A Combined Model Based on CEEMDAN, Permutation Entropy, Gated Recurrent Unit Network, and an Improved Bat Algorithm for Wind Speed Forecasting

TAO LIANG [ID][1], **GAOFENG XIE** [ID][1], **SHURUI FAN** [ID][2], **AND ZHAOCHAO MENG** [ID][1]

[1]School of Artificial Intelligence, Hebei University of Technology, Tianjin 300130, China
[2]School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China

Corresponding author: Tao Liang (liangtao@hebut.edu.cn)

ABSTRACT Accurate and reliable wind speed forecasting is crucial for wind farm planning and grid operation security. To improve the accuracy of wind speed forecasting, a novel combined model is proposed for wind speed forecasting in this article. First, the complete ensemble empirical mode decomposition adaptive noise (CEEMDAN) and permutation entropy (PE) are employed to decompose the original wind speed time series into the sub-series with obvious complexity different; To overcome the disadvantage of weak generalization ability of single deep learning method when facing diversiform data, a cluster of gated recurrent unit networks (GRUs) with different hidden layers and neurons are applied to capturing the unsteady characteristics and implicit information of each sub-series; The predictions of the GRUs of each sub-series are aggregated into a nonlinear-learning regression top-layer which is consisted of radial basis function neural network (RBFNN), and improved bat algorithm (IBA) is introduced to optimize the parameters of RBFNN; Lastly, the prediction values of each nonlinear-learning top-layer are superimposed to obtain the final prediction values. To validate the effectiveness of the model, 15-min and 1-h wind speed data from the wind farm in Zhangjiakou, China, are used as test cases. The experimental results demonstrate that the proposed combined model can achieve the best performance and stability compared to other models. Such as the performance evaluation indexes (RMSE = 0.3294, MAPE = 2.6169%) are smallest obtained from case study 1, and (RMSE = 0.5876, MAPE = 4.7875%) are smallest obtained from case study 2.

INDEX TERMS Wind speed forecasting, combined model, complete ensemble empirical mode decomposition adaptive noise, gated recurrent unit.

I. INTRODUCTION

Due to the rapid development of modern industry, the rapid consumption of fossil fuels, the problems of environmental degradation and traditional resource consumption have become some increasingly serious. Wind energy as an environmentally friendly renewable energy, has been widely recognized and promoted worldwide. According to the Global Wind Energy Council, as of the end of 2018, the total installed capacity of wind power has reached 591GW in the worldwide, 209533 MW was contributed by China [1]. However, the randomness, abruptness and uncertainty of wind resources can not only destroy the reliability and stability for the wind farms and power system operation, but also result in low efficiency and high operating cost. And these will bring great challenges to the safe and stable operation of wind farms and grids with large-scale wind power integration [2], [3]. Therefore, reliable and precise wind speed forecasting not only helps to maintain the stability of the wind turbine output and develop an optimal wind farm maintenance plan, but also can reduce the operating cost of the power system [4]. In recent years, with the increasing number

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaowei Zhao.

of wind farms, power generation companies have established corresponding wind farm centralized monitoring centers to facilitate centralized management and dispatch. And this allows the forecasting systems originally set up in the wind farms to be moved to the centralized monitoring center, and then the centralized control center distributes the wind speed and wind power prediction results obtained by the prediction system to the corresponding wind farms, thereby greatly saving costs. However, due to the scale, geographic location, climatic conditions and wind turbine models of each wind farm are different [5], it is difficult for the wind speed forecasting strategy of the same centralized control center to adapt to the characteristics of different wind farms, which gives the centralized control center management scheduling brings many problems. Therefore, it is necessary to propose a reliable and accurate wind speed forecasting method that is universally applicable to multiple wind farm environments under the centralized control system. Moreover, the centralized management of wind farms provides environmental support for us to build a wind speed forecasting system with more complex algorithms and stronger adaptability.

In recent years, many researchers have given much attention to the research of the methods of wind speed prediction. These methods can be classified into the following four types: Physical models, conventional statistical models, artificial intelligence models, and hybrid models. Physical models always use current geographic data and a variety of meteorological data to solve complex mathematical models such as temperature, pressure and humidity, and it has a large error in wind speed forecasting in short-term or ultra-short term. The widely used conventional statistical models include the ARIMA (autoregressive integrated moving average) model, the quantile-regression model (QR), and the Kalman-filter model. Masseran [6] has proposed a prediction model consisting of ARIMA and ARCH (Autoregressive Conditional Heteroskedasticity). Kavasseri and Seetharaman [7] proposed a fraction-ARIMA to predict the wind speed of one-day and two-day ahead in North Dakota. The conventional statistical model uses historical data to establish a predictive model, which can capture the linear relationship of the training data set well. However, due to the randomness and inherent complexity of the wind speed, the conventional statistical model has a low prediction accuracy. With the rapid development of artificial intelligence technology, artificial intelligence models have been successfully applied to time series prediction. Among them, back propagation neural network (BPNN), RBFNN, extreme learning machine (ELM), wavelet neural network (WNN) and other artificial neural networks, support vector machine (SVM) and expert systems have been widely used in wind speed prediction [8], [9]. Wang *et al.* [10] used RBFNN to predict wind speed. The experimental results show that the method has achieved certain improvement in accuracy. Santamaríabon *et al.* [11] used SVM to predict wind speed in the short and medium term. Jiang *et al.* [12] proposed a v-SVM hybrid short-term wind speed prediction model based on the cuckoo search algorithm optimization,

which has improved the nonlinear approximation ability of the prediction model.

At present, a large number of wind speed prediction methods have been proposed, which improve the accuracy of wind speed prediction to some extent. However, the original wind speed time series is noisy and unstable. Therefore, directly using the original wind speed time series to establish a prediction model will produce a large of mistakes [13], [14]. The research shows that the data analysis method for wind speed time series can reduce the influence of wind speed non-stationarity on the prediction results. The decomposition-reconstruction prediction model based on signal analysis method has received more and more attention [15]. Empirical Mode Decomposition (EMD) is an adaptive signal processing method that decomposes time series into an intrinsic mode functions (IMFs) and a residue [16], [17]. Reference [18] proposed a new method by combining the two technologies of EMD and Elman neural network (ENN). The model showed a better and more accurate prediction. However, EMD has many drawbacks, such as end effect, mode-mixing [19]–[21], and so on. As an extension of EMD, ensemble empirical mode decomposition (EEMD) decomposes signals by adding white Gaussian noise and treating the average result as a real final result [22]–[24], which improves the EMD and avoids the phenomenon of modal-aliasing. However, the white noise added by the EEMD method in practice has not been completely offset after multiple averaging. CEEMDAN is proved to be an important improvement progress on EEMD [25]. Ye *et al.* [26] combined EMD and its improved version, EEMD/CEEMD/CEEMDAN with Support Vector Regression (SVR), to establish a predictive model. The results show that CEEMDAN-SVR is the best method.

In recent years, deep learning has been growing rapidly, which has achieved obvious advantages in different application fields, and has received more and more attention. [27]. Among the many deep learning methods, GRU and LSTM has improved the recurrent neural network (RNN) with special memory structure and gate structure to enhance long-term memory [28], [29]. And they have compensated for the gradient explosion of RNN and the long-term dependence of time series. However, GRU uses a gated recurrent neural network structure, which has fewer training parameters than LSTM. Meanwhile GRU is superior to LSTM in forecasting accuracy and forecasting speed, and it shows more adaptability in the analysis of time series data [30], [31]. Therefore, the GRU is chosen as the principal forecasting engine in this article. In recent years, in order to obtain an advanced prediction approach for higher accuracy levels and wider forecast horizons, the approaches called as hybrid models and combined models have emerged [32]–[34]. Niu and Wang [35] proposed a including complete ensemble empirical mode decomposition with adaptive noise and a multi-objective grasshopper optimization algorithm based on a no-negative constraint theory. The advantages of each individual forecasting model are effectively utilized, the stability

and accuracy of the prediction results are considered, and this method achieves better prediction.

Based on the above analysis, a novel combined model, based on CEEMDAN, PE, GRU neural network, RBFNN and IBA is proposed in this study for wind speed forecasting. First, CEEMDAN is used to decompose non-stationary wind speed time series into different IMF components with corresponding frequencies. Next, PE was applied to analyze the complexity of each IMF component, and recombine the IMFs with similar entropy values. To solve the issues of over-decomposition and computational burden [36], [37]. Then, inspired by integrated learning technology, a cluster of GRUs with diverse hidden layers and neurons are applied to capture the unsteady characteristics and implicit information of each recombination sequence. In order to overcome the shortcomings of the traditional combination model linear representation, the predictions of GRUs are aggregated into a nonlinear-learning regression top-layer to give the final ensemble prediction of each recombinant sub-series in this article. The nonlinear learning top layer used in this article is composed of RBFNN, and the improved bat algorithm (IBA) is introduced to optimize the center vector, base width and weight of RBF neural network by which to overcome the problem of easy convergence to local optimum. Finally, the prediction results of the nonlinear-learning top layer of each recombination sequence are superimposed to obtain the final wind speed prediction value. To validate the accuracy of the proposed combined model, 15-min and 1-h wind speed data from the wind farm in Zhangjiakou, China, are used as cases. The results of experiments and discussions show that the proposed combined model can achieve a better forecasting performance and stability compared to other models.

The rest of the paper is organized in the following way. The related basic theory of nonlinear-learning ensemble of deep learning time series prediction for wind speed forecasting, CEEMDAN and PE are introduced in Section 2. In Section 3 BA and IBA are described. Section 4 presents the main structure of the proposed combined model. Then, in Section 5, two case studies are demonstrated, in which the prediction results of the proposed model and other involved models are evaluated. Finally, the conclusions are presented in Section 6.

II. RELATED METHODOLOGY
A. CEEMDAN

When EMD technology was proposed by Huang *et al.* in 1998 [16]. EMD can adaptively decompose the original signal into IMF components with different frequencies. EEMD is an improved method based on EMD, which mainly adds different white Gaussian noises to the original sequence multiple times, then separately performs EMD decomposition, and finally obtain the final result by the average of the obtained IMF components, avoiding the phenomenon of mode-mixing [38], [39]. However, in practice, the white noise added by the EEMD method does not be canceled completely

after a number of averaging. On the basis of EEMD, CEEMDAN obtains the IMF by adding adaptive white noise and calculating the unique margin signal to overcome the EEMD deficiency, so that the reconstructed signal is almost identical to the original.

Let $s(n)$ be the wind speed time series, $v^i(n)$ be the white Gaussian noise sequence added by the i-th experiment, and the i-th decomposition wind speed time series can be expressed as $s^i(n) = s(n) + v^i(n)$. The definitions of $E_k(.)$ and are the Kth modal components produced by EMD and CEEMDAN, respectively. The specific steps of the CEEMDAN algorithm are as follows:

(1) Like the EEMD method, CEEMDAN performs I-time decomposition for wind speed time series $s(n) + v^i(n)$. The first modal component is calculated by the EMD method:

$$IMF_1(n) = \frac{1}{I} \sum_{i=1}^{I} IMF_1^i(n) = \overline{IMF_1}(n) \qquad (1)$$

(2) Calculate the first residual sequence:

$$r_1(n) = s(n) - IMF_1(n) \qquad (2)$$

(3) Decompose $(r_1(n)) + \varepsilon_1 E_1(v^i(n))$. for $i(1, 2, \ldots I)$ times Calculate the second modal component:

$$IMF_2(n) = \frac{1}{I} \sum_{i=1}^{I} E_1(r_1(n)) + \varepsilon_1 E_1\left(v^i(n)\right) \qquad (3)$$

(4) Similarly, calculate the kth residue sequence in the remaining phase. Then, according to step (3), the (k +1)th modal component is obtained:

$$r_k(n) = r_{k-1}(n) - IMF_k(n) \qquad (4)$$

$$IMF_{k+1}(n) = \frac{1}{I} \sum_{i=1}^{I} E_1(r_k(n)) + \varepsilon_k E_k\left(v^i(n)\right) \qquad (5)$$

(5) When the residual sequence cannot be decomposed, that is, the maximum number of extremum points of the residual signal is no more than two, the algorithm terminates. At this point, k modal components are obtained, and the final result of the residual sequence is:

$$R(n) = s(n) - \sum_{k=1}^{K} IMF_k \qquad (6)$$

The original wind speed time series $s(n)$ can be expressed as: $s(n) = \sum_{k=1}^{K} IMF_k + R(n)$.

