Energy 160 (2018) 87–100

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A hybrid forecasting approach applied in wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm

Zhongshan Yang, Jian Wang*

School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China

ARTICLE INFO

Article history:
Received 13 April 2018
Received in revised form
4 June 2018
Accepted 3 July 2018
Available online 4 July 2018

Keywords:
Wind speed forecasting
Complementary ensemble empirical mode
Modified wind driven optimization
Broyden family
Hybrid model

ABSTRACT

Conducting the accurate forecasting of wind speed is a both challenging and difficult task. However, this task is of great significance for wind farm scheduling and safe integration into the grid. In this paper, the wind speed at 10 or 30 min is predicted using only historical wind speed data. The existing single models are not enough to overcome the instability and inherent complexity of wind speed. To enhance forecasting ability, a novel hybrid model based on complementary ensemble empirical mode decomposition (CEEMD) and modified wind driven optimization is introduced for wind speed forecasting in this paper. CEEMD is utilized to decompose the original wind speed series into several intrinsic mode functions (IMFs), and each IMF is forecasted by back propagation neural network (BP). A new optimization algorithm combined Broyden family and wind driven optimization is presented and applied to optimize the initial weights and thresholds of BP. Finally, all forecasted IMFs are integrated as final forecasts. The 10min and 30min wind speed from the province of Shandong, China, were used in this paper as the case study, and the results confirm that the proposed hybrid model can improve the forecasting accuracy and stability.

1. Introduction

With increased demand for energy along with the strengthening of environmental protection laws, the popularization and application of clean energy has become an inevitable trend. Wind power is the most cost-effective and rapidly growing renewable energy alternative to conventional power generation technology. By the end of 2015, China's total installed wind power capacity reached approximately 13.1 GW and is expected by 2020 to have an installed capacity of 21.0 GW or more [1]. Wind power is a clean and sustainable energy that has many advantages, but many problems remain to be solved. Unlike traditional power generation, wind is characteristically random and volatile, resulting in low wind power prediction accuracy.

Accurate forecasting of wind farm power is an important part of wind power operations management, and wind speed forecasting is the premise behind power forecasting [2]. There are many classifications of wind speed forecasting in wind farms. These classifications can be divided into four types based on different time horizons: long-term forecasting (1 day or more ahead), medium-term forecasting (6 h–1 day ahead), short-term forecasting (30 min–6 h ahead) and ultra-short-term forecasting (several seconds to 30 min ahead) [3]. All of the models can be grouped according to whether they are single model, a hybrid model or a combined model depending on the model's structure. The single model is an individual model used for wind speed prediction; the concept of a hybrid model is to integrate several models in order to provide an advanced forecasting model; and the combined model is to combine several forecasting models and assign a weighting coefficient to each method according to its forecasting performance [4]. The models can also be divided into three types: physical models, statistical models and artificial intelligence models.

Physical models are used to simulate changes to wind speed trends at a geographical position in three dimensions by considering wind speed drivers, such as temperature, air pressure, terrain roughness, and obstructions [5,6]. Physical models do not need to be trained using historical data, and when more detailed background information, such as time and geography, is considered, it is possible to obtain more accurate long-term forecasting values. Physical models have many advantages, but they are also very restrictive. To reflect the nature of atmospheric motion, numerical

* Corresponding author.
E-mail address: jianwang0826@163.com (J. Wang).

https://doi.org/10.1016/j.energy.2018.07.005
0360-5442/

List of abbreviations

WRF	weather research and forecast
HRM	high resolution model
COSMO	consortium for small scale modeling
MM5	mesoscale model 5
NWP	numerical weather prediction
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
ANN	artificial neural network
KF	Kalman filter
BP	back propagation neural network
ENN	Elman neural network
RBF	radical basis function
SVM	support vector machine
GA	genetic algorithm
SVR	support vector regression
WT	wavelet transform
SSA	singular spectrum analysis
EMD	empirical mode decomposition
EEMD	ensemble empirical mode decomposition
CEEMD	complementary ensemble empirical mode decomposition
IMFs	intrinsic mode functions
WDO	wind driven optimization
MWDO	modified wind driven optimization
DM DE	Diebold-Mariano value differential evolution
GRNN	generalized regression neural network

weather prediction (NWP) data must be used as input data. However, the horizontal resolution of NWP data provided by meteorological departments is too low. Even after reducing the scale of the mesoscale model, the resolution is still on the order of kilometers [7]. Statistical prediction models are based on the historical data of wind speed; the background information of wind speed signal is usually not taken into account. The statistical models commonly used in wind speed prediction are the autoregressive integrated moving average (ARIMA) [8,9], grey prediction models [10,11] and exponential smoothing. Compared with physical methods, the statistical methods are simpler and more economical; numerous studies on statistical models of wind speed prediction have been published. Kavasseri [12] proposed a fractional-ARIMA model to forecast wind speeds with day-ahead (24 h) and two-day-ahead (48 h) horizons. Erdem [13] introduced an improved autoregressive moving average (ARMA) model to forecast the tuple of wind speed and direction; the results show that the component model is better at predicting wind speed and direction than is the traditional-linked ARMA model. Jethro Dowell and Pierre Pinson [14] combined a parametric probabilistic framework based on the logit-normal distribution and fitting a sparse vector autoregressive model to generate very-short-term wind speed forecasts for smart grids. The results indicate this model is better than conventional vector autoregressive models.

Recent studies [15–17] have shown that artificial intelligence models are being used more often and achieve higher wind speed prediction accuracy. Many methods, such as back propagation neural network (BP) [18,19], Elman neural networks (ENN) [20,21], radial basis function network (RBF) [22] and support vector machine (SVM) [23–25] are used to forecast wind speed. Compared with statistical methods, neural networks perform better when fitting nonlinear curves. However, the single models also have their limitations; the initial thresholds and weights of the neural network are generated randomly and easily fall into the local optimum [26]. Therefore, optimization models have been proposed which optimize the parameters of the single model. These combined methods and hybrid models are considered advanced methods. Santamaría-Bonfil [27] proposed PSR-SVR$_{GA}$ to solve the short-term wind speed forecasting problem based on support vector regression (SVR) and genetic algorithm (GA). The model performs best when compared to other single models. However, the accuracy and performance depends largely on the quality of the historical data set. Therefore, it is necessary to preprocess the data set through such means as interpolation, decomposition, and seasonal adjustment [28]. De-noising techniques such as empirical mode decomposition (EMD) [29], wavelet transforms (WT) [30] and singular spectrum analysis (SSA) is often used to preprocess wind speed sequences to improve forecasting accuracy.

In this paper, a hybrid model which incorporates de-noising techniques, an intelligent optimization model and artificial neural networks is introduced to improve the accuracy of wind speed forecasting. Complementary ensemble empirical mode decomposition (CEEMD) plays a key role in data preprocessing. The original sequence is decomposed into several intrinsic mode functions (IMFs) which are arranged in order of frequency. The IMF with the highest frequency is then eliminated. BP neural network is used to forecast 10-min-ahead and 30-min-ahead wind speed. A modified wind driven optimization model (MWDO) is utilized to optimize the initial weights and thresholds of the BP network. The wind driven optimization algorithm (WDO) is a heuristic global optimization method and the core is to combine Newton's second law and the state equation for an ideal gas to study an air parcel in the atmosphere. Several single models and hybrid models are selected as benchmark models. Single models are used to verify that the hybrid models have better forecasting performance. To verify that CEEMD has better de-noising performance for improving forecasting ability compared with SSA and WT, the MWDO-SSA-BP and MWDO-WT-BP models were selected as benchmark models.

The contributions of this paper are summarized as follows:

1) Based on the Broyden family quasi-Newton method, a new modified wind driven optimization algorithm was developed to find the best air parcel. When the convergence criterion is increased, the original WDO algorithm tends to fall into local optimality, but the modified WDO algorithm (MWDO) can improve the optimization speed and optimization accuracy. To verify the performance and feasibility of the proposed algorithm, four test functions were used.
2) This new forecasting model, whose life is one day, is updated by reconstructing and training with new samples after providing a one-day prediction. The model will be continuously validated for one week.
3) More metrics were used to assess the predictive performance of the proposed model. The performance of the model can be measured based on accuracy and stability. The Diebold-Mariano test, Bias-Variance framework and four error criteria, including AE, MAE, MAPE and RMSE were involved in this paper.

The rest of the paper is organized as follow: Section 2 introduces the proposed models, including the de-noising model, optimization model and the test of MWDO. The experimental procedure and experimental results are presented in Section 3, Section 4 provides further discussion, and the study's conclusions are presented in Section 5.

2. Proposed approach

In this section, complementary ensemble empirical mode decomposition (CEEMD), modified wind driven optimization (MWDO) and back propagation neural network will be introduced briefly. The testing process and results of using the MWDO algorithm are also presented in this section.

