Wind Power Prediction Based on Variational Mode Decomposition and Feature Selection

Gang Zhang, Benben Xu, Hongchi Liu, Jinwang Hou, and Jiangbin Zhang

Abstract—Accurate wind power prediction can scientifically arrange wind power output and timely adjust power system dispatching plans. Wind power is associated with its uncertainty, multi-frequency and nonlinearity for it is susceptible to climatic factors such as temperature, air pressure and wind speed. Therefore, this paper proposes a wind power prediction model combining multi-frequency combination and feature selection. Firstly, the variational mode decomposition (VMD) is used to decompose the wind power data, and the sub-components with different fluctuation characteristics are obtained and divided into high-, intermediate-, and low-frequency components according to their fluctuation characteristics. Then, a feature set including historical data of wind power and meteorological factors is established, which chooses the feature sets of each component by using the max-relevance and min-redundancy (mRMR) feature selection method based on mutual information selected from the above set. Each component and its corresponding feature set are used as an input set for prediction afterwards. Thereafter, the high-frequency input set is predicted using back propagation neural network (BPNN), and the intermediate- and low-frequency input sets are predicted using least squares support vector machine (LS-SVM). After obtaining the prediction results of each component, BPNN is used for integration to obtain the final predicted value of wind power, and the ramping rate is verified. Finally, through the comparison, it is found that the proposed model has higher prediction accuracy.

Index Terms—Wind power prediction, feature selection, variational mode decomposition (VMD), max-relevance and min-redundancy (mRMR).

I. INTRODUCTION

WIND power has the characteristics of instability, which may lead to cascaded failure and certain shock in the power system. This brings severe challenges to the safe and stable operation of power system [1], [2]. Wind power prediction is a prerequisite for the grid-connected wind farm. Otherwise, wind energy resources will not be effectively utilized, which will restrict the effective installed capacity of wind farms. If the wind power could be predicted more accurately, the power system dispatching department can adjust the dispatching plan in time, and rationally formulate the control strategy. Consequently, this will reduce the rotating reserve capacity of the power grid, reduce the wind power ramping and power generation cost, and improve the safety of wind power generation [3].

At present, wind power prediction methods commonly used at home and abroad are mainly machine learning methods such as neural network method [4], [5], time series method [6], [7], support vector machine method [8], [9], and Kalman filtering method [10]. However, it is difficult to capture its characteristics using only one method due to the randomness and uncertainly of wind power. Therefore, some researchers first decompose the wind power data and then conduct the prediction.

For example, [11] uses a combination of empirical mode decomposition (EMD) and support vector machine (SVM) prediction model; [12] uses a combination of EMD and nonlinear autoregressive exogenous (NARX) neural network prediction model; and [13], [14] use the ensemble empirical mode decomposition (EEMD). Reference [15] uses a short-term wind power prediction model combining variational modal decomposition (VMD) and extreme learning machine (ELM), and [16] uses multi-frequency combined VMD decomposition to predict wind power. All these studies conduct the prediction on the sub-sequences after decomposing the wind power, and then integrate the predicted values of the sub-sequences. It can be seen from the studies that the decomposition of VMD is more thorough, and that the final prediction accuracy is higher. However, the wind power data will be affected by meteorological factors such as wind speed, air pressure, and temperature. The sub-sequences decomposed by various decomposition methods will naturally be affected by these factors. There is no doubt that the prediction accuracy will be reduced only by considering the study of wind power data and ignoring the meteorological factors.

In recent years, some researchers have considered the influence of some factors in the wind power prediction pro-

G. Zhang (corresponding author) is with the State Key Laboratory of Eco-Hydraulics in Northwest Arid Region, Xi'an University of Technology, Xi'an 710048, China (e-mail: zhanggang3463003@163.com).

B. Xu, H. Liu, J. Hou, and J. Zhang are with the School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China, and J. Zhang is also with The Institute of Water Resources and Hydroelectric Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: 18302980426@163. com; lhc2861145027@163. com; hjw20190124@163. com; jbzhang@xaut. edu.cn).

cess. For example, [17] extracts the features from wind power historical values. Reference [18] uses mutual information (MI) to extract the spatial correlation information between variables and target variables, and then uses conditional kernel density estimation methods for wind power prediction. Reference [19] uses the wind speed and wind direction data in different numerical weather prediction (NWP) data to train the model. At present, there are three common feature selection methods including filtering methods, packaging methods, and embedding methods. The accuracy of the packaging method and the embedding method are higher but are easy to lead to over-fitting, while the filtering method is less likely to cause over-fitting and could reduce the computation dimension [20]. The common filtering method is the MI. It arranges the variables according to the standard of MI, and then selects the top-ranked variables. This method can preprocess the variables before using the predictive model for prediction [21]. However, the MI only considers the correlation between the variable and the target variable but ignores the redundancy, which makes the input dimension larger and reduces the computation efficiency of the model. Comparatively, the max-relevance and min-redundancy (mRMR) based on MI considers not only the correlation between variables and target variables, but also its redundancy, so it could reduce the input dimension and improve the computation efficiency.

Based on the above discussion, this paper improves wind power prediction on the basis of [16], and proposes a prediction method of "decomposition-feature selection-prediction-integration". Firstly, the wind power data is decomposed using VMD to obtain the components with different fluctuation characteristics. These components are divided into high-, intermediate- , and low-frequency components according to their fluctuation characteristics. Then, a feature set containing wind power historical data and meteorological factors is established, and selects a feature set of each component from the set using mRMR. After that, each component and its corresponding feature set are taken as an input set. In the prediction process, the high-frequency input set is predicted using back propagation neural network (BPNN), and the intermediate- and low-frequency input sets are predicted using least squares support vector machine (LS-SVM). After obtaining the predicted values of each component, the BPNN is used to integrate the predicted values of each component to obtain the final predicted values of wind power. Finally, the prediction results are compared with those of other models. By analyzing the four indexes including mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and root mean square error (RMSE), it is shown that the results of the proposed prediction method are the closest to the actual values.

The rest of this paper is organized as follows. Section II briefly introduces the prediction model and prediction performance evaluation indicators. Section III conducts case studies and analyzes the results. The research conclusions are given in Section IV.

II. Prediction Method

According to the above discussion, the flow chart of the prediction method proposed in this paper is shown in Fig. 1. It can be seen that the prediction process is divided into four modules: ① module 1 is the decomposition, i. e., the wind power data is decomposed using VMD; ② module 2 is the feature selection, i.e., mRMR is used to select the input feature set of each component from the established feature set; ③ module 3 is the multi-frequency prediction, i. e., each component and its corresponding input feature set are combined into a new matrix in the prediction model for various frequencies to obtain the predicted value of various frequency components; ④ module 4 is the integration, i. e., BPNN is used to integrate the prediction results of component of various frequency components to obtain the final wind power predicted value and the predicted value of wind power ramping rate.

