RESEARCH ARTICLE

Distribution Field Construction and Prediction Method for Gas Leakage based on Kriging model and Gaussian Process

Chenglong Hou[1], **Yuhao Zha**[1], **Jun Yang**[1,*] **and Ning Wang**[2]

[1] School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[2] School of Cyber Science and Technology, Beihang University, Beijing 100191, China

Abstract

Gas leakage poses a significant hazard in chemical industry operations, where failure to respond rapidly to gas diffusion can lead to poisoning, fire, or explosion. Timely and accurate prediction of gas dispersion is therefore essential for emergency decision-making and operational safety. While existing methods such as computational fluid dynamics, spatiotemporal statistics, and surrogate models emphasize prediction accuracy, they often suffer from excessive computational delays—especially critical in leak scenarios where casualties can occur within minutes. To address this gap, this paper introduces a Gaussian process-Markov random field-Kriging (GP-MRF-K) model for fast and reliable prediction of gas concentration fields. The approach integrates Markov random field (MRF) neighborhood structures into Kriging-based spatial interpolation, reducing computational complexity from $O(n^3)$ to $O(n \cdot m^3)$, where n is the total grid points and m is the average neighbor count. Gas concentration time series are forecasted using Gaussian process regression (GPR), and the MRF-Kriging framework rapidly reconstructs the full concentration field. Validation with real ammonia concentration data from a warehouse-scale experimental setup confirms the feasibility and superiority of GP-MRF-K. With 150 training points and 10 prediction steps, the model achieves an MSE of 4660 and RMSE of 68.26, improving MSE by 67% over GPR-K (MSE=14003) and 87% over LSTM-K (MSE=36172), while attaining an R^2 of 0.9847. Computation time is reduced to 39.04 seconds, a 21.5% gain over GPR-K (49.72s) and a 98% reduction compared to LSTM-K (1990.85s), thereby meeting real-time emergency response requirements.

Keywords: gas leakage, field prediction, kriging model, markov random field, neighborhood structure, gaussian process.

1 Introduction

Chemical plants often need lots of industrial gases for production. These gases almost have toxic physical properties and active chemical properties [1, 2]. Once the external environment changes dramatically, it is easy to cause these gases to react [3]. Industrial gas leakage can lead to serious safety accidents such as personnel poisoning, fire, and explosion [4, 5]. On

Submitted: 07 November 2025
Accepted: 08 December 2025
Published: 03 February 2026

Vol. 2, **No.** 1, 2026.
10.62762/TSSR.2025.861997

***Corresponding author:**
✉ Jun Yang
tomyj2001@buaa.edu.cn

Citation
Hou, C., Zha, Y., Yang, J., & Wang, N. (2026). Distribution Field Construction and Prediction Method for Gas Leakage based on Kriging model and Gaussian Process. *ICCK Transactions on Systems Safety and Reliability*, 2(1), 11–25.

© 2026 ICCK (Institute of Central Computation and Knowledge)

August 31, 2013, the Shanghai Wengpai industry had an ammonia leakage accident, resulting in 15 deaths, 7 serious injuries, and 18 minor injuries, with a direct economic loss of 25 million yuan. It can be seen that gas leakage is one of the potential major hazard sources in the chemical industry, which seriously affects the safe production and public safety of the chemical plant [6]. In order to minimize the loss of life and property, managers must make emergency decisions quickly and accurately. Therefore, it is important to quickly describe and predict the changes of gas concentration after leakage.

At present, in order to reflect the gas distribution in the warehouse in real time, the warehouse storing gas in chemical plants is equipped with special gas concentration detectors to detect the gas concentration in the environment [7–9]. However, each detector has a limited detection range, which cannot effectively describe the gas distribution of the whole warehouse [10]. At the same time, the single detection method cannot effectively predict the future gas distribution in the warehouse. Therefore, this study focuses on predicting and constructing the future gas distribution field quickly from the detector data.

Many scholars have done a lot of research on how to construct and predict the distribution field of the physical quantities, such as computational fluid dynamics (CFD) [11, 13], Spatio-temporal statistical methods [14–16], surrogate models [17–19], and other models. Since the 1970s, with the popularization of computers and the continuous improvement of computing power, CFD-based on numerical calculation has been formed and developed vigorously [11]. Tominaga et al. [12] reviewed current modeling techniques in the CFD simulation of near-field pollutant dispersion in urban environments. For different flow Reynolds numbers from 6000 to 24000, Sharma et al. [13] presented a two-dimensional computational fluid dynamics analysis to evaluate the effect of discrete triangle wave corrugations on the absorber plate by constructing a temperature field. The Spatio-temporal statistical model is also the classical method for field construction and prediction. It is based on the regionalized variable theory with the characteristics of time and space distribution, and takes the variogram as the main tool to study the natural phenomena in time and space distribution [15]. Martinez et al. [16] smoothed the spatiotemporal trend of the spatiotemporal median, conducted the spatio-temporal Kriging interpolation on the residuals, and pointed out that the accuracy of precipitation estimation is higher than that of the ordinary spatiotemporal Kriging. The Spatio-temporal statistical model is widely applied to a wide range of regions, and it is also applicable to predict the gas concentration field in a small area such as a chemical plant warehouse. People have paid more attention to the surrogate models in field construction and prediction in recent years [17, 18].The rudiment of the surrogate models is the polynomial response surface, which is applied to the structural optimization design in the 1970s [20]. At the end of the 20th century, the surrogate models was introduced into the field of aerodynamic optimization design [21–23]. In terms of the surrogate models research, many surrogate models have been developed, including polynomial response surface [20], Kriging model [24, 25], radial basis functions (RBFs) [26], neural networks (NN) [27, 28], multivariable interpolation regression (MIR) [29]. Zhou et al. [30] obtained the temperature data of the boiler furnace, and combined the reflected S-shaped radial basis function with QR decomposition to build a temperature field, which had higher accuracy in both global and edge local areas. The calculation results of the surrogate models are very close to the original model, but the computing efficiency has been significantly improved.

However, existing methods mainly focus on the prediction accuracy, ignoring the calculation speed. In reality, the gas leakage and diffusion in chemical plants have very high requirements for the prediction speed. According to the liquefied ammonia leakage accident report of the Shanghai Wengpai industry, casualties were caused within 1 minute after the leakage, which showed that the liquefied ammonia leakage can fill the warehouse environment in a very short time. Therefore, in order to satisfy the actual need, the field construction and prediction methods should pay attention to both the prediction accuracy and the calculation speed. However, most methods based on the CFD and Spatio-temporal statistical models often take a long time to calculate, which does not meet the needs of emergency decision-making after gas leakage in chemical plants.

In order to grasp the gas distribution field in time after leakage for chemical plants to ensure production safety, the GP-MRF-K method is proposed to construct and predict the gas distribution field timely and sufficiently. The GP-MRF-K method is combined by Gaussian process regression and the Kriging model which combines the neighborhood structure of the

Markov random field.

1. To construct the gas concentration field timely, the neighborhood structure of the Markov random field is introduced to the Kriging spatial interpolation method, and the Markov Random Field-Kriging (MRF-K) method is proposed.

2. To predict the gas distribution field effectively, the Gaussian process regression is used to predict the future gas concentration at the detectors, and the MRF-K method is used to construct the global gas concentration field quickly and accurately.

The structure of this paper is organized as follows. In Section 2, the Kriging model combined with Markov random field neighborhood structure is introduced as the basis of field construction. In Section 3, the principle and establishment process of the GP-MRF-K method with fast prediction speed and high accuracy is introduced in detail. In Section 4, the data collected from the ammonia leakage scale model are carried out to verify the effectiveness of the proposed method. Section 5 summarizes the main conclusions and discusses the possible improvement of the proposed method.

2 Field Construction Method

2.1 Kriging Model

The kriging model in this paper is based on the MATLAB toolbox: Design and Analysis of Computer Experiments (DACE). This toolbox is a Kriging interpolation toolkit which generally includes model establishment, experimental design, sample information collection, and experimental parameter analysis [31]. The method proposed in this paper is based on the Kriging model and its derivation process is as follows.

