positive role in reducing network leakage, preventing pipe bursts, controlling leakage rebounds, delaying leak detection cycles, and maintaining leak detection results. Using pressure control to reduce leakage across the network is a cost-effective way to do so and can reduce the risk of more severe leaks through gentle pressure fluctuation control. It is essential to improve system efficiency and reduce costs in the water supply process [8]. An efficient water network system provides equivalent service and reduces energy consumption. One effective improvement is the management of network pipe system pressure, optimizing the least stressed network nodes, which also helps to improve other known leakage problems.

2. Related Works

Urban water supply pipelines have been used for a long time and have serious aging problems. The pipes are underground, making maintenance complex. The construction and maintenance of the network are lagging behind the overall world standard, which makes water resources seriously wasted and causes an increase in water supply costs and brings a lot of inconvenience to citizens’ lives, which is not conducive to the harmonious development of society [9]. It is necessary to test and evaluate the urban water supply pipelines in stages and then implement corresponding renovation measures. At present, the information management about the water supply network is slightly inadequate, and the evaluation of the health condition of the network is still in the development stage [10]. The methods to study the health condition of the water supply network are currently divided into the following: hierarchical analysis, multiple linear regression, fuzzy theory, and other methods to establish models. Tinelli and Juran use the existing urban water supply pipe burst phenomenon to collect information through multiple linear regression to develop a model, first, to make a relevant description of the above phenomenon, but also in working condition of the water supply pipe network reference comparison to give an assessment, set the water supply pipe network breakage variable adjustment, to obtain experimental data [11]. Hannan and Annala predicted the degree of influence of each variable on the water supply pipe with the support of experimental data of the water supply pipe network [12]. A combination of matrix improved hierarchical analysis and gray correlation was utilized to evaluate the water supply pipe network system considering the material of water supply pipes and the qualities of water resources. Based on the gray system approach, Wang L analyzed and diagnosed the city’s example pipe network by understanding the results of multivariate factors such as pipe material, connection method, underground depth, pipe age, and pipe radius that affect the pipe network accident, using a multilevel fuzzy integrated evaluation method [13].

In a study by Pagano et al., graphite corrosion was identified as the primary mechanism of leakage in gray cast iron pipe sections during long-term service [14]. In the presence of an electrolyte, the steel in the graphite-steel mixture is anodic for the graphite. It electrochemically forms
an iron oxide matrix that weakens the pipe strength of the gray cast iron pipe and reduces its resistance to risk. Zhao et al. defined several pressure-related influences; they experimentally confirmed that the maximum pressure in the line and the magnitude of pressure variation between day and night could significantly impact the occurrence of leakage in the pipe [15]. Jeihouni et al. compared the rainfall pattern in the study area with the timing of leakage events in the water supply network. They found that rainfall coincided with the time of leakage events 70% of the time and inferred that rainfall affects soil [16]. The results of the study found that the timing of rainfall events overlapped with the timing of leakage events by 70%.

Water consumption forecasting is necessary for the subsequent optimal scheduling and scientific management of pipeline networks, so many scholars have researched water consumption forecasting methods. There are already many water consumption forecasting methods available, and all of them are highly accurate. In terms of water consumption forecasting models, the most basic models in water consumption forecasting have been established, such as the average shift forecasting model, exponential smoothing forecasting model, and seasonal exponential smoothing forecasting model. These models have laid the foundation for the development of water consumption forecasting. Meanwhile, in terms of daily and hourly water consumption forecasting, scholars like Siddique MAB and Islam ARMT have established various models such as time series forecasting for forecasting and comparison, and the forecasting effect is good [17]. Regarding long-term water use prediction, scholars such as Shan B have established nonlinear regression models that combine impact factors for prediction, and the accuracy is high [18]. Although the scholars started late, they have also made excellent contributions to the study of water use prediction algorithms. They have conducted in-depth research on optimization modeling algorithms for predicting water consumption, such as BP neural network method, support vector machine method, and time series prediction model. These algorithms have achieved good results when applied to water consumption prediction. Li et al. adopted the adaptive combined dynamic modeling technique to predict water consumption, which solved the time series prediction model [19]. Su et al. proposed to segment water consumption forecasting according to the water consumption law. They used BP neural network algorithms to forecast, respectively, which significantly improved the efficiency of model forecasting. Various single items established for water consumption forecasting often have different advantages, so some scholars try to show combined forecasting models to integrate the benefits of single models [20]. For example, Dar et al. combined BP neural networks with fuzzy time series and gray models with BP neural networks to improve the accuracy of the models. It has been proved that these combined prediction models do have high prediction accuracy and stability [21].