It can be seen from Fig. 1 that the VMD, mRMR, BPNN, and LS-SVM algorithms are used in this paper. Among them, VMD is proposed in [22], which is essentially a set of adaptive Wiener filter banks. It adopts non-recursive mode decomposition that can simultaneously estimate the modalities of different center frequencies and could avoid the modal aliasing caused by empirical decomposition in EMD and EEMD. Therefore, it has been applied in many fields [15], [16], [23], [24]. The feature selection method of mRMR is proposed in [25]. Based on MI, the method analyzes the mRMR between each feature and target variable, which reduces the input dimension and enhances the prediction accuracy.

For the prediction methods of various frequencies, BPNN and LS-SVM are the mature methods that have been used in various fields including wind power prediction [26] and the study on wind power ramping [27]. Therefore, the principles of these two methods are not repeated here.

When evaluating the proposed model, this paper uses the MAPE to evaluate the prediction accuracy of intrinsic mode functions (IMFs). For comparison with other models, this paper uses three evaluation indicators including MAE, MSE, and RMSE. The calculation formulas of each indicator are as follows:

$$MAE = \frac{1}{M} \sum_{m=1}^{M} \left| Y_m - F_m \right| \tag{1}$$

$$MSE = \frac{1}{M} \sum_{m=1}^{M} \left(Y_m - F_m \right)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(Y_m - F_m \right)^2} \tag{3}$$

$$MAPE = \frac{1}{M} \sqrt{\sum_{m=1}^{M} \left| \frac{Y_m - F_m}{Y_m} \right|} \times 100\% \tag{4}$$

where Y_m is the actual value; F_m is the predicted value; and M is the sample capacity.
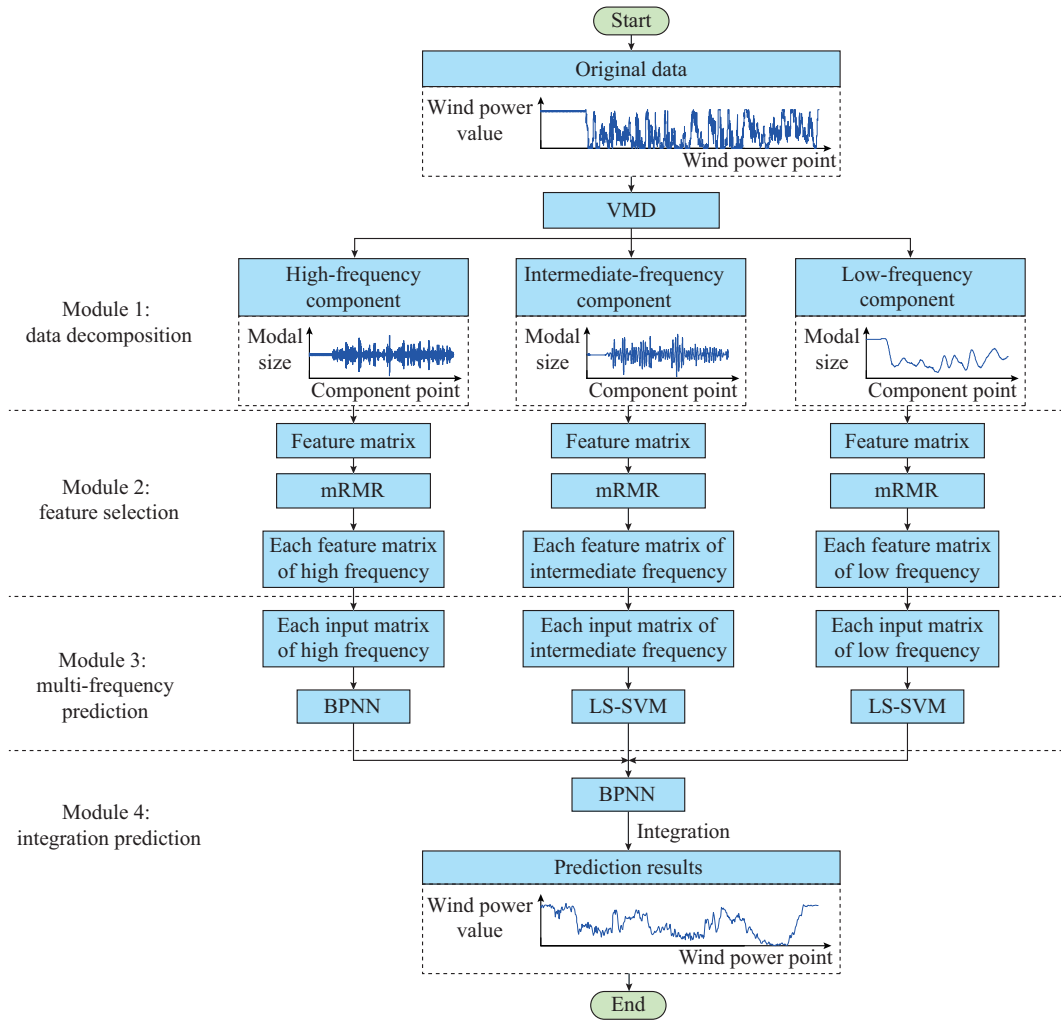
Fig. 1.   Flowchart of prediction process.

III. CASE STUDY

A. Decomposition of Wind Power Data

The wind power data in this paper is collected every 10 minutes, and each wind farm has 144 data points per day. In this paper, the data in October, 2009 in northern Shaanxi, China and the data of wind farm in July, 2010 in Yunnan, China will be selected for analysis. For the division of the training set and test set, this paper uses a variety of divisions. Considering the influencing factors, the training data is used to train the model and make predictions on the test set, and the final prediction result is evaluated by MAE. The comparison results are shown in Table I. It can be seen from Table I that when the training data is for 28 days and the test data is for 3 days, the error is the smallest. In this case, the occurrence of under-fitting and over-fitting can be effectively reduced [28]. Therefore, this paper uses 4464 data points for 31 days a month for each wind farm for modeling, 4032 data points from the 1st to the 28th of the month as training samples, and 432 data points from the 29th to the 31st of the month are used as test samples, respectively. Figure 2 illustrates the wind power data of the wind farm in northern Shaanxi in October, 2009.

TABLE I
PREDICTION ERROR FOR MULTIPLE SITUATIONS OF DIVISION

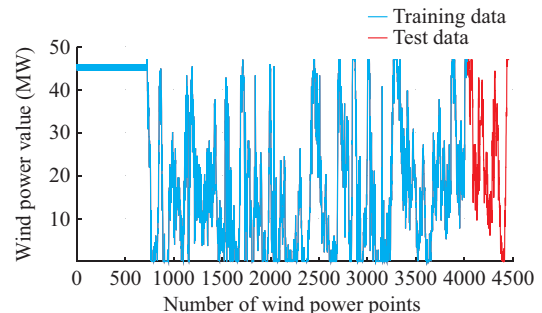| No. of training days	No. of test days	Error (MW)
25	6	2.8548
26	5	2.8244
27	4	2.7641
28	3	2.6767
29	2	2.6912
30	1	2.7324

Fig. 2.   Wind power data for wind farm in northern Shaanxi, China in October, 2009.