Suppose the independent variable of m observation points $X_m = [x_1, x_2, \ldots, x_m]^T$, $x_i \in \mathbb{R}^n$ and the corresponding response values $Y_m = [y_1, y_2, \ldots, y_m]^T$, $y_i \in \mathbb{R}^n$ are known. The Kriging models mostly rely on this decomposition.

$$Y(x) = f(x) + \varepsilon(x), \tag{1}$$

where $\varepsilon(x)$ is the error which ensures that the deviation value, and $\varepsilon(x)$ is independent and obeys the normal distribution, that is $\varepsilon(x) \sim N(0, \delta^2)$.

In multidimensional space, the Kriging model can be described as $Y(x) = f(x) + Z(x)$, which includes polynomial function $f(x)$ and deviation function $Z(x)$. It can be expressed as:

$$Y(x) = \sum_{j=1}^{p} \beta_j f_j(x) + Z(x), \tag{2}$$

where $f_j(x)$ is the basic function of the known regression function, β_j is the coefficient corresponding to the basic function, $Z(x)$ represents the uncertainty of $Y(x)$.

For $Z(x)$, we have:

$$E(Z(x)) = 0, \tag{3}$$

$$Cov(Z(\omega), Z(x)) = \sigma^2 R(\omega, x), \tag{4}$$

where σ^2 is the process variance, $R(\omega, x)$ represents the spatial correlation function.

Extend Equation (2) to the situation of m points.

$$Y = F\beta + Z, \tag{5}$$

where

$$\begin{aligned} Y &= [Y(x_1), Y(x_2), \ldots, Y(x_m)]^T, \\ F &= [f(x_1), f(x_2), \ldots, f(x_m)]^T, \\ f(x) &= [f_1(x), f_2(x), \ldots, f_p(x)]^T, \\ \beta &= [\beta_1, \beta_2, \ldots, \beta_p]^T, \text{ and} \\ Z &= [Z(x_1), Z(x_2), \ldots, Z(x_m)]^T. \end{aligned}$$

To ensure the optimal linearity and unbiasedness, the measurement error needs to be quantified to ensure the optimality of the variance, and constraint conditions are needed to ensure unbiasedness, which is finally transformed into a minimization problem of Lagrangian parameter constraint. Consider the linear predictor $\hat{Y}(x) = c^T Y$. The error is:

$$\begin{aligned} Y^*(x) - Y(x) &= c^\top Y - Y(x) \\ &= c^\top Z - Z(x) + (F^\top c - f(x))^\top \beta. \end{aligned} \tag{6}$$

To ensure unbiasedness of the estimation, the constraint $F^\top c = f(x)$ is needed. The MSE of estimation is:

$$\varphi(x) = E\left[(\hat{Y}(x) - Y(x))^2\right] = \sigma^2\left(1 + c^\top R c - 2c^\top r\right). \tag{7}$$

The Lagrangian multiplier λ is introduced to obtain the objective function with the principle of minimizing the variance of the error of estimation.

$$\min \sigma^2 \left(1 + c^\top Rc - 2c^\top r\right) - \lambda \left[F^\top c - f(x)\right]. \quad (8)$$

The gradient of Equation (8) with respect to c and λ are respectively.

$$2\sigma^2(Rc - r) - F\lambda = 0, \quad (9)$$

and

$$F^T c - f(x) = 0. \quad (10)$$

Represent Equation (9) and Equation (10) in matrix form.

$$\begin{bmatrix} R & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} c \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} r \\ f \end{bmatrix}, \quad (11)$$

where

$$\tilde{\lambda} = -\frac{\lambda}{2\sigma^2}.$$

The solution of Equation (11) as follows.

$$\begin{cases} \tilde{\lambda} = \left(F^T R^{-1} F\right)^{-1} \left(F^T R^{-1} r - f\right) \\ c = R^{-1}\left(r - F\tilde{\lambda}\right) \end{cases} \quad (9)$$

Substituting Equation (12) into $\hat{Y}(x) = c^T Y$, we obtain.

$$\begin{aligned} \hat{Y}(x) &= \left(r - F\tilde{\lambda}\right)^T R^{-1} Y \\ &= r^T R^{-1} Y - \left(F^T R^{-1} r - f\right)^T \left(F^T R^{-1} F\right)^{-1} F^T R^{-1} Y. \end{aligned} \quad (10)$$

According to Equation (5), the parameter β can be estimated through the generalized least squares method.

$$\hat{\beta} = \left(F^T R^{-1} F\right)^{-1} F^T R^{-1} Y. \quad (11)$$

Substituting Equation (12) into $\varphi(x) = \sigma^2\left(1 + c^T Rc - 2c^T r\right)$, and calculate the maximum likelihood estimate of σ^2 with respect to variance.

$$\hat{\sigma}^2 = \frac{1}{m}\left(Y - F\hat{\beta}\right)^T R^{-1}\left(Y - F\hat{\beta}\right). \quad (12)$$

2.2 Neighborhood Structure of Markov Random Field

A Markov random field can be described by an undirected graph $S = (V, E)$ where $V = \{1, 2, \ldots, n\}$ are the points and E represent the edges. V and E are related via a neighboring system.

$$N = \{N_i | \forall i \in V\}. \quad (16)$$

Here N_i is a set of points neighboring i. A neighborhood structure has the following properties [32].

$$i \notin N_i, \quad i \in N_j \iff j \in N_i; \quad i \neq j. \quad (17)$$

For a regular point, such a neighborhood structure can be described as [33].

$$N_i = \{j \in V | d(x_i, x_j) \leq r; j \neq i\}, \quad (18)$$

where x_i and x_j represent positions of point i and j and $d(\cdot, \cdot)$ is a distance between two points. Since the Euclidean distance represents the true distance between two points in the m-dimensional space and fits the reality in the warehouse of a chemical plant, this paper adopts Euclidean distance. The neighborhood structure used in this paper is shown in Figure 1, a system with regular grid points and $r = 2$ (left) and $r = 3$ (right) [34]. The red points are the detectors uniformly arranged on the **stope**. The blue lines are the edges, leading from the center point to its neighbors. The distance between the two closest points vertically and horizontally is 2.

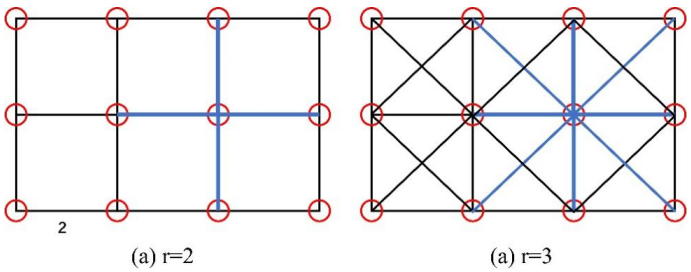


Figure 1. Schematic diagram of the neighborhood structure.

2.3 MRF-K field construction method

In general, using the DACE toolbox only needs to import the coordinates of the data and the corresponding gas concentration into MATLAB to realize the construction of the gas distribution field. However, when predicting the gas concentration at a certain coordinate, the DACE toolbox calls all the known data. If the known point is too far away from

the unknown point, it will make little contribution to the estimation of the gas concentration at the unknown point. In addition, the DACE toolbox needs to calculate the correlation function between points. Removing points which are far away can improve the calculation efficiency and reduce the calculation complexity.

The DACE toolbox is built on the neighborhood structure of Markov Random Field to complete MRF-K field construction method. The selection of neighborhood radius r is crucial for balancing computational efficiency and prediction accuracy. Set the parameter r and establish the neighborhood structure as shown in Equation (18). When predicting a certain point, we use the neighbors of the point within the distance r to establish a Kriging estimation model, as shown in Figure 2.

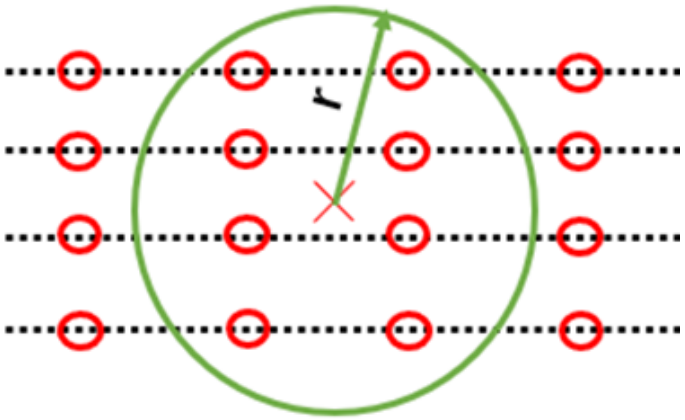


Figure 2. The neighborhood structure of gas field.

