J Electr Eng Technol.2018; 13(5): 1841-1851
http://doi.org/10.5370/JEET.2018.13.5.1841

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Short-term Wind Power Prediction Based on Empirical Mode Decomposition and Improved Extreme Learning Machine

Zhongda Tian†, Yi Ren* and Gang Wang*

Abstract – For the safe and stable operation of the power system, accurate wind power prediction is of great significance. A wind power prediction method based on empirical mode decomposition and improved extreme learning machine is proposed in this paper. Firstly, wind power time series is decomposed into several components with different frequency by empirical mode decomposition, which can reduce the non-stationary of time series. The components after decomposing remove the long correlation and promote the different local characteristics of original wind power time series. Secondly, an improved extreme learning machine prediction model is introduced to overcome the sample data updating disadvantages of standard extreme learning machine. Different improved extreme learning machine prediction model of each component is established. Finally, the prediction value of each component is superimposed to obtain the final result. Compared with other prediction models, the simulation results demonstrate that the proposed prediction method has better prediction accuracy for wind power.

Keywords: Short-term wind power, Prediction, Empirical mode decomposition, Improved extreme learning machine.

1. Introduction

With the rapid development of the world economy, the corresponding demand for energy has also increased greatly, and traditional fossil energy faces the threat of energy exhaustion. At the same time, caused by a mass of traditional fossil energy consumption and climate warming, as well as the increasingly serious problem of environmental pollution is becoming more and more serious, the ecological system, economy and human health are posing a serious threat [1]. In order to cope with the shortage of traditional fossil energy and the environmental pollution caused by traditional fossil energy, wind power has become the research direction of many countries [2]. However, the rapid development of wind energy also faces some serious problems [3]. Due to the instability of the wind, wind power has volatility and intermittency, which will seriously affect the power grid make a reasonable scheduling of wind power. Therefore, the accurate prediction of wind power is very meaningful to the power grid [4].

From time scales, wind power prediction includes short-term, medium-term, and long-term prediction. The time scale for short-term prediction is several hours to several days in advance. Short-term prediction can help the grid to carry out reasonable economic scheduling, unit combination operation and choose the right time to maintain the fan. The time scale for medium-term prediction is several days to several months in advance. The results of medium-term wind power prediction can help wind farms to make quarterly power generation plans and arrange maintenance activities. The time scale for long-term prediction is several months to several years in advance. Long-term wind power prediction can be used to evaluate the potential annual power generation of a region, mainly in the location of wind farms. Due to the complexity of the wind speed and other weather conditions in the short-term scale, the prediction accuracy of short-term wind power is more difficult to be guaranteed [5]. This paper focuses on the prediction of short-term wind power.

At present, the common wind power forecast includes numerical weather prediction (NWP) method and statistical method [6]. NWP prediction method requires accurate NWP data and detailed physical information on the area where the wind farm is located. At the same time, NWP data will be updated once every few hours. So NWP prediction method is more suitable for medium-term and long-term wind power prediction. It is not suitable for short-term wind power prediction. The prediction accuracy of NWP method relies on a high degree of accuracy and popularity of the NWP data. Meanwhile, NWP method requires a lot of parameters such as wind speed, wind direction, air pressure, etc. In many areas, NWP system has not been established. NWP data is affected by the environment and physical factors. At the same time, the wind power is proportional to the cube of the wind speed, the prediction error will be will be amplified. These factors greatly limit the prediction accuracy of NWP prediction

† Corresponding Author: College of Information Science and Engineering, Shenyang University of Technology, China. (tianzhongda@126.com)
* College of Information Science and Engineering, Shenyang University of Technology, China. ({940261475, 1365812827}@qq.com)
Received: December 15, 2017; Accepted: June 7, 2018

model of wind power [7].

The statistical methods include time series prediction method [8, 9], neural network prediction model [10-12], Kalman filtering prediction model [13], fuzzy clustering prediction model [14, 15], auto regressive moving average (ARMA) prediction model [16], support vector machine (SVM) [17, 18], least square support vector machine (LSSVM) [19, 20], etc,. According to the intrinsic statistical characteristics and development rules of system development, through the historical data of the past and statistical analysis, time series analysis can further speculate the future development trend of the system. Through the study of time series to understand the structural characteristics of the system (such as the cycle of fluctuations, amplitude, the type of trends, etc.); reveal its operating rules, and then used to predict and control their future behavior. Through the study of time series can grasp the nature of the system, so as to achieve the future prediction. However, time series analysis methods have different prediction results with the different model order. Although the neural network has good robustness, generalization and fault tolerance ability, its learning convergence speed is slow, easy to fall into local optimal value and difficult to achieve the global optimal value. Kalman filtering model is difficult to obtain statistical properties of noise. ARMA is only suitable for the prediction of a linear sequence. For wind power with the nonlinear characteristics; the prediction accuracy is difficult to be guaranteed. At the same time, the fuzzy rules in the fuzzy clustering model are too difficult to be determined. As a learning method based on statistical learning theory, SVM and LSSVM need very few samples and have strong generalization ability. But the optimal parameters of SVM or LSSVM are difficult to be determined. The inappropriate parameters directly affect the prediction effect of SVM or LSSVM. Therefore, these prediction methods based on statistical models also have their specific shortcomings.

At present, the wind power prediction accuracy obtained by single prediction method is low. The main reason is due to the intermittency, uncertainty of the wind speed and the limitations of the prediction methods. Therefore, the single prediction model has been unable to meet the needs of prediction accuracy. Combination prediction model is one of the development directions of wind power prediction [21]. According to the characteristics of each model, the combination prediction model establishes the combination model through the idea of complementary advantages, and improves the prediction accuracy.

In 2006, Huang et al proposed a new neural network - extreme learning machine extreme (ELM) algorithm [22]. The ELM algorithm uses the random mechanism to reduce the parameter setting and choice. It is one kind of simple feasible fast learning algorithm. Compared with other traditional neural network learning algorithms, SVM or LSSVM, etc., ELM algorithm has the advantages of fast learning speed and strong generalization ability [23]. The literature [24] points out that the computing time of ELM is usually several thousand times faster than BP neural network or SVM. Therefore, the ELM algorithm has been applied to prediction of many time series including wind power [25, 26]. But in the ELM algorithm, the value of the new and old training samples is equal, and the same weight is given to them, which can not highlight the role of the new training samples. The updating model of the network weights is not flexible, and it is easy to increase unnecessary computation. In this paper, the forgetting factor is introduced to weaken the influence of the old training samples. At the same time, based on the generalization ability, the output weights are selectively updated to improve the accuracy and speed of the algorithm.