The basic information of the wind power value is shown in Table II.

TABLE II
PREDICTION ERROR FOR MULTIPLE SITUATIONS OF DIVISION

Region	Maximum value (MW)	Minimum value (MW)	Average value (MW)	Standard deviation (MW)
Northern Shaanxi	47.03	0	21.38	16.82
Yunnan	47.02	0	12.30	13.18

The wind power values of the two wind farms are highly random. In addition, the difference between the average and standard deviation of data of wind farm in Yunnan in July, 2010 is 0.88 MW, and the difference between the average and standard deviation of data of wind farm in northen Shannxi in October, 2009 is 4.56 MW. The greater the difference between the average and the standard deviation is, the greater the dispersion of the data will be. Therefore, by comparison, the wind power fluctuation of wind farm in northern

Shaanxi in October, 2009 is stronger. In the following analysis, a specific explanation is conducted on the data of the wind farm in northern Shaanxi in October, 2009. The prediction process of the wind farm in Yunnan is consistent with that in northern Shaanxi. First, VMD is used to decompose the wind power data of the wind farm in northern Shaanxi in October, 2009 to better utilize its multi-frequency characteristics. The wind power data is decomposed by VMD, and the result is shown in Fig. 2. In the decomposition process, the parameters are set using the method in [23].

Each component, which is obtained by decomposing wind power data by the VMD, is shown in Fig. 3. It can be seen from Fig. 3 that the fluctuations of IMF1-IMF6 are strong and its frequency is large, the fluctuation after IMF7 starts to ease, and at IMF10, the fluctuation tends to be stable and IMF10 is also the most moderate of all components, that is, the fluctuation frequency is also the smallest. Therefore, IMF1-IMF6 are defined as high-frequency components, IMF10 is defined as low-frequency component, and the remaining components are defined as intermediate frequency components.
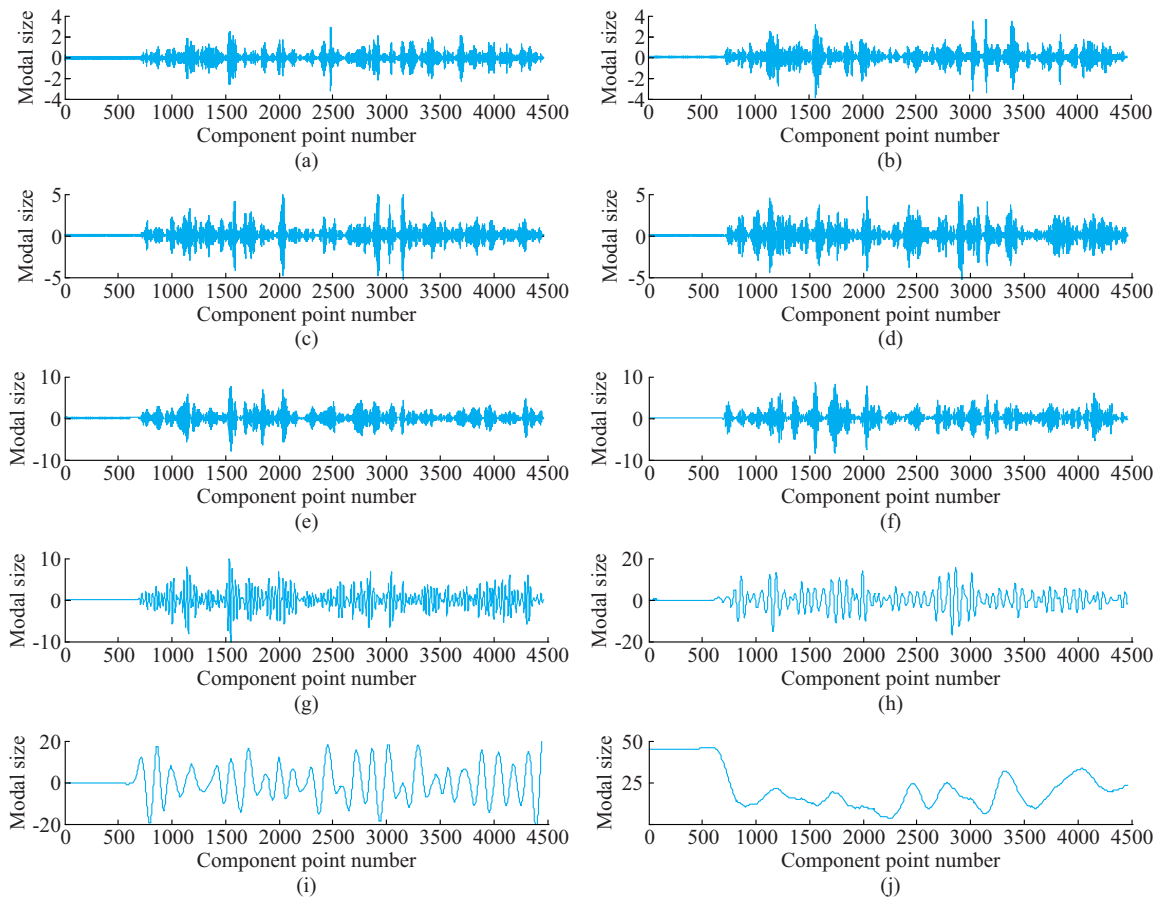

Fig. 3. Decomposition map and spectrogram. (a) IMF1. (b) IMF2. (c) IMF3. (d) IMF4. (e) IMF5. (f) IMF6. (g) IMF7. (h) IMF8. (i) IMF9. (j) IMF10.

B. Feature Selection

The wind power is affected by characteristic factors such as wind speed, temperature, and wind direction. Therefore, each sub-component obtained by decomposing the wind pow-

er using VMD is also affected by different characteristics. This sub-section will use mRMR to select the feature set of each component. The specific process is shown in Fig. 4, where N is the number of IMFs after VMD decomposition. It can be seen from Fig. 4 that the establishment of feature

set of each component mainly includes the following steps: ① establishing an initial feature set F, which includes influencing characteristics affecting wind power variation; ② using incremental search method to establish the candidate feature set J of each modal component from set F; ③ calculating the mRMR value of each feature in J, arranging them in descending order, and inputting into the error function one by one to calculate the error; ④ taking out the corresponding number of features when the error is minimized and establishing the set of component input features J_{IN}. The specific research process is as follows.
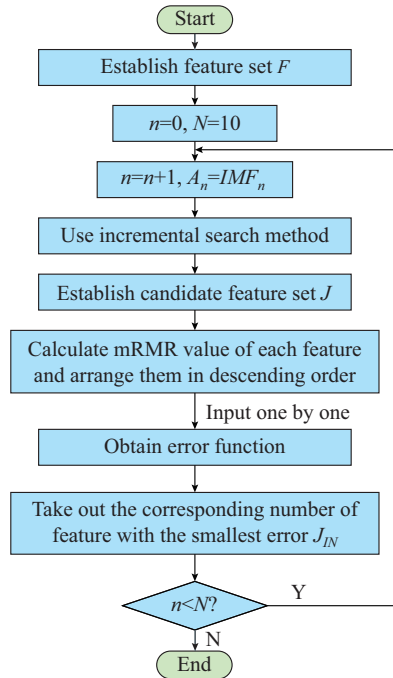

Fig. 4.   Flowchart of component feature selection.

