Future prospects research on offshore wind power scale in China based on signal decomposition and extreme learning machine optimized by principal component analysis

Dunnan Liu1,2 | Mingguang Liu1,2 | Xiaofeng Xu1,2 | Jiani Wang1,2 | Chenkai Fan1,2 | Xiaoyu Li1,2 | Jinshan Han1,2

1School of Economics and Management, North China Electric Power University, Beijing, China
2State Key Lab Alternate Elect Power System Renewable, North China Electric Power University, Beijing, China

Abstract
In recent years, China has promoted many new energy projects in order to meet the growing demand for electricity. Therefore, China's offshore wind power installed capacity has grown rapidly. China has a long coastline and abundant offshore wind energy resources. Offshore wind power is an important area for the development of renewable energy, which can promote wind power technology advancement and energy structure adjustment. Therefore, conducting effective research and forecast on the cumulative installed capacity of China's offshore wind power will help the government to rationally deploy and reduce the risk of investment in offshore wind power. In order to accurately predict the future prospects of offshore wind power in China, this paper firstly constructed a set of influencing factors and used gray correlation analysis to screen the main influencing factors. Then, this paper proposed a novel forecasting model named e-VMD-PCA-RELM. The algorithm is based on the traditional RELM (robust extreme learning machine) algorithm, which effectively processes the noise information through the PCA (principal component analysis) algorithm, and extracted the feature elements of the RELM hidden layer to reduce the information redundancy. At the same time, the e-VMD (variational mode decomposition optimized by entropy) algorithm is used to decompose the original time series to obtain multiple components. By comparing with the other forecasting algorithms, it is proved that the proposed forecasting model has strong generalization ability and has achieved good prediction result. Finally, the e-VMD-PCA-RELM model is used to predict the scale of offshore wind farms in China from 2019 to 2035. We find that the cumulative installed capacity of China's offshore wind power will exceed 60 GW in 2035, and the installed capacity will increase year by year. In 2030, there will be a large increase, with a relative growth rate of 20%.

KEYWORDS
extreme learning machine, offshore wind power scale forecasting, principal component analysis, signal decomposition
1 | INTRODUCTION

Wind energy has become one of the most mature, most widely used, and fastest-growing new energy sources. The increasing installed capacity of wind power worldwide has proved that it can be applied on a large scale. During the “Twelfth Five-Year Plan” period in China, the newly added wind power installed capacity is 210 million kW, with an average annual growth rate of 18.4%. It is the fastest-growing power resource in clean energy. In China, the offshore wind speed is more than 20% higher than that of land, and the power generation per unit area can be increased by more than 70%. In recent years, with the increase in energy demand and the deterioration of the global environment, wind power generation should become a key energy project. As of 2017, the installed capacity of the global wind power market has historically increased by 4334 MW, an increase of 95% compared to 2016. The cumulative installed capacity of offshore wind power in the world reached 18814 MW, a 30% increase from 2016 (14384 MW). Europe continues to maintain its position as the world’s largest offshore wind power market, with nearly 84% (15780 MW) of offshore wind power facilities located in the offshore waters of 11 European countries. China’s offshore wind power is during the breakthrough progress with a total of 11.6 MW new installed capacity and cumulative installed capacity reached 27.9 MW. According to the National Development and Reform Commission’s Energy Institute, the potential of wind power installed capacity in China’s offshore can reach 100-200 GW. Therefore, in the future, the development of offshore wind power projects has become a hot spot for the use of renewable clean energy.

In order to accurately predict the cumulative installed capacity of offshore wind power in China, this paper proposes a forecasting model based on e-VMD (variational mode decomposition information optimized by entropy) and PCA-RELM (robust extreme learning machine optimized by principal component analysis) as a combined model of cumulative installed capacity of offshore wind power in China. The research innovations of this paper are as follows:

1. The VMD algorithm is optimized by entropy, which solves the problem that the historical time series characteristics of the cumulative installed capacity of offshore wind power are weak and difficult to extract. The entropy can reflect the validity of signal decomposition. The smallest principle of entropy value is used to determine the value of $k$ and optimize $\alpha$ in VMD.

2. The RELM algorithm is improved by PCA. Through the PCA, the principal component features of the hidden layer can be extracted, and the influence of the number of hidden layer nodes on the accuracy of the model can be reduced. The quick setting of the number of layer nodes simplifies the operation flow.

3. This paper uses combined prediction algorithm to predict the scale of offshore wind power in China. Eighteen secondary influencing factors are selected according to energy resources, industrial technology, investment, and macroeconomics and screened by gray relationship analysis, thus ensuring the accuracy of the prediction of the overall combined algorithm.

The main contents of the paper are as follows: Section 2 summarizes the research status of offshore wind power and prediction algorithms development; Section 3 introduces the algorithm principle of the VMD, RELM etc; Section 4 summarizes the overall flow of the e-VMD-PCA-RELM model; Section 5 applies the proposed e-VMD-PCA-RELM model to forecast the cumulative installed capacity of offshore wind power in China. Based on the comparison with RELM, LSSVM, PCA-RELM and VMD-PCA-RELM prediction results, it proves the practicability and accuracy of the proposed model. The cumulative installed capacity of China’s offshore wind power from 2019 to 2035 is output by the model; Section 5 presents forward-looking conclusions.

2 | LITERATURE REVIEW

In recent years, many scholars have devoted themselves to studying the status of offshore wind energy resources in various countries, and conducted evaluation from multiple layers. The hottest research is the construction technology of wind power such as tower structure, unit aerodynamic performance, wind turbine blade optimization design. Ma et al. established the time domain coupling calculation model of floating wind turbine, and proposed the wind turbine blade combination optimization design method. Jalbi et al. studied the role of vibration mode in the design of WTG (wind turbine generator) support structure and proposed the symmetrical (equal) triangle and asymmetry (isosceles) which can help the designer to optimize the basic layout. pokhrel et al. simulated different degrees of structural damage to offshore wind turbines (OWT) caused by different natural climates and
disasters, and conducted vulnerability analysis in the form of a brittle curve. Yao et al.\textsuperscript{6} improved the output characteristics of offshore wind power and enhance the wind power regulation capability of the southern coast of China and proposed an optimized selection method for seawater pumped storage power station (SPSP). Castro et al.\textsuperscript{15} proposed an evaluation model to achieve a life cycle cost assessment of a combined or hybrid floating marine renewable energy system to better understand the technical solution; Shiau et al.\textsuperscript{16} studied the social impact assessment of offshore wind power. Salvador et al.\textsuperscript{17} studied the legal constraints imposed by the protection of the marine environment and other interests to determine areas with large wind density (WPD) and less conflict.