Based on the above discussion, according to the nonlinear and non-stationary characteristics of wind power series, a short-term combined prediction model of wind power based on empirical mode decomposition (EMD) and improved ELM is proposed in this paper. The fluctuations or trends of different scales in the wind power series are decomposed. Then, according to the characteristics of each component, different improved ELM models are established, and the most appropriate model is determined. Each component is predicted, and the prediction results of each component are superimposed to obtain the final wind power prediction value. Simulation results show that the prediction method proposed in this paper is effective.

The main contents of this paper are as follows. Section 2 introduces the preliminaries. Section 3 introduces the detailed implementation process of the proposed prediction method. The simulation results are provided in Section 4. The summary and prospects of the paper are summarized in Section 5.

2. The Preliminaries

2.1 The characteristics of wind power

In this study, the actual wind power data from a wind power plant in Liaoning Province, China is chosen. The data sets have 2400 group data. The sampling period of data is 1 hour. The wind power time series is as shown in Fig. 1. In order to observe the details of the wind power sequence, the first 100 sets of data are displayed separately in Fig. 2. It can be observed from Fig. 1 and Fig. 2, the fluctuation of wind power series is more intense, and a large number of wind power values have risen sharply and drastically declined. At the same time, the wind power difference between the two adjacent sampling points is very large. Wind power presents stochastic, non-stationary and nonlinear characteristics. Therefore, wind power has a very high demand for the prediction performance of the prediction model. It is necessary to choose suitable

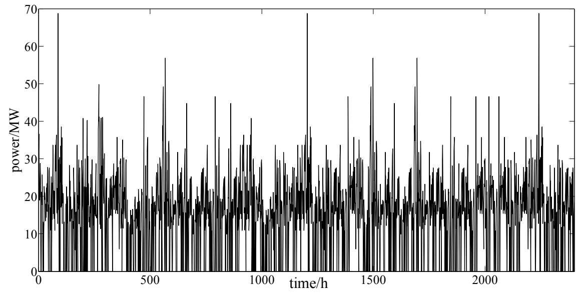

Fig. 1. Short-term wind power time series

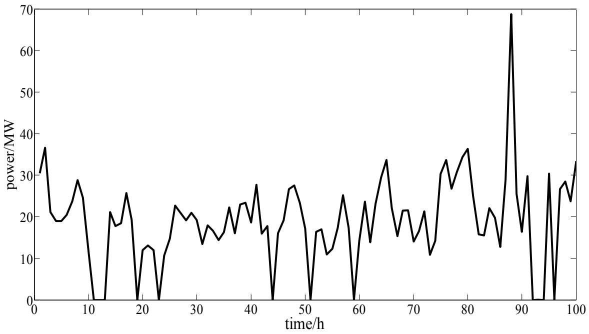

Fig. 2. The first 100 sets of short-term wind power time series

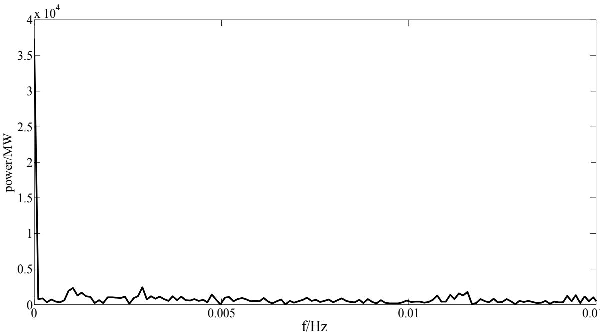

Fig. 3. Amplitude-frequency characteristic of wind power time series

forecasting models and forecasting methods for wind power prediction.

In order to analyze the characteristic of short-term wind power time series, this paper firstly analyzes the amplitude-frequency characteristics of wind power. For the wind power time series, fast Fourier transform method is used to transform the time domain to the frequency domain, and the result of amplitude-frequency characteristic is shown in Fig. 3.

Fig. 3 shows that short-term wind power time series have typical multi-time scale characteristics. The energy of short-term wind power time series is concentrated in the direct-current and low frequency parts (0 ~ 0.015 Hz). The energy of the high frequency part of the short-term wind power is very low. This means that the fluctuation

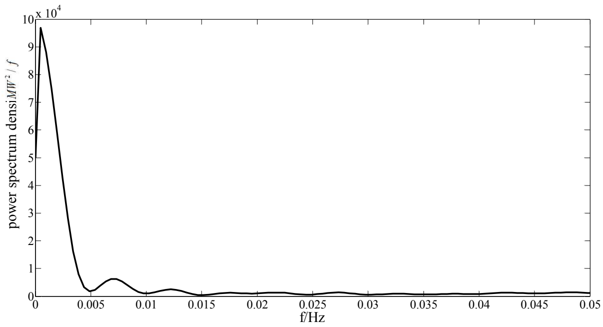

Fig. 4. The power spectrum density of short-term wind power time series

information caused by external factors, or the periodic information of the wind power system itself, is contained in the short-term wind power sequence with different frequency variations.

In order to further analyze the characteristics of short-term wind power, the power spectrum density is used to analyze the sequence. The power spectrum density of short-term wind power is as shown in Fig. 4.

It can be seen from Fig. 4, although the wind power is random, but based on the above analysis results, it can be determined that the wind power can be decomposed into many components with different frequency. At the same time, most of the wind power is slow change. The short-term wind power signal is nonlinear and non-stationary. Some decomposition methods can reduce the non-stationary of the signal and extract the information of different frequency signals. On this basis, combined prediction method can often improve the prediction effect. Therefore, the short-term wind power time series is processed first, and the components reflecting the different information of the sequence are resolved. Then, it is a feasible way to improve the prediction accuracy by establishing appropriate prediction models for different components. As a new target data analysis method, EMD is especially suitable for the decomposition and processing of nonlinear and non-stationary data. EMD adaptively decomposes the local feature information according to the signal itself, and does not need to set the parameters in advance, so it overcomes the problem of depending on the operator's subjective experience. Therefore, EMD method is used to decompose the wind power signal into high frequency component and low frequency component. The different components of the high and low frequencies are predicted by the appropriate improved extreme learning machine model. This reduces the influence of the instability of the wind power sequence on the prediction accuracy.

2.2 Empirical mode decomposition

EMD is a new signal processing method proposed by

Huang et al in 1998 [27]. Its essence is to smooth the time series, decompose the original signal into several intrinsic mode functions (IMF), and determine the basic intrinsic characteristics of the effective signals in the data according to the experiences. It can more effectively reflect the distribution of energy on the space (or time) scale in the physics process [28]. Through EMD decomposition, each IMF component of non-stationary complex signals is stable. Therefore, EMD is an effective method for non-stationary signal component decomposition. The low frequency component of IMF usually represents the original trend of signal. These IMFs need to meet the following conditions.

1. The number of extrema and the number of zero crossings must be equal or up to a difference of one.
2. The mean value of the envelope defined by the local maximum and the local minimum of the signal is zero at any point.

The main process of EMD is as follows.