This paper first establishes a feature set. The features and representation variables contained in the set are shown in Table III.

TABLE III
INFLUENCING FEATURES AND REPRESENTATION VARIABLES

Influencing feature	Representation variable
Time	t
Temperature	T
Air pressure	P
Wind direction	D
Wind speed	S
Historical value of wind speed	S_t
Historical value of wind power	P_t

The time scale of wind power in this paper is 10 minutes. Therefore, the time t in Table III is expressed by 0.0, 0.1, ..., 23.5. The historical values of wind speed and wind power are represented by S_t and P_t, respectively, and its meaning is the historical value t minutes ago. For example, S_{10} represents the wind speed value of the sample 10 minutes ago, and P_{50} represents the wind power value of the sample 50 minutes ago. In addition, the wind speed, air pressure, and temperature are the measured values.

After the feature matrix is established, the candidate feature set J is established for each component, respectively, by using the incremental search method. The size of the mRMR value of each feature in the candidate feature set J is calculated, and the features are arranged in descending order according to the magnitude of the mRMR value. The descending sorting results of candidate feature sets of each feature component are shown in Table IV.

TABLE IV
DESCENDING SORTING RESULTS OF CANDIDATE FEATURE SETS OF EACH FEATURE COMPONENT

Sorting serial number	Component									
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
1	$S_{10}=4.5055$	$P_{20}=4.4566$	$S=4.4215$	$S_{50}=3.8149$	$S_{30}=4.8252$	$S_{40}=4.5082$	$P_{40}=4.1146$	$P_{40}=3.4235$	$S_{30}=2.4146$	$T=4.7423$
2	$S=4.4861$	$P_{30}=4.4556$	$S_{30}=4.3936$	$S_{30}=3.7995$	$P_{40}=4.7868$	$S_{20}=4.4673$	$S_{30}=4.0923$	$S_{30}=3.4161$	$P_{40}=2.3898$	$D=4.5705$
3	$S_{40}=4.4154$	$P_{50}=4.4276$	$P_{50}=4.3124$	$S_{40}=3.7742$	$P_{20}=4.7505$	$P_{30}=4.4267$	$S_{20}=4.0188$	$P_{50}=3.3347$	$P_{50}=2.3721$	$P=3.0220$
4	$P_{20}=4.4033$	$P_{40}=4.4144$	$S_{20}=4.2961$	$P_{20}=3.7669$	$S=4.7401$	$S=4.4058$	$P_{50}=4.0062$	$P_{20}=3.3095$	$P_{20}=2.3610$	$P_{40}=1.9220$
5	$S_{30}=4.3394$	$S_{30}=4.3016$	$P_{60}=4.2310$	$S=3.6542$	$P_{60}=4.6367$	$P_{60}=4.2669$	$P_{60}=3.8442$	$P_{60}=3.2189$	$P_{60}=2.1652$	$S=1.9016$
6	$P_{50}=4.3295$	$S=4.2856$	$S_{40}=4.1917$	$P_{10}=3.5201$	$P_{30}=4.4955$	$P_{10}=4.2267$	$P_{10}=3.8134$	$P_{10}=3.1846$	$P_{10}=2.1594$	$P_{30}=1.8844$
7	$S_{50}=4.3053$	$S_{50}=4.2587$	$P_{30}=4.1322$	$P_{50}=3.4763$	$S_{40}=4.4939$	$P_{40}=4.2084$	$S_{40}=3.5759$	$P_{30}=2.6099$	$T=1.7645$	$P_{20}=1.7488$
8	$S_{20}=4.2983$	$S_{20}=4.2534$	$P_{10}=4.1113$	$P_{30}=3.4496$	$S_{50}=4.4870$	$S_{50}=4.1231$	$P_{30}=3.5545$	$S_{40}=2.5978$	$P_{30}=1.6750$	$P_{60}=1.7101$
9	$P_{60}=4.1642$	$S_{40}=4.2292$	$S_{50}=4.0662$	$S_{20}=3.4464$	$P_{10}=4.4829$	$P_{20}=4.1107$	$P_{20}=3.5013$	$S_{20}=2.5065$	$D=1.6618$	$P_{10}=1.5487$
10	$S_{60}=4.1627$	$P_{10}=4.1813$	$P_{20}=4.0446$	$P_{40}=3.4166$	$S_{20}=4.4445$	$S_{30}=4.1077$	$S_{50}=3.4504$	$S_{50}=2.4861$	$S_{40}=1.6366$	$S_{40}=1.4620$
11	$P_{10}=4.1576$	$S_{60}=4.1413$	$S_{60}=4.0288$	$S_{60}=3.2988$	$S_{60}=4.3846$	$S_{60}=3.9882$	$S_{60}=3.3236$	$S_{60}=2.3963$	$S_{50}=1.6128$	$S_{50}=1.4096$
12	$P_{40}=4.0882$	$S_{10}=4.0157$	$S_{10}=3.8723$	$S_{10}=3.2049$	$S_{10}=4.2217$	$S_{10}=3.9195$	$S_{10}=3.3095$	$S_{10}=2.3953$	$S_{20}=1.5961$	$S_{30}=1.4042$
13	$T=3.2907$	$P_{60}=3.2444$	$D=3.5070$	$T=2.3983$	$P_{50}=3.3580$	$P_{50}=3.2075$	$S=2.3902$	$S=1.3950$	$S_{10}=1.4392$	$S_{20}=1.2776$
14	$D=3.2476$	$D=3.1257$	$T=3.4850$	$D=2.3805$	$T=2.8679$	$T=2.1820$	$P=1.4513$	$P=0.5055$	$S_{60}=1.4149$	$S_{60}=1.2457$
15	$P_{30}=3.2409$	$T=2.9673$	$P_{40}=3.0523$	$P_{60}=2.3385$	$D=2.7589$	$D=2.1438$	$D=1.4262$	$T=0.2628$	$t=0.9956$	$S_{10}=1.1096$
16	$P=2.4770$	$P=2.3002$	$P=2.4283$	$P=1.9996$	$P=2.3385$	$P=1.8861$	$T=1.4166$	$t=0.1597$	$S=0.5254$	$t=0.4318$
17	$t=2.0774$	$t=2.1714$	$t=2.1435$	$t=1.3314$	$t=2.3357$	$t=1.8663$	$t=1.3126$	$D=0.1230$	$P=0.3966$	$P_{50}=0.3797$

Table IV shows the result of mRMR values in descending order. The corresponding numbers of the variables are the mRMR values of the feature. It can be seen that for each modal component, the mRMR values of wind speed, wind speed history value, and wind power history value are relatively in the front, while the feature time is relatively backward for each component. The size of the wind power value is not closely related to time. In addition, IMF10 is the only component where the temperature, wind direction, and air pressure have high rankings. It can be seen that the IMF10 is mainly affected by weather characteristics such as temper-ature, wind direction, and air pressure.