3. Random Matrix Theory Model Construction

A high-dimensional matrix is called a random matrix if all elements are random variables. RMT is a theory that takes large-dimensional random matrices as the object of study and analyzes the statistical distribution properties of their eigenvalues, singular values, and other related parameters. RMT has two basic concepts, the empirical spectral distribution function and the limiting spectral distribution function. For any $N \times N$-dimensional random matrix $A$ with real eigenvalues, the following process is said to be the observed spectrum distribution (ESD) of matrix $A$.

$$F(x) = \sum_{i=1}^{n} I\left(\frac{\lambda_i^A}{x} - 1\right),$$

where $N \times N$-dimensional real symmetric matrix $A_{n1}$, each element is a random variable independently identically distributed (IID) obeying the standard average distribution $N(0, 1)$. The diagonal elements are random variables with distribution $N(0, 2)$, then the empirical spectral distribution $FW$ of the Wigner matrix $W_{ni} = 1/\sqrt{N-1}A_{ni}$ obeys the semicircle law, then the probability density is

$$f_{sc}(x) = \left\{ \begin{array}{ll}
\frac{1}{2\pi} \sqrt{\frac{4}{x^2 - 1}} & \text{if } 1 - x,
0 & \text{otherwise}
\end{array} \right.$$

When the matrix $A_{ni}$ is not an asymmetric array, and each element obeys IID, all the eigenvalues of the matrix will converge in the unit circle of the complex plane. This distribution is called Girko’s integer circle law, as shown in Figure 1.

The random matrix $w = w_0 + jhH$, where the superscript $H$ denotes the conjugate transpose, $j$ represents a real diagonal array independent of the matrix $h$, $w_0$ is a deterministic Hermitian matrix, and an IID random matrix obeys a deterministic symmetric distribution of $N \times K$ dimensions. When $w_0 = 0$, $j = i$, $w = hhH$ hold, the singular value asymptotic spectral density function of $h$ is

$$q_{\infty}(x) = \sum_{i=1}^{n} \frac{1}{\pi} \sqrt{x^2 - 4}.$$  

Since the elements in the high-dimensional matrix $x$ of the input data are all real numbers, the eigenvalues can be mapped to the complex plane by processing the sample covariance moments of $x$ using your matrix $u$. The sample covariance matrix $x$ is singularized to obtain the equivalent matrix.

$$x_{uu} = \sum_{u=1}^{n} \sqrt{XX^H - U}.$$  

RMT also supports the simultaneous analysis of multiple matrices to analyze the system from multiple spatial and temporal perspectives by selecting multiple data source matrices $X(I = 1, \ldots, l)$ and taking the same steps to obtain the matrix accumulation $Z = \theta_i x_j$ for analysis, which has the elements of the equation:
Circular Law elements, and the MSR can also express the correlation significant number theorem and the central limit theorem, the values and reflects the trace of the matrix. From the significance (CPV) and the gravel plot methods. For the appropriate dimensionality reduction and direction of dimensionality reduction by extracting principal components. -he primary role of PCA is to achieve the effect of dimensionality reduction with the variance of this principal component. -he ratio of each eigenvalue to the sum of all eigenvalues is the contribution rate, and the cumulative contribution rate of the eigenvalues in descending order is called the cumulative variance contribution rate, which is calculated as follows:

\[ p = \sum_{i=1}^{l} \lambda_i + \sum_{i=1}^{m} (\lambda_i + m), \]  