Step 1: The two envelope curves of $x(t)$ are obtained by using two three spline curves of all maximum and minimum values of the original signal sequence. The mean value of two envelopes is calculated, which can be expressed as $m(t)$. Let $h(t)=x(t)-m(t)$.

Step 2: If $h(t)$ does not meet the requirements of the IMF, repeat Step 1 to calculate the new $h(t)$. If $h(t)$ satisfies the requirements of IMF, then $h(t)$ is the first IMF of $x(t)$, and the difference $r(t)$ between $x(t)$ and the IMF is obtained.

Step 3: $r(t)$ is used as the signal to be decomposed, and the process is repeated until the residual signal satisfies the given termination condition ($r(t)$ is small enough or is a monotonic function).

The final results of EMD can be expressed as.

$$x(t)=\sum_{i=1}^{n} c_i(t)+r_n(t) \tag{1}$$

where, $c_i(t)$ is the *ith* IMF component, r_n is residual component. Therefore, empirical mode decomposition can decompose the original signal $x(t)$ into the sum of n different frequencies of IMF and a trend item.

2.3 Improved extreme learning machine

The structure of ELM neural network algorithm is similar with the single-hidden layer feed forward neural networks (SLFNs). The different is ELM can randomly select its training parameters, and get a complete network training model only through the output weights obtained by the least squares method.

For a given training set $D=\{(\mathbf{x}_i,\mathbf{t}_i)\,|\,\mathbf{x}_i\in\mathbf{R}^n,\mathbf{t}_i\in\mathbf{R}^m,i=1,\dots,N\}$, activation function $f(x)$, the number of hidden nodes L. ELM regression model can be expressed as

$$\begin{aligned}&\sum_{i=1}^{L}\beta_i f\left(\mathbf{a}_i\mathbf{x}_1+b_i\right)=t_1\\&\sum_{i=1}^{L}\beta_i f\left(\mathbf{a}_i\mathbf{x}_2+b_i\right)=t_2\\&\quad\vdots\\&\sum_{i=1}^{L}\beta_i f\left(\mathbf{a}_i\mathbf{x}_k+b_i\right)=t_k\end{aligned} \tag{2}$$

where, $\mathbf{a}_i, i=1,\dots,L$ is the output weights, $b_i, i=1,\dots,L$ is bias, k is the numbers of the sample. Eq. (2) is rewritten as

$$\mathbf{T}_k=\mathbf{H}_k\boldsymbol{\beta}_k \tag{3}$$

where $\mathbf{H}_k$ is a neuron matrix and can be represented as

$$\mathbf{H}_k=\begin{bmatrix} f(\mathbf{a}_1\mathbf{x}_1+b_1) & \cdots & f(\mathbf{a}_L\mathbf{x}_1+b_L)\\ f(\mathbf{a}_1\mathbf{x}_2+b_1) & \cdots & f(\mathbf{a}_L\mathbf{x}_2+b_L)\\ \vdots & \ddots & \vdots\\ f(\mathbf{a}_1\mathbf{x}_k+b_1) & \cdots & f(\mathbf{a}_L\mathbf{x}_k+b_L)\end{bmatrix} \tag{4}$$

$\boldsymbol{\beta}_k$ is the output weights and can be expressed as

$$\boldsymbol{\beta}_k=\begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_k\end{bmatrix}^{\mathrm{T}} \tag{5}$$

$\mathbf{T}_k$ is the output weights and can be expressed as

$$\mathbf{T}_k=\begin{bmatrix}t_1 & t_2 & \cdots & t_k\end{bmatrix}^{\mathrm{T}} \tag{6}$$

The output weights can be obtained by solving Eq. (3).

$$\boldsymbol{\beta}_k=\left(\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k\right)^{-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{T}_k \tag{7}$$

Therefore, time series prediction model based on ELM can be obtained after training

$$t=\sum_{i=1}^{L}\beta_i f\left(\mathbf{a}_i\mathbf{x}+b_i\right) \tag{8}$$

However, ELM considers that the value of the new and old training samples is equal, and the equal weight of the training sample fails to highlight the role of the new training sample. Moreover, as soon as the new training samples are obtained, ELM updates the weights of the network. In order to solve this problem in ELM, this paper proposes an improved ELM with more effective sample updating mechanism. This paper considers that the new sample should be added to the training set after the initial network weights are calculated, and the corresponding network weights can be obtained on the basis of the initial network weights. At the same time, the new and old samples are given different weights. The effect of new training samples on the algorithm is further enhanced, which can further improve the regression prediction ability of ELM. The improved ELM algorithm can be described as follows.

Suppose that $\boldsymbol{\beta}_k$ in Eq. (7) is calculated by sampling sample $(\mathbf{x}_1,\mathbf{t}_1),(\mathbf{x}_2,\mathbf{t}_2),\cdots(\mathbf{x}_k,\mathbf{t}_k)$. When new sample $(\mathbf{x}_{k+1},\mathbf{t}_{k+1})$ is added into set, then $\boldsymbol{\beta}_{k+1}$ can be expressed as

$$\begin{aligned}\boldsymbol{\beta}_{k+1} &= \left(\begin{bmatrix}\mathbf{H}_k\\ \mathbf{h}_{k+1}\end{bmatrix}\begin{bmatrix}\mathbf{H}_k\\ \mathbf{h}_{k+1}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{H}_k\\ \mathbf{h}_{k+1}\end{bmatrix}\begin{bmatrix}\mathbf{T}_k\\ t_{k+1}\end{bmatrix}\\ &= (\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k + h_{k+1}^{\mathrm{T}}h_{k+1})^{-1}(\mathbf{H}_k^{\mathrm{T}}\mathbf{T}_k + \mathbf{h}_{k+1}^{\mathrm{T}}t_{k+1})\end{aligned} \tag{9}$$

where $\mathbf{h}_{k+1} = \left[f(\mathbf{a}_1\mathbf{x}_{k+1}+b_1)\ f(\mathbf{a}_2\mathbf{x}_{k+1}+b_2)\cdots f(\mathbf{a}_L\mathbf{x}_{k+1}+b_L)\right]$.