In order to verify the effectiveness and the advantage of the CEEMDAN, Fig. 2 shows the decomposition results of EMD, EEMD and CEEMDAN respectively applied to an original wind speed series. The original wind speed series is shown in Fig. 1. And the reconstruction error between the original wind speed series and the sum of its corresponding IMFs obtained by EMD, EEMD and CEEMDAN are shown in Fig. 3. It can be observed from Fig. 2a, the results of EMD decomposition has severe mode-mixing phenomenon (IMF3-IMF6 in Fig. 2a). Fig. 3 shows that EEMD method
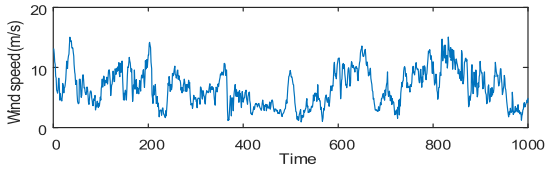
FIGURE 1. Original wind speed time series.

has the largest reconstruction error compared to other methods, it demonstrates the incomplete decomposition of EEMD algorithm. From Figs. 2 and 3, it can be seen that the CEEMDAN method show better decomposition performance and produces a negligible reconstruction error than other two methods. Therefore, the CEEMDAN method is adopted in this study for wind speed time series decomposition.

B. PERMUTATION ENTROPY
PE is a measure of the complexity of time series [37]. It has the advantages of simple calculation, strong anti-noise ability and stable calculation value. It possesses high sensitive to time series changes and good robustness and has been widely used in various time series. The wind speed data has certain randomness and non-stationarity, which makes the IMF component obtained by CEEMDAN decomposition more. Therefore, in order to reduce the computational scale, this article uses PE algorithm to carry out IMF reorganization, the algorithm of PE can be presented as follows:

(1) First, phase space reconstruction is performed on each IMF sequence $\{X(i), i = 1, 2, \ldots, N\}$ obtained by CEEMDAN to obtain a phase space matrix:

$$Y = [x(j), x(j + \tau), \ldots, x(j + (m - 1)\tau)] \quad (7)$$

where m and τ represent the embedding dimension and delay time, respectively; $j = 1, 2, \ldots, N - (m - 1)\tau$.

(2) Taking each row of the above matrix as a component, there are a total of K components, and there is a relationship $K = n - (m - 1)\tau$. Taking the j-th component $x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau)$ as an example, in ascending order according to the element value:

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau)$$
$$\leq \ldots \leq x(i + (j_m - 1)\tau) \quad (8)$$

where $j_1, j_2, \ldots, j_m$ represents the serial number of the component element. If there is, it is $x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau)$ sorted according to the size of j_1 and j_2 values. In summary, for each row vector of the phase space reconstruction matrix of any wind speed time series X_i, a set of sequences $S(l)$ can be obtained.

$$S(l) = \{j_1, j_2, \ldots, j_m\} \quad (9)$$

where, $l = 1, 2, \ldots, K$ and there are $k \leq m!$, m-dimensional spaces have a total of $m!$ different mapping symbol sequences.

(3) Calculate the probability $p_1, p_2, \ldots, p_k$ of occurrence of each symbol sequence $S(l)$. According to the form of Shannon entropy, the permutation entropy $H_p(m)$ of the k-th different symbol sequence of time series $X(i)$ can be defined as:

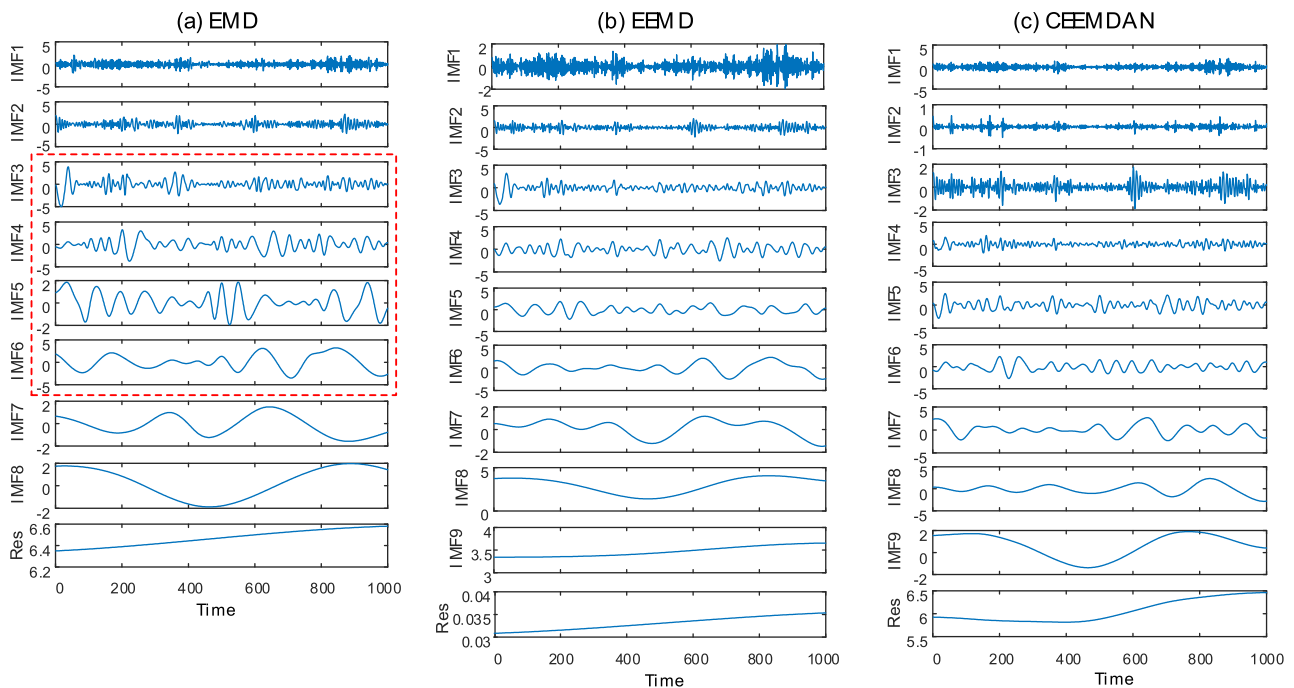$$H_p(m) = -\sum_{j=1}^{k} P_j \ln P_j \quad (10)$$

FIGURE 2. Decomposition results of wind speed by EEMD, EEMD and CEEMDAN.
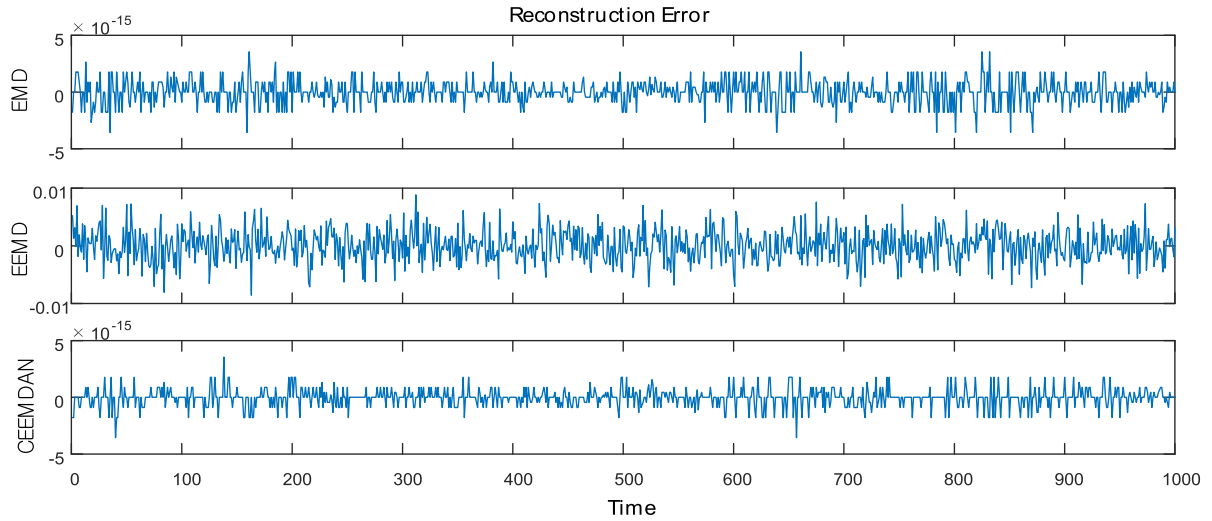
FIGURE 3. Error of reconstruction using EMD, EEMD and CEEMDAN algorithms for wind speed decomposition.

It can be seen from the above formula that when, $P_j = 1/m!$, $H_p(m)$ reaches the maximum value $\ln(m!)$. For the sake of convenience, standardize $H_p(m)$.

$$H_{\mathrm{PE}}(m) = H_{\mathrm{P}}(m)/\ln(m!) \tag{11}$$

where, $0 \le H_{\mathrm{PE}}(m) \le 1$ The size of the PE value represents the complexity and randomness of the time series. The larger the value, the stronger the randomness; on the contrary, the weaker the randomness. The embedding dimension m and the delay time τ are two important parameters of the PE algorithm. Studies have shown that when the sample size is small, the embedding dimension m is usually small; the delay time τ has less influence on the permutation entropy.

Through the above steps, each IMF arrangement entropy is calculated. According to the closeness of the arrangement entropy values, the IMF components are combined to obtain a recombinant subsequence.

C. GATED RECURRENT UNIT

Recurrent Neural Network (RNN) is a special neural network developed for processing time series data [27]. However, as the network layer deepens and the number of neurons increases, RNN often suffers from gradient disappearance and gradient explosion in actual training, which limits its application. To solve these problems, LSTM and GRU are proposed. They are similar in function. However, with the special gate structure, GRU is superior to LSTM in accuracy and speed of prediction and is widely used [30].

Fig. 4 shows the structure of GRU. It has two important gate structures, being r an update gate and z a reset gate. The following calculation formulas are used in the GRU output calculations:

(1) Reset Gate:
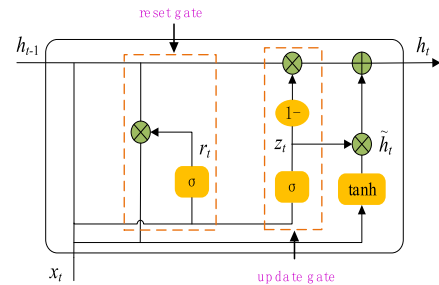
$$r_t = \sigma\left(\omega_r \cdot [h_{t-1}, x_t]\right) \tag{12}$$

FIGURE 4. Structure of a GRU cell.

(2) Update Gate:

$$z_t = \sigma\left(\omega_z \cdot [h_{t-1}, x_t]\right) \tag{13}$$
$$\tilde{h}_t = \tanh\left(\omega \cdot [r_t \cdot h_{t-1}, x_t]\right) \tag{14}$$
$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{15}$$

(3) Output:

$$y_t = \sigma\left(\omega_o \cdot h_t\right) \tag{16}$$

Where: h_t and h_{t-1} represent the historical information; $\tilde{h}_t$ represents the candidate activation value; x_t represents the current sequence input; σ and $\tanh$ represent the activation function; $\omega_r, \omega_z, \omega, \omega_o$ are training parameter matrix.

III. THE NONLINEAR-LEARNING ENSEMBLE OF GRU TIME SERIES PREDICTION FOR SHORT-TERM WIND SPEED FORECASTING

To achieve a better wind speed forecasting, in the structure of nonlinear-learning ensemble learning, the predicted values of a cluster of GRUs, for each recombination sequence, are input to a nonlinear-learning regression top-layer to obtain the final predicted values of each recombination sequence. Considering the RBF neural network can approximate any nonlinear function and the convergence rate is fast.

Thus, RBFNN is introduced as the nonlinear-learning top-layer. To improve the predictive ability of the nonlinear-learning regression top-layer, an improved bat algorithm, called IBA, is proposed to optimize the center vector c_i, base width σ_i and weight w_{ij} of RBF neural network. The basic concepts of RBFNN and IBA are described below.

A. PRINCIPLE OF RBF NEURAL NETWORK

The RBF neural network is a highly efficient three-layer neuron feedforward neural network, which can approximate any nonlinear function and has a fast convergence rate [40].