Wind power resources forecasting is a new research spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning. Literature\textsuperscript{21-31} studied the wind power resources calculation spot, which is important to capacity layout and planning.

So far, many scholars have proposed various prediction models for wind energy installation prediction and predictive control algorithm. The modified model reduces the output error by output feedback and error correction. García-Zhang et al.\textsuperscript{38} proposed a new method for predicting stock prices in combination with Firefly algorithm (FA). The experimental results showed that the proposed combined algorithm has superior performance compared with other algorithms.

3 | FORECASTING MODEL

3.1 | Robust extreme learning machine

RELM is a single hidden layer feedforward neural network proposed in recent years. RELM is different from other neural networks, which can randomly generate weight and threshold while we input data. Meanwhile, it can obtain output weight by generalizing inverse matrices without adjusting parameters. RELM learns faster and has better generalization performance. The algorithm process is described as follows.

Suppose there are different samples $(x_i, y_i)$, among them, $x_i = [x_i1, x_i2, \ldots, x_in]^T \in \mathbb{R}^n$, $y_i = [y_i1, y_i2, \ldots, y_im]^T \in \mathbb{R}^m$, the network has hidden layer nodes, then the RELM network output is

$$\theta(x_i) = \sum_{k=1}^{K} \beta_k g(W_k \cdot x_i + b_k)$$  \hspace{1cm} (1)

In the formula, $W_k = (w_{k1}, w_{k2}, \ldots, w_{kn})^T$ is the weight of the $k$th hidden layer neuron and the input layer neurons. $\beta_k = (\beta_{k1}, \beta_{k2}, \ldots, \beta_{km})^T$ is the weight of the $k$th hidden layer neuron and the output layer neuron. $b_k$ is the threshold of the $k$th hidden layer neuron. $W_k \cdot x_i$ is inner product of $W_k$ with $x_i$, $g(x)$ is the activation function.

RELM that owns $k$ implicit neurons and one activation functions $g(x)$ that can make the error of sample approximately equal to 0, $\sum_{i=1}^{l} (\theta(x_i) - y_i) \approx 0$. Then, we can get the equation as follows.

$$\sum_{k=1}^{K} \beta_k g(W_k \cdot x_i + b_k) = y_i, i = 1, 2, \ldots, J$$  \hspace{1cm} (2)
Above equations can be summarized as
\[ H\beta = Y \quad (3) \]

Among them, \( H \) is the output matrix for the hidden layer, \( \beta = (\beta_1, \beta_2, \ldots, \beta_K)_{K \times m} \), \( Y = (y_1^T, y_2^T, \ldots, y_m^T)_{m \times m} \). Then randomly determine input weights and thresholds, and use the least-squares solution to calculate output weights.

\[ \beta = H^+y \quad (4) \]

where \( H^+ \) is the generalized inverse matrix \( H \).

3.2 Principal component analysis algorithm

Principal component analysis (PCA) is the basic method of multivariate statistics.\(^{41}\) The basic idea of PCA is to find a new set of variables that replace the original variables. The new variables are linear combinations of the original variables. The new variables are irrelevant and carry the useful information of the original variables to the maximum extent. The principal component extracted by PCA can reflect the core information of the research object, remove the linear relationship and redundant information between high-dimensional data variables, thereby realizing data dimensionality reduction, and improving the efficiency of analysis problems while ensuring modeling accuracy.

Supposing a data set consists of \( m \) variables, \( n \) sets of values, this paper converted the data set into a matrix \( X, X \in \mathbb{R}^{mn} \), as shown in formula (5).

\[ X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (5) \]

The calculation formula of the steady-state point of the optimal problem is as follows.

\[ \max_{v \neq 0} \frac{v^T X^T X v}{v^T v} \quad (6) \]

where \( v \in \mathbb{R}^m \). Formula (6) can be calculated by singular value decomposition.

\[ \frac{1}{\sqrt{n-1}} X = U \sum V^T \quad (7) \]

where \( U \in \mathbb{R}^{mn} \), \( V \in \mathbb{R}^{mn} \), matrix \( \sum \in \mathbb{R}^{mn} \) are positive numbers that includes with successive descending along its main diagonal, and all other elements are 0. The load vector is an orthogonal column vector in the matrix \( V \), but the variance of the projection of the historical data set \( V \) along the \( i \)th column of the matrix \( V \).

Formula (7) is equivalent to formula (8), which means the decomposition of the sample covariance matrix \( S \).

\[ S = \frac{1}{n-1} X^T X = V \Lambda V^T \quad (8) \]

Diagonal matrix \( \Lambda = \sum \Lambda \in \mathbb{R}^{mn} \) has eigenvalues that are successively decremented, where the \( i \)th eigenvalue is equal to the square of the \( i \)th singular value.

Retaining the larger singular value corresponding to the load vector according to the cumulative variance contribution rate formula.

\[ \frac{\sum_{i=1}^{a}\lambda_i}{\sum_{i=1}^{m}\lambda_i} > 85\% \quad (9) \]

Select the column of the load matrix \( P \in \mathbb{R}^{mn} \) to make the columns of matrix \( T \) correspond to each other. Therefore, the matrix \( T \) contains the projection of the data set \( X \) in a low dimensional space.

\[ T = XP \quad (10) \]

3.3 RELM optimized by PCA

The change in installed capacity of offshore wind power is a complex system. The traditional RELM cannot effectively model the data of offshore wind power installation. Aiming at this problem, this paper proposes a novel robust extreme learning machine based on principal component extraction. By performing PCA on the RELM hidden layer, the principal component features of the data are extracted, and the linear correlation between variables is removed to simplify the research problem.

PCA can effectively process high-dimensional noise data. Principal feature extraction is performed on the elm hidden layer to reduce the data dimension and remove redundant information between data. Further learning of the neural network based on the extracted principal component features can reduce the influence of the number of hidden layer nodes on the accuracy, and realize the rapid setting of the number of nodes in the RELM hidden layer, so that RELM is robust.

Compared with the traditional ELM, the PCA-RELM model has the advantages of simple design, good robustness, and high precision, which can guide the prediction of the installed capacity of offshore wind power. The example shows that the novel model does not require high empirical data and multiple training to optimize the parameters, so that higher
precision prediction results can be obtained, and the machine learning design is simple and quick.

The PCA-RELM network structure diagram is shown in Figure 1.

Given network training sample set $U$ that includes $I$ group of multiple input data $U = \{ (X_i, Y_i) \mid i = 1, 2, \ldots, I; X_i \in \mathbb{R}^n, Y_i \in \mathbb{R} \}$. Establishing a PCA-RELM model based on training sample $U$. The steps are as follows.