$\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k$ and $\mathbf{H}_k^{\mathrm{T}}\mathbf{T}_k$ are given weights, the Eq. (9) can be rewritten as

$$\boldsymbol{\beta}_{k+1} = (\mu \mathrm{H}_k^{\mathrm{T}}\mathbf{H}_k + \mathbf{h}_{k+1})^{-1}(\mu\mathbf{H}_k^{T}\mathbf{T}_k + \mathbf{h}_{k+1}^{\mathrm{T}}t_{k+1}) \tag{10}$$

where μ , $0<\mu<1$ is weight coefficient. Let

$$\mathbf{P}_{k+1} = (\mu\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k + \mathbf{h}_{k+1}^{\mathrm{T}}\mathbf{h}_{k+1})^{-1} \tag{11}$$

The inverse of Eq. (11) can be obtained

$$\mathbf{P}_{k+1}^{-1} = \mu\mathbf{P}_k^{-1} + \mathbf{h}_{k+1}^{\mathrm{T}}\mathbf{h}_{k+1} \tag{12}$$

Eq. (12) is substituted into Equation (10), the next can be obtained.

$$\begin{aligned}\boldsymbol{\beta}_{k+1} &= \mathbf{P}_{k+1}(\mu\mathbf{H}_k^{\mathrm{T}}\mathbf{T}_k + \mathbf{h}_{k+1}^{\mathrm{T}}t_{k+1}) = \mathbf{P}_{k+1}(\mu\mathbf{P}_k^{-1}\boldsymbol{\beta}_k + \mathbf{h}_{k+1}^{\mathrm{T}}t_{k+1})\\ &= \mathbf{P}_{k+1}((\mathbf{P}_{k+1}^{-1} - \mathbf{h}_{k+1}^{\mathrm{T}}h_{k+1})\boldsymbol{\beta}_k + \mathbf{h}_{k+1}^{\mathrm{T}}t_{k+1})\\ &= \boldsymbol{\beta}_k + \mathbf{P}_{k+1}\mathbf{h}_{k+1}^{\mathrm{T}}(t_{k+1} - \mathbf{h}_{k+1}\boldsymbol{\beta}_k)\end{aligned} \tag{13}$$

When a new sample x_{k+1} is obtained, it is necessary to judge the change trend of the error. When the error value is greater than a threshold value ε , $\mathbf{P}_k$ is updated, otherwise $\mathbf{P}_k$ remains unchanged. The update mechanism is shown in the following formula.

$$\mathbf{P}_{k+1} = \begin{cases}(\mu\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k + \mathbf{h}_{k+1}^{\mathrm{T}}\mathbf{h}_{k+1})^{-1} & (E_N > \varepsilon)\\ \mathbf{P}_k & (E_N \le \varepsilon)\end{cases} \tag{14}$$

where, $E_N=\sqrt{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2/N}$, $\hat{x}_i$ is the predictive value of x_i .

To sum up, the implementation steps of improved ELM algorithm are as follows.

Step 1: The embedding dimension m of data samples is determined. The initial N sample data is $x_1,x_2,\cdots x_N$ is transformed into training sample set $(\mathbf{x}_1,\mathbf{t}_1),(\mathbf{x}_2,\mathbf{t}_2),\cdots$ $(\mathbf{x}_k,\mathbf{t}_k)$. $\mathbf{x}_i = \left[x_i\ x_{i+1}\cdots x_{i+m-1}\right]^{\mathrm{T}}$ is chosen as input, $t_i = x_{i+m}$ is chosen as output. where $k=N-m>L$.

Step 2: The initial output weight is calculated according to the next formula.

$$\boldsymbol{\beta}_k = \mathbf{P}_k\mathbf{H}_k^{\mathrm{T}}\mathbf{T}_k \tag{15}$$

Step 3: $\mathbf{x}_{k+1} = \left[x_{k-m+1}\ x_{k-m+2}\cdots x_k\right]^{\mathrm{T}}$ is chosen as input, input vector $\mathbf{h}_{k+1}$ is calculated. Then, one step prediction value of x_{k+1} is obtained.

$$\hat{x}_{k+1} = \mathbf{h}_{k+1}\boldsymbol{\beta}_k \tag{16}$$

Step 4: The matrix $\mathbf{P}_k$ is updated according to Eq. (14). Then, $\boldsymbol{\beta}_k$ is updated according to Eq. (13). Let k=k+1. Go to Step 3.

3. The Implementation Process of the Proposed Prediction Method

Fig. 5 is the structure of wind power prediction method proposed in this paper. The input wind power time series is decomposed by EMD to generate n IMF components and a residual component. IMF components and residual component are used as input training sets to model improved ELM prediction model respectively. When the model is established, the wind power in the future can be predicted. In particular, one step prediction mechanism is adopted in this paper.

The implementation process of the proposed prediction method can be described as the following.

Step 1: The parameters are initialized. These parameters include maximum number of hidden layer nodes L, activation function $f(x)$, embedding dimension of samples m, weight coefficient μ , error threshold ε and etc.

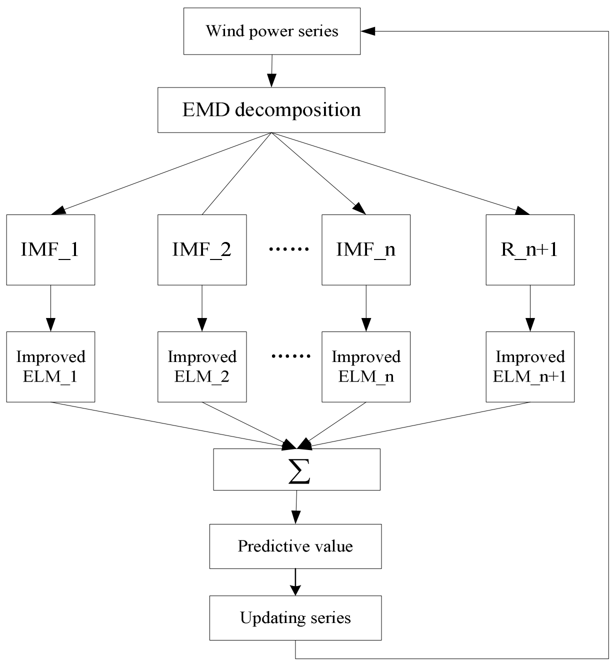

Fig. 5. The structure of wind power prediction method proposed in this paper

Step 2: The original wind power training sample time series is $w_1, w_2, \cdots, w_k$. It is decomposed by EMD, *n* IMF component and one residual component is obtained. The decomposed components can be expressed as $x_1^j, x_2^j, \cdots, \ x_k^j$, $j = 1, 2 \cdots, n+1$, j represents the number of the components. $\mathbf{x}_i^j = \left[x_i^j \ x_{i+1}^j \cdots x_{i+m-1}^j \right]^{\mathrm{T}}$ is chosen as input, $t_i^j = x_{i+m}^j$ is chosen as output. After training, *n*+1 improved ELM prediction models can be obtained.

Step 3: Suppose that the current sampling numbers of wind power time series are *k*, that is $w_1, w_2, \cdots, w_k$. $w_1, w_2, \cdots, w_k$ is decomposed by EMD. The decomposed components can be expressed as $x_1^j, x_2^j, \cdots, x_k^j$, $j = 1, 2 \cdots$, $n+1$, j represents the number of the components. $\mathbf{x}_{k+1}^j = \left[x_{k-m+1}^j \ x_{k-m+2}^j \cdots x_k^j \right]^{\mathrm{T}}$ is chosen as input, then predictive value $\hat{x}_{k+1}^j$ is obtained. Each predictive value $\hat{x}_{k+1}^j$ is added by the following Eq. (17), and then the final predictive value $\hat{w}_{k+1}$ is obtained.

$$\hat{w}_{k+1} = \hat{x}_{k+1}^1 + \hat{x}_{k+1}^2 + \cdots \hat{x}_{k+1}^{n+1} \tag{17}$$

Step 4：Updating wind power input time series. w_1 is removed. $\hat{w}_{k+1}$ is treated as the actual value and is inserted at the end of the wind power series. The length of wind power is still *k*. Go to Step 3, until the maximum prediction step is reached.