Because the average MSE obtained by cross-validation is the smallest, we choose the constant basis function as the regression function and the EXP model as the correlation function. The constant basis function as follows.

$$f_1(x) = 1, \ (p = 1). \tag{19}$$

The general form of the correlation function as follow.

$$R(\theta, w, x) = \prod_{j=1}^{n} R_j(\theta, w_j, x_j). \tag{20}$$

The form of the EXP model is:

$$R_j(\theta, d_j) = \exp(-\theta_j \cdot |d_j|), \ d_j = w_j - x_j. \tag{21}$$

Among various correlation functions available in the DACE toolbox, we selected the EXP (Exponential) model based on the following considerations:

1. **Physical appropriateness:** The exponential decay characteristic of the EXP model aligns with the diffusion behavior of gas concentration in space, where the influence of a point decreases exponentially with distance.
2. **Comparative analysis:** We compared four commonly used correlation functions (Gaussian, EXP, Linear, and Spherical) using 10-fold cross-validation. The EXP model achieves the best prediction accuracy while maintaining reasonable computational efficiency.
3. **Smoothness:** The EXP model provides C^1 continuity, suitable for describing the smooth spatial variation of gas concentration fields.

3 Field Prediction Method

In this paper, three representative methods of the traditional time series prediction model, machine learning, and deep learning will be selected to predict the gas concentration data at the detectors. The Autoregressive Integrated Moving Average (ARIMA) model, Gaussian Process Regression (GPR) model, and Long Short-Term Memory (LSTM) neural network model are combined with the MRF-K field construction method to find the best field prediction method.

3.1 ARIMA Model

The basic principle of the ARIMA model is to treat the data sequence formed by the prediction object over time as a random sequence, and use a certain mathematical model to approximately describe the sequence. Once the model is identified, it can predict the future value from the past and the present value of the time series. It is aimed at the stationary time series model. However, in real life, most time series are non-stationary. Therefore, the data needs to be differentiated to convert it into a stationary time series, so the ARIMA model can be used to predict the data. ARIMA model can make effective and accurate short-term prediction according to the correlation of data in different periods in the past. It makes up for the problem of too many parameters in AR and MA prediction, and has a wide range of applications in the field of short-term prediction. Its form is as follows:

$$\begin{aligned} Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} \\ + \xi_t + \theta_1 \xi_{t-1} + \theta_2 \xi_{t-2} + \cdots + \theta_q \xi_{t-q}, \end{aligned} \tag{13}$$

where c represents a constant, p represents the autoregressive order, q represents moving average order, and d represents the different order.

The modeling process of an ARIMA model is as follows.

Step 1: Test the stability of the data by ADF test. Nonstationary time series can be transformed into stationary time series by difference. If a time series has stationarity after a different operation, it is a different stationary time series, which can be analyzed by the ARIMA model.

Step 2: Determine the order of the ARIMA model. Generally, the values of p and q can be judged by the Autocorrelation Function (ACF) graph and Partial Autocorrelation Function (PACF) graph. However, this method requires manual judgment, but the manual judgment is uncertain in terms of time and accuracy, so it is not suitable for field prediction. In this paper, Akaike Information Criterion (AIC) is used to make the model order from low to high within the specified range. By calculating the AIC value respectively, we determine the order with the smallest AIC value as the appropriate order of the model.

$$AIC = 2k - 2\ln(L), \tag{14}$$

Step 3: Estimate the parameters of the model. Usually, correlation moment estimation, least square estimation and maximum likelihood estimation are used to estimate the parameters of an ARIMA model. Then verify the fitting effect of the model. If the model completely or basically explains the correlation of the data, the noise sequence of the model is white noise sequence. If the obtained model fails the test, the model should be refitted until it can pass the white noise test.

3.2 GPR Model

Due to the convenience of the Gaussian process and its kernel function, GPR has been widely used in time series analysis. Given the training set $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(n)}, y^{(n)})\}$, we assume that it follows the following GPR model.

$$y^{(i)} = f(x^{(i)}) + \sigma_y \quad \text{with} \quad \epsilon \sim N(0, 1), \tag{15}$$

where f represents a Gaussian Process (GP) and $f \sim gp(u, K)$, σ_y represents the noise. Let

$$X = [x^{(1)}, x^{(2)}, \ldots, x^{(n)}]^T, \tag{16}$$

$$Y = [y^{(1)}, y^{(2)}, \ldots, y^{(n)}]^T, \tag{17}$$

we have

$$Y = f(X) + \epsilon \sim N(u_x, K_{xy}), \tag{18}$$

$$u_x = [u(x^{(1)}), u(x^{(2)}), \ldots, u(x^{(n)})]^T, \tag{19}$$

$$K_{xx} = \begin{bmatrix} K(x^{(1)}, x^{(1)}) & K(x^{(1)}, x^{(2)}) & \cdots & K(x^{(1)}, x^{(n)}) \\ K(x^{(2)}, x^{(1)}) & K(x^{(2)}, x^{(2)}) & \cdots & K(x^{(2)}, x^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ K(x^{(n)}, x^{(1)}) & K(x^{(n)}, x^{(2)}) & \cdots & K(x^{(n)}, x^{(n)}) \end{bmatrix}, \tag{20}$$

$$K_{xy} = K_{xx} + \sigma_y^2 I_n, \tag{21}$$

where I_n represents the n-dimensional identity matrix. Then we can get the posterior prediction distribution on the test set as follows,

$$Y_*|X, X_*, Y \sim N(u_*, K_*), \tag{22}$$

$$u_* = u_{X_*} + K_{XX_*}^T (K_{XX} + \sigma_y^2 I_n)^{-1}(Y - u_X), \tag{23}$$

$$K_* = K_{X_*X_*} - K_{XX_*}^T (K_{XX} + \sigma_y^2 I_n)^{-1} K_{XX_*}. \tag{24}$$

When GPR is used for time series prediction, the values of the parameters are very important, which mainly consist of three components: the mean function $u(x, \theta^u)$, the covariance function $K(x, x', \theta^K)$, and the error term in the model σ_y. We use the maximum likelihood method to obtain the estimation of $\Theta = (\theta^u, \theta^K, \sigma_y)$.

The hyperparameters in the GPR model, including the length scale of the kernel function and the noise variance σ^2, are optimized using the maximum likelihood estimation (MLE) method. Specifically:

1. **Hyperparameter optimization:** The hyperparameters $\theta = \{l, \sigma^2, \sigma_n^2\}$ are learned by maximizing the log marginal likelihood.
2. **Optimization algorithm:** We employ the L-BFGS-B algorithm with multiple random restarts ($n = 10$) to avoid local optima.
3. **Validation:** The optimized hyperparameters are validated using a separate validation set to prevent overfitting.

3.3 LSTM Neural Network Model

LSTM neural network is essentially a special Recurrent neural network (RNN), which has more advantages in solving the problem of long sequence data [35]. LSTM and RNN are similar in neural network structure and parameter training. The main difference is in the nodes of circulating neurons. As shown in Figure 3, it is a typical LSTM neural unit $\tilde{C}$.

In the LSTM neural unit, C_t represents the internal state, h_t represents the external state, f_t, i_t and o_t represent the forget gate, input gate, and output gate value respectively. The mathematical expressions are

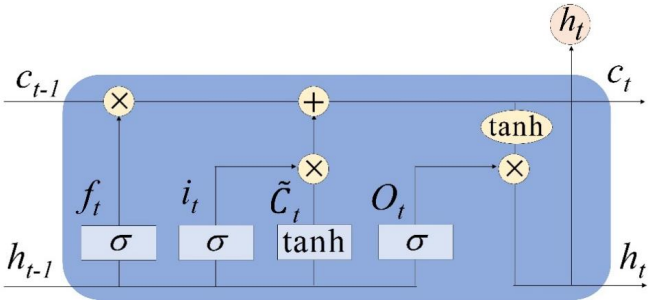


Figure 3. LSTM neural unit.

as follows.

$$f_t = \text{sigmoid}\left(W_f\left[h_{t-1}, x_t\right] + b_f\right), \tag{25}$$

$$i_t = \text{sigmoid}\left(W_i\left[h_{t-1}, x_t\right] + b_i\right), \tag{26}$$

$$o_t = \text{sigmoid}\left(W_o\left[h_{t-1}, x_t\right] + b_o\right), \tag{27}$$

where W_f, W_i and W_o are weights matrices, b_f, b_i and b_o are bias vectors.