The RBF neural network consists of an input layer, an implicit layer, and an output layer, in which the input layer is used to connect the network to the external environment; There are many forms of the radial basis function of the hidden layer. The Gaussian function has the advantages of radial symmetry, good smoothness and simple expression. Therefore, Gaussian function is chosen as the radial basis function to map the input vector from the input space to the high-dimensional space. The output layer is the linear output of the hidden unit.

The typical structure of an RBF neural network is shown in Fig. 5. Where, the input is: $x = [x_1, x_1, \ldots, x_n]^T$, the output of the hidden layer is: $h = [h_1, h_2, \ldots, h_n]^T$, h_1 represents the output of the i-th neuron in the hidden layer:

$$h_i = \exp(\frac{\|x - c_i\|^2}{2\sigma_i^2}) \qquad (17)$$

where $1 \leq i \leq t$, σ_i is the width of the Gaussian function of the i-th neuron of the hidden layer, c_i is the central value of the t-th neuron Gaussian function of the hidden layer. $w \in R^{t \times n}$ is the output weight matrix of the RBF network, $y = [y_1, y_2, \ldots, y_m]^T$ is the output of the RBF network, y_j represents the jth output of the hidden layer:

$$y_j = \sum_{i=1}^{m} w_{ij} h_i (\|x - c_i\|) \qquad (18)$$
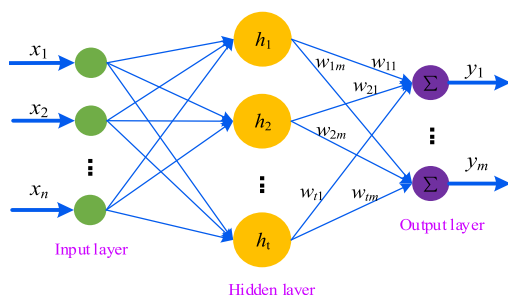
where $1 \leq j \leq m$.

FIGURE 5. The structure of RBF neural network.

B. PROPOSED IMPROVED BAT ALGORITHM (IBA)

BA is the most recent nature inspired optimization algorithm proposed by Professor Yang in 2010 to simulate bat hunting with ultrasound. It combines the advantages of intelligent algorithms such as particle swarm and ant colony to achieve a major breakthrough in convergence rate and optimization ability. The BA is based on the analysis of bat behavior characteristics, and establishes mathematical models with pulse frequency f, pulse intensity A, velocity v, and pulse emissivity BA. The basic steps of the BA are as follows:

(1) Parameter space initialization: Initialize the parameter space. That of the BA algorithm consists of the number of iterations $iter$, the position vector x, the velocity vector v, the pulse intensity A, the pulse emissivity r, and the pulse range $[f_{\min}, f_{\max}]$.

(2) Calculate the fitness value of the bat and find the current optimal solution. And record the location of the optimal solution bat.

(3) Iterative training: update the frequency f_i, speed v_i and position x_i of individual bats according to Eq. (19), Eq. (20) and Eq. (21)

$$f_i = f_{\min} + (f_{\max} - f_{\min}) \times \beta_i \qquad (19)$$
$$v_i^t = v_i^{t-1} + (x_i^t - x^*) \times f_i \qquad (20)$$
$$x_i^t = x_i^{t-1} + v_i^t \qquad (21)$$

where $\beta \in [0, 1]$ is a random vector that satisfies a uniform distribution, x^* is the current global optimal solution, and t is the current number of iterations. In addition, during the parameter initialization process, A indicates the frequency at which the i-th bat emits sound waves, $f_i \in [f_{\min}, f_{\max}]$.

(4) Generate a random number: $rand$. If the condition $rand > r_i$ is satisfied, according to the Eq. (22), search for the local new solution near the current global optimal solution

$$x_{new} = x_{old} + \varepsilon A^t \qquad (22)$$

where ε is the random number satisfying $[-1, 1]$, and A^t is the average pulse intensity of all bats.

(5) Generate a random number rand, if the new solution satisfies the following condition: $rand < A_i \& f(x_i) < f(x^*)$, the pulse intensity and pulse transmission frequency can be updated according to the Eq. (23) and Eq. (24).

$$A_i^{t+1} = \beta A_i^t \qquad (23)$$
$$r_i^t = r_i^0 [1 - \exp(-\gamma t)] \qquad (24)$$

where γ is the pulse transmission frequency enhancement coefficient, $\gamma > 0$ and β are the attenuation coefficients of the loudness, $\beta \in [0, 1]$.

(6) Evaluate the bat population, and search for the bat individual with the minimum fitness value.

(7) Determine whether the algorithm termination condition is met, and if yes, proceed to step (8), otherwise, returns to step (2) to continue the iterative training.

(8) The algorithm ends, outputting the position vector and fitness value of the optimal bat individual.

Because BA balances the global and local search process well, it has a performance advantage over traditional intelligent algorithms. However, as the number of digits in the optimization problem increases, its performance may degrade.

TABLE 1. Benchmark function for testing of proposed IBA algorithm.

Function name	Domain	Optimum value	Test function
Sphere	[-5.12,5.12]	0	$f(x) = \sum_{i=1}^{d} x_i^2$
Rosenbrock	[2.084,2.084]	0	$f(x) = \sum_{i=1}^{d-1}[100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2]$
Rastrigin	[-5.12,5.12]	0	$f(x) = \sum_{i=1}^{d}(x_i^2 - 10(2\pi x_i) + 10)$
Schaffer	[-5.12,5.12]	0	$f(x) = \dfrac{\sin^2 \sqrt{\sum_{i=1}^{d} x_i^2} - 0.5}{[1 + 0.001(\sum_{i=1}^{d} x_i^2)]^2} + 0.5$

In order to improve that, relevant literature has proposed some research. Nevertheless, the parameters of these modified BAs are obtained through trial and error methods, while introducing cumbersome computational problems. Based on the Eq. (21), this article introduces the dynamic inertia weight, increases the speed diversity of BA in the global search and local search process, and does not increase the excessive computational burden. The update is shown in Eq. (25):

$$v_i^{t+1} = \omega_i(t)v_i^t + (x_i^t - x^*).f_i \qquad (25)$$

where $\omega(t)$ is the dynamic inertia weighting factor, and its value can affect the convergence rates. In order to increase the diversity of the algorithm in speed, the expression of $\omega_i(t)$ is:

$$\omega_i(t+1) = \omega_i(t) + N(0,1) \qquad (26)$$

where $N(0,1)$ is a random variable subject to a normal distribution.

In order to enhance the diversity of the bat position, the loudness attenuation coefficient in the loudness Eq. (23) is changed to improve its adaptability.

The loudness attenuation coefficient A, when taken as a smaller value, will give the algorithm a local search capability, while the larger value will encourage the algorithm to search globally. In order to achieve an appropriate balance between global search and local search, the experiment found the following dynamic formula:

$$\beta^{new} = (\frac{1}{2t})^{1/2t}\beta^{old} \qquad (27)$$

C. BENCHMARK FUNCTIONS FOR VALIDATION OF THE IBA ALGORITHM

As shown in Table 1, four typical benchmark functions, namely, Sphere, Rosenbrock, Rastrigin and Schaffer are applied to validate the performance of IBA and BA algorithm. The test of all benchmark functions was run in MATLAB R2018b on Windows 10 with a 3.30 GHz Intel Core i5-4590 CPU and 8.00 GB of RAM.

The experiment parameters of IBA and BA are listed in Table 2. From Table 3, two points can be concluded:

(a) For the four test functions, the maximum, minimum, and average number of iterations of the IBA are less than that of BA. This means that IBA has successfully improved the convergence ability of BA.

TABLE 2. The experiment parameters of BA and IBA.

Experimental parameters	BA	IBA
Maximum generation	1000	1000
Population size	50	50
Con vergence tolerance	10^{-5}	10^{-5}
Maximum generation of IBA	-	20

(b) For the Rastrigin's function and the Schaffer's function, the convergence rates of BA didn't obtain 1. For these two test functions, the convergence rate of the IBA is 1. Therefore, the optimization performance of the IBA algorithm is also significantly improved compared to BA.

D. CONSTRUCT IBA-RBF ALGORITHM

The RBF neural network has a strong advantage in dealing with complex nonlinear mapping problems. However, the performance of an RBF neural network is highly dependent on its center vector, base width, and weight. This article introduces an improved bat algorithm to optimize the center vector, base width and weight of the RBF neural network to overcome its problem of easy convergence to local optimum. The objective function chosen by this article is:

$$f = (\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2)^{1/2} \qquad (28)$$

where N is the number of samples, y_n is the actual value, and $\hat{y}_n$ is the predicted value at time n.

The flow chart of IBA optimization RBFNN is shown in Fig. 6.

IV. MAIN STRUCTURE OF PROPOSED COMBINED MODEL

In this section, A novel combined model is proposed, which takes CEEMDAN-PE technique, a cluster of GRUs with diverse hidden layers and neurons and IBA optimize the parameters of RBFNN together. The main process of is demonstrated in Fig. 7. The detailed processes of the proposed combined model are as follows:

A. STEP1: DATA PREPROCESSING STRATEGY

Due to the intermittent characteristics and highly noisy of wind speed series seriously restrict the wind speed

TABLE 3. Test results of BA and IBA.

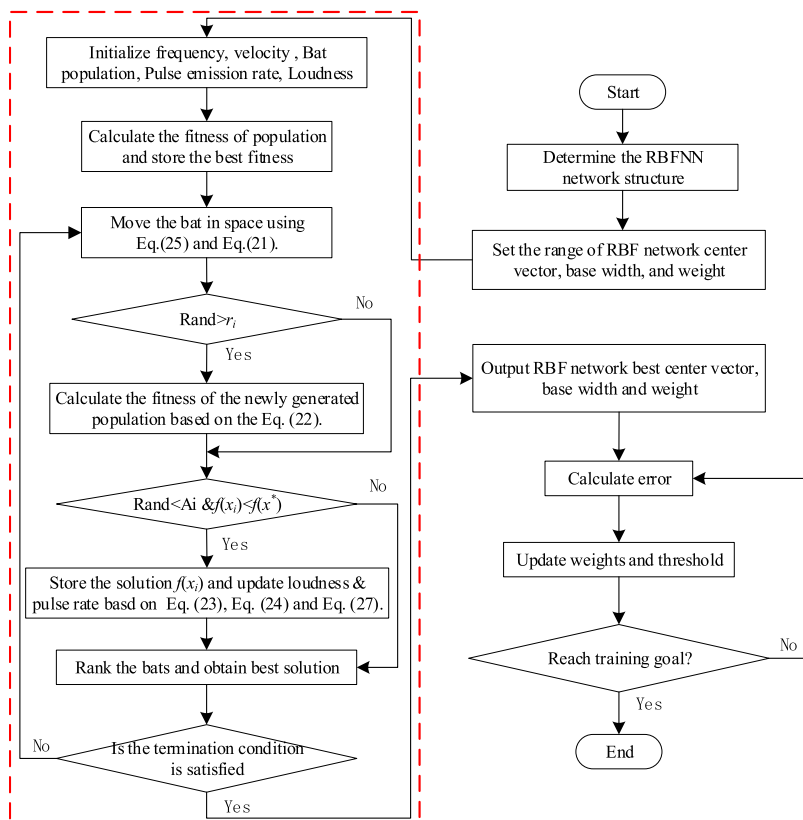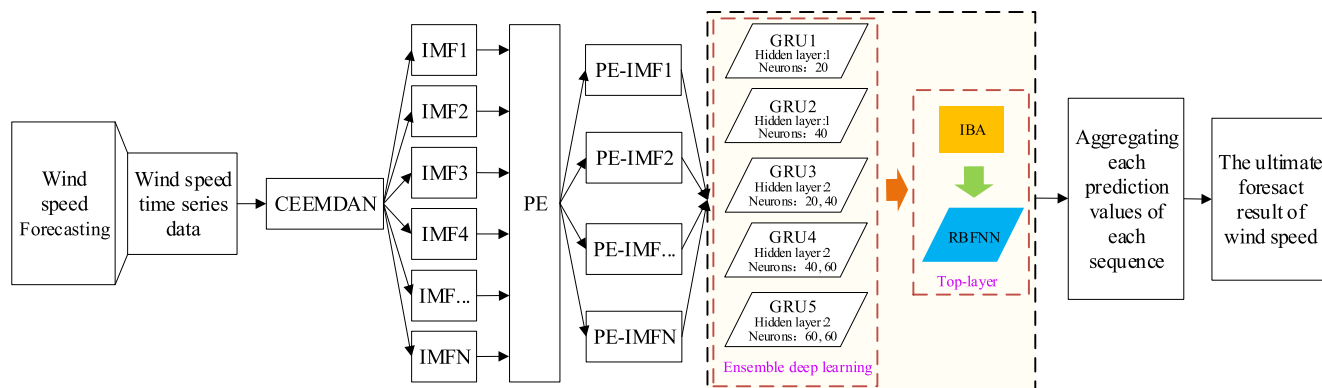| Test function	Algorithm	Max value of iteration	Min value of iteration	Average value of iteration	Convergence rate (%)
Sphere	BA	521	374	431	100
	IBA	210	101	133	100
Rosenbrock	BA	215	179	156	100
	IBA	126	90	74	100
Rastrigin	BA	432	351	374	83
	IBA	283	197	243	100
Schaffer	BA	1121	963	992	87
	IBA	409	151	254	100

FIGURE 6. The flowchart of IBA-RBFNN.