4. Simulation

The 2400 group wind power data is used as simulation data. The former 1920 group data is chosen as the training set. The latter 480 group data is chosen as the testing set. Through EMD, 1920 groups wind power data can be divided into 9 IMF components and 1 residual component. Fig. 6 shows the components after EMD decomposition. It can be seen from Fig. 6, except IMF 1, IMF 2 and IMF 3, the change of other components is relatively stable. Decomposition reduces the interference and coupling between different characteristic information. Therefore, EMD reduces the difficulty of prediction modeling and improves the prediction accuracy. Fig. 7 shows the autocorrelation function of the original wind power time series. Fig. 8 shows the autocorrelation function of components after EMD decomposition. It can be seen from the Fig. 7 that the autocorrelation function of the original wind power time series has not been reduced to 0 in all time. However, the results in the Fig. 8 show that the autocorrelation function of each IMF component after EMD quickly attenuates to zero for the first time. It indicates that the IMF components after EMD decomposition have the short correlation property. EMD can reduce the self-similarity of wind power time series. Compared with the long correlation model, the short correlation model has low complexity, so it can reduce the complexity of prediction after EMD processing.

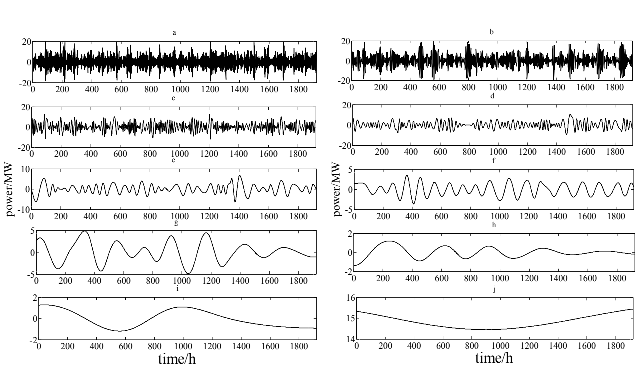

Fig. 6. The components after EMD decomposition (a: IMF1; b: IMF2; c: IMF3; d: IMF 4; e: IMF 5; f: IMF 6; g: IMF 7; h: IMF 8; i: IMF9; j: Residual)

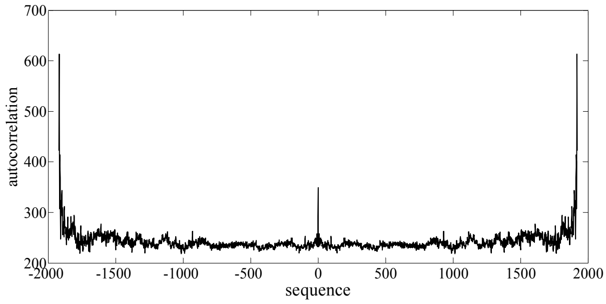

Fig. 7. The autocorrelation function of the original wind power time series

1920 group data of wind power after EMD decomposition is used to model 10 improved ELM prediction models. The activation function is chosen as Sigmoid function. The data embedding dimension *m* is determined as 48. The number of neurons in the hidden layer *L* is set to 100. The weight coefficient μ is set as 0.95. The threshold value ε is chosen as 0.2. Fig. 9 shows a comparison between the 480 groups of prediction value and the actual value of each IMF component and residual of the wind power time series after EMD process. The prediction values of each component are very close to the actual values. Each prediction model has a good prediction effect.

Fig. 10 shows the root square mean error (RMSE) of each component. From the Fig. 10, it can be seen that with the increase of the order of each component, RMSE decreases rapidly and the prediction accuracy becomes higher. The reason is that the randomness of the sequence decreases with the increase of the order of IMF, so the accuracy of prediction is improved.

When the prediction values of each component are obtained, the final prediction value of wind power can be obtained by superposition of the prediction values of each component. Fig. 11 gives the comparison of final prediction value and actual value of 480 groups wind power.

In order to further verify the prediction accuracy, the proposed method is compared with echo state network in [12], ARMA in [16], SVM in [18], LSSVM in [19] and ELM