Step 1: Data normalization

The experimental data are normalized to have a uniform dimension.

$$\begin{align*}
\bar{x}_i &= \frac{x_{\text{max}} - x_{\text{in}}}{x_{\text{max}} - x_{\text{min}}} (i = 1, 2, \ldots, I) \\
\bar{y}_i &= \frac{Y_i - Y_{\text{min}}}{Y_{\text{max}} - Y_{\text{min}}} (i = 1, 2, \ldots, I)
\end{align*} \quad (11)$$

Among them,

$$\begin{align*}
x_{\text{max}} &= \max \{ x_{1n}, x_{2n}, \ldots, x_{in} \} \\
x_{\text{min}} &= \min \{ x_{1n}, x_{2n}, \ldots, x_{in} \} \\
x_{\text{max}} &\neq x_{\text{min}} \quad (13)
\end{align*}$$

$Y_{\text{min}}$ and $Y_{\text{max}}$ are the minimum and maximum values in the output vector $Y$, respectively.

Step 2: Randomly generate input weight.

$$W_k = (w_{1k}, w_{2k}, \ldots, w_{nk})^T, k = 1, 2, \ldots, K \quad (14)$$

where $k$ is the number of hidden layer nodes.

Step 3: Randomly generated hidden layer neuron threshold $b_1, b_2, \ldots, b_k$, and calculate the hidden layer output matrix $H$ according to formula (3).

Step 4: Extract the main components of the hidden layer output matrix $H$.

Constructing PCA process on matrix $H$ and obtaining a new sample set $U' = \{ (P_i, Y_i) \mid i = 1, 2, \ldots, I; P_i \in \mathbb{R}^m, Y_i \in \mathbb{R} \}$. Among them, $P_i = (p_{i1}, p_{i2}, \ldots, p_{im})$ is the extracted principal component. Output matrix $H$ is transformed like the follows.

$$H' = \begin{bmatrix}
p_{11} & \cdots & p_{1m} \\
\vdots & \ddots & \vdots \\
p_{I1} & \cdots & p_{Im}
\end{bmatrix}_{IxM} \quad (15)$$

Step 5: Calculate the output weight matrix $\beta$ according to formula (4).

Step 6: Train and test the PCA-RELM model.

Based on the above steps, the general flow of the PCA-RELM model proposed in this paper is shown in Figure 2.

### 3.4 Variational mode decomposition

Variational mode decomposition (VMD) is a new adaptive signal processing method based on Wiener filtering. In the VMD algorithm, each IMF is called AM – FM Signal.
where, \( A_k(t) \) is the instantaneous amplitude of the signal; \( u_k(t) \) can be regarded as the harmonic signal.

Assuming that each modality is a finite bandwidth with a center frequency, and the center frequency and bandwidth are constantly updated during the decomposition process, the variational problem can be expressed as seeking \( k \) modal functions \( u_k(t) \), and make the estimated bandwidth of all modal functions are minimum. The sum of modalities is the input signal \( f \). Specific steps are as follows.

**Step 1.** Construct Hilbert transformation for each modal function \( u_k(t) \).

\[
\left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t)
\]  

**Step 2.** Mix the estimated center frequency of each modal analysis signal, and transfer the spectrum of each modality to the baseband.

\[
\left( \delta(t) + \frac{j}{\pi t} \right) \cdot u_k(t) \cdot e^{-j\omega_k t}
\]  

**Step 3.** Calculate the bandwidth of the signal after frequency shifting by using \( H^1 \) Gaussian smooth estimation, resulting in the following constrained variational problem.

\[
\min \left\{ \sum_k \left| \frac{\partial}{\partial t} \left( \delta(t) + \frac{j}{\pi t} \right) \ast u_k(t) e^{-j\omega_k t} \right|^2 \right\}
\]  

s.t. \( \sum_{k=1}^{K} u_k = f \)

where \( u_k \) is the obtained decomposition. \( \omega_k \) is the center frequency of each component.

**Step 4.** Solve the above variational problem.

Introduce a quadratic penalty factor \( a \) and Lagrange multiplication operators \( \lambda \). \( a \) can ensure the accuracy in the presence of Gaussian noise. \( \lambda \) can ensure the Guarantee the strictness of the constraints. Through repeated iterations, the “saddle point” of the Lagrange expression is found to solve the minimum value. The extended Lagrange expression is as follows.

\[
\Gamma(\{u_k\}, \{\omega_k\}, \lambda) = a \sum \left\| \frac{\partial}{\partial t} \left( \left( \delta(t) + \frac{j}{\pi t} \right) \ast u_k(t) e^{-j\omega_k t} \right) \right\|_2^2 + \left\| f - \sum u_k(t) \right\|_2^2 + \left\langle \lambda(t), f - \sum u_k(t) \right\rangle
\]  

where the optimal solution is intrinsic mode function \( \{u_k\} \) and their respective center frequencies \( \{\omega_k\} \).

### 3.5 VMD parameter optimization process based on entropy value

Aiming at the problem that the historical time series characteristics of the cumulative installed capacity of offshore wind power are weak and difficult to extract, a signal feature extraction method based on entropy theory is proposed.

Because the historical change of the cumulative installed capacity of offshore wind power can be decomposed into a variety of signals, based on,\(^{36}\) it can be considered that the more ordered the signal, the smaller the value of entropy. Therefore, entropy can be used as a kind of representation of the effectiveness of signal decomposition. When the information entropy value has the minimum value, it can be considered that it has a good VMD decomposition effect. Based on this, this paper proposes a method to determine the VMD parameters based on the principle of entropy.

The proposed method in this paper aims to extract the signal components including the fluctuation characteristics of the cumulative installed capacity of offshore wind power from the original signals, then achieving weak changes. Firstly, the original fault signal is subjected to VMD to obtain a set of IMF components. When optimizing the parameters, the IMF with the top three entropy values is removed, and the IMF component with smaller entropy value is selected as the effective IMF component. The envelope demodulation analysis is carried out to extract the fluctuation characteristic frequency of the historical time series of the cumulative installed capacity of the offshore wind power. The flow chart is shown in Figure 3.

### 4 APPLICATION OF E-VMD-PCA-RELM FORECASTING MODEL

Due to the insufficient historical data of the cumulative installed capacity of offshore wind power in China, it is difficult to achieve the data magnitude requirements of intelligent algorithms. Therefore, it is easy to cause the original data homogeneity of the influencing factors. Meanwhile, with the
FIGURE 3  VMD parameter optimization process based on entropy value
different scale information characteristics of the input data, the traditional machine intelligence algorithm cannot fully discover the time-frequency characteristics of the time series data, and thus greatly affect the performance of the forecasting model. The accuracy of the RELM algorithm cannot be achieved. Therefore, this paper firstly separates the original data by e-VMD and extracts the eigenmode components including the volatility characteristics of the cumulative installed capacity of the offshore wind power, thereby realizing the magnitude of the data, and realize the data requirements of the machine algorithm.