FIGURE 7. The structure of the proposed combined model.

prediction accuracy, the CEEMDAN technique was employed to decompose raw wind speed series into several IMF signals. Next, PE was employed to calculate the PE values of each IMF, and then to recombine the IMFs into a new subseries (PE-IMFs) to solve the problem of over-decomposition and computational burden.
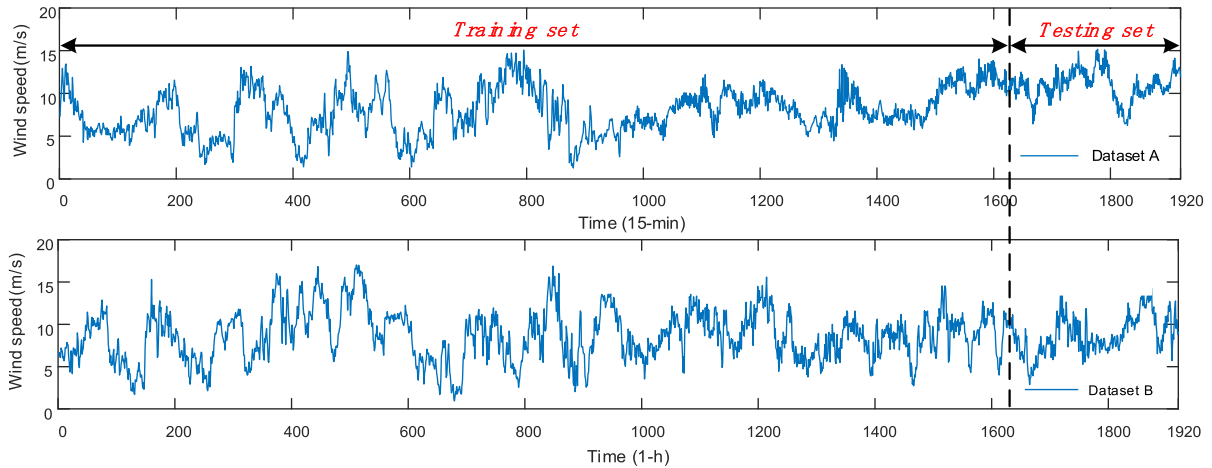
FIGURE 8. The original wind speed time series from Zhangjiakou wind farm.

B. STEP2: THE NONLINEAR-LEARNING ENSEMBLE OF DEEP LEARNING TIME SERIES PREDICTION FOR PE-IMFS

Since different signals have different characteristics, individual model can no longer adapt to all properties of the PE-IMFs. To this end, a cluster of GRUs with diverse hidden layers and neurons are employed separately is proposed for forecasting PE-IMFs with different characteristics.

It should be noted that the network structure of the GRU cannot be predefined for specific data. In the actual problem, the parameters are selected through trial and error to make a trade-off between the learning performance and the complexity of the model. After that, the GRU cluster used in the integrated learning part of this article consists of five different GRUs.

Five diverse GRUs are adopted based on trial and error: GRU1: 1 hidden layer with 20 neurons. GRU2: 1 hidden layer with 40 neurons. GRU3: 2 hidden layers with 20,40 neurons. GRU4: 2 hidden layers with 40,60 neurons. GRU5: 2 hidden layers with 60,60 neurons. Five different GRU networks are trained as a single predictive model to forecast the PE-IMFs on train and test dataset. The prediction of five GRUs models on training set is the input as train-feature into IBA-RBFNN to learn. The task of IBA-RBF is to learn the nonlinear relationship of five GRUs predictors, similar to solving multiple regression problems.

Final, the prediction results of the five GRU models on the test dataset are input to the well-trained IBA-RBFNN to obtain the prediction result of the recombinant sub-sequence PE-IMFs.

C. STEP3: ASSEMBLE FORECASTING RESULT

The combined forecasting result of each PE-IMF is obtained by step 2. Then, the forecasting results of which are superimposed to get the final result.

V. EXPERIMENTS AND ANALYSIS
A. WIND SPEED DATA DESCRIPTION

The Zhangjiakou area has a rich reserve of wind energy resources, especially in the Bashang area. The Zhangjiakou Bashang area is a rare wind energy resource gathering area in China. The city's wind energy resource reserves in Zhangjiakou reach more than 20 million KW (70% of the land wind resource reserves in Hebei Province).

In this article, two historical short-time wind speed datasets i.e. dataset A with a 15-min interval, and dataset B with a 1-hour interval are regarded as two illustrative examples to test the forecasting performance of the proposed combined model. The wind speed time series is displayed in Fig. 8. Dataset A: The wind speed data were sampled at an interval of fifteen-min from November 21, 2013 to December 10, 2013, 1920 data points in total. Dataset B: The wind speed data were sampled at an interval of one-hour from January 22, 2013 to April 10, 2013, 1920 data points in total. Each dataset is classified into a training set (the first 1632 values) and a testing set (the last 288 values). Specifically, the first 816 values of training set, were applied to train five GRUs models; And the last 816 values of training set, were applied to train the nonlinear-learning regression top-layer model. The statistical information of the above dataset is showed in Table 4. And the proposed combined model was implemented though mixed language programming based on Python and MATLAB. And the GRU Neural Networks was operated by using the "Keras" deep learning package.

TABLE 4. The statistical information of wind speed in two case studies.

Dataset	Samples	Number	Statistical Indicator (m/s)			
			Mean	Max	Min	Std.
Dataset A	All samples	1920	8.3695	15.78	1.27	2.7714
	Training	1632	8.1587	15.02	1.27	2.8832
	Testing	288	10.9567	15.78	5.47	1.8255
Dataset B	All samples	1920	8.66849	17.01	0.95	3.0726
	Training	1632	8.3485	17.01	0.95	3.2524
	Testing	288	9.8851	15.10	5.28	2.5102

TABLE 5. Four evaluation criteria recombined by PE value of CEEMDAN subsequence.

	IMFn	$H_{PE}(n)$	Recombination	New IMFn		IMFn	$H_{PE}(n)$	Recombination	New IMFn
Dataset A	IMF1	0.9212	IMF1	PE-IMF1	Dataset B	IMF1	0.9305	IMF1	PE-IMF1
	IMF2	0.8084	IMF2&IMF3	PE-IMF2		IMF2	0.7629	IMF2&IMF3	PE-IMF2
	IMF3	0.7719				IMF3	0.6818		
	IMF4	0.4889	IMF4	PE-IMF3		IMF4	0.4715	IMF4	PE-IMF3
	IMF5	0.3188	IMF5& IMF6	PE-IMF4		IMF5	0.3034	IMF5& IMF6&IMF7	PE-IMF4
	IMF6	0.2524	IMF7& IMF8&IMF9 & IMF10& IMF11	PE-IMF5		IMF6	0.2325		
	IMF7	0.1728				IMF7	0.1874		
	IMF8	0.1441				IMF8	0.1387	IMF8&IMF9&IMF10 & IMF11& IMF12	PE-IMF5
	IMF9	0.1244				IMF9	0.1217		
	IMF10	0.1139				IMF10	0.1101		
	IMF11	0.0727				IMF11	0.1108		
						IMF12	0.0645		

B. EVALUATION OF FORECASTING PERFORMANCE

Four evaluation criteria are considered to check the forecast performance of the proposed models [42]. And they are defined as shown in Table 6.

TABLE 6. Four evaluation criteria.

Metric	Equation	Definition
MAE	$MAE = \frac{1}{N}\sum_{n=1}^{N}\|y_n - \hat{y}_n\|$	Mean absolute error of N forecasting results
RMSE	$RMSE = (\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2)^{1/2}$	Square root of average of error squares
MAPE	$MAPE = \frac{1}{N}\sum_{n=1}^{N}\|\frac{y_n - \hat{y}_n}{y_n}\|\times 100\%$	Average of N absolute percentage error
SSE	$SSE = \sum_{n=1}^{N}(y_n - \hat{y}_n)^2$	The square sum of the error

Here y_n and $\hat{y}_n$ represent the actual value and predicted value at time n, respectively. N is the sample size.

C. DATA DECOMOPOSITION BY CEEMDAN-PE

CEEMDAN decomposition was performed on the wind speed time series, and I = 500 groups of white noise signals with a non-standard deviation of 0.2 were added to obtain IMF components. The permutation entropy values of the IMF components are calculated, and the results are shown in Table 6. In order to avoid excessive decomposition and reduce the calculation scale, the IMF components with similar entropy values are superimposed to obtain the recombination component PE-IMFs, according to the permutation entropy values shown in Table 5. The recombinant sequences are shown in Fig. 6 and Fig. 7.

D. CASE STUDY 1: 15-min TIME-SCALE WIND SPEED FORECASTING

In this case study, the wind speed data of Dataset A is used to perform 15-min ahead utmost short-term wind speed forecasting. The wind speed is displayed in Fig.8.

1) EXPERIMENT 1: COMPARISON WITH MODELS USING DIFFERENT DATA DECOMPOSITION TECHNIQUES

Experiment I was designed to compare proposed combined model with three models based on different data decomposition techniques, namely WPD based model, EMD based model and EEMD based model. Fig. 11 shows the forecasting performances comparison of proposed combined model and three models. The comparison results are shown in Table 7. From the table, the proposed combined model achieves the best forecasting performance, with the MAE, RMSE, MAPE and SSE of 0.2778, 0.3294, 2.6169% and 31.2580. Then, the models from highest to lowest based on the forecasting accuracy are EEMD based model, EMD based model and WPD based model with MAPE values of 2.8926%, 3.1513% and 3.2723%.