The forget gate controls the contribution of the internal state at the $t-1$ moment, the input gate controls the contribution of the candidate state at the t moment, and the output gate controls the contribution of the internal state at the t moment to the external state. The calculation process of the LSTM is as follows.

Step 1: the values of the forget gate, input gate, output gate, and candidate state are calculated based on the external state h_{t-1} and the input at the t moment.

Step 2: the internal state C_{t-1} and the last step is used to calculate the value of forget gate, input gate, and candidate state to update the internal state C_t.

Step 3: the information is transported to the external state h_t through the current internal state and output gate.

3.4 Construction of Field Prediction Method

The establishment process of the field prediction method is as follows.

Firstly, this paper selects the representative methods of traditional time series prediction, machine learning, and deep learning to predict the gas concentration at the detection point, which are the ARIMA model, GPR model, and LSTM model respectively.

Then, we use the predicted gas concentration data at the detectors to interpolate the spatial values through the MRF-K field construction method. Finally, we depict the change trend of the future field through the predicted and interpolated data, so as to provide emergency decision support for safety managers.

The above is the establishment process of the field prediction method, which is composed of the time series prediction model and field construction method. The effect of the combination of the three time-series prediction models and the MRF-K field construction method will be described in detail in the next chapter through the case analysis of a scaling model.

4 Case Analysis

A liquid ammonia gas leakage scale model is established to obtain the changing data of the gas concentration field. As shown in Figure 4, a total of 12 ammonia concentration detectors (the black dots) are placed in the scale model. Ammonia gas is transmitted to the model through the pipeline, and a regulating valve is provided to control the speed of gas input to the model.

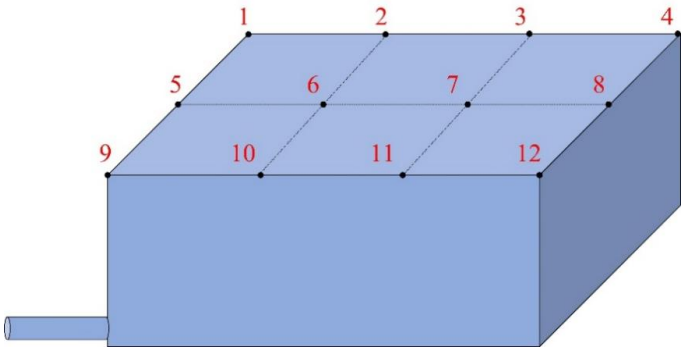


Figure 4. A liquid ammonia gas leakage scale model.

The detectors set in the scaled model detect and record the data of the concentration of the gas at the frequency of 100Hz. The data collected through this model is shown in Table 1.

Table 1. The data collected by the detectors.

Location	t=0.01s	t=0.02s	t=0.03s	t=0.04s	…	t=10.00s
1	1873.38	1869.02	1864.90	1873.15	…	1902.35
2	2638.59	2638.07	2619.90	2640.67	…	2651.72
3	1201.66	1199.01	1206.36	1205.34	…	1208.30
4	1421.35	1418.59	1431.76	1431.97	…	1570.90
5	640.62	631.05	627.00	618.92	…	650.12
6	763.27	777.06	772.90	764.03	…	778.10
7	1262.24	1263.47	1278.57	1267.19	…	1234.21
8	1343.96	1350.87	1352.13	1366.20	…	1258.86
9	1292.46	1298.69	1278.36	1291.63	…	1292.04
10	1251.30	1251.51	1257.08	1254.19	…	1258.02
11	1261.41	1258.93	1283.54	1257.90	…	1269.88
12	1219.05	1217.21	1218.84	1213.52	…	1220.89

The data at 12 locations at the same time is called as a set of data. Next, the collected data will be used to verify the rationality and superiority of the field prediction model.

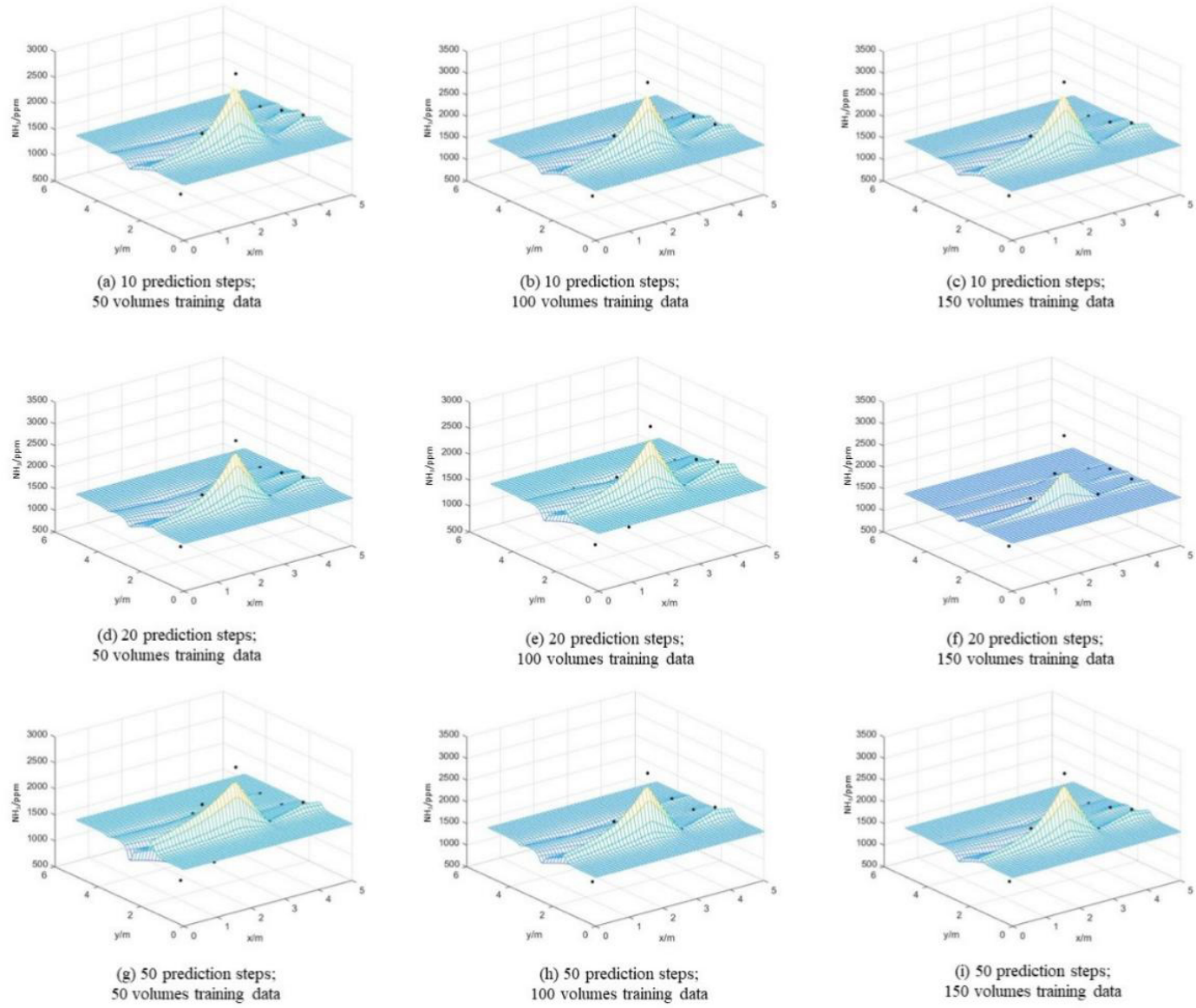


Figure 5. The field prediction of the LSTM-MRF-K method.