2.1. Complementary ensemble empirical mode decomposition

Huang [31] proposed a time-frequency analysis method (Hilbert-Huang Transform) for non-linear and non-stationary signals, the core technology of which is empirical mode decomposition (EMD). EMD can decompose a complex signal into a series of intrinsic mode functions (IMFs). At the same time, EMD has the characteristics of orthogonality, completeness and self-adaptability, which has great advantages for non-stationary signal analysis. Orthogonality means that each of the IMF is independence and does not impact on others, completeness means that the original signal can be obtained by adding all the obtained IMF and self-adaptability means that the frequency range of IMF is determined by the characteristics of the original signal. The details of EMD are shown below:

Step 1: Find all the maximum and minimum values of the original signal$x(t)$, and obtain the upper envelope $x_h(t)$ and the lower envelope $x_l(t)$ of the series using a cubic spline function. The mean envelope of the original series is calculated as:

$$m_1(t) = \frac{x_h(t) + x_l(t)}{2} \tag{1}$$

Step 2: Calculate the first intrinsic mode function $h_1(t)$.The first IMF is generally not a smooth series:

$$h_1(t) = x(t) - m_1(t) \tag{2}$$

Step 3: Repeat step 1 and step 2 k times for $h_1(t)$ until $h_1(t)$ accords with the requirement of IMF that the mean value tends to zero. Calculate the first component of IMF as follows:

$$h_1^{k-1}(t) - m_1^k(t) = h_1^k(t) \tag{3}$$

$$c_1(t) = h_1^k(t) \tag{4}$$

Step 4: $c_1(t)$ is removed from the original signal to obtain a signal $r_1(t)$ without a high frequency component:

$$r_1(t) = x(t) - c_1(t) \tag{5}$$

Step 5: Regard $r_1(t)$ as the original series and repeat the above process to obtain a second IMF component$c_2(t)$. Repeat several times until the termination condition is reached**.** The original series is decomposed into n IMFs and a residual series.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{6}$$

However, EMD has a drawback of mode mixing. Mode mixing means that the obtained IMF consisting of oscillations of dramatically disparate scales, often caused by intermittency of the driving mechanisms [32]. To resolve this drawback in EMD decomposition, Wu and Huang [32] proposed an improved EMD method, called ensemble empirical mode decomposition (EEMD). EEMD incorporates anti-noise characteristics into the decomposition process by adding different Gaussian white noise. However, decomposition and reconstruction in the EEMD process consumes too much time and reconstruction of the signal from the originated IMF includes residual noise [33]. An improved EEMD method, called complementary ensemble empirical mode decomposition (CEEMD), was introduced by Yeh and Huang [34] to solve the drawbacks of EMD and EEMD. CEEMD eliminates residual noise in the reconstructed signal and improves computational efficiency by adding positive and negative Gaussian white noise. The CEEMD decomposition is based on EMD decomposition and consists of three steps:

Step 1: Initialize the number of realizations and the standard deviation of the noise. Let $i = 1$ and the maximum number of decompositions is I.

Step 2: Perform EMD decomposition of the signal after adding i times white noise;

a) Add white noise with same amplitude and opposite phase to the signal ***x(t)***, to obtain the signals:

$$\begin{cases} \mathrm{P}_i(t) = \mathrm{x}(t) + \mathrm{n}_i(t) \\ \mathrm{N}_i(t) = \mathrm{x}(t) - \mathrm{n}_i(t) \end{cases} \tag{7}$$

where $\mathrm{n}_i(t)$ is the white noise added for the i-th decomposition, $P_i(t)$and $N_i(t)$ represent the series with white noise.

b) EMD decomposes the signal $P_i(t)$ and $N_i(t)$ to obtain q IMFs:

$$\begin{cases} \mathrm{P}_i(t) = \sum_{j=1}^{q} c_{j,i}^1(t) \\ \mathrm{N}_i(t) = \sum_{j=1}^{q} c_{j,i}^2(t) \end{cases} \tag{8}$$

where $c_{j,i}^1(t)$ and $c_{j,i}^2(t)$ denote the j-th IMF from the i-th decomposition, and q is the number of IMFs.

c) If $i < I$, return to steps a) and b), and add different white noise for each experiment.

Step 3: Calculate the average of the IMFs after EMD decomposition I times:

$$c_j = \frac{1}{2I}\sum_{i=1}^{I}\left(c_{j,i}^1(t) + c_{j,i}^2(t)\right) \tag{9}$$

where c_j represents the j-th IMF obtained by CEEMD decomposition.

2.2. Back propagation neural network (BP)

BP neural network is a multi-layer feedforward neural network. The main characteristics of the network is the signal forward transmission and error back propagation [35]. The network has three layers: an input layer, a hidden layer and an output layer. The network adjusts the network weights and thresholds to minimize the prediction error such that the BP network prediction output will approach the expected output.

The non-linear space expression of BP neural network is

$$H_j = f\left(\sum_{i=1}^{n} \omega_{ij} x_i - a_j\right) j = 1, 2, \cdots, l \tag{10}$$

$$O_k = \sum_{j=1}^{l} H_j \omega_{jk} - b_k k = 1, 2, \cdots, m \tag{11}$$

$$e_k = Y_k - O_k k = 1, 2, \cdots, m \tag{12}$$

where i, j and k denote input node, hidden node and output node, respectively. The a_j and b_k denote the hidden layer thresholds and the output layer thresholds, respectively. ω_{ij} is the weight of the connection from the input layer to the hidden layer and ω_{jk} is the weight of the connection from the hidden layer to the output layer. H_j and O_k denote hidden layer output and output layer output, respectively. Y_k is the expected output and e_k is the error.

Next, the weights and thresholds of the network are updated based on the prediction errore_k. The related formulas are shown below:

$$\omega_{ij} = \omega_{ij} + \eta H_j(1 - H_j)x(i)\sum_{k=1}^{m}\omega_{jk}e_k \; i = 1, 2, \cdots, n; j = 1, 2, \cdots, l \tag{13}$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \quad j = 1, 2, \cdots, l; k = 1, 2, \cdots, m \tag{14}$$

$$a_j = a_j + \eta H_j(1 - H_j)\sum_{k=1}^{m}\omega_{jk}e_k \quad j = 1, 2, \cdots, l \tag{15}$$

$$b_k = b_k + e_k \quad k = 1, 2, \cdots, m \tag{16}$$

where n, l and m are the number of input nodes, hidden nodes and output nodes, respectively, and η is the learning rate.

2.3. Modified wind driven optimization (MWDO)

It is necessary to use the optimization model to generate the weights and thresholds of the neural network. In this section, a new optimization model (MWDO) which combines Broyden family and wind driven optimization (WDO) is introduced.

2.3.1. Broyden family

Newton's method [36] is one of the best methods for solving unconstrained optimization problems. However, a drawback of Newton's method is the second-order derivative of the objective function, i.e., the Hessian matrix, is calculated during each iteration. For large scale problems, the computation time for calculating the Hessian matrix is very large, or the matrix cannot be calculated. To overcome the shortcomings of Newton's method, a Quasi-Newton method is proposed to use a matrix $\boldsymbol{B}_k$ to replace the Hessian matrix $\boldsymbol{H}_k$. $\boldsymbol{B}_k$ satisfies the quasi-Newton equation:

$$\mathrm{B}_{k+1}s_k = y_k \tag{17}$$

where s_k is the displacement and y_k denotes gradient difference.

Broyden's [37] method is recognized as one of the most commonly used and efficient quasi-Newton method. Broyden family algorithm is a combination of two Quasi-Newton methods , which are Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) [38] and Davidon-Fletcher-Powell algorithm (DFP) [39] respectively; its correction formula for $\boldsymbol{B}_K$ is

$$\begin{aligned}\mathbf{B}_{k+1}^{\varphi} &= \varphi_k\mathbf{B}_{k+1}^{DFP} + (1 - \varphi_k)\mathbf{B}_{k+1}^{BFGS} \\ &= \mathbf{B}_k - \frac{\mathbf{B}_k s_k s_k^T \mathbf{B}_k}{s_k^T\mathbf{B}_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \varphi_k\left(s_k^T\mathbf{B}_k s_k\right)v_k v_k^T\end{aligned} \tag{18}$$

$$v_k = \frac{y_k}{y_k^T s_k} - \frac{\mathbf{B}_k s_k}{s_k^T\mathbf{B}_k s_k} \tag{19}$$

where $\mathbf{B}_{k+1}^{DFP}$ is the $\boldsymbol{B}_K$ obtained by the DFP method and $\mathbf{B}_{k+1}^{BFGS}$ is the $\boldsymbol{B}_K$ obtained by the BFGS method, and $\varphi_k \in [0, 1]$ is a real parameter.

Obviously, when$\varphi_k = 0$, Broyden family algorithm is the BFGS algorithm; when$\varphi_k = 1$, Broyden family algorithm is the DFP algorithm.