The prediction steps of the overall algorithm of e-VMD-PCA-RELM model are as follows.

Step 1: Time series data decomposition. Using e-VMD to decompose \( x(t) \) and obtain \( n \) IMF components and one residual \( r_n \).

Step 2: Based on the principle of entropy, the first three IMF components with larger entropy value are eliminated, and the final \( n - 3 \) IMF components and one residual \( r_n \) are obtained.

Step 3: Build training and test sample sets of the IMF component and the \( r_n \). The input component and the output of each component are constructed.

Step 4: Construct PCA-RELM training for each component. In the PCA-RELM algorithm, the number of nodes of any hidden layer is selected. Establish PCA-RELM model to obtain prediction results for each component.

Step 5: Output prediction results. For each IMF component and \( r_n \), the predicted results are summed to obtain the final predicted results of the cumulative installed capacity of offshore wind power in China.

In summary, the overall forecasting flow chart of scale of offshore wind power in China is shown in Figure 4.

5 | EMPIRICAL ANALYSIS

5.1 | Data processing

A total of 18 impact indicators such as electricity consumption are used as preliminary indicators for the factors which are shown in Table 1. The classification and specific data of preliminary indicators are from the China Statistical Yearbook and the official website of the China Energy Administration. Among them, the offshore wind energy resource data are derived from the statistical data of the height of 50 m in the China Sea. Industrial technology and investment data are derived from China’s Shanghai offshore wind power statistics. Since the cumulative installed capacity of offshore wind power is a complex nonlinear problem with many influencing

![Overall forecasting flow chart of scale of offshore wind power in China](image)

**FIGURE 4** Overall forecasting flow chart of scale of offshore wind power in China
factors, the relationship between each index and the cumulative installed capacity is vague. Therefore, this paper uses the gray correlation method to define the influencing factors. The effective screening of the preliminary impact indicator set can effectively improve the running speed and accuracy of the forecasting model.

### 5.1.1 Normalization of indicators

Normalize all corrected data by using Z-score data standardization method.

\[ y_i = \frac{x_i - \overline{x}}{\sigma} \]  \hspace{1cm} (21)

where \( y_i \) is the normalized data, \( \overline{x} \) is the average of the original series, and \( \sigma \) is the data standard deviation of the original number.

Since the average water area roughness and the average investment payback period are negative indicators, in order to achieve the data forward, the inverse index forward processing method is adopted.

\[ X_i = \frac{Y_i - \max}{\min - \max} \]  \hspace{1cm} (22)

where \( X_i \) is the data after the forwardization, \( \max \) is the maximum value of the normalized data, and \( \min \) is to the minimum value of the normalized data.

### 5.1.2 Gray correlation method

In this paper, the normalized and forwarded data of the preliminary impact indicators from 2000 to 2018 are input to the gray correlation algorithm. The gray correlation is shown in Table 2.

| Primary indicator                  | Secondary indicators                  | Indicator description                                                                 |
|-----------------------------------|---------------------------------------|----------------------------------------------------------------------------------------|
| Offshore wind energy resources    | Average wind power density (W/m³)     | The power per unit area of the wind perpendicular to the wind direction                  |
|                                   | Offshore wind turbulence intensity (%)| Ratio of standard wind speed deviation to average wind speed                             |
|                                   | Annual average wind speed (m/s)       | The average time of the instantaneous wind speed over a full year interval               |
|                                   | Average water roughness               | The smaller pitch of the surface in the water area and the unevenness of the tiny peaks |
|                                   | Offshore wind speed vertical shear    | The change of wind speed in the plane perpendicular to the wind direction                |
| Industrial technology             | Annual total power generation (kWh)   | Annual total amount of offshore wind power generation                                   |
|                                   | Offshore wind power abandonment rate (%)| The ratio of the amount of electricity that the partial wind turbine stops generating electricity to the total installed capacity due to insufficient grid acceptance capacity |
|                                   | Average injection harmonic pass rate (%)| The ratio of the number of times the voltage level bus voltage harmonic voltage to the total number of injections |
|                                   | Average unit life (y)                 | Average life of offshore wind turbines                                                |
|                                   | Average power generation (kW)         | Annual average power generation of offshore wind turbines                                |
|                                   | Total equipment utilization hours (h)  | Total annual utilization hours of offshore wind turbines                                 |

| Investment                        | Average investment payback period (y) | Annual average payback period for offshore wind turbines                                |
|                                   | Average investment yield (%)          | Annual average investment yield for offshore wind turbines                                |
|                                   | Average net profit (million)          | Annual average net profit of offshore wind turbines                                      |
|                                   | Average internal rate of return (%)   | Annual average internal rate of return for offshore wind turbines                        |

| Macroeconomics                    | Whole society electricity consumption (kWh) | China’s annual total social electricity demand |
|                                   | GDP (yuan)                                | Chinese annual GDP                        |
|                                   | Secondary industry electricity consumption (kWh) | Total annual electricity consumption of China’s secondary industry |
The four indicators below 0.5500 are excluded,33 that is, the average waters roughness, the sea abandonment rate, the average investment payback period, and the average investment return rate, and then, final influencing factors of installed capacity of offshore wind power in China are obtained. The normalized index values (2000-2018) of main influencing factors are shown in Table A1 in Appendix A.

5.2 Application of e-VMD-PCA-RELM model

Decomposing the original historical load sequence by e-VMD. Firstly, input the time series signal of the cumulative installed capacity into the e-VMD algorithm. In the combined forecasting algorithm, the cumulative installed capacity e-VMD model of China’s offshore wind power from 1990 to 2018 is first loaded. 4 IMFs and one residual are obtained. The decomposition results are shown in Figure 5.

From Figure 5, it can be found that each IMF showed more obvious characteristics of periodic changes. IMF1 has the highest frequency and a shorter wavelength while IMF2 to Residual decrease in order. The time series of the cumulative installed capacity has obvious multi-scale features, and the 4 IMF components present information of high and low variations with different fluctuation scales. Among them, the high frequency of IMF1 can reflect the random noise information of the original time series data, the residual frequency is low, and its change is stable. In fact, it represents the average trend of the original data sequence of the cumulative installed capacity of offshore wind farms in China. The original data sequence is independent of each other, and the individual IMF components are mutually orthogonal.

The influencing factor indicator set and each component are used as input of the training set and testing set of the PCA-RELM model.

The training parameters unified in this paper are shown in Table 3.

This paper reconstructs the prediction series to obtain the prediction results and relative errors of the components and residuals of the time series decomposition of the cumulative installed capacity of offshore wind farms in China since 1990. The predicted results for each series are shown in Figure 6.