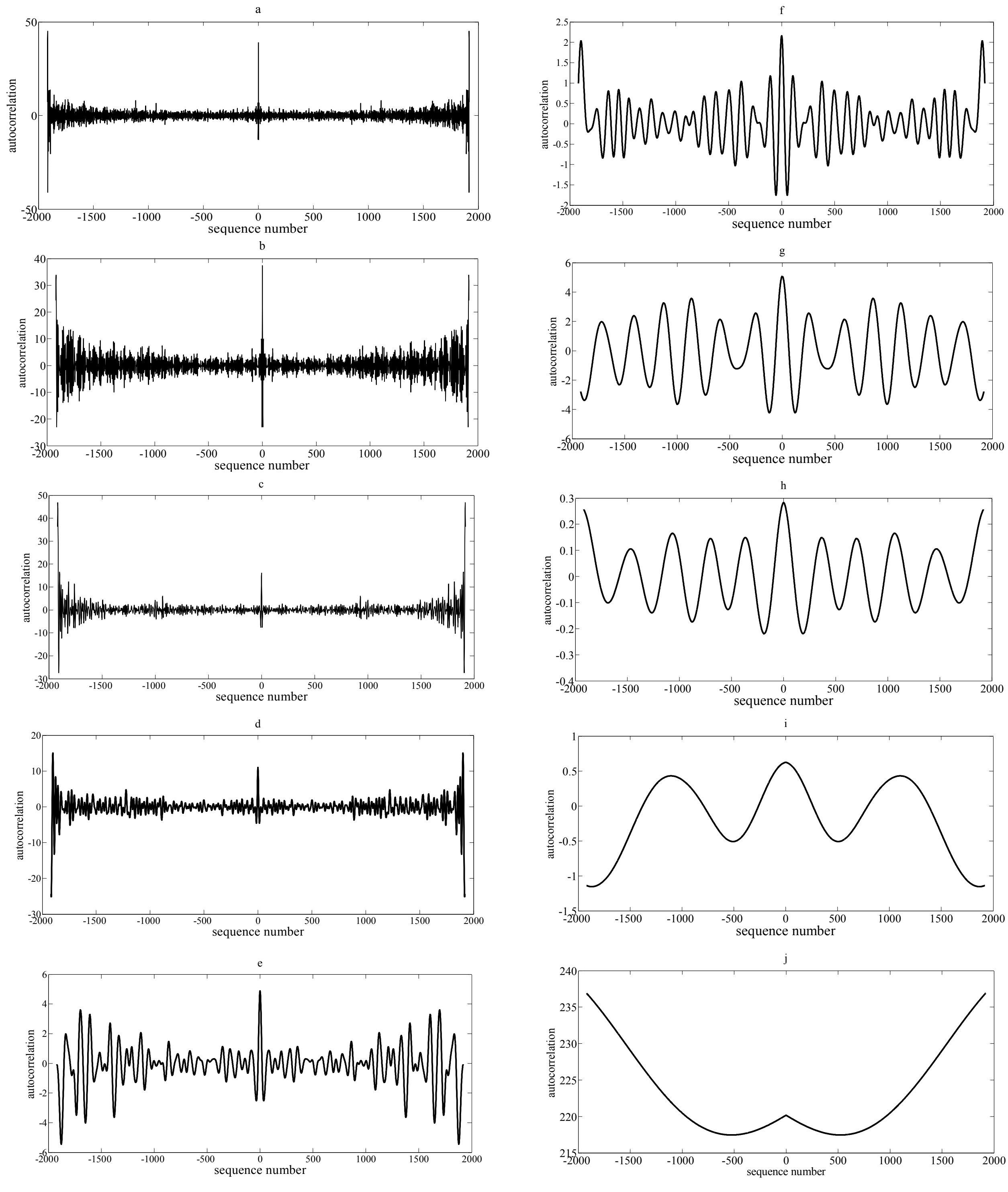

Fig. 8. The autocorrelation function of the componentsafter EMD decomposition (a: IMF1; b: IMF2; c: IMF3; d: IMF 4; e: IMF 5; f: IMF 6; g: IMF 7; h: IMF 8; i: IMF 9; j: Residual)

in [25]. The parameters of echo state network are $SR=0.622$, $N=126$, $IS=0.475$, $SD=0.173$. The parameters of ARMA are $p=8$, $q=6$ by AIC criterion. The parameters of SVM are obtained by cross validation method in the SVM toolbox, that is $\varepsilon=42.368$, $C=9.636$. The parameters of LSSVM are obtained by cross validation method in the LSSVM toolbox, that is $\gamma=24.268$, $\sigma^2=11.361$. The parameters of ELM are the number of hidden layer nodes is 48 obtained by the method of trial and error, activation functions of ELM is selected ass Sigmoid function with λ is 1. Fig. 12 is the comparison curve between actual and prediction value of these prediction methods.

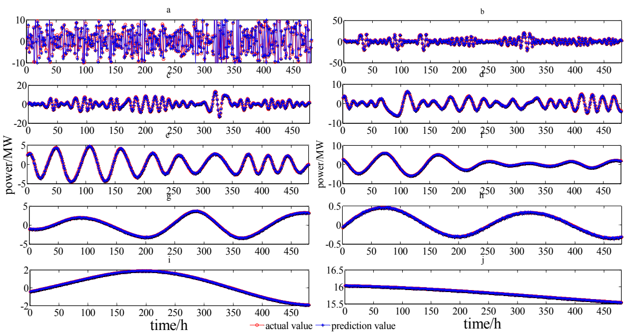

Fig. 9. The comparison between the 480 groups of prediction value and the actual value of each component (a: IMF1; b: IMF2; c: IMF3; d: IMF 4; e: IMF 5; f: IMF 6; g: IMF 7; h: IMF 8; i: IMF 9; j: Residual)

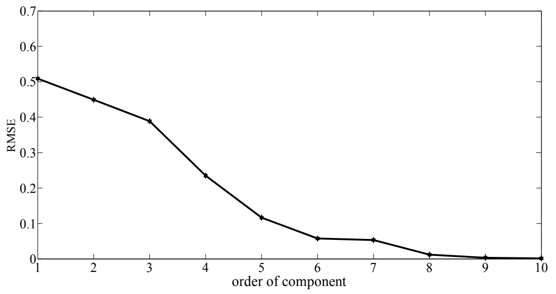

Fig. 10. Root square mean error of each component

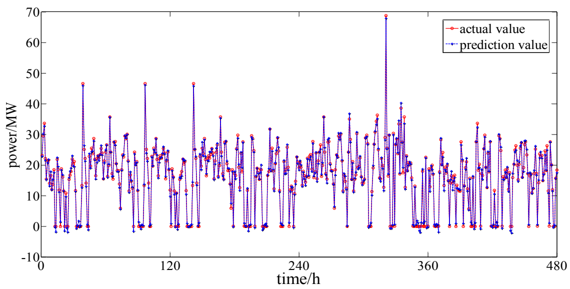

Fig. 11. The comparison of final prediction value and actual value of 480 groups wind power

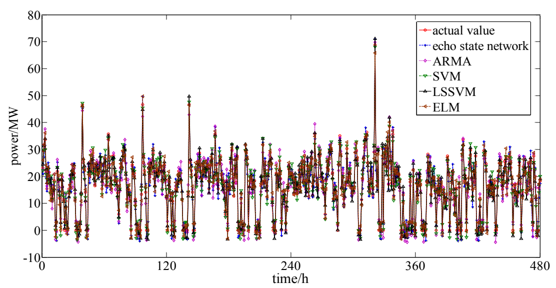

Fig. 12. The comparison curve between actual and prediction value of five prediction methods

It can be observed from Fig. 11 and Fig. 12, the predictive effect of the prediction method in this paper is better than echo state network, ARMA, SVM, LSSVM,

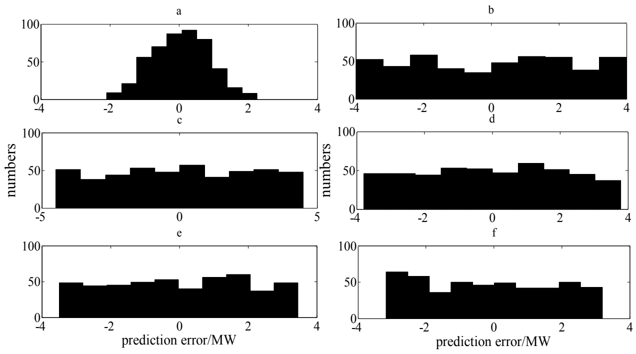

Fig. 13. The prediction error distribution histogram of the prediction methods (a: proposed method; b: echo state network; c: ARMA; d: SVM; e: LSSVM; f: ELM)

and ELM.

Fig. 13 is the prediction error distribution histogram of the prediction methods mentioned in this paper. From Fig. 13 can be known, because the prediction error is smaller and prediction error distribution is more uniform, the prediction effect and performance in this paper is better.

In this paper, the next five performance indexes are adopted. These performance indexes include root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), reliability, and R Square. The definitions of these performance indexes are as follows.