A rudimentary Broyden family algorithm is outlined as follows:

Algorithm: Broyden family

Parameters:
ε– convergence tolerance.
k –current iteration number.
maxk – the maximum number of iterations
1:/* Initialize the parameters and generate the initial point x_0*/
2: Setting convergence tolerance ε, $k \leftarrow 0$
3: Evaluate the inverse of Hessian matrix
4: **WHILE** $k < maxk$ **DO**
5:/* Calculate the gradient.*/
6: $g_k = \nabla f(x_k)$
7: **IF** $\|g_k\| \le \varepsilon$ **BREAK; END**
8:/* Calculate the search direction.*/
9: $d_k = -H_k * g_k$, $m \leftarrow 0$
10: **WHILE** $m < maxm$ **DO**
11: **IF** $f(x_k + \delta^m d_k) \le f(x_k) + \sigma\delta^m g_k^T d_k$ **Then**
12: $\alpha_k = \delta^{mk}$
13: $x_{k+1} = x_k + \alpha_k d_k$
14: **END IF**
15: **END WHILE**
16:/*Broyden family correction.*/
17: **IF** $s_k^T g_k \le 0$ **THEN**
18: $H_{k+1} = H_k$
19: **ELSE IF** $H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \varphi_k(s_k^T H_k s_k)v_k v_k^T$
20: **END IF**
21: **END WHILE**
22: **RETURN** x_k

2.3.2. Wind driven optimization

The wind driven optimization algorithm is a heuristic global optimization method proposed by Bayraktar et al., in 2010 [40]. Its core is to combine Newton's second law and the state equation for an ideal gas to study an air parcel in the atmosphere. Each air parcel in the WDO algorithm represents a candidate solution, which is developed by the velocity update equation and the position update equation. The term “pressure value” in reference to atmospheric dynamics indicates the fitness value of the air parcels. WDO has strong global search ability, faster convergence speed and higher searching efficiency. It can be used to solve multi-dimensional and multi-modal problems, and continuous and discrete optimization problems.

For the simplified air particle, the updating equation for position is

$$\mathrm{x}_{ik}^{t+1} = \mathrm{x}_{ik}^{t} + \left(\mathrm{u}_{ik}^{t+1}\Delta t\right) \tag{20}$$

where x_{ik}^{t+1} denotes the updated air particle position, and u_{ik}^{t+1} is the updated air particle velocity. To simplify the equation, let the time interval$\Delta t = 1$.

The air particle velocity update equation is:

$$\mathrm{u}_{ik}^{t+1} = (1 - \alpha)\mathrm{u}_{ik}^{t} - gx_{ik}^{t} + \left[RT\left|\frac{1}{j} - 1\right|\left(x_{gbest} - x_{ik}^{t}\right)\right] + \frac{cu_{iotherdim}^{t}}{j} \tag{21}$$

where u_{ik}^t is the speed of the i-th air particle at the k-th dimension during the t-th iteration, α is the coefficient of friction, g is the acceleration of gravity, R is the ideal gas constant, T represents temperature, c is the comprehensive coefficient and $c = -2\Omega RT$, Ω is the earth curl vector, $\boldsymbol{x}_{\mathrm{gbest}}$ is the global optimal position of the

whole population and $cu^{t}_{iotherdim}$ denotes the influence on velocity from another randomly chosen dimension of the same air parcel.

The wind driven optimization algorithm flow is as follows:

Step 1: Initialize the population size of WDO, set related parameters, search boundaries, and define pressure functions (fitness functions).

Step 2: Randomly initialize the air particles and randomly assign the initial velocities and positions.

Step 3: Calculate the pressure values of the air particles in the current iteration. The population is rearranged according to the pressure values.

Step 4: Update the velocity of the air particles using Equation 21.

Step 5: Update the position of the air particles using Equation 20.

Step 6: If the termination condition is reached, output x_{gbest}; otherwise, return to step 3.

The pressure value at the last iteration is recorded as the optimal result. The termination condition is generally a sufficiently good pressure value or a preset maximum iteration number.

2.3.3. MWDO

Although the WDO model has a strong ability to search, shortcomings are obvious such as it is easier to fall into local optimization, premature convergence, and loss of diversity at a late stage. To overcome these shortcomings, a new optimization algorithm is proposed. The Broyden family algorithm is used when WDO updates solutions during an iteration to find a local optimal solution, and thereby enhance the local optimization ability and the speed of the local convergence of the algorithm.

The pseudo-code for modified wind driven optimization algorithm is as follows.

Algorithm: MWDO

Input: $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^T$ – a sequence of raw data.

Output: $\boldsymbol{x}_b$ – the value of $\boldsymbol{x}$ with the best fitness value in population of air particles

Parameters:

n – the population size

g – the gravitational constant

α – the coefficient of friction

G_{max} – the maximum number of iterations

RT, c –RT coefficient and Coriolis effect coefficient

1: /*Set the parameters of WDO and Broyden family.*/
2: /* Initialize population, position and velocity.*/
3: **FOR EACH** $1 \le i \le n$ **DO**
4: | Evaluate the corresponding fitness function $\boldsymbol{P}_i$.
5: **END FOR**
6: /*Finding best air particle in the initial population.*/
7: $\boldsymbol{P}_{gbest}=\boldsymbol{P}_{pbest}$
8: $\boldsymbol{x}_{gbest}=\boldsymbol{x}_{pbest}$
9: /*Start iterations.*/
10: **FOR EACH** $iter \le G_{max}$ **DO**
11: | /*Update the velocity.*/
12: | $$\boldsymbol{u}_{ik}^{t+1}=(1-\alpha)\boldsymbol{u}_{ik}^{t}-g\,\boldsymbol{x}_{ik}^{t}+\left[RT\left|\frac{1}{j}-1\right|(\boldsymbol{x}_{gbest}-\boldsymbol{x}_{ik}^{t})\right]+\frac{c\,\boldsymbol{u}_{iotherdim}^{t}}{j}$$
13: | /*Update air parcel positions.*/
14: | $$\boldsymbol{x}_{ik}^{t+1}=\boldsymbol{x}_{ik}^{t}+(\boldsymbol{u}_{ik}^{t+1}\Delta t)$$
15: | Use **Broyden family** to update the new location of air particles $\boldsymbol{x}_i$.
16: | /*Finding best particle in population.*/
17: | **FOR EACH** $1 \le i \le n$ **DO**
18: | | Evaluate the corresponding fitness function $\boldsymbol{P}_i$.
19: | **END FOR**
20: | **IF** $\boldsymbol{P}_{pbest} < \boldsymbol{P}_{gbest}$ **THEN**
21: | | $\boldsymbol{x}_{gbest}=\boldsymbol{x}_{pbest}$
22: | **END IF**
23: **END FOR**
24: **RETURN** $\boldsymbol{x}_{gbest}$

2.3.4. Test of MWDO

To compare the performance of the WDO algorithm and the MWDO algorithm, four benchmark functions were used. The benchmark functions are shown in Table 1. Each benchmark function was independently run 100 times, and records the number of iterations when the stop condition is reached. Parameter settings are necessary in testing. For the WDO algorithm and the MWDO algorithm, the population size is 40; the dimension of an air particle is 30; the acceleration of gravity is 0.2; the coefficient of friction is 0.4; and the *RT* coefficient and the Coriolis coefficient are 3 and 0.4, respectively. For MWDO algorithm, the iterations of the Broyden family is 10.

The results, including the maximum, minimum, average, and standard deviations of number of iterations when the stop condition is reached, are shown in Table 2. For each benchmark model, the modified wind driven optimization algorithm has better search capability than the original wind driven optimization algorithm.

Table 2
Test results of WDO and MWDO.

Function name	Algorithm	Min	Max	Mean	Std
Sphere	WDO	776	884	828.28	26.95
	MWDO	**2**	**5**	**4.05**	**0.6416**
Rosenbrock	WDO	669	1837	1144.58	215.50
	MWDO	**3**	**10**	**4.07**	**1.09**
Rastrigin	WDO	72	27654	1445.21	4379.12
	MWDO	**29**	**335**	**94.35**	**59.02**
Griewank	WDO	191	266	229.76	13.19
	MWDO	**31**	**51**	**42.99**	**4.04**

2.4. Framework of the proposed hybrid model

The flow chart of the proposed model MWDO-CEEMD-BP is shown in Fig. 1. The detailed steps of the proposed hybrid model are shown below:

Step 1: Decompose the original wind speed series into several time series IMFs by complementary ensemble empirical mode decomposition, and remove the highest frequency IMF. The IMFs are sorted from low frequencies to high frequencies and all IMFs after the fifth IMF form a residue IMF.

Step 2: BP neural network optimized by MWDO is used to forecast each IMF. For each IMF, the data for the IMF is the input of the neural network, and then the corresponding forecasted values are obtained.

Step 3: Integrate all of the predicted IMFs as a final predictor of wind speed.