Through the analysis of Figure 6, this paper finds that the e-VMD-PCA-RELM prediction model proposed in this paper has high prediction accuracy for each IMF and the residual. The average error of each component is −3.368%, −0.789%, −2.211%, 1.895%, 2.211%, and the total average relative error is −0.453%. Among them, the predicted average relative error of the residual signal timing sequence does not exceed 2.211%, indicating that the residual frequency is low and the change is stable, reflecting the trend information of the cumulative installed capacity of China’s offshore wind power. Therefore, it is proved that the e-VMD-PCA-RELM prediction model proposed in this paper is highly scientific and applicable.

5.3 Predictive effect analysis

In order to compare the superiority of the proposed model, this paper also inputs the data of the main influencing factors into LSSVM, RELM, PCA-RELM, and VMD-PCA-RELM models. Finally, the results of the five prediction models are shown in Figure 7. Since the difference between 2000 and 2016 cannot be distinguished in (A), the partial enlargement of 2000-2016 is shown in (B).

In order to objectively compare the accuracy of a variety of models, common statistical indicators including rmse, \( r^2 \), mre are adopted and the index calculation formula is as follows:

\[
\text{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{q}_i - q_i)^2} \quad (23)
\]

\[
\text{r}^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{q}_i - \bar{q})^2}{\sum_{i=1}^{n} (q_i - \bar{q})^2} \quad (24)
\]

\[
\text{mre} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{q}_i - q_i|}{q_i} \quad (25)
\]

| Table 2 | Gray correlation |
|---------|------------------|
| Influencing factor | Gray correlation |
| Average wind power density | 0.9348 |
| Offshore wind turbulence intensity | 0.9021 |
| Annual average wind speed | 0.9485 |
| Average water roughness | 0.5117 |
| Offshore wind speed vertical shear | 0.9386 |
| Annual total power generation | 0.8829 |
| Offshore wind rate | 0.3862 |
| Injection harmonic pass rate | 0.8715 |
| Average power generation | 0.9470 |
| Average unit life | 0.5324 |
| Total equipment utilization hours | 0.6448 |
| Average payback period | 0.4223 |
| Average return on investment | 0.4982 |
| Average net profit | 0.5853 |
| Average internal rate of return | 0.6769 |
| Whole society electricity consumption | 0.9022 |
| GDP | 0.8955 |
| Secondary industry electricity consumption | 0.9320 |
where \( \hat{q}_i \) is the predicted value, \( \bar{q}_i \) is the sample mean, and \( n \) is a sample number.

The calculation results for the five forecasting models are compared in Table 4.

According to the calculation results of the statistical indicators, we conclude the following points:

1. The accuracy of the signal decomposition prediction algorithm is higher than that of the single algorithm.
2. The prediction accuracy of the model under the combined algorithm is higher than that of the single algorithm.
3. The prediction effect of the RELM algorithm itself is not as good as that of the LSSVM but the combination of algorithms can greatly improve the its prediction effect.
4. The values of the statistical indicators show good performance of e-VMD-PCA-RELM model, indicating that signal decomposition plays an important role in the combined prediction algorithm.

5.4 | Final prediction output

The paper applied GM(1,1) to predict the value of main influencing factors from 2019 to 2035, which will be used as input data of e-VMD-PCA-RELM forecasting model. Finally, the paper got the amount of cumulative installed capacity of China offshore wind power from 2019 to 2035, which has been shown in Figure 8.

6 | CONCLUSION

Promoting the construction and access of China’s offshore wind power projects is a key part of the current energy planning work. It is significant for the expansion of offshore wind power installed capacity to update energy structure and decrease carbon emission. This paper uses the e-VMD-PCA-RELM forecasting model to successfully predict the cumulative capacity of offshore wind power installed in China in the future. It can be found that the compound growth rate reaches 16.90%, and it shows a trend of increasing year by year. Through comparison and verification of various models, it is found that the combined prediction model used in this paper has the highest prediction accuracy. The forecasting results show that by the end of 2035, the cumulative installed capacity of China’s offshore wind power will reach more than 60.924 GW. Therefore, it can be preliminarily concluded that by 2035, the proportion of China’s offshore wind

**TABLE 3** Main parameters of the e-VMD-PCA-RELM model

| Parameter                                          | Value       |
|----------------------------------------------------|-------------|
| Modal value \( k \)                                | 9           |
| Modal number \( \alpha \)                         | 286         |
| Number of feature dimensions after PCA dimension reduction | 35         |
| Copy                                               | True        |
| Whiten                                             | False       |
| SVD solver                                         | Randomized  |
| RELM node parameter                               | \( 2^{12} \) |

**FIGURE 5** The results of e-VMD signal decomposition
FIGURE 6  The forecasting results of all training and testing sets: (A) the predicted results of IMF1; (B) the predicted results of IMF2; (C) the predicted results of IMF3; (D) the predicted results of IMF4; (E) the predicted results of Residual
power generation to the total power generation in the country will increase to important level. Considering that China’s offshore wind power installed capacity reached 3.630 GW by the end of 2018, according to the forecasting results, the next 15 years will be the peak stage of China’s offshore wind power development, and there is still room for offshore wind power development. Based on the results of this study, the targeted forward-looking conclusions are as follows:

### TABLE 4  The calculation results of five models

| Forecasting model | rmse   | $r^2$ (100%) | mre  |
|------------------|--------|--------------|------|
| RELM             | 520 787.78 | 96.71        | 0.220 |
| LSSVM            | 357 197.02  | 96.00        | 0.170 |
| PCA-RELM         | 389 062.91  | 96.92        | 0.121 |
| VMD-PCA-RELM     | 49 382.73   | 99.98        | 0.074 |
| e-VMD-PCA-RELM   | 23 462.59   | 99.99        | 0.047 |

### FIGURE 7  Comparison of prediction results for the 5 models; (A) forecasting results from 2000-2018; (B) Partial enlargement from 2000-2016

### FIGURE 8  Forecasting results of e-VMD-PCA-RELM forecasting model (2019-2035)
1. According to GM (1,1), this paper has achieved a gray forecasting value for offshore power generation costs. In the next 15 years, the comprehensive life rate of unit life will reach 19 297.37 hours, and the comprehensive average internal rate of return will reach 9.54%. The increase in internal rate of return is mainly due to the fact that wind power affiliated facility costs and management costs will gradually decrease with the enrichment of technology and management experience. Although the cost of offshore wind power generation is higher than that of onshore wind power generation, the cost of offshore wind power generation is much faster than that of onshore wind power, reflecting the stronger development prospects of offshore wind power.