where \( \lambda_i \) are the eigenvalues obtained from the C-feature decomposition and ranked from largest to smallest, \( l \) represents the number of principal components, and \( p \) is the set value, usually chosen as 85%. By drawing a scatter plot of \( \lambda_i \) about \( i \), we can distinguish a cut-off point between large and small eigenvalues, and this cut-off point is generally the number of principal components we need. We can determine a cut-off point between large and small eigenvalues, usually the number of prominent features we need. For example, suppose that there is a set of data to find out the eigenvalues, as shown in Figure 2. In that case, we can find that from the 6th eigenvalue, the size of the eigenvalue is very different from the previous one, so we can determine the cut-off point is the 6th eigenvalue, and we can select the principal components corresponding to the first 5 eigenvalues as the ones we need and discard the later ones, that is, the number of prominent features is set as 5. Compared with the CPV method, this method is more intuitive. Still, it is difficult to accurately determine the number of prominent features in some cases where the eigenvalues are uniformly decreasing.

PCA-based process monitoring determines whether a process fails by monitoring two multivariate statistics, \( T^2 \) and \( Q \), where the \( T2 \) statistic is defined as follows:

\[ T2 = x_i'(p^{1/2} - p^{1/2}x_i). \]  

The control limits \( D_C \) and \( Q_C \) for the \( T^2 \) and \( Q \) statistics are calculated in PCA.

\[ D_C = \frac{l(n^2 - 1)}{n(n - 1)}, \]  

where \( n \) is the number of modeled data samples, \( l \) is the number of principal components retained in the main element, \( a \) is the significance level, and the critical value of the \( F \) distribution under the conditions of degrees of freedom \( l, n - l \) can be found through the statistical table.

\[ Q_a = \left[ \frac{\theta_0 h_0 - \frac{1}{\theta^2}}{\theta^2} + \sqrt{C_a(2 \theta - h_0)} \right], \]  

where \( C_a \) represents the critical value of the normal distribution at significance level \( a \) and represents the smaller of several eigenroots in the data covariance matrix. Dynamic modeling of water supply pipe system is an important tool for water companies to achieve scientific management to improve efficiency and service level.
5. Multivariate Statistical Model Design for Water Leakage of Urban Water Supply Network

The most basic and essential work before establishing the pipe network model is to collect information and data on the pipe network, which mainly includes municipal design and planning drawings, daily operation data records of water plants, monitoring data of pipe network monitoring points, accident maintenance records of pipe network, and revenue records of the pipe network [23]. The main types of data collected are GIS system management data of pipe network attributes, data of regular operation of pipe network, data that need to be confirmed and measured on-site, and derived data of pipe network system, etc.

The first step of modeling is to collect and process the pipe network data, and the processed static and dynamic information data of the pipe network need to be converted to the format. The converted format data are imported into the modeling software to establish the preliminary topological relationship model of the pipe network. Most of the initial modeling data are pipe network attribute data, i.e., GIS system pipe network data, which generally has a more standard format, and GIS system and modeling software have typically a particular input and output docking module, which not only improves efficiency but also reduces the error of manual input. In contrast, manual information is required for pipe network data not available in the GIS system. After the initial data are imported, the hydraulic calculation cannot be carried out directly due to many problems in the established network topology. The topology of the pipe network includes isolated nodes, isolated pipe sections, duplicate pipe sections, etc. The pipe network can form an actual topology only after the preliminary arrangement. The simplification of the pipe network is based on the principle of simplifying the water supply pipe with slight head loss, including smaller diameter branch pipes, interval distance of fewer than 2m connecting line, and the same diameter of the pipe "T" connecting pipe. After finishing and simplifying the pipe network, a preliminary hydraulic parity calculation can be carried out.