TABLE 7. Forecasting performance comparison table of combined model and models using different data decomposition techniques in case study 1.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.2778**	**0.3294**	**2.6169**	**31.2580**
WPD based model	0.3471	0.4173	3.2723	49.7932
EMD based model	0.3312	0.4031	3.1513	46.8145
EEMD based model	0.3032	0.3708	2.8926	38.6086

2) EXPERIMENT 2: COMPARISON WITH MODELS USING SAME DATA DECOMPOSITION TECHNIQUE

In order to verify the forecasting performance of the proposed combined model, compare it with five models only based on CEEMDAN-PE, namely CEEMDAN-PE-GRU, CEEMDAN-PE-LSTM, CEEMDAN-PE-SVR, CEEMDAN-PE-BP and CEEMDAN-PE-RBF. Fig. 12 visualizes the forecasting results and residual errors of the proposed combined model and other models on Dataset A. The forecasting performance of a combined model and the comparison with other models are listed in Table 8. Fig. 13 visualizes the comparison
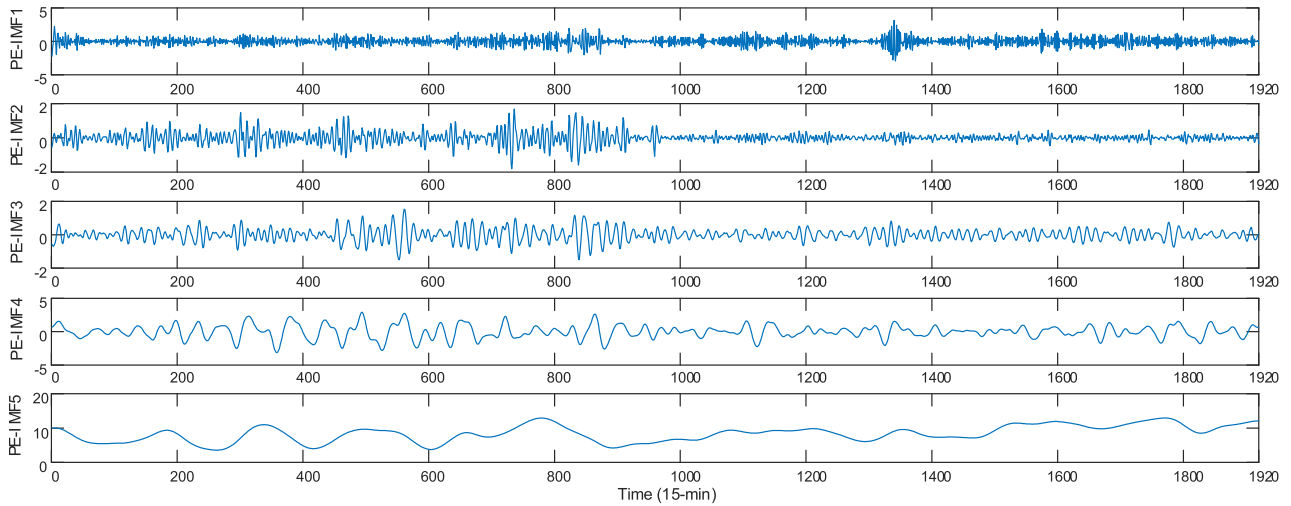
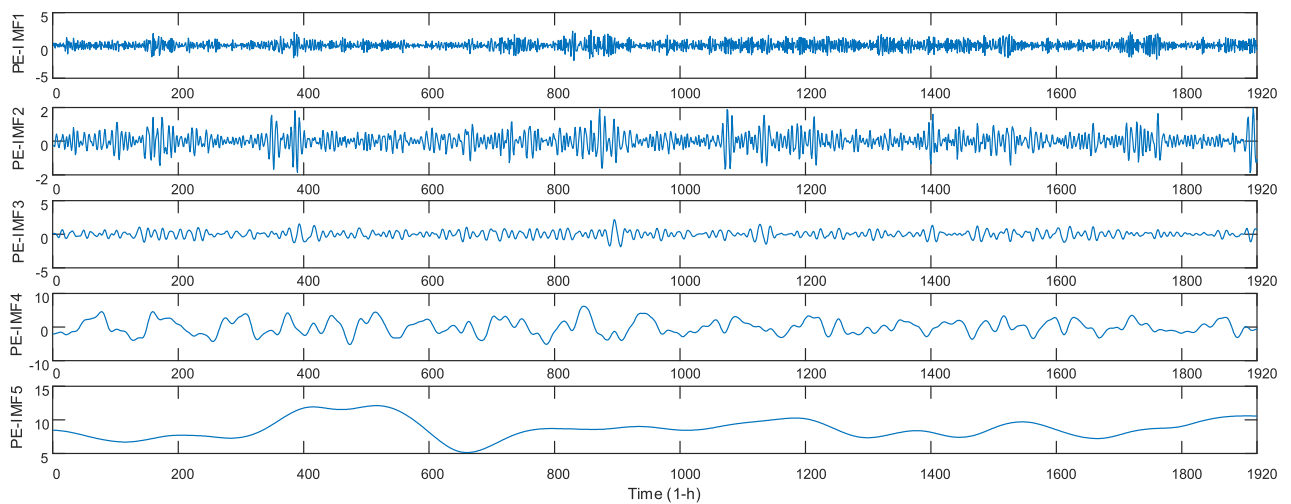**FIGURE 9.** Recombination results for the dataset A decomposed using CEEMDAN-PE.

FIGURE 10. Recombination results for the dataset B decomposed using CEEMDAN-PE.

TABLE 8. Forecasting performance comparison table of combined model and models using same data decomposition technique in case study 1.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.2778**	**0.3294**	**2.6169**	**31.2580**
CEEMDAN-PE-SVR	0.5157	0.6854	4.9521	106.2218
CEEMDAN-PE-BP	0.4924	0.6541	4.6432	98.5844
CEEMDAN-PE-RBF	0.4735	0.5981	4.4607	93.0497
CEEMDAN-PE-LSTM	0.4014	0.4720	3.7818	64.1748
CEEMDAN-PE-GRU	0.3954	0.4774	3.5795	65.6517

of forecasting performance. The lowest MAE, RMSE, MAPE and SSE values among all comparison models are marked in bold. Clearly, the proposed combined model shows the highest accuracy.

Moreover, from the analysis of Figs. 12 and 13, a comparison of the forecasting results indicates that the proposed combined model has the best accuracy and stability when compared with other models, shows apparently a better curve fitting of the actual wind speed time series and smaller residual errors. Meanwhile, it can be clearly seen that compared with the CEEMDAN-PE-GRU, the ensemble of GRUs has improved the forecasting performance of single GRU.

3) EXPERIMENT 3: COMPARISON WITH MODELS USING DIFFERENT NONLINEAR-LEARNING TOP-LAYER

In experiment 3, in order to analysis the impacts of the nonlinear-learning top-layer on prediction performance, three forecasting models of different nonlinear-learning top-layers are used for comparison with the proposed combined model, namely CEEMDAN-PE-GRU-SVR (using SVR model as nonlinear-learning top-layer), CEEMDAN-PE-GRU-BP (using BP model as nonlinear-learning top-layer) and CEEMDAN-PE-MeanGRU (the ensemble learning is an average forecasting result). Fig. 14 shows the forecasting
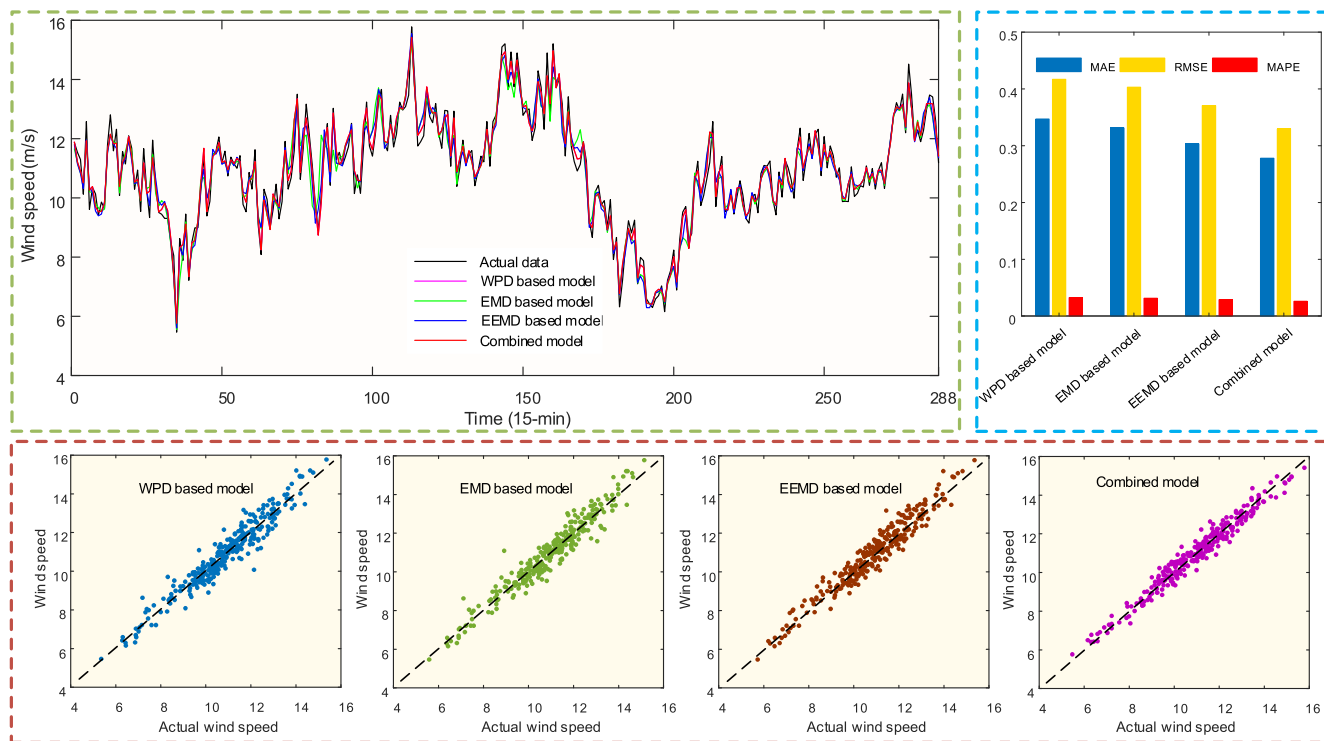
FIGURE 11. Forecasting results of combined model and models employing diverse data decomposition techniques in case study 1.

performances comparison of proposed combined model and three models. The forecasting results of each model are shown in Table 9. From Fig. 14 and Table 9. It can be seen that the proposed combined model can achieve better forecasting performance, which indicates the superiority of nonlinear-learning top-layer composed by IBA-RBF model.

TABLE 9. Forecasting performance comparison table of combined model and models using different ensemble learning top-layer in case study 1.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.2778**	**0.3294**	**2.6169**	**31.2580**
CEEMDAN-PE-GRU-SVR	0.3083	0.3889	2.9561	40.7945
CEEMDAN-PE-GRU-BP	0.3212	0.3986	3.0628	43.2564
CEEMDAN-PE-MeanGRU	0.2932	0.3506	2.8621	36.0231

E. CASE STUDY 2: 1-H TIME-SCALE WIND SPEED FORECASTING

As described above, the proposed combined model has shown its superiority in 15-min time-scale wind speed forecasting. In order to investigate the potential of the model in 1-h time-scale wind speed forecasting, in case study 2 three experiments are designed to validate the accuracy of the proposed combined model.

1) EXPERIMENT 1: COMPARISON WITH MODELS USING DIFFERENT DATA DECOMPOSITION TECHNIQUES

Experiment I was designed to compare proposed combined model with three models using different data decomposition

techniques, namely WPD based model, EMD based model and EEMD based model. Fig. 15 shows the forecasting performances comparison of proposed combined model and three models. The comparison results are shown in Table 10.

From Table 10 and Fig. 15, the proposed combined model achieves the best forecasting performance, with the MAE, RMSE, MAPE and SSE of 0.4453, 0.5876, 4.7875% and 96.9548. Then, the models from highest to lowest based on the forecasting accuracy are EEMD based model, EMD based model and WPD based model with MAPE values of 4.9072%, 5.2324 % and 5.4861%.