2. While promoting the absorption of offshore wind power, the government needs to subsidize the on-grid tariff of offshore wind power to ensure that offshore wind power has priority for power generation, and the corresponding price policy can provide price support for future offshore wind power generation, which is beneficial to attract investors into the offshore wind power industry.

ACKNOWLEDGMENTS
The completion of this paper has been helped by many teachers and classmates. We would like to express our gratitude to them for their help and guidance.

REFERENCES
1. GWEC. Annual Report on the Status of Global Wind Power Industry in 2017 (with download). Internet Data Information Center-199IT; Chinese Internet Data Research Information Center-199IT; nd. http://www.199IT.com/archives/716405.html. Accessed 5.28.19.
2. Sun Y. Application of F-ANP in risk evaluation of offshore wind power projects, 2019. J Liaoning Tech Univ (Nat Sci). 2011;01:96-99.
3. Ma Y, Zhang A, Yang L, Hu C, Bai Y. Investigation on optimization design of offshore wind turbine blades based on particle swarm optimization. Energies. 2019;12:1972.
4. Jalbi S, Bhattacharya S. Minimum foundation size and spacing for jacket supported offshore wind turbines considering dynamic design criteria. Soil Dyn Earthq Eng. 2019;123:193-204.
5. Pokhrel J, Seo J. Natural hazard vulnerability quantification of offshore wind turbine in shallow water. Eng Struct. 2019;192:254-263.
6. Yao W, Deng C, Li D, Chen M, Peng P, Zhang H. Optimal sizing of seawater pumped storage plant with variable-speed units considering offshore wind power accommodation. Sustainability. 2019;11:1939.
7. Bhardwaj U, Teixeira AP, Soares CG. Reliability prediction of an offshore wind turbine gearbox. Renew Energy. 2019;141:693–706.
8. Qian X, Zhang Y, Gendeel M. State rules mining and probabilistic fault analysis for 5 MW offshore wind turbines. Energies. 2019;12:2046.
9. Yang H, Chen J, Pang X. Wind turbine optimization for minimum cost of energy in low wind speed areas considering blade length and hub height. Appl Sci. 2018;8:1202.
10. Jing L, Zhenzhen X, Xiaobo C. Risk assessment of operational period of offshore wind power project based on SVM. J Eng Manag. 2013;20:51-55.
11. Jing L, Yasheng S. Application of fuzzy network analysis in risk assessment of offshore wind power projects. J Liaoning Tech Univ Sci. 2011;08:96-99.
12. Celik AN. A techno-economic analysis of wind energy in southern turkey. Int J Green Energy. 2007;4:233-247.
13. Astariz S, Perez-Collazo C, Ahanades J, Iglesias G. Co-located wave-wind farms: economic assessment as a function of layout. Renew Energy. 2015;83:837-849.
14. Jiang FC, Guo YB, Ma QD. Comprehensive evaluation of the offshore wind power project navigation safety. Adv Mater Res. 2014;986-987:259-262.
15. Castro-Santos L, Filgueira-Vizoso A, Álvarez-Feal C, Carral L. Influence of size on the economic feasibility of floating offshore wind farms. Sustainability. 2018;10:4484.
16. Shiau T-A, Chuen-Yu J-K. Developing an indicator system for measuring the social sustainability of offshore wind power farms. Sustainability. 2016;8:470.
17. Salvador S, Costoya X, Sanz-Larruga FJ, Gimeno L. Development of offshore wind power: contrasting optimal wind sites with legal restrictions in Galicia, Spain. Energies. 2018;11:731.
18. Marjan A, Shafiee M. Evaluation of wind resources and the effect of market price components on wind-farm income: a case study of Ørland in Norway. Energies. 2018;11:2955.
19. Lian J, Cai O, Dong X, Jiang Q, Zhao Y. Health monitoring and safety evaluation of the offshore wind turbine structure: a review and discussion of future development. Sustainability. 2019;11:494.
20. Castro-Santos L, Martins E, Guedes Soares C. Cost assessment methodology for combined wind and wave floating offshore renewable energy systems. Renew Energy. 2016;97:866-880.
21. Rusu E. A 30-year projection of the future wind energy resources in the coastal environment of the Black Sea. Renew Energy. 2019;139:228-234.
22. Akdag SA, Dinler A. A new method to estimate Weibull parameters for wind energy applications. Energy Convers Manag. 2009;50:1761-1766.
23. Ganea D, Mereuta E, Rusu E. An evaluation of the wind and wave dynamics along the European coasts. J Mar Sci Eng. 2019;7:43.
24. Onea F, Rusu E. An evaluation of the wind energy in the northwest of the Black Sea. Int J Green Energy. 2014;11:465–487.
25. Kim H-G, Kim J-Y, Kang Y-H. Comparative evaluation of the third-generation reanalysis data for wind resource assessment of the southwestern offshore in South Korea. Atmosphere. 2018:9:73.
26. Onea F, Raileanu A, Rusu E. Evaluation of the wind energy potential in the coastal environment of two enclosed seas. Adv Meteorol. 2015;08:67-69.
27. Piasecka I, Tomporowski A, Flizikowski J, Kruzelniczka W, Kasner R, Mrzoński A. Life cycle analysis of ecological impacts of an offshore and a land-based wind power plant. Appl Sci. 2019;9:231.
28. Hasager CB, Mouche A, Badger M, et al. Offshore wind climateology based on synergetic use of Envisat ASAR, ASCAT and QuikSCAT. Remote Sens Environ. 2015;156:247-263.
29. Marx J, Schreiber A, Zapp P. Response to ‘Life-cycle green-house gas emissions of onshore and offshore wind turbines’. J Clean Prod. 2019;219:33-34.
30. Badger M, Badger J, Nielsen M, Hasager CB, Pena A. Wind class sampling of satellite SAR imagery for offshore wind resource mapping. J Appl Meteorol Climatol. 2010;49:2474-2491.
31. Baseer MA, Meyer JP, Rehman S, Alam MM. Wind power characteristics of seven data collection sites in Jubail, Saudi Arabia using Weibull parameters. Renew Energy. 2017;102:35-49.
32. Vahidzadeh M, Markfort CD. Modified power curves for prediction of power output of wind farms. Energies. 2019;12:1805.
33. Huan J, Cao W, Qin Y. Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. Comput Electron Agric. 2018;150:257-265.
34. Kim D, Hur J. Stochastic prediction of wind generating resources using the enhanced ensemble model for Jeju Island’s wind farms in South Korea. Sustainability. 2017;9:817.
35. Niu X, Wang J. A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. Appl Energy. 2019;241:519-539.
36. Li C, Tao Y, Ao W, Yang S, Bai Y. Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. Energy. 2018;165:1220-1227.
37. Muñiz C. Modeling algal atypical proliferation using the hybrid DE–MARS–based approach and M5 model tree in La Barca reservoir: a case study in northern Spain. Ecol Eng. 2019;130:198-212.
38. García-Nieto PJ, García-Gonzalo E, Alonso Fernández JR, Díaz Muñíz C. Modeling algal atypical proliferation using the hybrid DE–MARS–based approach and M5 model tree in La Barca reservoir: a case study in northern Spain. Ecol Eng. 2019;130:198-212.
39. Rodríguez F, Markou I, Pereira F. Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach. Inf Fusion. 2019;49:120-129.
40. García-Heller V, Espinasa R, Paredes S. Forecast study of the supply curve of solar and wind technologies in Argentina, Brazil, Chile and Mexico. Renew Energy. 2016;93:168-179.
41. Menglu L, Wei W, Gejirifu D, Xionghua J, Zhongfu T. Forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on Grey prediction theory and extreme learning machine optimized by support vector machine algorithm. Energies. 2018;11:2475.
42. Shao Z, Wakil K, Usak M, Amin Heidari M, Wang B, Simoes R. Kriging Empirical Mode Decomposition via support vector machine learning technique for autonomous operation diagnosing of CHP in microgrid. Appl Therm Eng. 2018;145:58-70.
43. Yue C-D, Liu C-C, Tu C-C, Lin T-H. Prediction of power generation by offshore wind farms using multiple data sources. Energies. 2019;12:700.
44. Zhang J, Teng Y-F, Chen W. Support vector regression with modified firefly algorithm for stock price forecasting. Appl Intell. 2019;49:1658-1674.