After obtaining the candidate feature sets of the respective components, the features are input into the prediction model one by one to calculate the prediction error, and the number of input features with the smallest error is taken as the final input feature set J_{IN}. The MAPE is used for evaluation. The relationship between the error of each component and the number of input features is shown in Fig. 5. When the number of input eigenvalues is equal to that of eigenvalues corresponding to the blue bar, the error is the smallest.
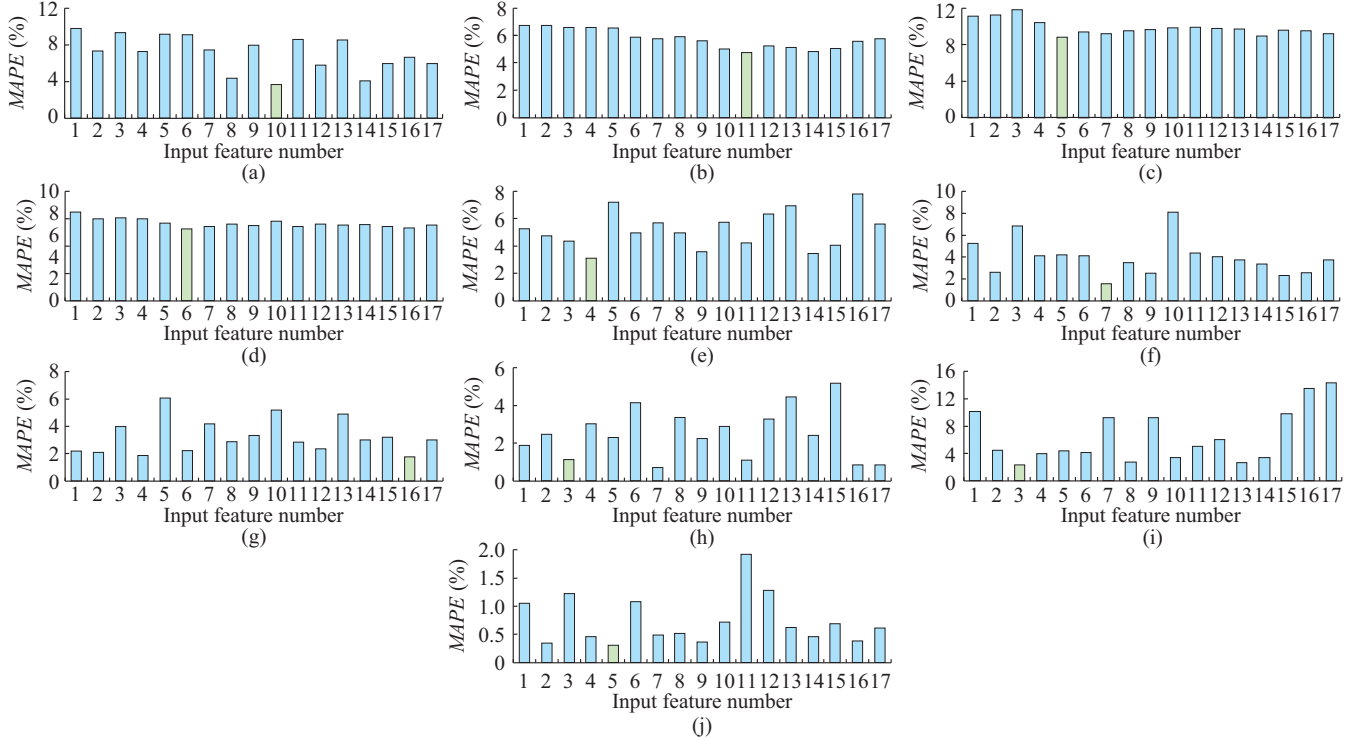

Fig. 5. Relationship between number of each component feature and error. (a) IMF1. (b) IMF2. (c) IMF3. (d) IMF4. (e) IMF5. (f) IMF6. (g) IMF7. (h) IMF8. (i) IMF9. (j) IMF10.

It can be seen from Fig. 5 that the relationship diagram is mainly divided into two categories.

1) Continuous fluctuation. We will get a minimum error in the fluctuation finally, for example, Fig. 5(a), (e) - (j). There will be cases where several points are with less error such as Fig. 5(i). When the numbers of features are 3, 8, and 13, respectively, the errors of the three points are very close, but increasing the input quantity will lead to the increase of running time of the program and reduce the running speed. Therefore, this paper will select the number of input features corresponding to the point with the smallest input quantity as the input feature matrix of the component.

2) Small fluctuation. The error first decreases and then undergoes a period of smooth fluctuation such as Fig. 5(b)-(d), where the point with the smallest error occurs in the process of the error drop.

Through the above analysis, the input feature set J_{IN} of each modal component is selected out, as shown in Table V.

TABLE V
INPUT FEATURE SET OF EACH COMPONENT

Component	Input feature set
IMF1	$S_{10}, S, S_{40}, P_{20}, S_{30}, P_{10}, S_{50}, S_{20}, P_{60}, S_{60}$
IMF2	$P_{20}, P_{30}, P_{50}, P_{40}, S_{30}, S, S_{50}, S_{20}, S_{40}, P_{10}, S_{60}$
IMF3	$S, S_{30}, P_{50}, S_{20}, P_{60}$
IMF4	$S_{50}, S_{30}, S_{40}, P_{20}, S, P_{10}$
IMF5	$S_{30}, P_{40}, P_{20}, S$
IMF6	$S_{40}, S_{20}, P_{30}, S, P_{60}, P_{10}, P_{40}$
IMF7	$P_{40}, S_{30}, S_{20}, P_{50}$
IMF8	P_{40}, S_{30}, P_{50}
IMF9	S_{30}, P_{40}, P_{50}
IMF10	T, D, P, P_{40}, S

It can be seen from Table V that the input feature set of IMF1-IMF9 mainly includes the wind speed, wind speed historical values, and wind power historical values. Among

them, J_{IN} of IMF1-IMF6 contains more wind speeds and historical values of wind speeds. Since wind speeds are uncertain and intermittent, these components mainly include the randomness and volatility of wind power due to the uncertainty of wind speed. By contrast, J_{IN} of IMF10 mainly includes meteorological features such as temperature, wind direction, and air pressure. The generation of wind speeds and directions is affected by the temperature difference of the environment. In addition, the pressure gradient force is the direct cause of wind formation. It can be seen that IMF10 mainly contains the factors of wind generation, and the mag-

nitude of the wind power is directly affected by this factor of wind generation. Therefore, it is considered that IMF10 mainly includes the changing trend of wind power.

C. Prediction of Wind Power

As described above, BPNN is used to predict high-frequency components, and LS-SVM is used to predict intermediate- and low-frequency components. Before adding the influencing factors, BPNN and LS-SVM are used to directly predict various frequency components of VMD decomposition, and MAPE is used to represent the error. The results are shown in Table VI.