TABLE 10. Forecasting performance comparison table of combined model and models using different data decomposition techniques in case study 2.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.4453**	**0.5876**	**4.7875**	**96.9548**
WPD based model	0.5031	0.6788	5.4861	113.5632
EMD based model	0.4935	0.6356	5.2324	111.1425
EEMD based model	0.4801	0.6188	4.9072	104.1894

2) EXPERIMENT 2: COMPARISON WITH MODELS USING SAME DATA DECOMPOSITION TECHNIQUE

Experiment 2 shows a comparison between the proposed combined model and five models only based on CEEMDAN-PE, namely CEEMDAN-PE-GRU,CEEMDAN-PE-LSTM, CEEMDAN-PE-SVR, CEEMDAN-PE-BP and CEEMDAN-PE-RBF. The specific forecasting results are listed in Table 11. Fig. 16 provide the bar graph of prediction
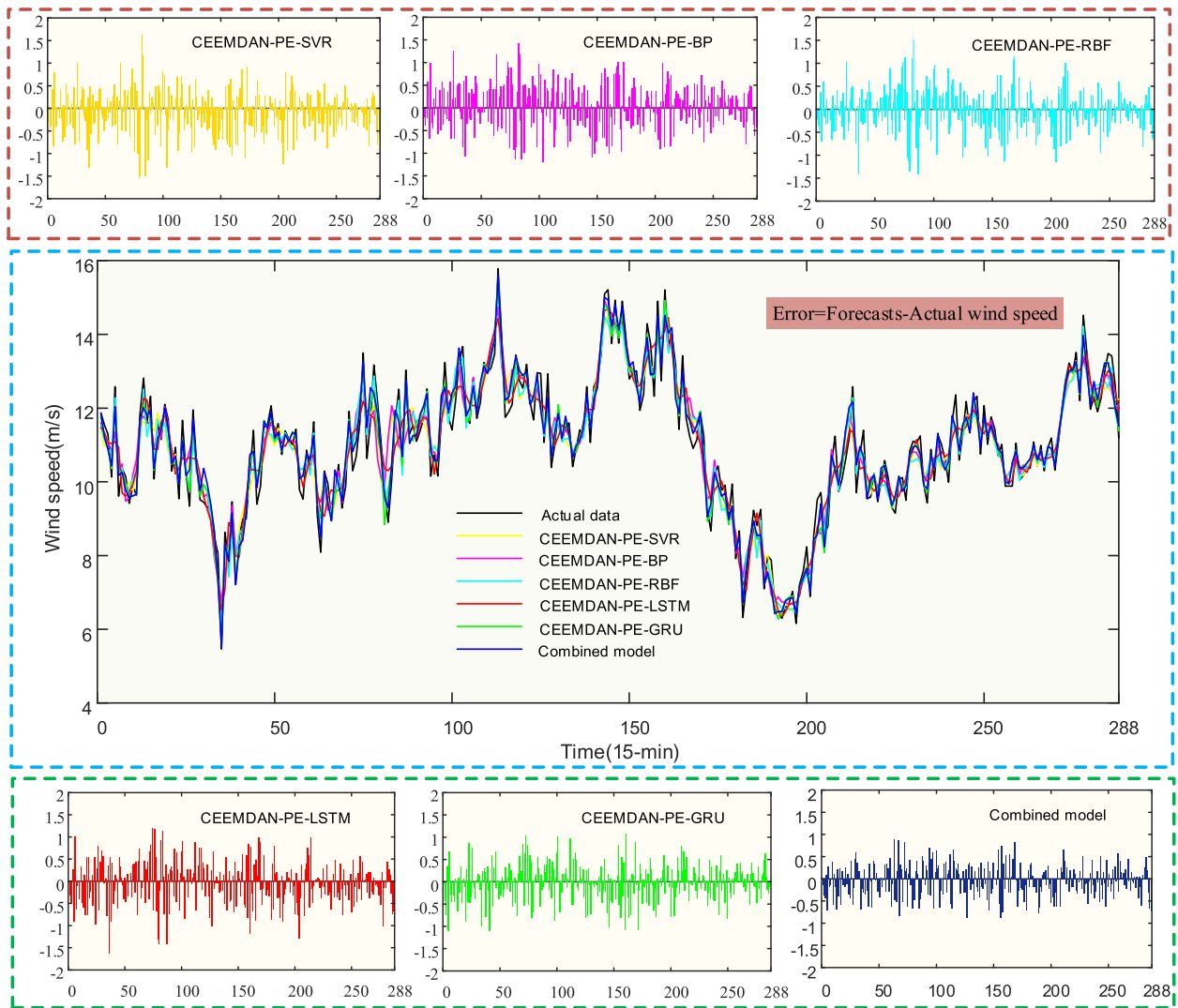
FIGURE 12. Forecasting results of combined model and models using same data decomposition technique in case study 1.
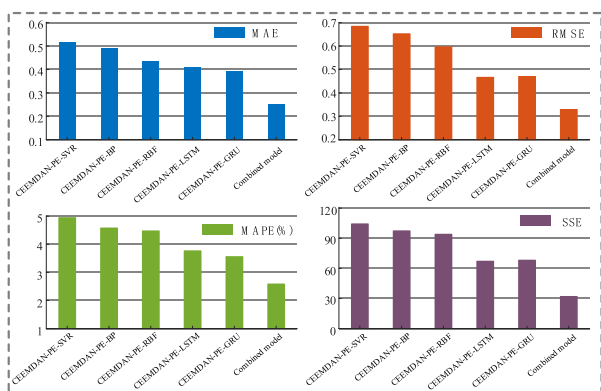

FIGURE 13. The forecasting performance comparison of combined model and models using same data decomposition technique in case study 1.

TABLE 11. Forecasting performance comparison table of combined model and models using same data decomposition technique in case study 2.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.4453**	**0.5876**	**4.7875**	**96.9548**
CEEMDAN-PE-SVR	0.6106	0.7856	6.8826	157.1289
CEEMDAN-PE-BP	0.59737	0.7559	6.5704	148.2902
CEEMDAN-PE-RBF	0.5839	0.7529	6.4221	145.4651
CEEMDAN-PE-LSTM	0.5344	0.7216	5.8783	125.9659
CEEMDAN-PE-GRU	0.5079	0.6809	5.4663	114.3327

residual errors for different forecast methods and visualizes the comparison of forecasting results. Fig. 17 visualizes the comparison of forecasting performance.

From Tables 11 and Fig. 16, it can be clearly seen that models only based on CEEMDAN-PE are far away from true data, and our proposed combined model performs superior wind speed forecasting. Next, the models from highest to lowest based on the forecasting accuracy are CEEMDAN-PE-GRU, CEEMDAN-PE-LSTM, CEEMDAN-PE-RBF, CEEMDAN-PE-BP and CEEMDAN-PE-SVR with MAPE values of
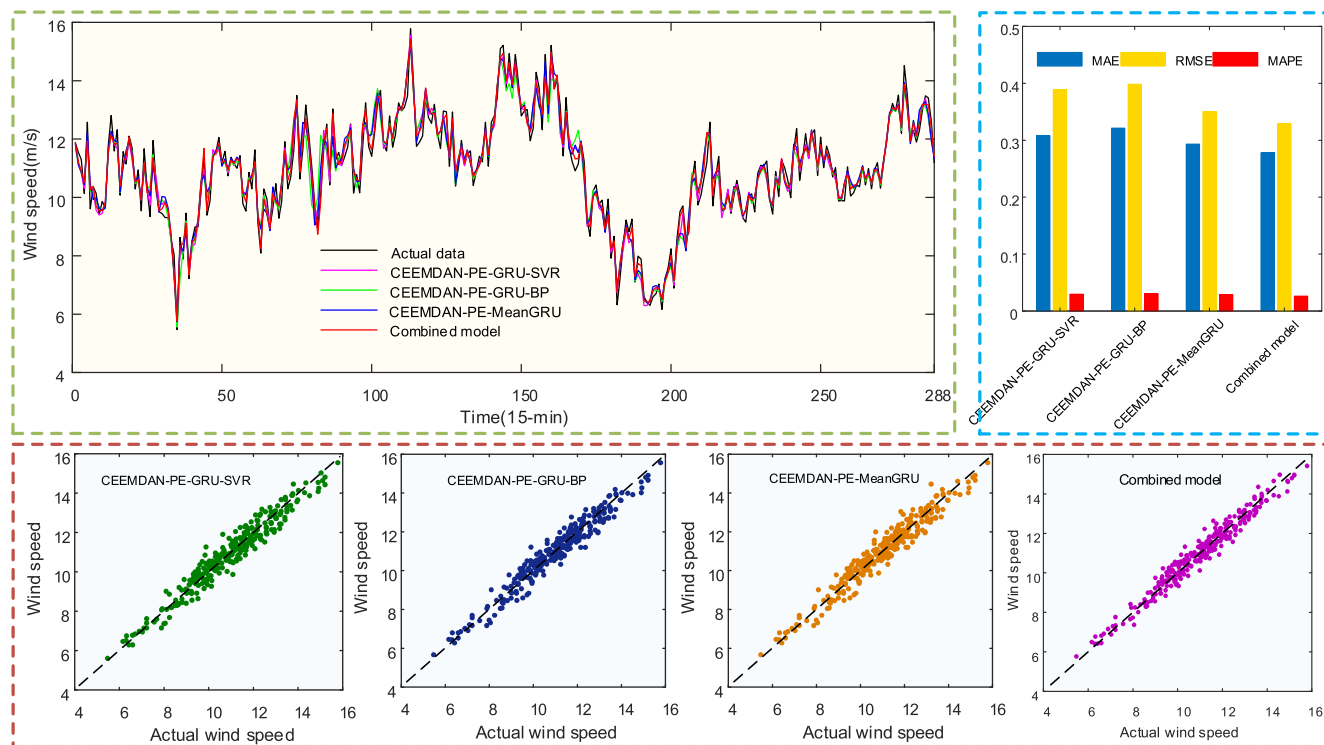
FIGURE 14. Forecasting results of combined model and models employing different nonlinear-learning top-layer in case study 1.
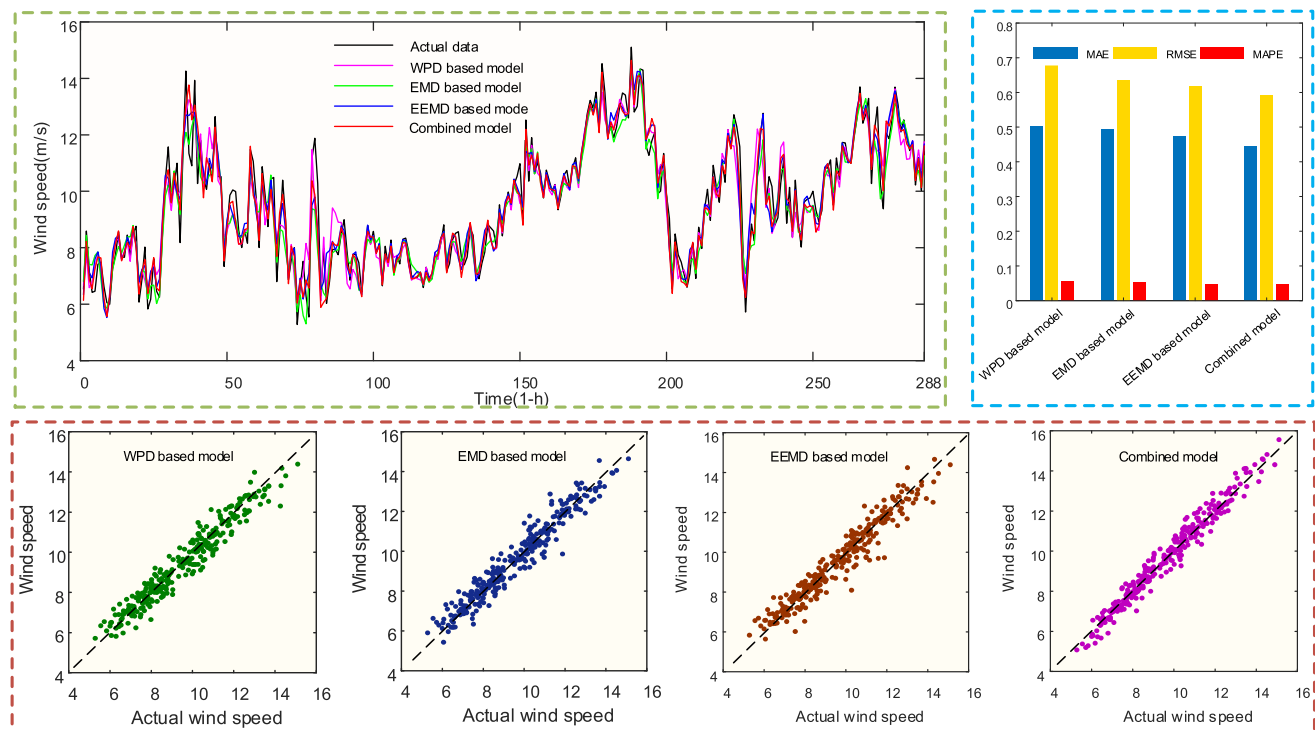

FIGURE 15. Forecasting results of combined model and models employing different nonlinear-learning top-layer in case study 2.

3.5795%, 3.7818%, 4.4607%, 4.6432% and 4.9521%. Similar to case study 1, compared with the CEEMDAN-PE-GRU, the ensemble of GRUs has improved the forecasting performance of single GRU.