After establishing the pipe network topology, it is necessary to establish the fundamental equations of pipe network calculation, i.e., the equation of state of the pipe network. The equation of state of the pipe network mainly includes continuity equation, energy equation, and head equation of the pipe network, and the establishment of the equation of state of the pipe network is the basis for the hydraulic calculation of the microscopic model of the pipe network. After the network topology and network state equations are established, the preliminary network hydraulic calculation can be carried out. After the analysis, we can get the information on the network conditions, including nodal water pressure, pipe section flow, pipe section head loss, pipe network water quality condition, water plant discharge pressure, and water plant discharge flow, etc., and we can use the modeling software to draw the corresponding parameter images for observation [24]. The preliminary hydraulic calculation of the pipe network working condition information can be compared with the existing pipe network working condition information. Calculate the error and the accuracy of the pipe network verification. Pipe network calibration generally requires two stages, namely static calibration and dynamic calibration. The static calibration is mainly the correctness check of the pipe network property data.

In contrast, the dynamic calibration mainly calculates the pipe coefficient and pump characteristic curve. Until the error is within the allowable range of the model calculation, the model can be used for practical applications, such as optimal scheduling and pipe burst analysis. After the model is built and calibrated to ensure accuracy, the model can be used for thematic applications, such as pipe network optimization and scheduling, pipe network transformation, and pipe network accident analysis. Since the GIS data and SCADA monitoring data of the pipe network are constantly changing, the model needs to be updated and maintained continuously during the application process to ensure the accuracy of the hydraulic model of the pipe network application. The specific method of pipe network modeling is shown in Figure 3.

The microscopic pipe network model is a mechanical model expressed by the abstraction and simplification of pipe network engineering entities to obtain the topological with relevant data and hydraulics equations. The topological graph is based on the basic principle of graph theory. The nodes are connected by pipes, and the water flows from the upstream node to the downstream node inside the tube. Thus, the geometric diagram of the pipe network can be abstracted and understood as a directed diagram. The microscopic hydraulic model of the water supply network is based on the steady-state hydraulic calculation theory of the pipe network, and the hydraulic model of the water supply network is established by detailed information about the pipe network. The microscopic modeling of the pipe network is shown in Figure 4.

The pipe network graphic and element simulation export the urban water supply model’s graphic data and element attribute data through the GIS system to form a graphic containing nodes and pipelines in the water supply network.
The pipe network parameter simulation can also be obtained partly through the GIS system, and part needs to be determined empirically or through actual measurements. The state simulation needs to be determined by analyzing the exact measurements of the SCADA system or the values reported by each water plant.

The pump curve represents the relationship between the pump head and the flow rate. The head is the head of the feed water transmitted through the pump and decreases as the flow rate becomes higher. The curve is usually presented in a three-point format, i.e., the pump curve is defined by three operating points: the low flow point (flow and head at common or zero flow conditions), the design flow point (discharge and charge at the desired operating point), and the maximum flow point. EPANET attempts to define the entire pump curve through three points, fitting a continuous function of the following form:

$$h_g = \sum (Bq^c - A). \tag{11}$$

The variation of water consumption (i.e., nodal flow) is the critical factor affecting the flow rate of the pipeline network pipes when the water supply network is operating under specific operating conditions. In this paper, the effect of the change of node flow rate on the flow rate of all pipes is expressed by the influence coefficient. If a specific urban water supply network has \(N\) nodes and \(M\) pipes, the value of nodal flow variation at one operating condition node \(i\) is \(q_i\). The flow variation of pipe \(l\) is \(Q_l\), then the magnitude of the influence of flow variation at node \(i\) on the flow variation of pipe \(l\) is expressed as \(Q_l/q_i\). Since the importance of the flow at each node varies widely, to avoid the lack of comparability of \(Q_l/q_i\) due to the significant difference in the magnitude of the flow at different nodes, a certain percentage of the base value of the flow at each node is taken as \(q_i\). Therefore, the coefficient of influence of the change in flow rate at node \(i\) on the change in flow rate at pipe \(l\) is

$$X_{li} = \sum \frac{Q_l}{q_i} - \left(\frac{Q_l + q_i}{q_i - Q_l}\right). \tag{12}$$