TABLE VI
PREDICTION RESULTS

Model	MAPE (%)									
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
BPNN	4.00	3.44	4.75	1.67	1.47	2.36	3.12	5.42	1.66	0.030
LS-SVM	4.23	5.57	6.19	5.02	2.53	2.87	3.06	0.83	0.21	0.005

It can be seen from Table VI that for high-frequency components IMF1-IMF6, the accuracy of prediction results using BPNN is significantly higher than those using LS-SVM. For the intermediate- and low-frequency components, however, the prediction accuracy is lower than that of BPNN, especially after IMF8, it becomes more and more obvious.

To sum up, the prediction performance of BPNN is better than LS-SVM for high-frequency components, and the prediction performance of LS-SVM is better than BPNN for low- and intermediate-frequency components.

In the training model, the input data is extracted from the new matrix composed of the frequency components and the corresponding J_{IN}. When the input quantity of BPNN is trained, three points of high-frequency components and their corresponding features are used as the input of the neural network, which is also called the input layer. For example, the number of input layer nodes of IMF5 is 15. In addition, the number of iterations is set to be 1000, the learning speed is set to be 0.1, and the expected error is set to be 0.0004. When LS-SVM is trained, the kernel function is the radial basis function (RBF), and the particle swarm optimization (PSO) algorithm is used to optimize the kernel parameters and regularization parameters. After the model of each frequency component is trained, and the wind power prediction is done, the prediction results of the various frequency components are integrated using BPNN to obtain the final predicted values. The results and distribution map of predicted points are shown in Fig. 6.

It can be seen from Fig. 6 that the curve from the 50th wind power data point to the 100th wind power data point has larger error to the actual curve compared with the curve of other intervals, because the data has strong fluctuation in the above interval. The remaining prediction curves are basically close to the actual curve, and the error is small. In order to clearly analyze the prediction method in this paper, the distribution figure of each predicted point is drawn and shown in Fig. 6(b).
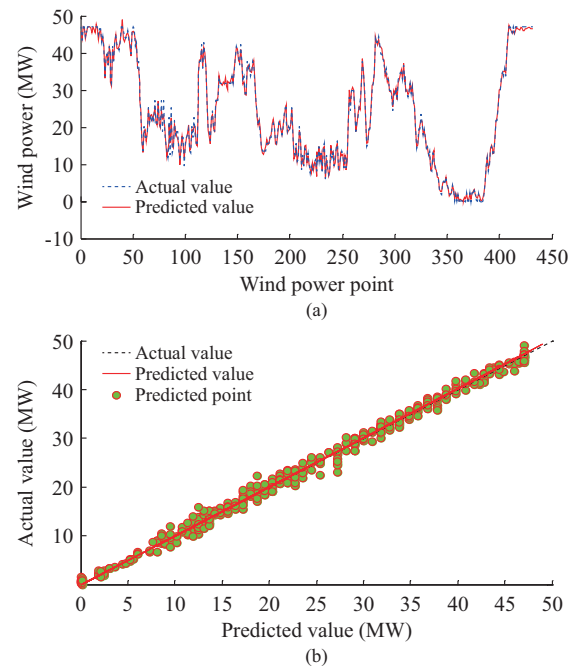

Fig. 6.   Wind power prediction results. (a) Prediction and actual wind power. (b) Distribution of predicted points.

In Fig. 6, if the green point is closer to the black line, the closer the predicted value is to the actual value, and the higher the prediction accuracy will be. It can be seen from Fig. 6(a) that in the interval where the predicted value is 25-30, there are some points that are farther away from the black line; the value of wind power is between 20-30 in the range of 50-100 where the error is relatively obvious. The distribution shown in Fig. 6(b) coincides with the analysis in Fig. 6(a). In addition, it can be seen from Fig. 6(b) that the predicted value of wind power is the most accurate in the interval of 0-10, and if some points are ignored in other intervals, most of the predicted points can be close to the actual value line of black. Therefore, the red line is also very close

to the black line.

D. Verification of Wind Power Ramping Rate

Wind power ramping is likely to cause imbalances to the active power of the system, disrupt the frequency stability, and even lead to large-scale load shedding, which severely threatens the safe, stable, and economic operation of power system. Therefore, after predicting the wind power value, the wind power ramping rate still needs to be predicted. Wind power ramping rate refers to the rate of change of wind farm power caused by the random nature of wind, i.e., the power ramping rate (PRR), which is calculated as:

$$PRR = \frac{\Delta P_r}{\Delta t} \qquad (5)$$

where ΔP_r is the amplitude change value of wind power; and Δt is the duration of power fluctuation. The key to the definition of wind power ramping rate is the selection of Δt. Generally, Δt has 3 reference values, 15 min, 30 min, and 60 min [29]. This paper will choose $\Delta t = 30$ min to study the wind power ramping rate. Different countries have different requirements for wind power ramping rate. In China, when $P_N < 150$ MW, the maximum power change does not exceed $33\% P_N$ for 10 min and $10\% P_N$ for 1 min; when $P_N > 150$ MW, the maximum power change is 50 MW for 10 min and 15 MW for 1 min, where P_N is the rated installed capacity. In Section III-C, 432 wind power data points in three days are predicted, and 144 predicted values of wind power ramping rate can be further calculated. The prediction results are shown in Fig. 7.
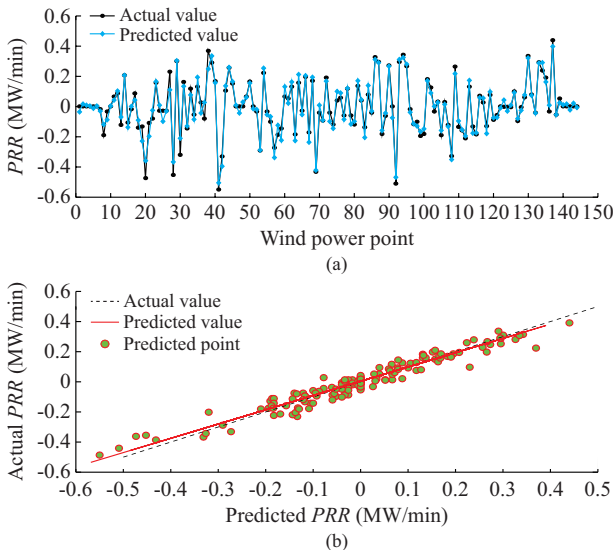

Fig. 7.   Prediction results of wind power ramping rate. (a) Predicted and actual PRR. (b) Distribution of predicted points.

It can be seen from Fig. 7 that with the exception of individual points, most of the predicted values are very close to the actual values. The upward ramping rate and downward ramping rate of wind power ramping rate is less than 0.5 MW/min, which is less than $10\% P_N$ (4.703 MW/min) and meets the standard. It can be seen from the Fig. 7(b) that the distribution of the predicted points is always near the actual

value, and basically, the fitted predicted value also coincides with the actual value, which indicates that the prediction error is small and the accuracy is high. This shows that the method of this paper, which first predicts the value of wind power and then predicts the wind power ramping rate, is effective in prediction. Therefore, we can make targeted adjustments to wind power according to the prediction results, and reduce the probability of ramping events effectively. As a result, the operation of power systems with wind power will become safer and more economical, and the power supply and grid planning will be more reasonable.