4.1 Construction of ARIMA-MRF-K Model

When the ARIMA model is used for prediction, ADF should be carried out in advance to check whether the time series is stationary, and the order of the model should be judged by AIC criteria. According to the actual calculation results of the program, under the training of the 100 sets of data, the ARIMA model needs a lot of time to predict the data of 12 detectors in the next 1 second (100 data), which has no application value in actual production. Therefore, the ARIMA model is abandoned as a component of the field prediction method.

4.2 Construction of LSTM-MRF-K Model

In order to combine the LSTM model and the MRF-K field construction method to form a field prediction method with practical application value, we plan to adopt the dynamic LSTM network training strategy and three training prediction schemes to verify the feasibility of the method.

In this paper, the training data of 500, 1000, and 1500 volumes are selected to train the LSTM neural network. Under the condition of each data volume, the ammonia gas concentration at the detectors in the future 10, 20, and 50-time units is predicted respectively. Each prediction simulates that the detectors are constantly receiving new data and the above prediction steps will repeat 100 times. To reduce the running time of the program, the trained network structure remains unchanged before each update of training data, so as to meet the actual needs of production.

The field prediction of the LSTM-MRF-K method is shown in Figure 5. It can be seen that the gas

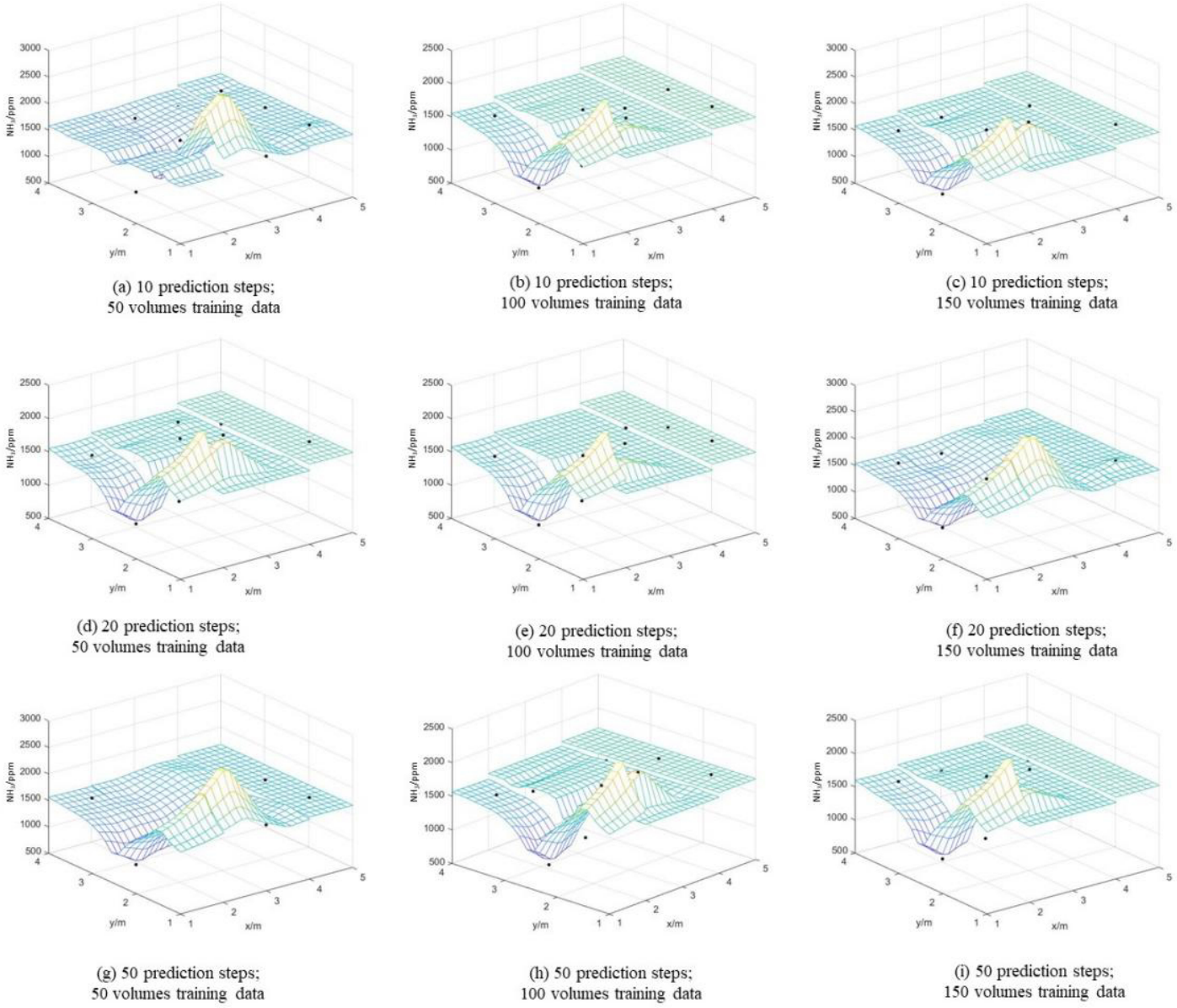


Figure 6. The field prediction of the GPR-MRF-K method.

concentration near the pipeline increases rapidly, but with the further diffusion of the gas, the gas concentration in other areas gradually decreases and tends to be stable. With the increase in step length, the gas concentration also shows an upward trend. Besides, with the continuous improvement of training data, the prediction effect is getting closer and closer to the actual value.

4.3 Construction of GPR-MRF-K Model

The GPR model is also trained with the same prediction scheme as the LSTM model, where 500, 1000, and 1500 volumes of training data are selected respectively to predict the ammonia gas concentration at the detectors of 10, 20, and 50-time units in the future. Each prediction scheme is executed 100 times to simulate the real-time update data of the detectors and predict the distribution of the future gas concentration field.

The field prediction of the GPR-MRF-K method is shown in Figure 6. The gas concentration near the pipeline increases rapidly, but with the further diffusion of the gas, the gas concentration in other areas gradually decreases and tends to be stable. Besides, with the continuous improvement of training data, the prediction effect is getting closer and closer to the actual value.

4.4 Comparison of Related Methods

Firstly, we also used LSTM-K (LSTM-Kriging) and GPR-K (GPR-Kriging) field prediction methods to construct ammonia concentration distribution fields.

Table 2. Comparison of prediction effects of four methods.

Training Data Volume	Prediction Step	Methods	MSE	RMSE	R^2	time (s)
150	10	LSTM-K	36172	190.19	0.8811	1990.854
		LSTM-MRF-K	32464	180.18	0.8933	1820.421
		GPR-K	14003	118.33	0.954	49.718
		GPR-MRF-K	**4660**	**68.26**	**0.9847**	**39.036**
	20	LSTM-K	36425	190.85	0.8802	2221.346
		LSTM-MRF-K	30417	174.4	0.9	1859.954
		GPR-K	13511	116.24	0.9556	49.772
		GPR-MRF-K	**4176**	**64.62**	**0.9863**	**40.546**
	50	LSTM-K	38295	195.69	0.8741	2322.784
		LSTM-MRF-K	23961	154.79	0.9212	1940.305
		GPR-K	17215	131.21	0.9434	49.746
		GPR-MRF-K	**6161**	**78.49**	**0.9797**	**40.794**
100	10	LSTM-K	37135	192.7	0.8779	1749.03
		LSTM-MRF-K	31923	178.67	0.895	1727.374
		GPR-K	14931	122.19	0.9509	47.181
		GPR-MRF-K	**6064**	**77.87**	**0.9801**	**35.857**
	20	LSTM-K	40133	200.33	0.868	1761.947
		LSTM-MRF-K	39021	197.54	0.8717	1745.176
		GPR-K	14475	120.31	0.9524	45.856
		GPR-MRF-K	**5829**	**76.35**	**0.9808**	**36.29**
	50	LSTM-K	47333	217.56	0.8444	1862
		LSTM-MRF-K	41289	203.2	0.8642	1856.88
		GPR-K	17879	133.71	0.9412	45.579
		GPR-MRF-K	**7879**	**88.76**	**0.9741**	**42.876**
50	10	LSTM-K	39197	197.98	0.8711	1706.729
		LSTM-MRF-K	35817	189.25	0.8822	1697.938
		GPR-K	13979	118.23	0.954	44.624
		GPR-MRF-K	**6027**	**77.63**	**0.9802**	**32.988**
	20	LSTM-K	42587	206.37	0.86	1733.195
		LSTM-MRF-K	40027	200.07	0.8684	1713.018
		GPR-K	14555	120.64	0.9521	42.649
		GPR-MRF-K	**6424**	**80.15**	**0.9789**	**33.064**
	50	LSTM-K	49027	221.42	0.8388	1828.985
		LSTM-MRF-K	45265	212.76	0.8512	1823.879
		GPR-K	1824	42.71	0.994	42.523
		GPR-MRF-K	**7894**	**88.85**	**0.974**	**33.288**

The prediction results are shown in Figures 7 and 8 respectively. Then, in order to compare the prediction effects of the four methods, we plan to use MSE and program running time to evaluate, in which MSE is used to measure the prediction accuracy and program running time is used to measure the prediction speed, the comparison results are shown in Table 2.