The influence coefficient matrix calculated by the above is \(X_{\text{mean}}\), in which the number of elements in the level of difference is relatively large and can not be used for direct calculation because it will amplify the impact of data of larger order on the clustering, so it is necessary to change the size of the original data. Still, it does not affect the order of the data. The following is the formula for the extreme value normalization method.

$$X_{li} = \frac{x_{li} - x_{\text{min}}}{\sqrt{x_{\text{max}} - x_{\text{min}}}}. \tag{13}$$

![Figure 3: Pipe network modeling specific process.](image)

Given the monitoring criterion \(\lambda\), when the influence coefficient matrix \((l, i)\) is greater than or equal to the criterion, it can be assumed that the flow variation of pipe \(l\) has a monitoring effect on node \(i\); conversely, it cannot be effectively monitored. The size of the detection standard \(\lambda\) value determines the number of flow measurement points; the larger the value of \(\lambda\) taken, the more flow monitoring points need to be arranged the higher the accuracy of monitoring, while it will cause increased costs and operating expenses; conversely, the smaller the number of arrangements, the lower the corresponding accuracy [25]. The value of \(\lambda\) in the actual engineering calculation can be considered according to the requirements of accuracy and cost. When
the elements in the matrix $X_{mn}'$ are greater than or equal to $\lambda$, it is recorded as 1, and all the remaining parts are recorded as 0 to obtain the effective monitoring matrix.

The simplification of the pipe network consists of three general steps: decomposition, merging, and omission of the pipe network. The connection can be disconnected directly for a pipe network connected by only one pipeline, thus breaking the original network into two relatively independent pipe networks. For the branch network connected by two lines, if the branch network is located at the end of the network and the flow rate and direction of the connected lines can be determined, then it can also be decomposed; the original network, after decomposition, will be transformed into several independent pipe network, they can be hydraulically calculated. For some, smaller pipe diameters, close to each other, parallel, and close to the pipeline’s location, can be combined. Pipeline omission operation is omitted when the hydraulic conditions change when the impact of smaller channels is generally relatively small-diameter pipelines in the network. In general, after the above simplification operation, the error of the calculation results is within a tolerable range [26]. The existing pipe network must be considered integrated before modeling according to its actual situation. Three heuristic rules as the basis, an ideal simplification scheme is proposed, and its straightforward process is as follows: the node with minor water consumption of the pipeline is merged with the end of the channel; part of the pipeline with a smaller diameter is ignored, and the ignored users use one node instead; the nodes with similar water pressure in the adjacent nodes are merged into one node. Brandon’s proposed distribution implementation scheme makes it possible to achieve the requirement that the combined pipe network calculation is close to the actual operating conditions and dramatically simplifies the modeling workload.

6. Analysis of Results

6.1. Multivariate Statistical Model Analysis of Water Leakage of Urban Water Supply Network. In the analysis and comparison of the flow data of the pipe network without leakage and pipe leakage at the edge of the pipe network, it is first necessary to determine whether the two groups of data can be put together for analysis and comparison. The solution is to conduct an F-hypothesis test for the two data groups and a chi-square test for the variance between the data groups to determine whether a subsequent in-depth study of the data can be performed. Next, the variance chi-square test of the two working conditions of the flow data at each node is performed, and the default value of significance in the test process is $\alpha = 0.05$. If the $P$ value of the $F$-test is greater than or equal to 0.05, the variance between the two groups of data can be considered as chi-square, and the subsequent data analysis can be performed; if the $P$ value is less than 0.05, the corresponding two groups of data are not chi-square. Due to many pipes in the water supply network, only some lines are analyzed in detail below, and the $P$ values obtained from the overall $F$-test are summarized to draw scatter plots. The distribution of $F$-test $P$ values for all pipeline flows without and with marginal leakage is shown in Figure 5.