How to cite this article: Liu D, Liu M, Xu X, et al. Future prospects research on offshore wind power scale in China based on signal decomposition and extreme learning machine optimized by principal component analysis. Energy Sci Eng. 2020;8:3513–3530. https://doi.org/10.1002/ese3.761

APPENDIX A

| TABLE A1 | Index values of the main influencing factors |
|----------|-------------------------------------------|
| **Year** | **Annual average wind speed (m/s)** | **Average power generation (kW)** | **Offshore wind speed vertical shear** | **Average wind power density (W/m²)** | **Second industry electricity consumption (billion kWh)** | **Electricity consumption in the whole society (100 million kWh)** |
|----------|-----------------------------------------|---------------------------------|--------------------------------------|----------------------------------|--------------------------|----------------------------------|
| 2000    | 0.272                                   | 0.255                           | 0.300                                | 0.550                            | 0.550                     | 0.550                            |
| 2001    | 0.222                                   | 0.254                           | 0.238                                | 0.543                            | 0.531                     | 0.539                            |
| 2002    | 0.550                                   | 0.234                           | 0.050                                | 0.450                            | 0.504                     | 0.524                            |
| 2003    | 0.355                                   | 0.219                           | 0.363                                | 0.381                            | 0.463                     | 0.499                            |
| 2004    | 0.433                                   | 0.214                           | 0.550                                | 0.318                            | 0.435                     | 0.473                            |
| 2005    | 0.348                                   | 0.197                           | 0.425                                | 0.309                            | 0.407                     | 0.446                            |
| 2006    | 0.237                                   | 0.184                           | 0.363                                | 0.279                            | 0.371                     | 0.413                            |
| 2007    | 0.352                                   | 0.164                           | 0.425                                | 0.245                            | 0.328                     | 0.375                            |
| 2008    | 0.217                                   | 0.141                           | 0.488                                | 0.238                            | 0.315                     | 0.358                            |
| 2009    | 0.050                                   | 0.111                           | 0.550                                | 0.231                            | 0.300                     | 0.336                            |
| 2010    | 0.367                                   | 0.091                           | 0.425                                | 0.159                            | 0.246                     | 0.291                            |
| 2011    | 0.333                                   | 0.068                           | 0.363                                | 0.138                            | 0.198                     | 0.245                            |
| 2012    | 0.433                                   | 0.050                           | 0.300                                | 0.125                            | 0.181                     | 0.220                            |
| 2013    | 0.402                                   | 0.010                           | 0.425                                | 0.109                            | 0.150                     | 0.180                            |
| 2014    | 0.398                                   | −0.026                          | 0.363                                | 0.091                            | 0.132                     | 0.160                            |
| 2015    | 0.237                                   | −0.079                          | 0.425                                | 0.076                            | 0.128                     | 0.145                            |
| 2016    | 0.220                                   | −0.138                          | 0.488                                | 0.069                            | 0.114                     | 0.115                            |
| 2017    | 0.333                                   | −0.192                          | 0.363                                | 0.056                            | 0.085                     | 0.098                            |
| 2018    | 0.352                                   | −0.245                          | 0.300                                | 0.050                            | 0.050                     | 0.050                            |
**APPENDIX B**

The specific calculation steps of the gray correlation analysis are as follows:

**Step 1: Determine the analysis series.**

Determine the reference and the comparison series that reflects the behavior of the system series. A sequence of data that reflects the behavioral characteristics of the system, called the reference sequence. A sequence of data that is a component of factors affecting system behavior, called a comparison series. Assume that the reference number is \( Y = \{ y(k) \mid k = 1, 2, \ldots n \} \), and compare series as

\[
X_i = \{ x_i(k) \mid k = 1, 2, \ldots n \}, i = 1, 2, \ldots, m
\]

**Step 2: Dimensionlessness of variables**

Since the data in the various factor columns in the system may be different in dimension, it is not convenient to compare or it is difficult to get a correct conclusion when comparing. Therefore, in the gray correlation analysis, the data are generally dimensionless:

\[
x_i(k) = \frac{X_i(k)}{X_i(l)}, k = 1, 2, \ldots n; i = 1, 2, \ldots, m
\]

**Step 3: Calculate correlation coefficient**

Calculate correlation coefficient of \( x_0(k) \) and \( x_i(k) \):

\[
\xi_i(k) = \frac{\min \min | y(k) - x_i(k) | + \rho \max \max | y(k) - x_i(k) |}{| y(k) - x_0(k) | + \rho \max \max | y(k) - x_0(k) |}
\]

\((B3)\)

Let \( \Delta_i(k) = | y(k) - x_i(k) | \), and then

\[
\xi_i(k) = \frac{\min \min \Delta_i(k) + \rho \max \max \Delta_i(k)}{\Delta_i(k) + \rho \max \max \Delta_i(k)}
\]

\((B4)\)

\(\rho \in (0, \infty)\) is called the resolution coefficient. The smaller \(\rho\), the greater the resolution. Generally, the value range of \(\rho\) is \((0, 1)\), and the specific value may depend on the situation. When \(\rho \leq 0.55\), the resolution is best, and we take \(\rho = 0.55\) in this paper.