It can be seen from Table 2 that the GPR-MRF-K method performs better than the other three methods in the MSE and the program running time. The program calculation time meets the requirements of emergency decision-making in case of ammonia leakage, and the GPR-MRF-K can timely and effectively predict the future state of the field.

After determining the best choice of the field prediction method, this paper selects the traditional Spatio-temporal statistical method for effective comparison. The spatiotemporal statistical method is often used in some large-scale scenarios such as ocean temperature, national air quality, global ozone concentration, etc., but it is also applicable to small-scale scenarios such as ammonia leakage.

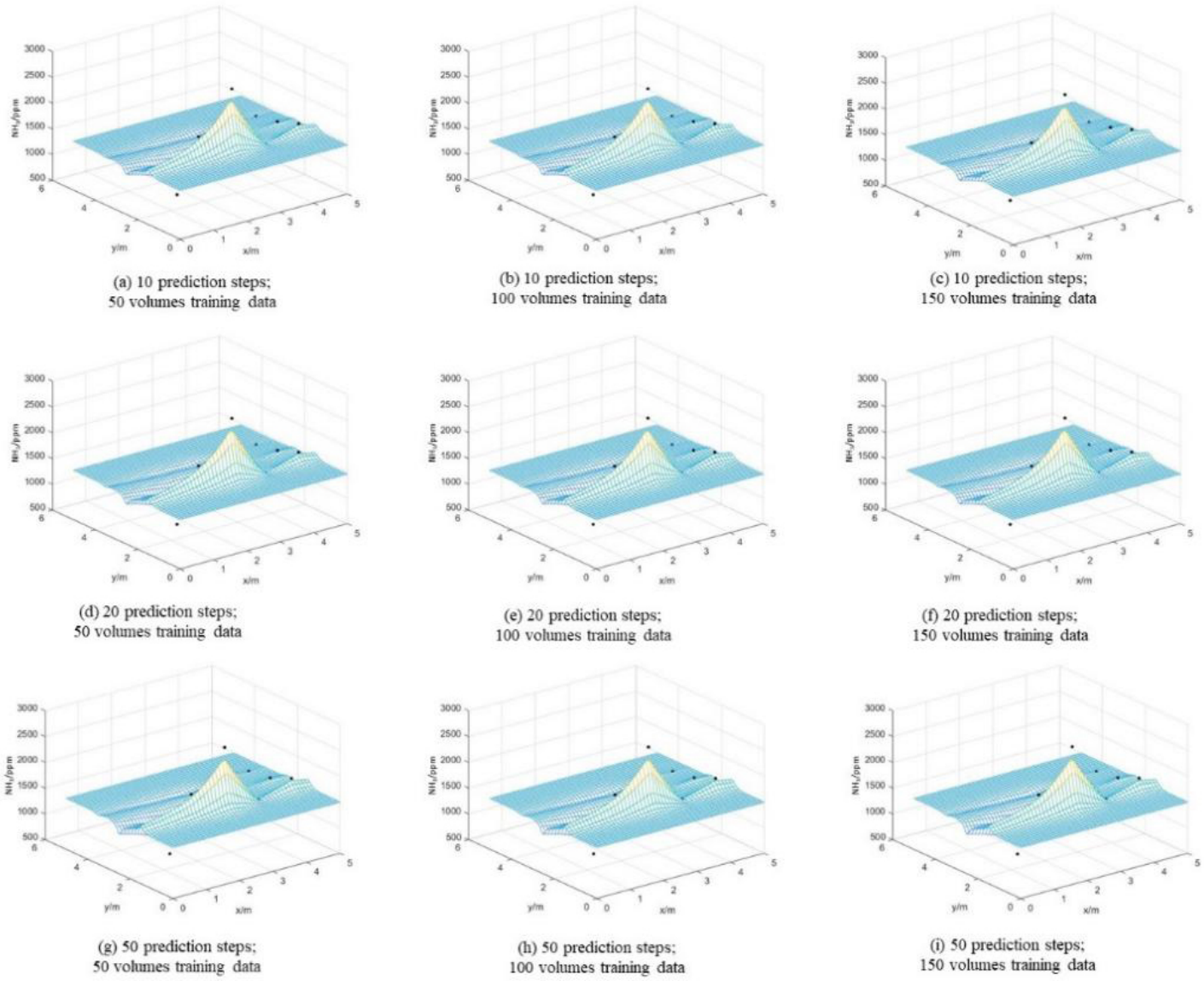
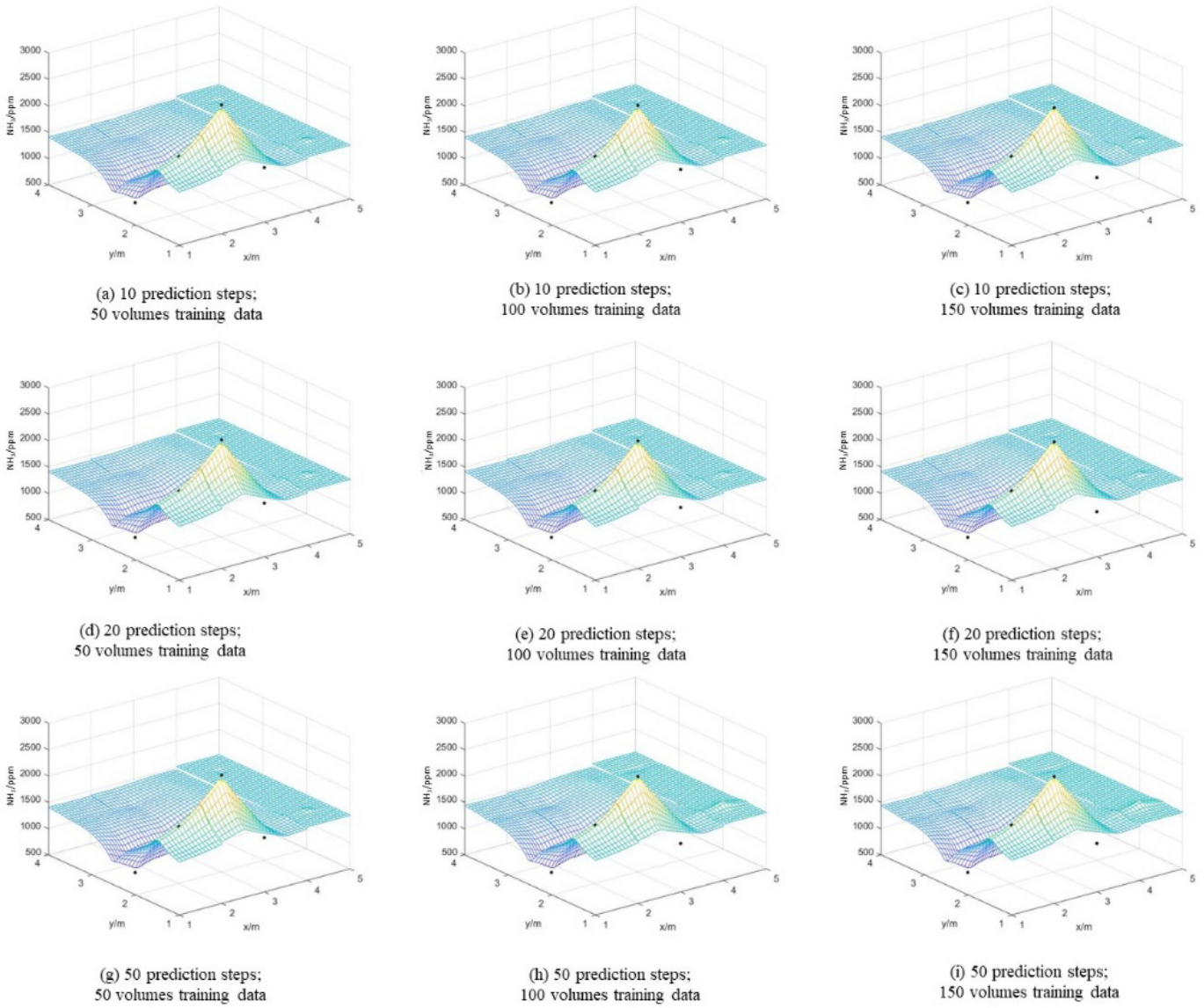


Figure 7. The field prediction of the LSTM-K method.