3. Empirical study

This section explains the process and results of the proposed hybrid model called MWDO-CEEMD-BP to predict wind speed. Several single models and hybrid models were used to verify the forecasting performance of the proposed model.

3.1. Evaluation criteria

The forecasting performance of a model is reflected in forecasting accuracy and stability. To verify the predictive performance of a model, a systematic standard is used in this paper. Four statistical error criteria are used in this paper: mean error (AE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). In addition, the Diebold-Mariano test and the Bias-Variance framework are also used.

The four metrics are:

$$AE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right) \tag{22}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right| \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2} \tag{24}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\% \tag{25}$$

where $\widehat{y}$ and y denote the forecasted value and real value, respectively.

3.1.1. Diebold-Mariano test

The Diebold-Mariano test, a comparison test, was proposed by Diebold and Mariano [41] to evaluate forecasting accuracy. A set of time series data $\{y_t\}$ is acquired, and n forecasting models are established to obtain n sets of predicted values $\widehat{y}_t^i (i = 1, 2, \cdots, n)$. The corresponding sequence of prediction errors is $e_t^i = y_t - \widehat{y}_t^i (i = 1, 2, \cdots, n)$ and the loss function is $L(e_t^i)$.

Loss functions come in many forms. The most commonly used is the square error loss, which is the method is used in this paper:

$$L\left(e_t^i\right) = \sum_{t=1}^{T}\left(e_t^i\right)^2 \tag{26}$$

followed by the absolute error loss:

$$L\left(e_t^i\right) = \sum_{t=1}^{T}\left|e_t^i\right| \tag{27}$$

To determine whether one model's predictive ability is better than that of another model, we test the following null hypothesis:

$$H_0 : E\left[L\left(e_t^i\right)\right] = E\left[L\left(e_t^j\right)\right] \tag{28}$$

The alternative hypothesis is:

Table 1
Test functions.

Function name	Benchmark function	Variable domain
Sphere	$f_{Sp}(\vec{x}) = \sum_{i=1}^{n} x_i^2$	$\vec{x} = [-5.12, 5.12]$
Rosenbrock	$f_{Rn}(\vec{x}) = \sum_{i=1}^{n-1}[100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2]$	$\vec{x} = [-2.084, 2.084]$
Rastrigin	$f_{Ra}(\vec{x}) = \sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i) + 10)$	$\vec{x} = [-5.12, 5.12]$
Griewank	$f_{Gk}(\vec{x}) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$\vec{x} = [-8, 8]$

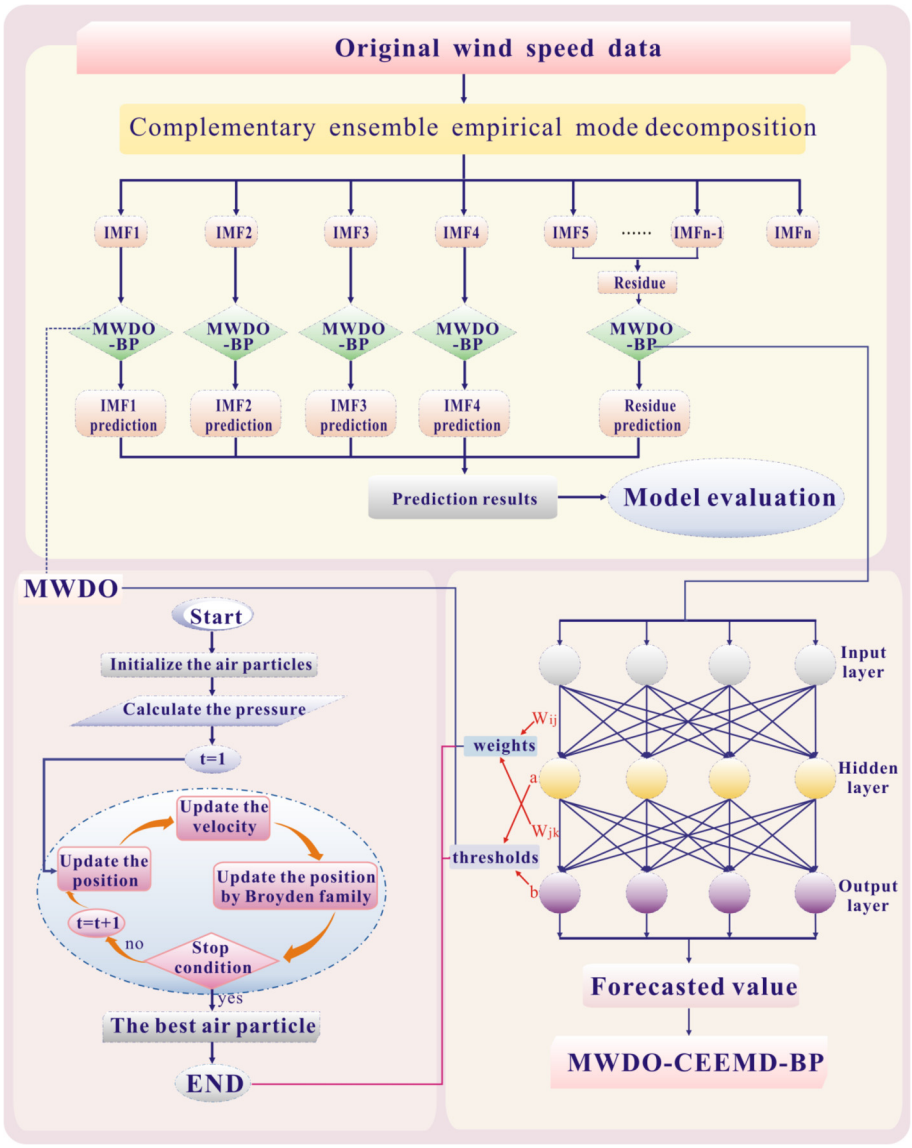

Fig. 1. The flow chart of the proposed model MWDO-CEEMD-BP.

$$H_1 : E\left[L\left(e_t^i\right)\right] \neq E\left[L\left(e_t^j\right)\right] \tag{29}$$

where $E(\cdot)$ represents mathematical expectation.

The test statistic DM is asymptotically $N(0,1)$ distributed, where the DM is:

$$DM = \frac{\sum_{t=1}^{T}\left[L(e_t^i) - L\left(e_t^j\right)\right]/T}{\sqrt{S^2/T}} S^2 \tag{30}$$

where S^2 is an estimator of the variance of $d_t = L(e_t^i) - L(e_t^j)$. The value of the DM statistic is compared with the critical value of the standard normal distribution. If the DM statistic is large, the original hypothesis is rejected, and the difference between the two models is significant; otherwise, the forecasting ability of the two models is the same.

3.1.2. Bias-variance framework

The Bias-variance framework [41] is a highly popular method for evaluating the stability and accuracy of a forecasting model. Bias measures the accuracy or quality of the model, and variance measures the precision or specificity of the model. A high bias or a high variance means the test model has a weak forecasting ability. Intuitively, the final goal is to minimize both the bias and variance simultaneously.

The bias-variance decomposition of the generalization error in the least squares sense is as follows:

$$E\left(\hat{y}\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \tag{31}$$

$$E(y) = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{32}$$

$$E\left(\hat{y}-y\right)^2 = E\left[\hat{y}-E\left(\hat{y}\right)+E\left(\hat{y}\right)-y\right]^2$$
$$= E\left(\hat{y}-E\left(\hat{y}\right)\right)^2 + \left[E\left(\hat{y}\right)-y\right]^2 \quad (33)$$
$$= Var\left(\hat{y}\right) + Bias^2\left(\hat{y}\right)$$

where $\hat{y}$ and y denote the forecasted value and real value, respectively.

3.2. Experimental data and parameters of related models

To verify the performance of the proposed hybrid model, wind speed data of 10 min and 30 min duration, collected from three observation sites in the city of Yantai, China, were used as sample data. The 10 min datasets from April 7 to 20, 2011 and the 30 min datasets from March 7 to April 3, 2011 were randomly selected to test the forecasting ability of the proposed model.

The simulation datasets were divided into a training set and a verification or test set. Data for an entire day was used for verification. The lifespan of each simulation was one day, with continuous verification of 7 days. For the 10 min wind speed simulation, the total number of data points per simulation was 1148, including 1004 training data points and 114 test data points. For the 30 min simulation, the total number of data points for each simulation was 1052 including 1004 training data points and 48 test data points. Fig. 2 shows the geographical location of Yantai, and the 10 min and 30 min wind speed datasets from three observation sites. The parameters of the proposed MWDO-CEEMD-BP model are shown in Table 3.