**Step 4: Calculate relevance**

Because the correlation coefficient is the degree of correlation between the comparison series and the reference sequence at each moment (ie, each point in the curve), it has more than one number, and the information is too scattered to

| Year | Wind speed (m/s) | Annual average Index values of the main influencing factors | Average power generation (10000 kW) | Average unit life (h) | Average internal rate of return | Average net profit |
|------|-----------------|-------------------------------------------------------------|--------------------------------------|-----------------------|---------------------------------|-------------------|
| 2010 | 0.367           | 0.091                                                       | 0.425                                | 0.159                 | 0.246                           | 0.291             |
| 2008 | 0.217           | 0.141                                                       | 0.488                                | 0.238                 | 0.315                           | 0.358             |
| 2007 | 0.352           | 0.164                                                       | 0.425                                | 0.245                 | 0.328                           | 0.375             |
| 2006 | 0.237           | 0.184                                                       | 0.363                                | 0.279                 | 0.371                           | 0.413             |
| 2005 | 0.348           | 0.197                                                       | 0.425                                | 0.309                 | 0.407                           | 0.446             |
| 2004 | 0.433           | 0.214                                                       | 0.550                                | 0.318                 | 0.435                           | 0.473             |
| 2003 | 0.355           | 0.219                                                       | 0.363                                | 0.381                 | 0.463                           | 0.499             |
| 2002 | 0.550           | 0.234                                                       | 0.050                                | 0.450                 | 0.504                           | 0.524             |
| 2001 | 0.222           | 0.254                                                       | 0.238                                | 0.543                 | 0.531                           | 0.539             |
| 2000 | 0.272           | 0.255                                                       | 0.300                                | 0.550                 | 0.550                           | 0.550             |
| 2013 | 0.402           | 0.010                                                       | 0.425                                | 0.109                 | 0.114                           | 0.115             |
| 2014 | 0.237           | 0.079                                                       | 0.425                                | 0.076                 | 0.128                           | 0.145             |
| 2015 | 0.237           | 0.079                                                       | 0.425                                | 0.076                 | 0.128                           | 0.145             |
| 2016 | 0.220           | 0.138                                                       | 0.488                                | 0.069                 | 0.114                           | 0.115             |

2018 0.352 −0.245 0.300 0.050 0.050 0.050
2017 0.333 −0.192 0.363 0.056 0.085 0.098
2016 0.220 −0.138 0.488 0.069 0.114 0.115
2015 0.237 −0.079 0.425 0.076 0.128 0.145
2014 0.398 −0.026 0.363 0.091 0.132 0.160
2013 0.402 0.010 0.425 0.109 0.150 0.180
2012 0.433 0.050 0.300 0.125 0.181 0.220
2010 0.367 0.091 0.425 0.159 0.246 0.291
2008 0.217 0.141 0.488 0.238 0.315 0.358
2007 0.352 0.164 0.425 0.245 0.328 0.375
2006 0.237 0.184 0.363 0.279 0.371 0.413
2005 0.348 0.197 0.425 0.309 0.407 0.446
2004 0.433 0.214 0.550 0.318 0.435 0.473
2003 0.355 0.219 0.363 0.381 0.463 0.499
2002 0.550 0.234 0.050 0.450 0.504 0.524
2001 0.222 0.254 0.238 0.543 0.531 0.539
2000 0.272 0.255 0.300 0.550 0.550 0.550
facilitate the overall comparison. Therefore, it is necessary to concentrate the correlation coefficients at various moments (ie, points in the curve) into one value, that is, to find the average value thereof, as the quantity representation of the degree of correlation between the comparison series and the reference series. The formula of relevance \( r_i \) is as follows:

\[
r_i = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k), \quad k = 1, 2, \ldots, n
\]  

(B5)

Step 5: Relevance ranking

Sort the relevance by size. If \( r_1 < r_2 \), the reference sequence \( y \) is similar to the comparison sequence \( X_2 \). After calculating the correlation coefficient between the \( X_i(k) \) sequence and the \( Y(k) \) sequence, the average value of each type of correlation coefficient is calculated, and the average value \( r_i \) is called the degree of association between \( Y(k) \) and \( X_i(k) \).

The mathematical model of GM(1,1) forecasting model is shown:

The specific calculation steps of the gray model (1,1) are as follows:

Assuming a raw sequence:

\[
x^0 = (x_1^0, x_2^0, \ldots, x_n^0)
\]  

(B6)

where \( x_k^0 \geq 0, k = 1, 2, \ldots, n \)

Adding raw arrays:

\[
x^1 = (x_1^1, x_2^1, \ldots, x_n^1)
\]  

(B7)

where \( x_k^1 = \sum_{i=1}^{k} x_i^0 \)

Make an equal-valued equal-weight generation sequence of \( x^1 \):

\[
z^1 = (z_1^1, z_2^1, \ldots, z_n^1)
\]  

(B8)

where \( z_k^1 = \frac{1}{2}(x_k^1 + x_k^1_{(k-1)}) \)

Therefore, the original differential equation of GM(1,1) forecasting model is:

\[
x_k^0 + az_k^1 = b
\]  

(B9)

where \( a \) is a development coefficient and \( b \) is ash action, so the whitening differential equation of Equation (B9) is:

\[
\frac{dx^1}{dt} + ax^1 = b
\]  

(B10)

The least-squares parameter estimates for \( a \) and \( b \) are as follows:

\[
\hat{a} = (a, b)^T = (B^T, B)^{-1}B^TY
\]  

(B11)

where:

\[
B = \begin{bmatrix}
-z_2^1 & 1 \\
-z_3^1 & 1 \\
\vdots & \vdots \\
-z_n^1 & 1
\end{bmatrix}, Y = \begin{bmatrix}
x_2^0 \\
x_3^0 \\
\vdots \\
x_n^0
\end{bmatrix}
\]  

(B12)

The time response sequence of the differential Equation (B9) in the GM(1,1) model is:

\[
\hat{x}^1_{k+1} = \left(x_1^1 - \frac{b}{a}\right)e^{(-ak)} + \frac{b}{a}
\]  

(B13)

Similarly, the whitening response function of the whitening equation is:

\[
\hat{x}^1_{k+1} = \left(x_1^1 - \frac{b}{a}\right)e^{(-ak)} + \frac{b}{a}
\]  

(B14)

Substituting initial conditions \( \hat{x}^1_1 = x_1^1 = x_1^0 \) to get:

\[
\hat{x}^1_{k+1} = \left(x_0^1 - \frac{b}{a}\right)e^{(-ak)} + \frac{b}{a}
\]  

(B15)

The GM(1,1) forecasting model of the original sequence \( x^0 \) can be obtained according to the cumulative reduction by Equation (B13):