Perhaps the simplest way to perform Spatio-temporal prediction would be to follow Tobler's law and simply average the data in such a way as to give more weight to the nearest observations in space and time. The most obvious way to do this is through inverse distance weighting (IDW). The Spatio-temporal basis function is also an important part of Spatio-temporal statistical models. It uses the fitting and coupling of time and spatial basis functions to predict the future spatiotemporal field. The following is the comparison between the two methods and the GPR-MRF-K method proposed in this paper.

The IDW method needs to calculate the distance between the interested node and the currently known node in time and space. Because the spatial distance is certain, if the time distance is too long, the time weight will be too large, and the predicted value will tend to be equal to the known node value. The MSE will quickly increase with the increase of the number of prediction steps. It can be seen from Figure 9 that IDW cannot accurately describe the spatial change of the field in the long prediction steps.

The effect of the next 10 steps predicted by the Spatio-temporal basis function model is shown in Figure 10. The MSE reached 483 after a cross-validation. But its problem is the same as that of the IDW method: due to the existence of the time basis function, when predicting long steps, the time basis function takes too much weight, which affects the accuracy of long step prediction. At the same time, the spatial temporary basis function model takes a long time to calculate, and the prediction can reach 52s, which is not enough to meet the needs of the emergency scenario of ammonia leakage.



(a) 10 prediction steps; 50 volumes training data

(b) 10 prediction steps; 100 volumes training data

(c) 10 prediction steps; 150 volumes training data

(d) 20 prediction steps; 50 volumes training data

(e) 20 prediction steps; 100 volumes training data

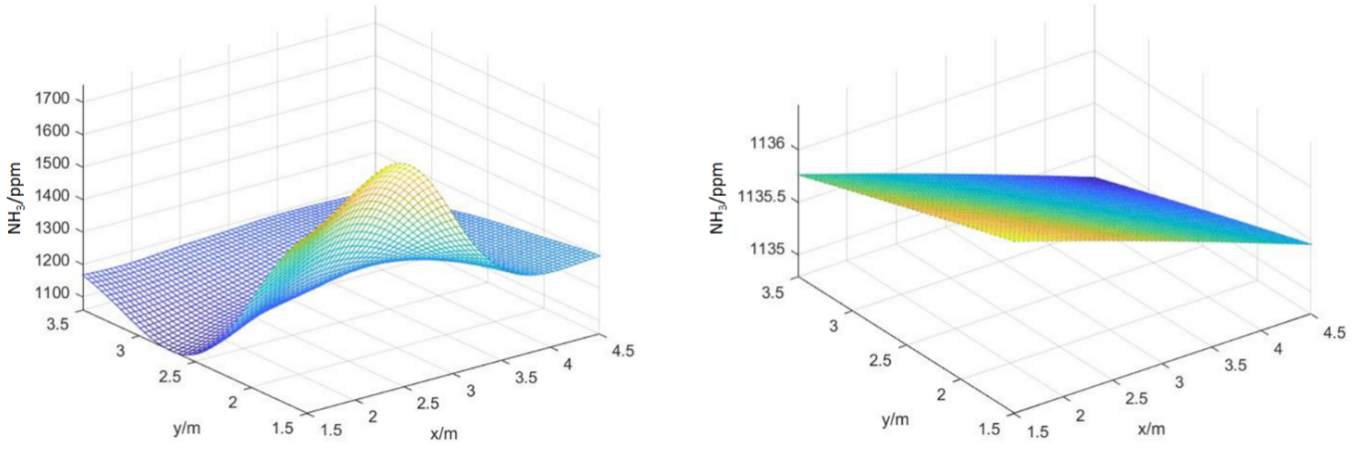
(f) 20 prediction steps; 150 volumes training data

(g) 50 prediction steps; 50 volumes training data

(h) 50 prediction steps; 100 volumes training data

(i) 50 prediction steps; 150 volumes training data

Figure 8. The field prediction of the GPR-K method.



(a) 10 Prediction Steps

(b) 50 Prediction Steps

Figure 9. The field prediction of the IDW method.

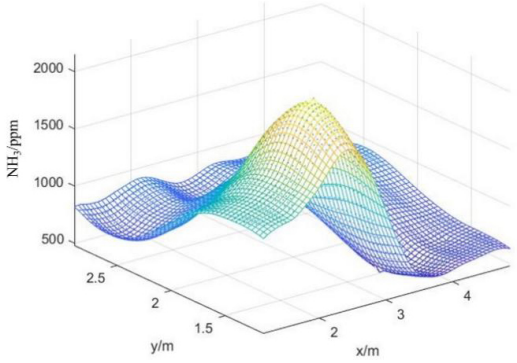


Figure 10. The field prediction of the Spatio-temporal basis function method.

5 Conclusion

In this paper, the MRF-K field construction method and the GPR-MRF-K field prediction method are proposed, which can quickly and accurately describe and predict the field of gas distribution. Firstly, the neighborhood structure of Markov random field is introduced into the Kriging model to reduce the computational complexity. Then, this paper combines three time series models: ARIMA, LSTM and GPR with the Kriging model and the MRF-K field construction method in order to select the best choice of field prediction methods. Next, through a scaled model case of ammonia leakage, the superiority of the proposed GPR-MRF-K method in calculation accuracy and calculation speed is verified. Finally, in the scenario of physical quantity monitoring and prediction, we recommend using the GPR-MRF-K method proposed in this paper.

In this study, the construction and prediction of gas distribution field are studied by improving and combining the existing methods. However, surrogate models and time series prediction models are developing rapidly. The method proposed in this paper can be combined with the latest models to obtain better results. Next, the optimization of model hyperparameters is not discussed in detail in this paper. The optimization method of hyperparameters may bring a new breakthrough to the computational efficiency. Finally, the method proposed in this paper can be extended to other fields, such as the temperature field in a confined space, the smoke concentration field in the fire scene, the temperature and humidity field in the drug warehouse and so on to verify the universality of the GPR-MRF-K method.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ignac-Nowicka, J. (2018). Application of the FTA and ETA method for gas hazard identification for the performance of safety systems in the industrial department. *Management Systems in Production Engineering*. *26*(1), 23–26. [CrossRef]

[2] Hou, J., Gai, W. M., Cheng, W. Y., & Deng, Y. F. (2020). Statistical analysis of evacuation warning diffusion in major chemical accidents based on real evacuation cases. *Process Safety and Environmental Protection, 138*, 90-98. [CrossRef]

[3] Feng, J. R., Gai, W. M., & Yan, Y. B. (2021). Emergency evacuation risk assessment and mitigation strategy for a toxic gas leak in an underground space: The case of a subway station in Guangzhou, China. *Safety science, 134*, 105039. [CrossRef]

[4] Qi, X., Wang, H., Liu, Y., & Chen, G. (2019). Flexible alarming mechanism of a general GDS deployment for explosive accidents caused by gas leakage. *Process Safety and Environmental Protection, 132*, 265–272. [CrossRef]

[5] Mei, Y., & Jian, S. (2022). Research on natural gas leakage and diffusion characteristics in enclosed building layout. *Process Safety and Environmental Protection, 161*, 247–262. [CrossRef]