3) EXPERIMENT 3: COMPARISON WITH MODELS USING DIFFERENT NONLINEAR-LEARNING TOP-LAYER

The aim of experiment 3 is to compare the proposed combined model with models using different nonlinear-learning
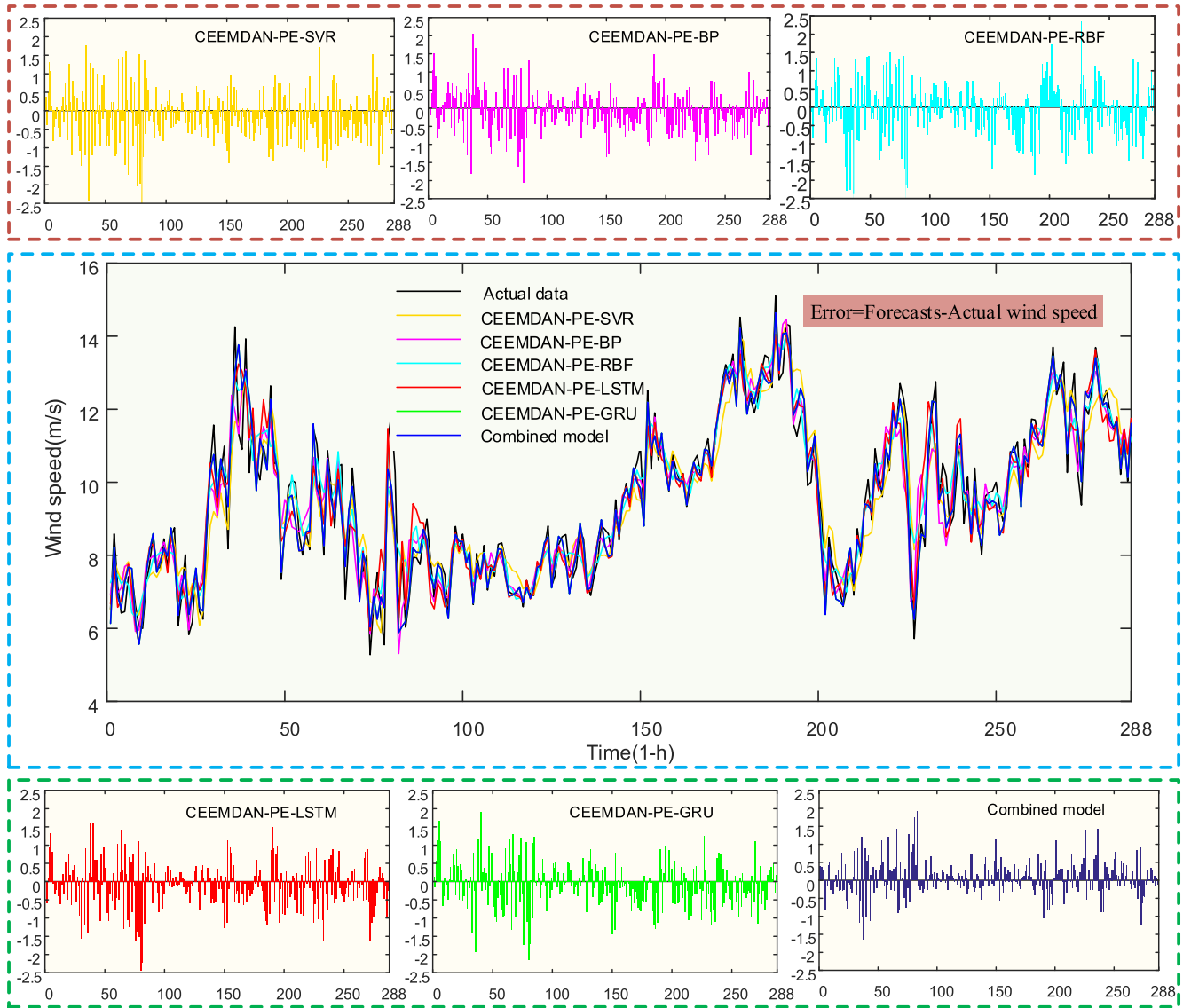
FIGURE 16. Forecasting results of combined model and models using same data decomposition technique in case study 2.

top-layer. Similar to case study 1, three forecasting models of different top-layers are used to analysis of impacts of the ensemble learning top-layer on prediction performance. The forecasting results of each model are shown in Table 12. Fig. 18 shows the forecasting performances comparison of proposed combined model and other models.

TABLE 12. Forecasting performance comparison table of combined model and models using different ensemble learning top-layer in case study 1.

Model	MAE	RMSE	MAPE(%)	SSE
Combined model	**0.4453**	**0.5876**	**4.7875**	**96.9548**
CEEMDAN-PE-GRU-SVR	0.4981	0.6702	5.2942	111.8856
CEEMDAN-PE-GRU-BP	0.4854	0.6488	5.0857	108.5641
CEEMDAN-PE-MeanGRU	0.4665	0.6196	4.9865	101.4685

It can be observed from Table 12 and Fig.18 that our proposed combined model performs superior wind speed forecasting when compared with CEEMDAN-PE-GRU-SVR, CEEMDAN-PE-GRU-BP and CEEMDAN-PE-MeanGRU, which is benefited from nonlinear-learning ensemble top-layer of IBA-RBF.

Remark: Compared with Case Study 1, it can be seen that when the forecasting time is extended from fifteen minutes to one hour, influenced by the volatility and abruptness of wind resources, the forecasting errors of each model also become correspondingly larger. 1-h time-scale wind speed forecasting is more complicated and difficult than 15-min time-scale wind speed forecasting. However, the accuracy and stability of prediction of the combined model proposed in this article are still optimal compared to other models.
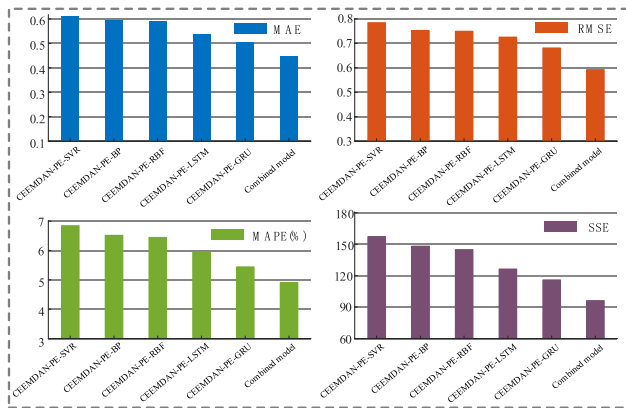
FIGURE 17. The forecasting performance comparison of combined model and models using same data decomposition technique in case study 2.

F. DISCUSSION

1) DIEBOLD-MARIANO (DM) TEST

In order to further assessing and discussing the forecasting performance of the proposed combined model, the DM test is conducted to examine the effectiveness of the proposed combined model from a statistical perspective [42]. The specific steps are presented are as follows:

The null hypothesis is that the predictions of the two models are no significant differences. In contrast, the alternative hypothesis is that a significant difference in the forecasting performances of the two models. The null hypothesis and alternative hypothesis of the DM test are expressed

as follows:

$$H_0 : E[F(e_t^1)] = E[F(e_t^2)] \qquad (29)$$

$$H_1 : E[F(e_t^1)] \neq E[F(e_t^2)] \qquad (30)$$

where e_t^1 and e_t^2 are the forecasting errors between real values and forecasted values of the proposed combined model and another compared model, and F represents the loss function for forecasting errors.

Further, the statistics of DM test can be expressed by:

$$DM = \frac{\sum_{i=1}^{n} (F(e_t^1) - F(e_t^2))/n}{\sqrt{S^2/n}} S^2 \qquad (31)$$

where S^2 is the estimated variance of $d = F(e_t^1) - F(e_t^2)$.

After obtaining values of the calculated DM and the critical value $Z_{\alpha/2}$, if the DM statistic is greater than $Z_{\alpha/2}$ or less than $-Z_{\alpha/2}$, the null hypothesis will be rejected, and it is deemed that the distinctions between our proposed combined model and the compared model are evident.

The values of the DM test for two time-scales are shown in Table 13. From the results given in Table 13, it can be clearly seen that the proposed combined model is a significant difference from all the other models at a 1% significance level. In addition, the smallest value of DM for two time-scales are separately 2.8906 and 2.9546. Thus, the null hypothesis could be rejected at a 1% significance level. Therefore, the proposed combined model significantly outperforms the other models, and this verifies the effectiveness of our proposed combined model for wind speed forecasting.
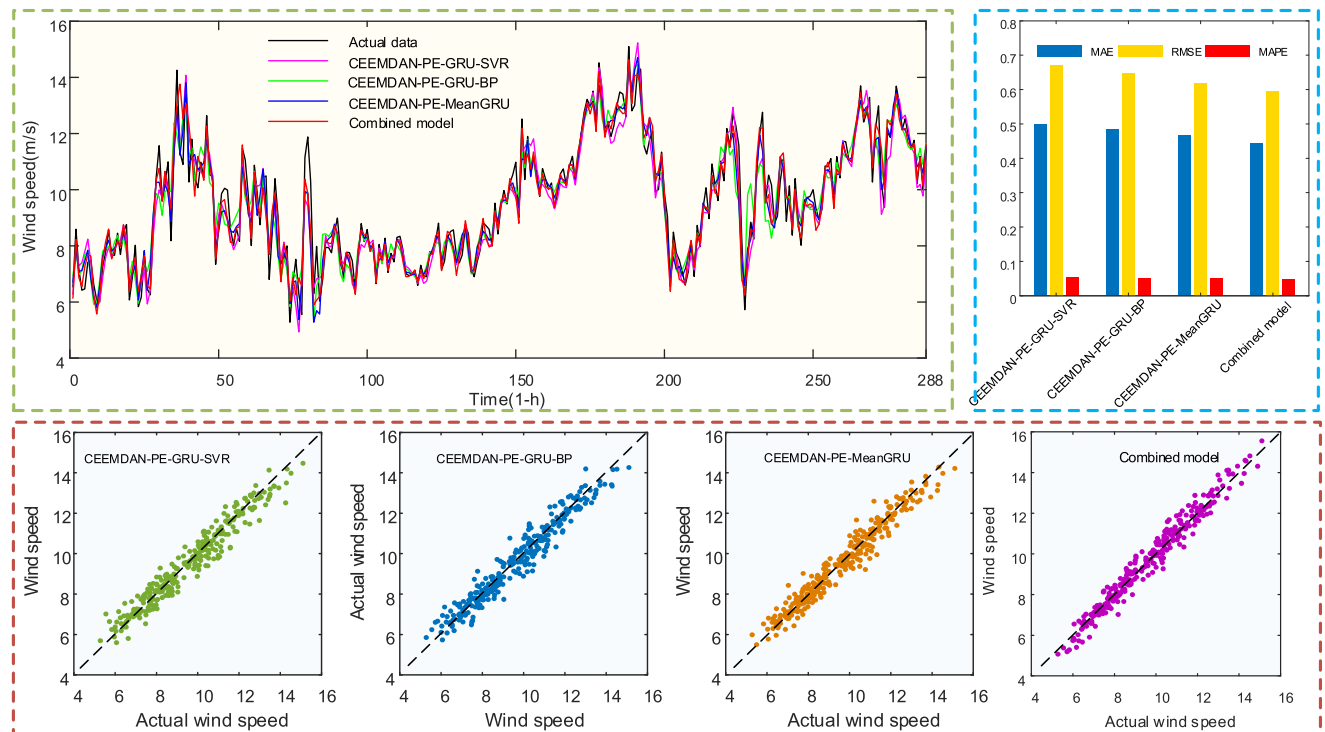

FIGURE 18. Forecasting results of combined model and models employing different nonlinear-learning top-layer in case study 2.

TABLE 13. The DM test results of different models.

Model	15-min	1-h
CEEMDAN-PE-SVR	5.0365*	5.1026*
CEEMDAN-PE-BP	4.9524*	4.8956*
CEEMDAN-PE-RBF	4.9863*	4.8362*
CEEMDAN-PE-LSTM	4.5637*	4.6321*
CEEMDAN-PE-GRU	4.5268*	4.4236*
WPD based model	4.0321*	4.1298*
EMD based model	3.9862*	3.9162*
EEMD based model	3.5265*	3.5136*
CEEMDAN-PE-GRU-SVR	3.5024*	3.5024*
CEEMDAN-PE-GRU-BP	3.2034*	3.3025*
CEEMDAN-PE-MeanGRU	2.8906*	2.9546*

*1% significance level.

2) COMPUTATIONAL EFFICIENCY: RUN TIME

In order to analysis the computational efficiency of the proposed combined model. Table 14 presents the average computation time of the proposed combined model and the other models with two time-scales wind speed forecasting. Specifically, the average computation time of the proposed combined model for two time-scales are separately are separately 382.56 and 386.98, which are less than the values from models using other data decomposition technologies, and models using different nonlinear-learning top-layer except the CEEMDAN-PE-MeanGRU model. Moreover, relative to the models only based on CEEMDAN-PE, the computation time of the proposed combined model is longer, which is reasonable because it has a better forecasting performance. And in practical implementation, the time consumed is acceptable. Moreover, the run time can be further improved by using a high-performance computer.