E. Comparison and Analysis

In order to visually analyze the proposed prediction model, we compare the predicted point distribution maps of LS-SVM, BPNN, long short-term memory (LSTM), deep belief network (DBN), EMD combination prediction model, EEMD combination prediction model, and VMD combination prediction model considering influencing factors by using MI with the model proposed in this paper. Among them, the EMD combination prediction model, the EEMD combination prediction model, and the VMD combination prediction model firstly use EMD, EEMD, or VMD to decompose wind power data, and then use BPNN to predict high-frequency components and LS-SVM to predict intermediate- and low-frequency components, and finally, use BPNN for integration. For the VMD combination prediction model considering the influencing factors by using MI, after using VMD to decompose the wind power data, we use MI to consider the influencing characteristics of each modal component, BPNN to predict the high-frequency components, and LS-SVM to predict the intermediate- and low-frequency components. Finally, BPNN is used for integration.

In order to analyze the above model prediction results more intuitively, we compare each model using the evaluation indicators mentioned above. The calculation results are shown in Table VII.

TABLE VII
PREDICTION PERFORMANCE INDICATORS OF EACH MODEL IN WIND FARM IN NORTHERN SHAANXI, CHINA

Model	Evaluation index (MW)		
	MSE	RMSE	MAE
VMD+mRMR+BPNN+LS-SVM	0.8537	0.9270	0.7044
VMD+MI+BPNN+LS-SVM	0.9240	0.9612	0.7478
VMD+BPNN+LS-SVM	1.3148	1.1466	0.9088
EMD+BPNN+LS-SVM	6.7821	2.6042	1.7741
EEMD+BPNN+LS-SVM	4.0752	2.0187	1.5600
LS-SVM	12.6100	3.5511	2.5979
BPNN	12.8671	3.5871	2.6844
LSTM	11.4221	3.3797	2.4910
DBN	11.1682	3.3419	2.3872

It can be clearly seen from Table VII that the prediction accuracy of the multi-frequency combination prediction model is significantly higher than that of the single prediction

model for the wind farm in northern Shaanxi, China. For example, for MSE, the prediction accuracy of the EMD combination prediction model is higher than those of the LS-SVM, BPNN, LSTM, and DBN by 46.2%, 47.3%, 40.6%, and 39.3%, respectively. The prediction accuracy of the EEMD combination prediction model is higher than that of the LS-SVM, BPNN, LSTM, and DBN by 67.7%, 68.3%, 64.3%, and 63.5%, respectively. The prediction accuracy of the VMD combination prediction model is higher than that of the LS-SVM, BPNN, LSTM, and DBN by 89.6%, 89.8%, 88.5%, and 88.2%, respectively. The prediction accuracy of the VMD combination prediction model, which considers the influencing factors by using MI, is higher than those of the LS-SVM, BPNN, LSTM, and DBN by 92.7%, 92.8%, 91.9%, and 91.7%, respectively. The prediction model in this paper is higher than those of the LS-SVM, BPNN, LSTM, and DBN by 93.2%, 93.4%, 92.5%, and 92.4%, respectively.

The combination prediction model proposed in this paper has the highest accuracy. For example, for MAE, the prediction accuracy in this paper is higher than those of the EMD combination model, EEMD combination model, and VMD combination model by 60.3%, 54.8%, and 22.5%, respectively. It can be seen that the wind power prediction accuracy is significantly improved considering the influencing factors such as meteorology.

Similarly, the prediction situation of wind farm in Yunnan is basically consistent with that of wind farm in northern Shaanxi. It can be seen from Table VIII that the prediction accuracy of the combination prediction model is higher than that of the single prediction model. For the combination prediction model, the improved EEMD combination prediction model based on EMD is more accurate than EMD. The accuracy of VMD combination prediction model that improves the "end-point effect" is higher than that of EEMD, and the accuracy of the VMD combination prediction model that considers the influencing factors is higher. However, the accuracy of the VMD combination prediction model that considers the influencing factors on the basis of MI and mRMR is higher than that of the VMD combination prediction model that considers the influencing factors only by using the MI.

TABLE VIII
PREDICTION PERFORMANCE INDICATORS OF EACH MODEL IN WIND FARM
IN YUNNAN, CHINA

Model	Evaluation index (MW)		
	MSE	RMSE	MAE
VMD+mRMR+BPNN+LS-SVM	2.2005	1.4834	0.8732
VMD+MI+BPNN+LS-SVM	2.5064	1.5831	0.9752
VMD+BPNN+LS-SVM	2.5506	1.5971	1.1283
EMD+BPNN+LS-SVM	4.4517	2.1099	1.4100
EEMD+BPNN+LS-SVM	2.7341	1.6535	1.1378
LS-SVM	11.8814	3.4469	1.9348
BPNN	12.1493	3.4856	2.2827
LSTM	11.5147	3.3933	1.8237
DBN	11.4712	3.3869	1.7184

IV. CONCLUSION

In this paper, the combination of decomposition method and feature selection method considers not only the multi-frequency of wind power data, but also the influence of wind speed and temperature on wind power. In order to avoid the modal aliasing and false components of EMD and EEMD when decomposing wind power data, this paper uses VMD to decompose wind power, whose principle is completely different from those of EMD and EEMD, aiming to make better use of the multi-frequency of wind power and improve the prediction accuracy. In addition, wind power data is affected by wind speed, direction, and other characteristics, so the components of various frequencies contain the above information. To this end, this paper uses mRMR for feature selection, which aims to select the features that have a greater impact on the component of various frequencies. When selecting features, a feature matrix composed of wind speed, wind direction, temperature and other features is firstly established, and the incremental search method is used to establish candidate feature sets of each component. The features are then arranged in the candidate feature set in descending order of mRMR, and are input into the prediction model one by one to calculate the error. Finally, the number of features is taken with the smallest error as the input feature of the component. It can be seen from the case study that after considering the influence of features on each component, the model prediction accuracy is significantly improved. It turns out that the prediction model proposed in this paper has higher accuracy.

In addition, when the input feature set is selected after the candidate feature set is established, the features of the candidate feature set are input into the prediction model one by one, and the calculated error is taken as the decisive factor. The workload is relatively large although this will greatly improve the prediction accuracy of each component and reduce the input dimension. Therefore, it is the next research direction of this paper to find a better way to choose the input feature set or develop a better feature selection method.

REFERENCES

[1] M. F Tahir, H. Y. Chen, A. Khan et al., "Optimizing size of variable renewable energy sources by incorporating energy storage and demand response," IEEE Access, vol. 7, pp. 103115-103126, Sept. 2019.

[2] M. F Tahir, H. T. Hassan, K. Mehmood et al., "Optimal load shedding using an ensemble of artificial neural networks," International Journal of Electrical and Computer Engineering Systems, vol. 7, no. 2, pp. 39-46, Aug. 2016.