The $P$ values of the $F$-tests for all the pipeline flows without and marginal leakage conditions are more significant than the default value of significance $\alpha = 0.05$, i.e., the criterion of variance chi-squared is satisfied, and the subsequent analysis between the data groups can be performed. The above chi-square test between the different working condition flow data groups shows that the analytical treatment between the data groups meets the criteria [27]. The prerequisites for the independent sample $t$-test are that the two data groups are independent of each other. The chi-square of the two data groups and the two
data groups come from a normal distribution overall. Therefore, before performing the $t$-test analysis and analyzing whether the variance of the two groups of data is chi-squared, it is also necessary to perform a (P-P) test for the normality of the data. The graph of the normality test results for the pipes is shown in Figure 6. There is a lot of background noise in the existing urban water supply network, caused by the measurement error of the sensor itself and the random fluctuation of the monitored object itself; there is also a specific error between the real-time leakage simulation model of the pipe network and the existing pipe network, and if all the monitoring points in the DMA area are used as the target nodes of the calibration will lead to the accumulation of multiple errors, thus bringing a lot of uncertainty to the optimal calibration accuracy positioning, as the calibration parameters. The selection of target nodes as calibration parameters should be representative (negative pressure wave is apparent).

The actual data and the predicted data from the water demand prediction method were grouped by sliding time windows. Every 15 min as a sampling point and every 4 groups of corresponding sampling points as a set of time series. For each group of sliding time series, the DTW distance was calculated, and the DTW distance at the same time under normal working conditions after removing abnormal data was used as the original data. The part of $-3$ is judged as leakage and alarm. The specific number of notices is shown in Figure 7.

To verify the method’s effectiveness in this paper, a DMA zonal leakage detection method based on water use pattern learning proposed by Nai-Fu Zhu is reproduced for comparison. The method is a classification-based leakage detection method. Still, it simulates leakage by adding volume to the no-leakage flow data set as a training set without requiring much historical monitoring data with labels. It has some practical application significance. This study also shows a higher detection rate and lower false alarm rate than several standard methods in the current research field, such as the Kalman filter estimation and statistical process analysis methods.

The analysis revealed that the leakage detection experiments were conducted by adding fixed-time leakage amounts to the no-leakage flow data set after removing abnormal patterns to simulate leakage events and using these leakage data for training and testing, with good results. However, the data for this trial used actual water use data as the test set, which included disturbances caused by influences such as temperature changes, weather changes, and changes in water use patterns. The method did not consider such regular water use changes and therefore had higher false alarms [28]. At the same time, the leakage flow rate at average leakage loss is mostly from small to large and eventually relatively stable. The method uses a simulated leakage event with a fixed leakage amount added as a label to train the classifier, making it difficult to detect leakage during increasing leakage flow from small to large, resulting in a relatively late detection time, even later than the time of reporting. Therefore, it is essential to analyze the changing trend of leakage loss flow during normal leakage to train a more accurate leakage loss recognition model.

6.2 Water Leakage of Urban Water Supply Network Multivariate Statistical Accurate Model Present. For the simulation of the pipe network without leakage and pipe leakage conditions at the edge of the pipe network, the $t$-test $P$ value of pipe 24 is equal to 0.05. The $P$ values of pipes 18, 26, and 27 are 0.04, 0.03, and 0.02 are less than 0.05, i.e., these three pipes are possible leaky pipes, and the size of the $t$-test $P$ value can rank the possibility of leakage. The possibility of leakage of pipe 27 is the largest because pipe 18 and pipe 27 are the pipes connected to pipe 26, so it will produce the $P$ value of the $t$-test is less than the significant default factor. For the simulation of no leakage in the pipe network and the middle pipe leakage condition in the pipe network, the $t$-test $P$ value of pipe 8 is equal to 0.05, and the $P$ values of pipe 9 and pipe 11 are 0.04 and 0.02, respectively; both of which are less than 0.05, i.e., these two pipes are possible leaky pipes, and the possibility of leakage can rank the size of the $t$-test $P$ value. The possibility of leakage of pipe 11 is the largest because pipe 9 is the pipe connected to the pipe. Since pipe 9 is the pipe connected to pipe 11, the $P$ value of the $t$-test for pipe 9 is less than the default significance factor. The leakage pressure measurement point graph is shown in Figure 8.