3.3. Study case

Two simulation experiments, 10 min wind speed prediction and 30 min wind speed forecasting, are discussed in this section. To verify the forecasting accuracy and stability of the MWDO-CEEMD-BP model, six single models include ARIMA, BP, GRNN, ENN, RBF and SVR and five hybrid models including DE-BP, WDO-BP, MWDO-BP, MWDO-SSA-BP and MWDO-WT-BP were used as benchmark models. The single models were used to verify the other hybrid models have better forecasting ability. The benchmark models, DE-BP, WDO-BP and MWDO-BP, are used to verify the proposed optimization model MWDO has better optimal performance compared with the DE and WDO models. The MWDO-SSA-BP and MWDO-WT-BP models were utilized to verify that CEEMD has better denoising ability compared with SSA and WT. Study case one is the simulation of 10 min wind speed forecasting and study case two is the simulation of 30 min wind speed forecasting.

3.3.1. Study case one

The 10 min wind speed simulation results were evaluated using the four statistical errors metrics. The DM test and the bias-variance framework are described in this section. For each simulation, a full day of data was used as a validation set with a week of continuous verification. The simulation results of the first day of wind speed forecasting at observation site A are shown in Fig. 3. The four statistical errors are shown in part A, and the DM values are shown in part B. The forecasting relation and the forecasting curves are shown in part C and D, respectively.

The simulation results for all the models evaluated using the four statistical error metrics are shown in Table 4. The average results were evaluated using the DM test and bias-variance framework and are shown in Table 5. Intuitively, the smallest errors

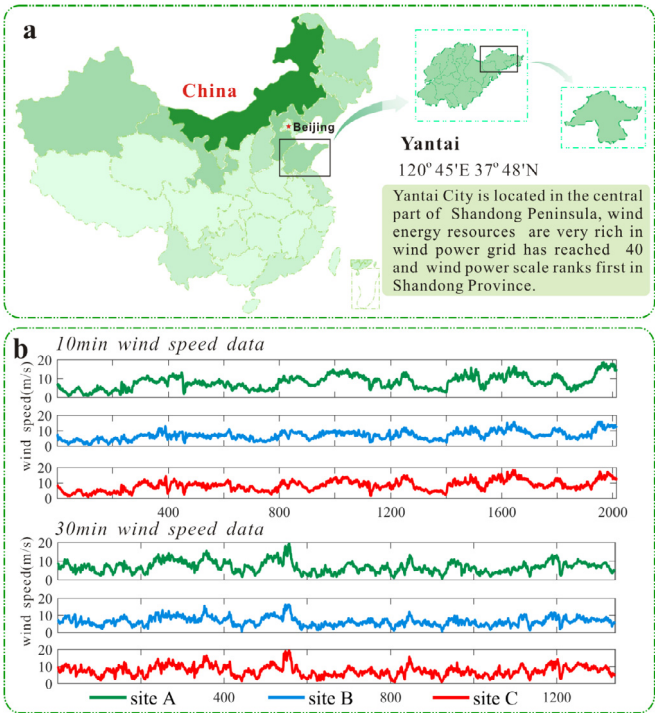

Fig. 2. Description of original wind speed data. (a) Location of the study sites; (b) Original wind speed series from three observation sites.

represent the best forecasting accuracy and stability. The values in bold indicate the smallest values of AE, RMSE, MAPE, MAE, DM value, bias and variance.

Tables 4 and 5 illustrate the following:

(a) For single models, we see that
  (1) Compared with SVR, ENN, GRNN and RBF, BP has the best performance based on the metrics AE, MAE, RMSE, and MAPE for the 10 min wind speed series from the three observation sites. The average AE, RMSE, MAE and MAPE of BP are −0.0709, 0.8465, 0.6164 and 7.23%, respectively.
  (2) For the bias-variance framework, the minimum variance belongs to the ARIMA where the value is 0.6455, and the

Table 3
Parameters of MWDO-CEEMD-BP

Model	Parameters	Default value
MWDO	Maximum number of iterations	100
	population size	20
	Dimension	31
	Coriolis effect	0.4
	RT coefficient	3
	coefficient of friction	0.4
	gravitational constant	0.2
	dimension boundaries	[-5,5]
	speed boundaries	[-0.3,0.3]
	Number of iterations of Broyden family	10
CEEMD	noise standard deviation	0.2
	number of realizations	50
	maximum number of sifting iterations allowed	500
BP	input layer	4
	hidden layer	5
	output layer	1
	iteration	100
	learning rate	0.1
	training requirement accuracy	0.000001

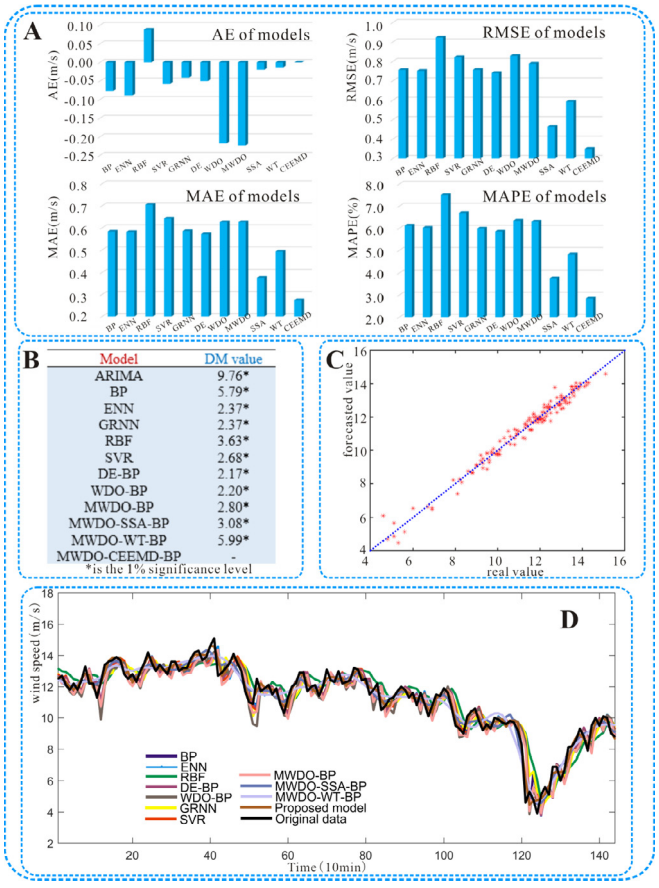

Fig. 3. The simulation results of the first day of wind speed forecasting at site.

minimum bias also belongs to the ARIMA where the value is 0.0632. BP also performed well based on variance and bias in all the single models. The values are 0.7127 and 0.0651, respectively; the differences between ARIMA are not large.

(3) For the DM test, the DM values of all single models are larger than the upper limits at the 1% significance level. The differences in forecasting wind speed series between MWDO-CEEMD-BP and single models are significant.

Remark: When comparing the different single forecasting models, it is observed that ARIMA has the best forecasting ability. However, ARIMA is not easy to optimize and its forecasting ability is much worse than the other proposed models in this paper. For other single models, BP has the smallest error based on the four statistical errors followed by ENN; the difference between the two models is not large. For the bias-variance framework, BP has the minimum bias, which indicates that it has the best forecasting accuracy. ENN has the minimum variance, which indicates that it has the best forecasting stability. BP and ENN are qualified single forecasting models for 10 min wind speed prediction. In this paper, BP was selected as the forecasting model.

Table 5
The results evaluated by DM test and bias-variance framework.

Model	Bias-Variance framework		D-M value
	Variance	Bias	
ARIMA	0.6455	0.0632	12.19[a]
BP	0.7127	0.0651	12.71[a]
ENN	0.7066	0.1228	12.92[a]
GRNN	1.3418	0.2036	11.91[a]
RBF	1.1685	0.1907	12.23[a]
SVR	1.4964	0.2572	8.74[a]
DE-BP	0.7030	0.1157	13.11[a]
WDO-BP	0.6943	0.0948	12.91[a]
MWDO-BP	0.6633	0.0732	13.31[a]
MWDO-SSA-BP	0.2497	**0.0090**	8.07[a]
MWDO-WT-BP	0.4054	0.0211	12.91[a]
MWDO-CEEMD-BP	**0.1499**	0.0140	–

[a] is the 1% significance level.

(b) For DE-BP, WDO-BP, and MWDO-BP models, we can see that

(1) Compared with DE-BP and WDO-BP, MWDO-BP has the smallest statistical errors. The AE, RMSE, MAE and MAPE of MWDO-BP are −0.0853, 0.8247, 0.6067 and 7.11%, respectively. Compared with DE-BP, MAPE of MWDO-BP decreased by 0.70%. Compared with WDO-BP, the value of MAPE decreased by 0.42%.

(2) For the bias-variance framework, the variances of DE-BP, WDO-BP and MWDO-BP are 0.7030, 0.6943 and 0.6633, respectively. The bias of DE-BP, WDO-BP and MWDO-BP are 0.1157, 0.0948 and 0.0732, respectively. For the DM test, the values of these three models are larger than the upper limits at the 1% significance level.