[3] L. Zhang, J. Lu, Y. Mei et al., "Wind power prediction based on different optimization criteria," Electric Power Automation Equipment, vol. 35, no. 5, pp. 139-145, May 2015.

[4] Y. Ju, L. Qi, and S. Liu, "Short-term wind power forecasting based on improved crow search algorithm and ESN neural network," Power System Protection and Control, vol. 47, no. 4, pp. 58-64, Feb. 2019.

[5] Y. Ju, G. Sun, Q. Chen et al., "A model combining convolutional neural network and lightGBM algorithm for ultra-short-term wind power forecasting," IEEE Access, vol. 7, pp. 28309-28318, Apr. 2019.

[6] J. Cao, R. Zhou, X. Deng et al., "Wind power forecast considering differential times of optimal ARIMA model," Proceedings of the CSU-EPSA, vol. 31, no. 1, pp. 105-111, Jan. 2019.

[7] X. Zhu and Y. Liu, "Wind power forecasting using time series model based on robust estimation," Proceedings of the CSU-EPSA, vol. 24, no. 3, pp. 107-110, Jun. 2012.

[8] D. Wu and C. Gao, "Short-term wind power generation forecasting

based on the SVM-GM approach," *Electric Power Components and Systems*, vol. 46, pp. 1250-1264, Jul. 2018.

[9] A. Liu, Y. Xue, J. Hu *et al*., "Ultra-short-term wind power forecasting based on SVM optimized by GA," *Power System Protection and Control*, vol. 43, no. 2, pp. 90-95, Jan. 2015.

[10] P. Salgado, G. Igrejas, and P. Afonso, "Multi-Kalman filter to wind power forecasting," in *Proceedings of 2018 13th APCA International Conference on Automatic Control and Soft Computing (CONTROLO)*, Ponta Delgada, Portugal, Jun. 2018, pp. 110-114.

[11] Y. Jiang, X. Yang, L. Chen *et al*., "Super-short-term wind power combination forecasting based on support vector machine optimized by EMD-SC and AGSA," *Chinese Journal of Engineering Design*, vol. 24, no. 2, pp. 187-195, Apr. 2017.

[12] Z. Zhang, C. Ma, J. Xu *et al*., "Novel total-power combinational forecasting method of wind farm based on EMD and NARX neural network," *Computer Engineering and Applications*, vol. 52, no. 2, pp. 265-270. Apr. 2016.

[13] Q. Cheng, L. Chen, Y, Cheng *et al*., "Short-term wind power forecasting method based on EEMD and LS-SVM model," *Electric Power Automation Equipment*, vol. 38, no. 5, pp. 27-35, May 2018.

[14] B. Tian, Z. L. Park, and D. Guo *et al*., "Wind power ultra short-term model based on improved EEMD-SE-ARMA," *Power System Protection and Control*, vol. 45, no. 4, pp. 72-79, Feb. 2017.

[15] A. A. Abdoos, "A new intelligent method based on combination of VMD and ELM for short term wind power forecasting," *Neurocomputing*, vol. 203, pp. 111-120, Aug. 2016.

[16] G. Zhang, H. Liu, J. Zhang *et al.*, "Wind power prediction based on variational mode decomposition multi-frequency combinations," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 2, pp. 281-288, Mar. 2019.

[17] Y. Xue, C. Yu, K. Li *et al*., "Adaptive ultra-short-term wind power prediction based on risk assessment," *CSEE Journal of Power and Energy Systems*, vol. 2, no. 3, pp. 59-64, Sep. 2016.

[18] J. Wang, X. Han, Y. Ma *et al*., "Short-term wind power probabilistic forecasting considering spatial correlation," in *Proceedings of 2017 IEEE Conference on Energy Internet and Energy System Integration*, Beijing, China, Nov. 2017, pp. 1-6.

[19] S. Buhan and I. Çadırcı, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1231-1242, Oct. 2015.

[20] A. Sharma, K. K. Paliwal, S. Imoto *et al*., "A feature selection method using improved regularized linear discriminant analysis," *Machine Vision & Applications,* vol. 25, pp. 775-786, Apr. 2014.

[21] H. Zhao and F. Magoulès, "Feature selection for predicting building energy consumption based on statistical learning method," *Journal of Algorithms and Computational Technology*, vol. 6, no. 1, pp. 59-78, Mar. 2012.

[22] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531-544, Mar. 2014.

[23] X. He, J. Luo, G. Zuo *et al*., "Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural net-works," *Water Resources Management*, vol. 33, pp. 1571-1590, Mar. 2019.

[24] Y. Wu, "Research on fault diagnosis of wind turbine transmission system based on variational modal decomposition," Ph. D. dissertation, School of Control and Computer Engineering, North China Electric Power University, Beijing, China, 2016.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1226-1238, Aug. 2005.

[26] Y. Zhang, Y. Wang, H. Deng *et al*., "IAFSA-BPNN for wind power probabilistic forecasting," *Power System Protection and Control*, vol. 45, no. 7, pp. 58-63, Apr. 2017.

[27] D. Zhang, Y. Yue, X. Zhang *et al*., "Review and prospect of research on wind power ramp events," *Power System Technology*, vol. 42, no. 6, pp. 1783-1792, Mar. 2018.

[28] M. F. Tahir, Tehzeeb-ul-Hassan, and M. A. Saqib. "Optimal scheduling of electrical power in energy-deficient scenarios using artificial neural network and Bootstrap aggregating," *International Journal of Electrical Power and Energy Systems*, vol. 83, pp. 49-57, Dec. 2016.

[29] A. Kusiak and H. Zheng. "Data mining for prediction of wind farm power ramp rates," in *Proceedings of 2008 IEEE International Conference on Sustainable Energy Technologies*, Singapore, Singapore, Nov. 2008, pp. 1099-1103.

Gang Zhang received his Ph.D. degree from Xi'an University of Technology, Xi'an, China, in 2013. He is currently working in the Institute of Water Resources and Hydro-electric Engineering, Xi'an University of Technology. His research interests include renewable energy prediction and power system dispatching.

Benben Xu is currently pursuing the Ph.D. degree in the School of Electrical Engineering, Xi'an University of Technology, Xi'an, China. His research interests include wind power prediction model and fault diagnosis of wind turbine.

Hongchi Liu is currently pursuing the Ph.D. degree in the Institute of Water Resources and Hydro-electric Engineering, Xi'an University of Technology, Xi'an, China. His research interests include wind power prediction model and power system analysis and control.

Jinwang Hou is currently pursuing the Ph.D. degree in the School of Electrical Engineering, Guangxi University, Nanning, China. His research interest includes the optimal dispatching of power system.

Jiangbin Zhang received his Ph.D. degree in water resources and hydrology from Xi'an University of Technology, Xi'an, China, in 2006. He is currently working in the Institute of Water Resources and Hydroelectric Engineering, Xi'an University of Technology. His research interests include control and adjustment of wind turbines.