1. RMSE

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(w_k - \widehat{w}_k)^2} \tag{18}$$

2. MAE

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|(w_k - \widehat{w}_k)| \tag{19}$$

3. MAPE

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}|(w_k - \widehat{w}_k)| \times 100\% / w_k \tag{20}$$

4. Reliability

$$R^{(1-a)} = [\frac{\xi^{(1-a)}}{N} - (1-a)] \times 100\% \tag{21}$$

5. R^2 (R Square)

$$R^2 = 1 - (\sum_{k=1}^{N}(w_k - \widehat{w}_k)^2 / \sum_{k=1}^{N}(w_k - \overline{w})^2) \tag{22}$$

where, N is the length of wind power sequence, w_k is the actual value of wind power, $\widehat{w}_k$ is the predictive value of wind power. $\xi^{(1-a)}$ is the numbers of confidence intervals in which the actual value falls under the

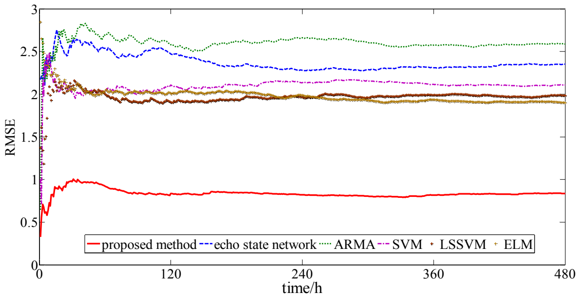

Fig. 14. RMSE comparison of the prediction methods

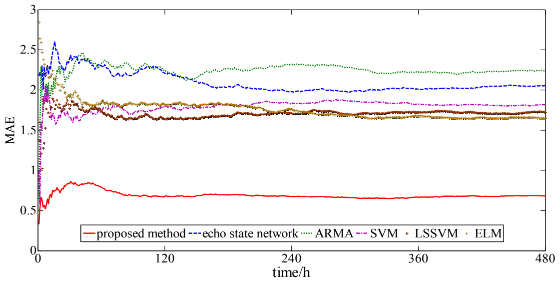

Fig. 15. MAE comparison of the prediction methods

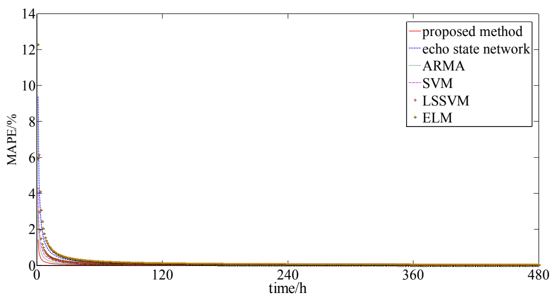

Fig. 16. MAPE comparison of the prediction methods

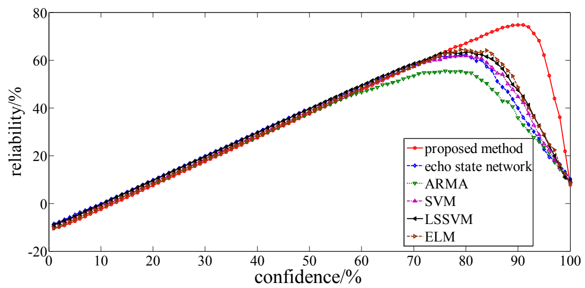

Fig. 17. The reliability and confidence distribution of the prediction methods

confidence level $1-\alpha$, $\overline{w}$ is the mean value of wind power. Fig. 14 gives RMSE comparison of the prediction methods mentioned in this paper. Fig. 15 gives MAE comparison of the prediction methods mentioned in this paper. Fig. 16 gives MAPE comparison of the prediction methods mentioned in this paper. From the performance indexes comparison results in Figs. 14 to 16, the prediction accuracy of the proposed prediction method is better than

Table 1. Comparison of performance indicators

Prediction methods	RMSE	MAE	MAPE	R Square
Proposed method	0.8356	0.6817	0.0029	0.9933
Echo state network	2.3511	2.0522	0.0195	0.9469
ARMA	2.5869	2.2322	0.0058	0.9357
SVM	2.1063	1.8150	0.0090	0.9574
LSSVM	1.9830	1.7246	0.0124	0.9622
ELM	1.8993	1.6447	0.0256	0.9654

the others.

Fig. 17 is the reliability and confidence distribution of the prediction methods mentioned in this paper. It can be observed from this graph that the prediction method in this paper has higher reliability under the same confidence level. It can be known that the reliability of wind power prediction method in this paper is better than other prediction methods.

Table 1 gives RMSE, MAE, MAPE and R^2 comparison of the prediction methods mentioned in this paper. From the performance indexes comparison results in Table 1, RMSE, MAE and MAPE of the proposed method is smaller than the other prediction methods. At the same time, R^2 value of the proposed method is closer to 1 than the other prediction methods. The closer the value of R^2 is to 1, the better the regression prediction performance of the model is. Therefore, the prediction accuracy of the proposed prediction method is better than the other methods.

In summary, from the above prediction contrast curve and prediction error distribution, performance indexes and reliability, the proposed wind power prediction method is better than other prediction methods. The one main reason of prediction accuracy improvement is the introduction of EMD. After EMD processing, the wind power time series is changed from long correlation sequence to short correlation, which highlights the change rule of each component, so the prediction difficulty and complexity are both reduced. The other reason is that the improved ELM prediction model has better regression prediction performance compared with other models.

5. Conclusion

1. A short-term wind power prediction method based on EMD and improved ELM is proposed in this paper. EMD can decompose the real-scale fluctuation or trend of the same scale in wind power time series step by step, and smooth the non-linear and non-stationary characteristic of series. It produces a series of data sequences with the same characteristic scale.

2. For each stationary sequence with approximate characteristics, the improved ELM prediction models are respectively established, so as to reduce the influence of nonlinearity and non-stationary on the wind power prediction results.

3. The simulation results of actual wind power data show that the prediction method proposed in this paper can better track the change rules of wind power, and effectively improve the prediction accuracy of short-term wind power.

4. In this paper, the future research work is to introduce more types of activation functions into improved ELM model, and find the better activation function which can improve the prediction accuracy. At the same time, the optimization of prediction parameters is further studied.

Acknowledgements

This work is supported by the Science Research Project of Liaoning Education Department (Grant No.LGD2016009), and the Natural Science Foundation of Liaoning Province (Grant No. 20170540686).

References

[1] Y. Wang, Z. X. Xie, Q. H. Hu, and S. H. Xiong. "Correlation aware multi-step ahead wind speed forecasting with heteroscedastic multi-kernel learning," *Energy Conversion and Management*, vol. 163, pp. 384-406, May. 2018.

[2] W. Y. Y. Cheng, Y. B. Liu, A. J. Bourgeois, Y. H. Wu, and S. E. Haupt, "Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation," *Renewable Energy*, vol. 107, pp. 340-351, Jul. 2017.