[6] Hou, J., Gai, W. M., Cheng, W. Y., & Deng, Y. F. (2021). Hazardous chemical leakage accidents and emergency evacuation response from 2009 to 2018 in China: A review. *Safety science, 135*, 105101. [CrossRef]

[7] Wang, F., Chang, J., Zhang, Z., Sun, J., Zhang, Q., Zhu, C., & Wang, Z. (2019). Distributed gas detection utilizing Fourier domain optical coherence based absorption spectroscopy. *Results in Physics, 13*, 102104. [CrossRef]

[8] Jiang, Y., Xu, Z., Wei, J., & Teng, G. (2020). Fused CFD-interpolation model for real-time prediction of hazardous gas dispersion in emergency rescue. *Journal of Loss Prevention in the Process Industries, 63*, 103988. [CrossRef]

[9] Klein, L. J., Van Kessel, T., Nair, D., Muralidhar, R., Hinds, N., Hamann, H., & Sosa, N. (2017, December). Distributed wireless sensing for fugitive methane leak detection. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4583-4591). IEEE. [CrossRef]

[10] Wang, N., Gao, Y., Li, C. Y., & Gai, W. M. (2021). Integrated agent-based simulation and evacuation risk-assessment model for underground building fire: A case study. *Journal of Building Engineering, 40*, 102609. [CrossRef]

[11] Liu, J., Zhu, S., Kim, M. K., & Srebric, J. (2019). A review of CFD analysis methods for personalized ventilation (PV) in indoor built environments. *Sustainability, 11*(15), 4166. [CrossRef]

[12] Tominaga, Y., & Stathopoulos, T. (2013). CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques. *Atmospheric environment, 79*, 716-730. [CrossRef]

[13] Sharma, V. R., S, S. S., Fernandes, D. V., & MS, M. (2022). Numerical analysis of heat transfer enhancement of solar air heater using discrete triangle wave corrugations. *Cogent Engineering, 9*(1), 2051312. [CrossRef]

[14] Espinosa, R., Jiménez, F., & Palma, J. (2022). Multi-objective evolutionary spatio-temporal forecasting of air pollution. *Future Generation Computer Systems, 136*, 15-33. [CrossRef]

[15] Feng, J., Yan, L., & Hang, T. (2019). Stream-flow forecasting based on dynamic spatio-temporal attention. *IEEE Access, 7*, 134754-134762. [CrossRef]

[16] Martínez, W. A., Melo, C. E., & Melo, O. O. (2017). Median Polish Kriging for space–time analysis of precipitation. *Spatial statistics, 19*, 1-20. [CrossRef]

[17] Herman, E., Stewart, J. A., & Dingreville, R. (2020). A data-driven surrogate model to rapidly predict microstructure morphology during physical vapor deposition. *Applied Mathematical Modelling, 88*, 589–603. [CrossRef]

[18] Wang, J., Peng, X., Chen, Z., Zhou, B., Zhou, Y., & Zhou, N. (2022). Surrogate modeling for neutron diffusion problems based on conservative physics-informed neural networks with boundary conditions enforcement. *Annals of Nuclear Energy, 176*, 109234. [CrossRef]

[19] Jeong, M., & Koo, H. (2025). Evaluating Spatio-Temporal Kriging with Machine Learning Considering the Sources of Spatio-Temporal Variation. *ISPRS International Journal of Geo-Information, 14*(6), 224. [CrossRef]

[20] Schmit, L. A., & Farshi, B. (1974). Some approximation concepts for structural synthesis. *AIAA Journal, 12*(5), 692–699. [CrossRef]

[21] Kim, C., Lee, H., Kim, K., Lee, Y., & Lee, W. B. (2018). Efficient process monitoring via the integrated use of Markov random fields learning and the graphical lasso. *Industrial & Engineering Chemistry Research, 57*(39), 13144-13155. [CrossRef]

[22] Jeong, S., Murayama, M., & Yamamoto, K. (2005). Efficient optimization design method using kriging model. *Journal of aircraft, 42*(2), 413-420. [CrossRef]

[23] Na, J., Jeon, K., & Lee, W. B. (2018). Toxic gas release modeling for real-time analysis using variational autoencoder with convolutional neural networks. *Chemical Engineering Science, 181*, 68-78. [CrossRef]

[24] Zhang, D., Liang, Y., Cao, L., Liu, J., & Han, X. (2022). Evidence-theory-based reliability analysis through Kriging surrogate model. *Journal of Mechanical Design, 144*(3), 031701. [CrossRef]

[25] Liu, X., Zhao, W., & Wan, D. (2022). Multi-fidelity Co-Kriging surrogate model for ship hull form optimization. *Ocean engineering, 243*, 110239. [CrossRef]

[26] Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural computation, 3*(2), 246-257. [CrossRef]

[27] Song, D., Lee, K., Phark, C., & Jung, S. (2021). Spatiotemporal and layout-adaptive prediction of leak gas dispersion by encoding-prediction neural network. *process safety and Environmental Protection, 151*, 365-372. [CrossRef]

[28] Cho, S., Kim, Y., Kim, M., Cho, H., Moon, I., & Kim, J. (2022). Multi-objective optimization of an explosive waste incineration process considering nitrogen oxides emission and process cost by using artificial neural network surrogate models. *Process Safety and Environmental Protection, 162*, 813-824. [CrossRef]

[29] Ma, Y., He, Y., Wang, L., & Zhang, J. (2022). Probabilistic reconstruction for spatiotemporal sensor data integrated with Gaussian process regression. *Probabilistic Engineering Mechanics, 69*, 103264. [CrossRef]

[30] Zhou, X., Dong, C., Zhao, C., & Bai, X. (2020). Temperature-field reconstruction algorithm based on reflected sigmoidal radial basis function and QR decomposition. *Applied Thermal Engineering, 171*, 114987. [CrossRef]

[31] Liao, Z., Wang, B., Xia, X., & Hannam, P. M. (2012). Environmental emergency decision support system based on Artificial Neural Network. *Safety Science, 50*(1), 150-163. [CrossRef]

[32] Picka, J. D. (2006). Gaussian Markov random fields: theory and applications. [CrossRef]

[33] Wang, D., Liu, K., & Zhang, X. (2022). A spatiotemporal prediction approach for a 3D thermal field from sensor networks. *Journal of Quality Technology, 54*(2), 215-235. [CrossRef]

[34] Xu, L., & Huang, Q. (2012). Modeling the interactions among neighboring nanostructures for local feature characterization and defect detection. *IEEE transactions*

on automation science and engineering, *9*(4), 745-754. [CrossRef]

[35] Li, S., Deng, J., Li, Y., & Xu, F. (2022, May). An intermittent fault severity evaluation method for electronic systems based on LSTM network. In *2022 Prognostics and Health Management Conference* (*PHM-2022 London*) (pp. 224-227). IEEE. [CrossRef]



Chenglong Hou is a second-year Master's student at the School of Reliability and Systems Engineering, Beihang University, Beijing, China. His research focuses on network reliability. (Email: hcl1255@163.com)



Yuhao Zha is a graduate of the School of Reliability and Systems Engineering, Beihang University, Beijing, China, where he completed his Master's degree. (Email: zhayuhao0523@163.com)



Jun Yang is a professor of School of Reliability and Systems Engineering, Beihang University, China. He got his Ph.D. from Academy of Mathematics and Systems Science, Chinese Academy of Sciences(CAS) in 2006. He is interested in developing statistical theory and method for data analysis. His research fields are mainly on reliability statistics, statistical process control and survey sampling. Now, he has published three books and more than 40 papers, included 10 papers indexed by SCI and 21 papers indexed by EI. Meanwhile, he presides Grant 11001005 from the National Natural Science Foundation of China. In addition, He is a member of Editorial Board of "Applications and Applied Mathematics: An International Journal". (Email: tomyj2001@buaa.edu.cn)



Ning Wang received his Ph.D. from the School of Reliability and Systems Engineering, Beihang University, Beijing, China. He is currently a Postdoctoral Researcher at the School of Cyber Science and Technology, Beihang University. His research focuses on network reliability. (Email: wangning1226@buaa.edu.cn)