(3) Compared with neural network single models, these three hybrid models have better performance when

Table 4
The results of 10min wind speed forecasting.

		BP	ENN	RBF	SVR	GRNN	ARIMA	DE-BP	WDO-BP	MWDO-BP	MWDO-SSA-BP	MWDO-WT-BP	MWDO-CEEMD-BP
Site A	**AE**	−0.0751	−0.1278	−0.2382	0.3317	−0.2595	−0.0011	−0.1293	−0.0845	−0.1026	**−0.0137**	−0.0409	−0.0148
	RMSE	0.8068	0.8373	1.6188	1.4514	1.2731	0.7652	0.8393	0.8135	0.8118	0.4656	0.6013	**0.3770**
	MAE	0.5935	0.6105	0.8669	0.8124	0.8088	0.5550	0.6092	0.5917	0.5964	0.3462	0.4584	**0.2735**
	MAPE	6.67	6.73	9.22	7.88	8.37	6.35	6.65	6.65	6.63	3.99	5.24	**3.11**
Site B	**AE**	−0.0920	−0.1506	−0.2231	0.1925	−0.1674	0.0030	−0.1243	−0.0870	−0.1101	−0.0139	**−0.0232**	−0.0236
	RMSE	0.8639	0.8446	1.1070	0.9571	0.9973	0.7966	0.8344	0.8285	0.8167	0.4967	0.6640	**0.3767**
	MAE	0.6220	0.6151	0.7490	0.6663	0.7140	0.5858	0.6099	0.6053	0.6028	0.3711	0.5130	**0.2811**
	MAPE	7.87	7.74	9.36	8.14	8.81	7.52	7.70	7.73	7.66	4.79	6.67	**3.57**
Site C	**AE**	−0.0457	−0.0989	−0.1595	0.2558	−0.1544	−0.0015	−0.1026	−0.1237	−0.0431	−0.0082	−0.0109	**0.0024**
	RMSE	0.8688	0.8672	1.4275	1.2941	1.0029	0.8433	0.8655	0.8733	0.8456	0.5341	0.6442	**0.4075**
	MAE	0.6336	0.6318	0.8391	0.7714	0.7229	0.6151	0.6361	0.6355	0.6208	0.4048	0.4957	**0.3022**
	MAPE	7.14	7.07	9.44	7.98	8.03	6.88	7.12	7.03	7.03	4.56	5.56	**3.39**
Mean	**AE**	−0.0709	−0.1258	−0.2069	0.2600	−0.1938	0.0002	−0.1187	−0.0984	−0.0853	−0.0120	−0.0250	**−0.0120**
	RMSE	0.8465	0.8497	1.3844	1.2342	1.0911	0.8017	0.8464	0.8384	0.8247	0.4988	0.6365	**0.3871**
	MAE	0.6164	0.6191	0.8183	0.7500	0.7486	0.5853	0.6184	0.6108	0.6067	0.3740	0.4890	**0.2856**
	MAPE	7.23	7.18	9.34	8.00	8.40	6.91	7.16	7.14	7.11	4.45	5.82	**3.36**

evaluated based on RMSE, MAPE and the bias-variance framework.

RemarkCompared with single neural network models, hybrid models have better performance, and the WDO-BP has the smallest values of MAE, AE, RMSE, MAPE, bias and variance. The forecasting errors of MWDO-BP are smaller than those of WDO-BP. It is observed that the improved optimization model proposed in this paper can effectively avoid having the BP model falling into local optimization and improve the forecasting accuracy and stability.

(c) For MWDO-SSA-BP, MWDO-WT-BP and MWDO-CEEMD-BP, we see that
  (1) These three de-noising optimization forecasting models outperform other optimization algorithms using all four evaluation metrics, and the MAPE of these three hybrid models are 4.45%, 5.82 and 3.36%, respectively. Of all the models, the proposed model has the best forecasting results based on the four statistical errors: AE, RMSE, MAE and MAPE are −0.0120, 0.3871, 0.2856 and 3.36%, respectively.
  (2) For the bias-variance framework, the proposed model has the smallest variance among all the models, with a value of 0.1499. The smallest bias belongs to MWDO-SSA-BP, followed by MWDO-CEEMD-BP; the difference between the two models is not large.
  (3) For the DM test, the smallest DM value is 8.07 for all the models, which is far larger than the upper limits at the 1% significance level. The values indicate that the forecasting differences between MWDO-CEEMD-BP and other benchmark models are significant.

Remark: The average values of four statistical metrics indicate the proposed model has the best forecasting accuracy. In addition, the DM values of all benchmark models are far larger than the upper limits at the 1% significance level, which shows that the forecasting ability of these models is different. Combining the four statistical errors and the DM values, we can see that the proposed hybrid model has better forecasting ability. The bias of the proposed hybrid model is the second smallest and the variance is the smallest. The bias-variance framework indicates the model has the best forecasting accuracy and stability. We can see that the proposed model has the best forecasting ability for 10 min wind speed.

3.3.2. Study case two

Study case two is designed to test the 30 min wind speed forecasting ability of the proposed hybrid model by comparing it with other benchmark models, including single models and hybrid models. The simulation results of the first day of 30 min wind speed forecasting at observation site B are shown in Fig. 4. The results evaluated by four statistical metrics are shown in part A, the DM values are shown in part B and the forecasting curves are shown in part C. The results for the statistical metrics are shown in Table 6 and the average values of DM, bias and variance are shown in Table 7. The forecasting accuracy and stability of models were evaluated using these performance metrics.

From Tables 6 and 7, it could be found that:

(1) The resulting statistical metrics AE, MAE, RMSE and MAPE are compiled and presented in Table 6. The values in bold represent the AE, MAE, RMSE, and MAPE values that are the lowest among all the forecasting models at the three observation sites. It can be seen that the proposed hybrid model has the highest accuracy at all wind farm sites with average values of RMSE, MAE and MAPE of 0.4377, 0.3299 and 5.45%,

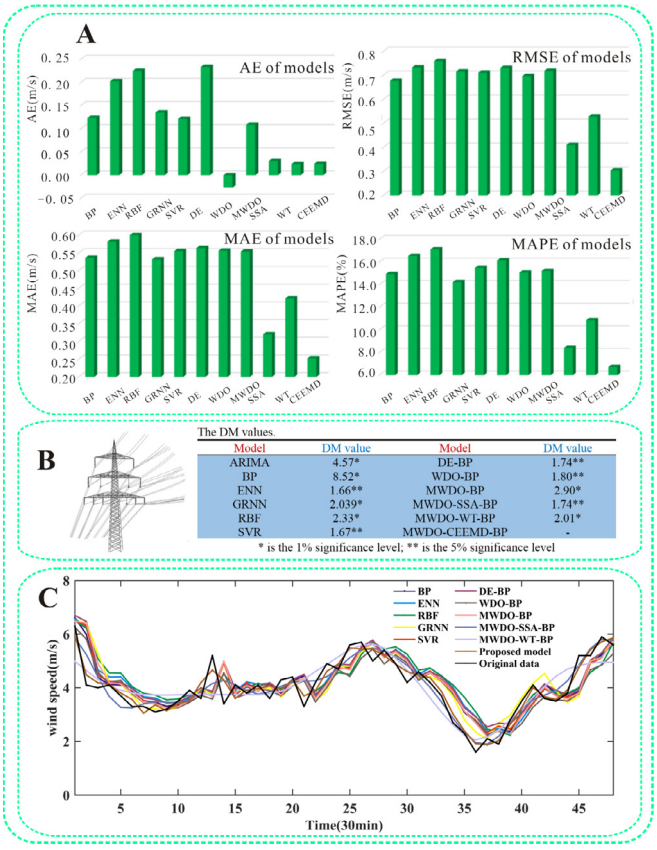

Fig. 4. The simulation results of the first day of wind speed forecasting at site B.

respectively. Although the minimum AE value does not belong to the proposed model, the differences between minimum AE value and the value of MWDO-CEEMD-BP are not significant

(2) For the bias-variance framework, the proposed model has the best bias and variance with values of 0.0227 and 0.1944, respectively. Compared with the best single models, the bias and variance of the proposed model is decreased by 45.04% and 77.92%, respectively, which proves the proposed model has better forecasting accuracy and stability than the single models. When compared with the MWDO-SSA-BP model, the bias and variance is seen to decrease by 24.58% and 36.82%, respectively.

(3) For the DM test, the DM values of all the models are far larger than the upper limits at the 1% significance level.

Remark: The MWDO-SSA-BP model has the best AE. The best MAE, RMSE and MAPE belong to the proposed model, we can see that the model has the best forecasting accuracy based on the statistical metrics. The DM values of all the models show that the differences between the forecasted series of the proposed model and the forecasted series of other models are significant. The bias and variance of the proposed model indicates that the model has the best forecasting accuracy and stability among all models. From the results evaluated by these performance metrics, we can see that the model has the best 30 min wind speed forecasting ability.