TABLE 14. Computation time comparison of the proposed combined model and other models.

Model	Computation time (s)	
	15-min	1-h
CEEMDAN-PE-SVR	181.26	179.65
CEEMDAN-PE-BP	172.65	177.63
CEEMDAN-PE-RBF	176.65	182.34
CEEMDAN-PE-LSTM	202.32	206.24
CEEMDAN-PE-GRU	191.45	193.51
WPD based model	415.27	418.25
EMD based model	405.82	406.37
EEMD based model	443.15	452.83
CEEMDAN-PE-GRU-SVR	397.24	399.51
CEEMDAN-PE-GRU-BP	392.15	395.34
CEEMDAN-PE-MeanGRU	365.25	368.71
Combined model	382.56	386.98

VI. CONCLUSION

Wind energy, as a crucial type of renewable energy, has aroused widespread interest and research enthusiasm. However, the intermittent characteristics and highly noisy of wind speed series increases forecasting difficulty to a great extent. In addition, due to the scale, geographic location, climatic conditions and wind turbine models of each wind farm are different, it is difficult for the wind speed forecasting strategy of the same centralized control center to achieve the desired accuracy across different wind farms. Motivated by recent developments in combined model forecasting, we developed a novel combined model for wind speed forecasting, which can not only achieve precise forecasting, but also be applicable to various wind farms environments under the centralized control system. In this article, two wind speed series collected from the wind farm in Zhangjiakou, China are adopted as cases to validate the accuracy and stability of the proposed combined model.

The experimental results demonstrate that the advantages of our proposed combined model mainly lie in the following aspects. (1) When comparing between the proposed combined model and models using different data decomposition techniques. The MAE, RMSE, MAPE and SSE values of our proposed combined model in two time-scales wind speed forecasting are all ranked as the best, which indicates the superiority of the proposed data decomposition technique. (2) Similarly, compared with models only based on CEEMDAN-PE in two time-scales wind speed forecasting, the lowest MAPE values for two time-scales are obtained from the proposed combined model, with values of 2.6169% and 4.7875%, which are reduced by 2.3352% and 2.0951% compared with maximum MAPE values obtained from the CEEMDAN-PE-SVR, which indicates the proposed combined model has excellent forecasting accuracy relative to models only based on CEEMDAN-PE. (3) When comparing between the proposed combined model and models using different nonlinear-learning top-layer. The proposed model can achieve better forecasting accuracy and more stable forecasting performance, which indicates the superiority of nonlinear-learning top-layer composed by IBA-RBF model. (4) In two case studies, generalized from all the models that, the proposed combined model has the best forecasting performance. In 15-minute ahead wind speed forecasting, the MAE, RMSE, MAPE and SSE of the proposed combined model are 0.2778, 0.3294, 2.6169%, 31.2580; compared with other models, the CEEMDAN-PE-MeanGRU has the second best forecasting performance, with the MAE, RMSE, MAPE and SSE of 0.2932, 0.3506, 2.8621%, 36.0231. Thus, the proposed combined model has a better performance, which MAE, RMSE, MAPE and SSE is separately 5.2524%, 6.0468%, 8.5671% and 13.2279% lower than the CEEMDAN-PE-MeanGRU model; In 1-h ahead wind speed forecasting, the MAE, RMSE, MAPE and SSE of the proposed combined model are 0.4453, 0.5876, 4.7875%, 96.9548; compared with other models, the CEEMDAN-PE-MeanGRU has the second best forecasting performance, with the MAE, RMSE, MAPE and SSE of 0.4665, 0.6196, 4.9865%, 101.4685. Thus, the proposed combined model has a better performance, which MAE, RMSE, MAPE and SSE is separately 4.5445%, 5.1646%, 3.9908% and 4.4645% lower than the CEEMDAN-PE-MeanGRU model. (5) Through the

DM test and computational efficiency analysis, we confirm that the proposed combined model has higher forecasting performance and lower cost in the proposed combined model.

Overall, we can conclude that the proposed combined model successfully enhances the forecasting capacity for wind speed, and provides a valuable reference the management of large wind farms.

REFERENCES

[1] T. Guo, L. Zhang, Z. Liu, and J. Wang, "A combined strategy for wind speed forecasting using data preprocessing and weight coefficients optimization calculation," *IEEE Access*, vol. 8, pp. 33039–33059, 2020.

[2] V. Khare, S. Nema, and P. Baredar, "Solar-wind hybrid renewable energy system: A review," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 23–33, May 2016.

[3] M. D. A. Al-falahi, S. D. G. Jayasinghe, and H. Enshaei, "A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system," *Energy Convers. Manage.*, vol. 143, no. 1, pp. 252–274, Jul. 2017.

[4] Z. Liu, P. Jiang, L. Zhang, and X. Niu, "A combined forecasting model for time series: Application to short-term wind speed forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114137.

[5] R. Li and Y. Jin, "A wind speed interval prediction system based on multi-objective optimization for machine learning method," *Appl. Energy*, vol. 228, pp. 2207–2220, Oct. 2018.

[6] N. Masseran, "Modeling the fluctuations of wind speed data by considering their mean and volatility effects," *Renew. Sustain. Energy Rev.*, vol. 54, pp. 777–784, Feb. 2016.

[7] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renew. Energy*, vol. 34, no. 5, pp. 1388–1393, May 2009.

[8] C. Wu, J. Wang, X. Chen, P. Du, and W. Yang, "A novel hybrid system based on multi-objective optimization for wind speed forecasting," *Renew. Energy*, vol. 146, pp. 149–165, Feb. 2020.

[9] P. Jiang, H. Yang, and J. Heng, "A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting," *Appl. Energy*, vol. 235, pp. 786–801, Feb. 2019.

[10] J. Wang, W. Zhang, J. Wang, T. Han, and L. Kong, "A novel hybrid approach for wind speed prediction," *Inf. Sci.*, vol. 273, pp. 304–318, Jul. 2014.

[11] G. Santamaría-Bonfil, A. Reyes-Ballesteros, and C. Gershenson, "Wind speed forecasting for wind farms: A method based on support vector regression," *Renew. Energy*, vol. 85, pp. 790–809, Jan. 2016.

[12] P. Jiang, Y. Wang, and J. Wang, "Short-term wind speed forecasting using a hybrid model," *Energy*, vol. 119, pp. 561–577, Jan. 2017.

[13] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A review on hybrid empirical mode decomposition models for wind speed and wind power prediction," *Energies*, vol. 12, no. 2, pp. 254-1–254-42, Jan. 2019.

[14] H. Bo, X. Niu, and J. Wang, "Wind speed forecasting system based on the variational mode decomposition strategy and immune selection multi-objective dragonfly optimization algorithm," *IEEE Access*, vol. 7, pp. 178063–178081, 2019.

[15] N. Bokde, A. Feijóo, and K. Kulat, "Analysis of differencing and decomposition preprocessing methods for wind speed prediction," *Appl. Soft Comput.*, vol. 71, pp. 926–938, Oct. 2018.

[16] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[17] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, Feb. 2004.

[18] J. Wang, W. Zhang, Y. Li, J. Wang, and Z. Dang, "Forecasting wind speed using empirical mode decomposition and elman neural network," *Appl. Soft Comput.*, vol. 23, pp. 452–459, Oct. 2014.

[19] Y. Peng, J. Cheng, Y. Yang, and B. Li, "Adaptive sparsest narrow-band decomposition method and its applications to rotor fault diagnosis," *Measurement*, vol. 91, pp. 451–459, Sep. 2016.

[20] D. S. Singh and Q. Zhao, "Pseudo-fault signal assisted EMD for fault detection and isolation in rotating machines," *Mech. Syst. Signal Process.*, vol. 81, pp. 202–218, Dec. 2016.

[21] P. Lu, L. Ye, B. Sun, C. Zhang, Y. Zhao, and J. Teng, "A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA," *Energies*, vol. 11, no. 4, p. 697, Mar. 2018.

[22] M. Santhosh, C. Venkaiah, and D. M. V. Kumar, "Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction," *Energy Convers. Manage.*, vol. 168, pp. 482–493, Jul. 2018.

[23] Y. Huang, S. Liu, and L. Yang, "Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM," *Sustainability*, vol. 10, no. 10, p. 3693, Oct. 2018.

[24] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renew. Energy*, vol. 94, pp. 629–636, Aug. 2016.

[25] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[26] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015.

[27] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, and K.-D. Lu, "Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization," *Energy Convers. Manage.*, vol. 165, pp. 681–695, Jun. 2018.

[28] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.

[29] G. Dai, C. Ma, and X. Xu, "Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU," *IEEE Access*, vol. 7, pp. 143025–143035, 2019.

[30] H. Liu, X. Mi, Y. Li, Z. Duan, and Y. Xu, "Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression," *Renew. Energy*, vol. 143, pp. 842–854, Dec. 2019.

[31] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690–76698, 2019.

[32] Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 82–91, Oct. 2015.

[33] L. Xiao, W. Shao, M. Yu, J. Ma, and C. Jin, "Research and application of a combined model based on multi-objective optimization for electrical load forecasting," *Energy*, vol. 119, pp. 1057–1074, Jan. 2017.

[34] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 136, pp. 439–451, Mar. 2017.

[35] X. Niu and J. Wang, "A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting," *Appl. Energy*, vol. 241, pp. 519–539, May 2019.

[36] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[37] L. Zunino, M. C. Soriano, I. Fischer, O. A. Rosso, and C. R. Mirasso, "Permutation-information-theory approach to unveil delay dynamics from time-series analysis," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 82, no. 4, pp. 565–590, Oct. 2010.

[38] Y. Huang, L. Yang, S. Liu, and G. Wang, "Multi-step wind speed forecasting based on ensemble empirical mode decomposition, long short term memory network and error correction strategy," *Energies*, vol. 12, no. 10, p. 1822, May 2019.

[39] M. Santhosh, C. Venkaiah, and D. M. V. Kumar, "Short-term wind speed forecasting approach using ensemble empirical mode decomposition and deep Boltzmann machine," *Sustain. Energy, Grids Netw.*, vol. 19, Sep. 2019, Art. no. 100242.

[40] G. W. Chang, H. J. Lu, Y. R. Chang, and Y. D. Lee, "An improved neural network-based approach for short-term wind speed and power forecast," *Renew. Energy*, vol. 105, pp. 301–311, May 2017.

[41] L. Zhang, Y. Dong, and J. Wang, "Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy," *IEEE Access*, vol. 7, pp. 176000–176023, 2019.

[42] Z. He, Y. Chen, Z. Shang, C. Li, L. Li, and M. Xu, "A novel wind speed forecasting model based on moving window and multi-objective particle swarm optimization algorithm," *Appl. Math. Model.*, vol. 76, pp. 717–740, Dec. 2019.

SHURUI FAN received the Ph.D. degree from the Hebei University of Technology, Tianjin, China, in 2011. He is currently an Associate Professor with the School of Electronical Information, Hebei University of Technology. He was a Visiting Post-Doctoral with the School of Information Technology and Electrical Engineering, The University of Queensland, from March 2015 to September 2016. His research interests include embedded system artificial intelligence, wireless sensor networks, gas sensor array, autonomous search, and data visualization analysis method.

TAO LIANG was born in Shijiazhuang, Hebei, China, in 1975. He received the Ph.D. degree in power electronics and power transmission from the Hebei University of Technology, in 2007. He is currently a Professor with the School of Artificial Intelligence, Hebei University of Technology. His research interests include new energy power generation, data mining, fault diagnosis, and artificial intelligence prediction theory and method.

GAOFENG XIE was born in Xuchang, Henan, China, in 1994. He received the B.S. degree from the School of Physics and Electrical Engineering, Anyang Normal University, China, in 2017. He is currently pursuing the M.S. degree with the School of Artificial Intelligence, Hebei University of Technology. His research interests include data mining, machine learning technologies, and the application of deep learning in the wind energy.

ZHAOCHAO MENG was born in Hengshui, Hebei, China, in 1995. He received the B.S. degree from the Weifang University of Science and Technology, China, in 2018. He is currently pursuing the M.S. degree with the School of Artificial Intelligence, Hebei University of Technology. His research interests include data mining, machine learning, and artificial intelligence techniques.

● ● ●