[3] Z. D. Tian, S. J. Li, Y. H. Wang, and X. D. Wang. "Wind power prediction method based on hybrid kernel function support vector machine," *Wind Engineering*, vol. 42, no. 3, pp. 252-264, Jun. 2018.

[4] U. Meyyappan. "Wavelet neural network-based wind speed forecasting and application of shuffled frog leap algorithm for economic dispatch with prohibited zones incorporating wind power," *Wind Engineering*, vol. 42, no. 1, pp. 3-15, Jan. 2018.

[5] I. Colak, S. Sagiroglu, and M. Yesilbudak. "Data mining and wind power prediction: A literature review," *Renewable Energy*, vol. 46, pp. 241-247, Oct. 2012.

[6] T. H. Ouyang, X. M. Zha, L. Qin, Y. Xiong, and T. Xia. "Wind power prediction method based on regime of switching kernel functions," *Journal of Wind Engineering & Industrial Aerodynamics*, vol. 153, pp. 26-33, Jun. 2016.

[7] S. Al-Yahyai, Y. Charabi, and A. Gastli. "Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment," *Renewable & Sustainable Energy Reviews*, vol. 14, no. 9, pp. 3192-3198, Dec. 2010.

[8] J. H. Li, J. M. Li, J. Y. Wen, S. J. Cheng, H. L. Xie, and C. Y. Yue. "Generating wind power time series based on its persistence and variation characteristics," *Science China Technological Sciences*, vol. 57, no. 12, pp. 2457-2486, Dec. 2014.

[9] C. Liu, C. Li, Y. H. Huang, and Y. F. Wang. "A novel *stochastic modeling method of wind power time series* considering the fluctuation process characteristics," *Journal of Renewable & Sustainable Energy*, vol. 8, no. 3, 033304, May. 2016.

[10] C. Wang, H. L. Zhang, W. H. Fan, and X. C. Fan. "A new wind power prediction method based on chaotic theory and Bernstein Neural Network," *Energy*, vol. 117, pp. 259-271, Dec. 2016.

[11] J. D. Wang, K. J. Fang, W. J. Pang, and J. W. Sun. "Wind power interval prediction based on improved PSO and BP neural network," *Journal of Electrical Engineering & Technology*, vol. 12, no. 3, pp. 989-995, May. 2017.

[12] R. R. B. de Aquino, O. N. Neto, R. B. Souza, M. M. S. Lira, M. A. Carvalho, T. B. Ludermir, and A. A. Ferreira. "Investigating the use of echo state networks for prediction of wind power generation," *IEEE on Computational Intelligence for Engineering Solutions*, 2015, pp. 148-154.

[13] C. D. Zuluaga, M. A. Álvarez, and E. Giraldo. "Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison," *Applied Energy*, vol. 156, pp. 321-330, Oct, 2015.

[14] J. R. Yang, X. C. Wang, X. F. Luo, and C. Jiang. "Intelligent combined prediction of wind power based on numerical weather prediction and fuzzy clustering," *IFAC - Papers Online*, vol. 48, no. 28, pp. 538-543, Aug. 2015.

[15] X. K. Wang, D. S. Luo, and H. Y. He. "An improved feature weighted fuzzy clustering algorithm with its application in short-term prediction of wind power," *6th Chinese Conference on Pattern Recognition*, 2014, pp. 575-584.

[16] Y. X. Liu and Y. Y. Zhang. "A rolling ARMA method for ultra short term wind power prediction," *13th IEEE Conference on Automation Science and Engineering*, 2017, pp. 1232-1236.

[17] X. B. Kong, X. J. Liu, R. F. Shi, and K. Y. Lee. "Wind speed prediction using reduced support vector machines with feature selection," *Neurocomputing*, vol. 169, pp. 449-456, Dec. 2015.

[18] Y. C. Xiao, C. Y. Li, and P. Wang. "Wind power prediction based on improved grey theory and SVM," *Journal of Information & Computational Science*, vol. 11, no. 16, pp. 5937-5944, Nov. 2014.

[19] Q. L. Wu and C. Y. Peng. "Wind power grid connected capacity prediction using LSSVM optimized by the bat algorithm," *Energies*, vol. 8, no. 12, pp. 14346-14360, Dec. 2015.

[20] J. L. Lou, H. Cao, B. Song, and J. Z. Xiao. "An output power prediction method for multiple wind farms under energy internet environment," *International*

Journal of Grid and Distributed Computing, vol. 9, no. 11, pp. 273-284, Nov. 2016.
[21] T. H. Ouyang, X. M. Zha, and L. Qin. "A combined multivariate model for wind power prediction," *Energy Conversion & Management*, vol. 144, pp. 361-373, Jul. 2017.
[22] G. B. Huang, Q. Y. Zhu, and C. K. Siew. "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, Dec. 2006.
[23] Y. Lan, Y. C. Soh, and G. B. Huang. "Two-stage extreme learning machine for regression," *Neurocomputing*, vol. 73, no. 16, pp. 3028-3038, Oct. 2010.
[24] G. B. Huang, H. M. Zhou, X. J. Ding, and R. Zhang. "Extreme learning machine for regression and multiclass classification," *IEEE Transactions Systems Man and Cybernetics Part B-Cybernetics*, vol. 42, no. 2, pp. 513-529, Apr. 2012.
[25] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong. "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033-1044, May. 2014.
[26] G. Y. Zhang, Y. G. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H. C. Lu. "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2706-2715, Sep. 2015.
[27] N. E. Huang, Z. Shen, S. R. Long, M. L. C. Wu, H. H. Shih, Q. N. Zhang, N. C. Yen, and C. C. Tung. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Mathematical Physical & Engineering Sciences*, vol. 454, no. 1971, pp. 903-995, Mar. 1998.
[28] S. C. Du, T. Liu, D. L. Huang, and G. L. Li. "An optimal ensemble empirical mode decomposition method for vibration signal decomposition," *Journal of Vibration & Acoustics*, vol. 139, no. 3, 031003, Jun. 2017.

Zhongda Tian He received the Ph. D degree in Control Theory and Control Engineering from Northeastern University, China in 2013. He is currently an Associate Professor in College of Information Science and Engineering, Shenyang University of Technology, China. His research interests include predictive control, delay compensation and scheduling for networked control system, time series prediction.

Yi Ren He received his B. Eng degree from Liaoning University of Science and Technology in 2016. He is currently a Master student of Control Theory and Control Engineering, College of Information Science and Engineering, Shenyang University of Technology, China. His interests mainly include time series prediction.

Gang Wang He received his B. Eng degree from Shanxi Datong University in 2016. He is currently a Master student of Control Theory and Control Engineering, College of Information Science and Engineering, Shenyang University of Technology, China. His interests mainly include time series modeling and forecasting.