To compare the forecasting ability of the model more intuitively, the average results for 10 min and 30 min wind speed forecasting evaluated by AE, RMSE, MAE, MAPE, the DM value and the bias-variance framework are compiled and presented in Table 8. Based on the statistical errors, bias-variance framework and DM test, it is

Table 6
The results of 30min wind speed forecasting.

		BP	ENN	RBF	SVR	GRNN	ARIMA	DE-BP	WDO-BP	MWDO-BP	MWDO-SSA-BP	MWDO-WT-BP	MWDO-CEEMD-BP
Site A	**AE**	−0.0104	0.0209	0.0150	0.0178	−0.0409	0.0252	−0.0079	−0.0176	**0.0000**	0.0010	0.0003	−0.0016
	RMSE	0.8190	0.8097	1.0768	0.8213	0.8784	0.8013	0.8120	0.8064	0.8131	0.4731	0.6661	**0.3713**
	MAE	0.6072	0.5999	0.7554	0.5964	0.6329	0.5955	0.5916	0.6013	0.5949	0.3459	0.5023	**0.2772**
	MAPE	10.95	10.96	14.16	10.80	11.74	10.65	10.76	10.91	10.79	6.11	8.91	**4.78**
Site B	**AE**	−0.0084	−0.0067	0.0133	0.0083	0.0027	0.0051	0.0489	−0.0314	−0.0351	**0.0022**	0.0056	0.0072
	RMSE	0.9875	0.9944	1.1073	0.9955	1.0698	0.9916	1.0136	0.9916	0.9770	0.5772	0.7801	**0.4620**
	MAE	0.7359	0.7346	0.7688	0.7410	0.7869	0.7364	0.7517	0.7347	0.7317	0.4478	0.6109	**0.3533**
	MAPE	13.34	13.47	14.05	13.54	14.80	13.25	14.00	13.30	13.25	7.82	10.74	**6.14**
Site C	**AE**	−0.0053	0.0047	−0.0173	0.0646	−0.0737	0.0208	−0.0040	0.0103	−0.0435	**0.0008**	0.0034	−0.0111
	RMSE	1.0069	1.0074	1.2599	1.0270	1.0588	1.0046	0.9979	1.0016	0.9982	0.6029	0.7891	**0.4799**
	MAE	0.7553	0.7550	0.8130	0.7658	0.8079	0.7525	0.7463	0.7542	0.7495	0.4613	0.6091	**0.3590**
	MAPE	11.66	11.68	12.51	11.68	12.61	11.57	11.47	11.63	11.49	6.85	9.20	**5.43**
Mean	**AE**	−0.0080	0.0063	0.0037	0.0302	−0.0373	0.0170	0.0124	−0.0129	−0.0262	**0.0013**	0.0031	−0.0018
	RMSE	0.9378	0.9371	1.1480	0.9480	1.0023	0.9325	0.9412	0.9332	0.9294	0.5511	0.7451	**0.4377**
	MAE	0.6995	0.6965	0.7791	0.7010	0.7426	0.6948	0.6965	0.6967	0.6920	0.4183	0.5741	**0.3299**
	MAPE	11.99	12.04	13.57	12.01	13.05	11.82	12.08	11.95	11.84	6.93	9.62	**5.45**

Table 7
The results evaluated by DM test and bias-variance framework.

Model	Bias-Variance framework		D-M value
	Variance	Bias	
ARIMA	0.8804	0.0460	7.22[a]
BP	0.8892	0.0505	7.85[a]
ENN	0.8889	0.0492	7.85[a]
GRNN	1.0130	0.0431	7.89[a]
RBF	1.1512	0.0565	7.09[a]
SVR	0.9080	0.0423	7.89[a]
DE-BP	0.8960	0.0413	7.45[a]
WDO-BP	0.8811	0.0461	8.09[a]
MWDO-BP	0.8722	0.0391	7.91[a]
MWDO-SSA-BP	0.3077	0.0301	4.44[a]
MWDO-WT-BP	0.5600	0.0405	7.83[a]
MWDO-CEEMD-BP	**0.1944**	**0.0227**	–

[a] Is the 1% significance level.

seen that the proposed model can effectively improve the accuracy and stability of wind speed forecasting, and the improvement is significant.

4. Discussion

Several factors related to predictive modeling, including the parameter settings of neural networks and the are seasonal pattern are discussed in this section.

4.1. Neural network

Many types of neural networks are widely used for forecasting nonlinear trends. These models have achieved great success in time series forecasting. Our experiments show that neural networks have good performance in the wind speed prediction. However, the forecasting results of neural networks have instabilities which depend on parameter settings. In this paper, we selected the BP neural network as the forecasting model. To achieve better forecasting accuracy, we need to determine suitable parameters, set better initial weights and thresholds and determine more suitable input layers and hidden layers.

To identify suitable parameters, we tested many parameter combinations, but found it was difficult to find a relationship between the parameter combinations and forecasting accuracy. The initial weights and thresholds of the neural network are randomly generated, which can lead to the neural network reaching a local optimum. In addition the convergence rate is too slow. To overcome these drawbacks, a modified wind driven optimization algorithm was utilized to find better initial weights and thresholds. The experimental results show that the optimization method can effectively improve the forecasting accuracy and stability of the model.

For neural networks, the choice of input layers and hidden layers has a great impact on forecasting accuracy. If the number of input layer nodes and hidden layer nodes is too small, a BP neural network cannot establish a complex mapping relationship, which leads to large forecasting errors. However, if the number of nodes is too large, the learning time of the neural network increases, and over-fitting phenomenon may appear. Therefore, it is very important and necessary to select the appropriate number of nodes for the input layer and hidden layer. In this paper, an exhaustive method was used to select the best input and hidden layers, and the

Table 8
The average results of 10min and 30min wind speed forecasting.

	AE	RMSE	MAE	MAPE	Bias-variance		DM value
					Variance	Bias	
ARIMA	0.0086	0.8671	0.6401	9.365	0.7630	0.0546	9.96[a]
BP	−0.0395	0.8922	0.6579	9.61	0.8009	0.0578	10.28[a]
ENN	−0.0597	0.8934	0.6578	9.61	0.7977	0.0860	10.39[a]
GRNN	−0.1155	1.0467	0.7456	10.73	1.1774	0.1233	9.90[a]
RBF	−0.1016	1.2662	0.7987	11.46	1.1599	0.1236	9.66[a]
SVR	0.1451	1.0911	0.7255	10.00	1.2022	0.1498	8.31[a]
DE-BP	−0.0532	0.8938	0.6575	9.62	0.7995	0.0785	10.28[a]
WDO-BP	−0.0557	0.8858	0.6538	9.54	0.7877	0.0704	10.50[a]
MWDO-BP	−0.0557	0.8771	0.6493	9.48	0.7678	0.0562	10.61[a]
MWDO-SSA-BP	−0.0053	0.5249	0.3962	5.69	0.2787	0.0195	6.26[a]
MWDO-WT-BP	−0.0109	0.6908	0.5316	7.72	0.4827	0.0308	10.37[a]
MWDO-CEEMD-BP	**−0.0069**	**0.4124**	**0.3077**	**4.40**	**0.1722**	**0.0183**	–

[a] is the 1% significance level.

results are shown in Table 9. We observed that when the number of input nodes and hidden nodes is 4 and 5, respectively, the BP network has the best forecasting performance, and the MAPE is 9.61%.

Aside from the discussion above, a training and verification ratio is also required to be set. We set the training and verification ratio to be 4:1, 8:1, 12:1, 16:1 and 20:1. However, the experimental results show that the relationship between the different training and verification ratios and the forecasting results is not obvious. Although neural networks have many limitations, these models have good wind speed forecasting accuracy.

4.2. Multiple seasonal patterns

The statistical properties of the wind speed exhibit a strong seasonal pattern. To further verify the forecasting performance of the proposed model, multiple seasonal patterns are used to reduce interferences from the original data. The simulation of 10min and 30min wind speed forecasting on other three seasons at three observation sites are added in this paper. The datasets of July 30 to August 26 (summer), September 26 to October 23 (autumn), and January 8 to February 4 (winter) were randomly selected to test the forecasting ability of the proposed model.

The average values of simulation results evaluated by four metrics at three observation sites are shown in Table 10. The values in bold indicate the smallest values of AE, RMSE, MAE and MAPE. The MAE, RMSE and MAPE of the proposed model are the smallest in all related model. The AE values of proposed model are not the smallest, but the difference between the minimum value is not large. The results evaluated by DM test and bias-variance framework are shown in Table 11. The DM values of all models shown that the differences between forecasting series of the model and the forecasting series of other models are significant. The bias and variance of the proposed model indicate that the model has the best forecasting accuracy and stability among all models. The simulation results further show that the proposed model can effectively improve the accuracy and stability of 10min and 30min wind speed forecasting in all seasons.