According to the leakage condition model of the edge section of the water supply network, the design leakage point is 150 m from node 16 on pipe sections 16–20. The dichotomous simulation is used to find out the best fit of the leakage location under different leakage coefficient conditions, the location of the leakage point in any of the states from 0 to 2, and the size of the fit analysis shows that when the damage level at the leakage of the pipe network is 0.5, the actual value of the pressure at the node of the leakage condition of the water supply pipe network fits the simulated value optimally, and the distance between the leakage point...
and node 16 based on the dichotomous search simulation under the damage level condition is 150.94 m. The distance between the searched leakage points and node 16 is 150.94 m, and the size of the error is controlled at 1-2 m, so the simulation based on the dichotomous method is of great reference significance for the accurate point location of the known leaky pipe section of the network [29]. The operating conditions analyzed above are the working conditions of the leaking pipe section at the edge of the pipeline network. The same method can determine the amount of leakage and locate the leakage point for the middle section of the pipeline network. The results of leakage area identification are shown in Figure 9.

Under different working conditions, the proposed leakage area identification method can more accurately identify the suspected leakage areas in the pipe network. Comparing the leakage identification results under other working conditions, it can be found that the leakage identification effect is slightly better when the network is in low load hydraulic condition. There are two reasons for this phenomenon: under the low load hydraulic condition, most of the nodes (including the simulated residential and commercial water) are in a joint or even zero water consumption condition, which reduces the uncertainty of the node flow, which is equivalent to lowering
the observation noise of the hydraulic model of the pipe network; under the low load hydraulic condition, the pressure of the nodes in the pipe network will generally increase, resulting in a higher leakage flow than under the increased load hydraulic condition. Under the low load hydraulic state, the pressure of the nodes in the network will generally increase, resulting in the leakage flow being higher than the leakage flow under the increased load hydraulic state, which is conducive to the identification of leakage areas. Therefore, when applying the proposed method to identify leakage areas in the natural water supply network, the nighttime monitoring data of the overall network at low load should be given priority to identify the suspected leakage areas. Alternatively, the nighttime common load pressure monitoring data can provide information about those leakage areas with smaller leakage flow that other time monitoring data cannot give.

7. Conclusion

The optimized arrangement of pressure and flow monitoring points of urban water supply network is carried out to select the optimized performance of monitoring points suitable for water supply network leakage positioning. By analyzing and comparing various hydraulic monitoring point optimization arrangement methods, fuzzy cluster analysis is selected for the optimization arrangement of pressure monitoring points. The effective monitoring matrix method is used to optimize flow monitoring points. This paper summarizes constructing a random matrix using water leakage data from the urban water supply network. The expanded random matrix with expanded data sources is built for the data that do not meet the arbitrary matrix theory analysis requirements. The augmented random matrix with changed weights highlights the influence of the critical focus on the link on the water leakage system of the urban water supply pipe network. The expanded random matrix for correlation study is constructed for the analysis of different data correlation of urban water supply pipe network. Bently modeling software Water GEMS establishes the topology of the pipe network, and the microscopic hydraulic model of the pipe network is set for the hydraulic calculation of the pipe network. By comparing the simulated values of flow, pressure, and other hydraulic indexes with the monitored values after the model calculation, the working conditions of the pipe network calculated by the pipe network model established by Water GEMS are consistent with the actual situation. Through the simulation of no leakage of the pipe network and the middle pipe leakage of the pipe network, it can be analyzed and compared from the obtained flow data: when the $t$-test $P$ value of the pipe is less than the significant default factor of 0.05, it is considered that the pipe is a possible leakage pipe, and the size of the $t$-test $P$ value can rank the possibility of leakage, the more minor the $P$ value, the greater the chance of leakage. Reducing the leakage of the water supply network can save water and effectively alleviate this problem and improve the economic efficiency of water supply enterprises.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the "Intelligent Management of Water Supply System under Real-Time Monitoring of Intelligent Water Meter Based on Data Visualization and Time Series Model," a Scientific Research Project of Education Department of Shaanxi Provincial Government (21JK0923).

References