Table 9
Comparison of BP with different numbers of input and hidden nodes (MAPE).

Input	Hidden layer									
layer	3	4	5	6	7	8	9	10	11	12
3 (%)	9.71	9.69	9.66	9.83	9.91	9.68	9.64	9.82	9.68	9.83
4 (%)	9.85	9.75	**9.61**	9.79	9.69	9.79	9.64	9.82	9.73	9.90
5 (%)	9.72	9.91	9.63	9.67	9.73	9.66	9.81	9.71	9.74	9.72
6 (%)	9.76	9.83	10.00	9.83	9.72	9.94	9.91	9.80	9.87	9.78
7 (%)	9.65	9.64	9.80	9.77	9.84	9.86	9.83	9.82	9.76	9.70
8 (%)	9.86	9.81	9.72	9.90	9.89	9.90	9.98	9.73	9.93	9.72

5. Conclusions

Accurate and stable forecasting of complex wind speed sequences is a difficult and challenging task that is necessary to ensure the safety of the power grid. To predict the features of wind speed sequences, a new hybrid model, MWDO-CEEMD-BP, combining modified wind driven optimization, complementary ensemble empirical mode, and a BP neural network is proposed for 10 min and 30 min wind speed forecasting in this paper. CEEMD as a very effective means of de-noising and plays a highly important role in removing noise from raw data. This model is able to extract characteristics of the data, after which several intrinsic mode functions (IMFs) are obtained by CEEMD decomposition. Then, the forecast of each IMF with BP optimized by modified wind driven optimization individually. The optimization algorithm, namely, MWDO, proposed in this paper combines the WDO model and the Broyden family. It was then utilized to optimize the initial thresholds and weights of the BP neural network. Finally, all of the forecaste IMFs results were integrated to generate the final predicted value. To verify the forecasting accuracy and stability of the proposed model, 10 min and 30 min wind speed data at three observation sites in Yantai, Shandong Province, China, were acquired. The simulation results, evaluated by four statistical metrics, the bias-variance framework and the DM test, show that the proposed model has the best forecasting ability when compared with other single and hybrid models. The MAPE of the proposed model

Table 10
The average values of simulation results at three observation sites.

		BP	ENN	RBF	SVR	GRNN	ARIMA	DE-BP	WDO-BP	MWDO-BP	MWDO-SSA-BP	MWDO-WT-BP	MWDO-CEEMD-BP
10min													
summer	**AE**	0.0838	0.0682	0.0737	−0.0308	0.0867	0.0695	0.0533	0.0650	0.0473	**0.0094**	0.0234	0.0103
	RMSE	0.4331	0.4127	0.2535	0.4030	0.4695	0.4330	0.4040	0.4217	0.4079	0.2509	0.3515	**0.1943**
	MAE	0.3314	0.3175	0.3829	0.3096	0.3613	0.3307	0.3112	0.3231	0.3120	0.1934	0.2719	**0.1502**
	MAPE	11.42	10.85	13.55	10.51	13.01	11.34	10.44	11.14	10.34	6.31	9.38	**4.90**
autumn	**AE**	0.0379	0.0206	0.0437	−0.0068	0.0304	0.0322	0.0080	0.0179	0.0160	**0.0027**	0.0098	0.0118
	RMSE	0.4783	0.4783	0.3522	0.4719	0.5159	0.4953	0.5025	0.4927	0.4947	0.2793	0.3806	**0.2177**
	MAE	0.3576	0.3556	0.4494	0.3531	0.3907	0.3631	0.3662	0.3606	0.3612	0.2130	0.2910	**0.1640**
	MAPE	10.84	10.64	14.82	10.58	12.17	10.82	10.67	10.70	10.71	6.26	8.63	**4.77**
winter	**AE**	0.0291	0.0189	0.0237	0.0058	0.0039	0.0257	0.0058	0.0356	0.0100	**0.0030**	0.0051	0.0038
	RMSE	0.5735	0.5707	0.4821	0.5767	0.6219	0.5734	0.5772	0.5844	0.5719	0.3615	0.4639	**0.2898**
	MAE	0.4236	0.4218	0.5305	0.4238	0.4685	0.4250	0.4237	0.4310	0.4211	0.2700	0.3557	**0.2064**
	MAPE	8.39	8.36	10.88	8.33	9.49	8.42	8.33	8.55	8.26	5.35	7.20	**3.94**
30min													
summer	**AE**	0.0736	0.0661	0.0411	−0.0036	0.0531	0.1133	0.0575	0.0402	0.0509	**0.0075**	0.0353	−0.0077
	RMSE	0.6165	0.6200	0.5482	0.6099	0.6662	0.6286	0.6220	0.6182	0.6087	0.4043	0.5395	**0.2988**
	MAE	0.4854	0.4911	0.5900	0.4755	0.5332	0.4957	0.4924	0.4904	0.4776	0.3148	0.4219	**0.2369**
	MAPE	16.88	16.99	21.48	16.06	19.25	17.45	17.13	16.77	16.59	10.66	14.21	**7.92**
autumn	**AE**	0.0913	0.0695	0.0585	−0.0145	0.0552	0.0932	0.0541	0.0737	0.0672	**0.0093**	0.0235	0.0129
	RMSE	0.7421	0.7300	0.7642	0.7262	0.7760	0.7372	0.7369	0.7232	0.7196	0.4567	0.6148	**0.3388**
	MAE	0.5629	0.5511	0.5754	0.5480	0.5819	0.5608	0.5570	0.5483	0.5466	0.3505	0.4808	**0.2646**
	MAPE	16.89	16.59	17.73	16.16	17.80	17.22	16.59	16.53	16.41	10.38	14.75	**7.77**
winter	**AE**	0.0912	0.0844	0.1013	−0.0612	0.0990	0.1061	0.0791	0.0877	0.0863	0.0109	0.0109	**0.0026**
	RMSE	0.8855	0.8812	0.9591	0.8896	0.9129	0.8935	0.8869	0.8864	0.8814	0.5217	0.7419	**0.4410**
	MAE	0.6739	0.6647	0.7531	0.6687	0.6884	0.6759	0.6697	0.6708	0.6650	0.3953	0.5747	**0.3329**
	MAPE	13.96	13.83	16.21	13.93	14.67	14.27	13.83	13.97	13.86	8.02	11.82	**6.63**

Table 11
The DM values, variance and bias of related models.

model	summer			autumn			winter		
	variance	bias	DM value	variance	bias	DM value	variance	bias	DM value
10min									
BP	0.1815	0.0824	16.42[a]	0.2275	0.0343	15.04[a]	0.3287	0.0222	13.25[a]
ENN	0.1667	0.0669	16.64[a]	0.2285	0.0167	14.63[a]	0.3258	0.0119	13.13[a]
RBF	0.2475	0.0717	16.45[a]	0.3506	0.0393	16.00[a]	0.4821	0.0102	15.28[a]
SVR	0.1619	0.0281	16.43[a]	0.2228	0.0094	14.97[a]	0.3329	0.0175	13.33[a]
GRNN	0.2142	0.0852	17.25[a]	0.2655	0.0255	15.46[a]	0.3874	0.0130	13.88[a]
ARIMA	0.1853	0.0681	16.26[a]	0.2445	0.0267	13.34[a]	0.3286	0.0181	13.36[a]
DE-BP	0.1616	0.0510	16.47[a]	0.2526	0.0092	13.53[a]	0.3336	0.0132	13.04[a]
WDO-BP	0.1721	0.0636	16.48[a]	0.2421	0.0217	13.41[a]	0.3404	0.0285	12.92[a]
MWDO-BP	0.1677	0.0453	16.63[a]	0.2245	0.0287	15.30[a]	0.3274	0.0131	13.21[a]
MWDO-SSA-BP	0.0634	**0.0049**	8.97[a]	0.0783	**0.0084**	9.16[a]	0.1308	0.0109	6.92[a]
MWDO-WT-BP	0.1239	0.0198	14.70[a]	0.1448	0.0167	13.34[a]	0.2155	0.0131	11.79[a]
MWDO-CEEMD-BP	**0.0378**	0.0127	–	**0.0474**	0.0120	–	**0.0841**	**0.0078**	–
30min									
BP	0.3787	0.0558	10.54[a]	0.5440	0.0809	9.34[a]	0.7779	0.0749	8.72[a]
ENN	0.3872	0.0487	10.64[a]	0.5300	0.0570	9.32[a]	0.7723	0.0692	8.43[a]
RBF	0.5468	0.0511	10.73[a]	0.5811	0.0616	9.07[a]	0.9512	0.0851	9.59[a]
SVR	0.3732	0.0318	10.17[a]	0.5293	0.0361	9.32[a]	0.7910	0.0387	8.26[a]
GRNN	0.4447	0.0371	10.43[a]	0.6032	0.0353	8.09[a]	0.8260	0.0839	8.77[a]
ARIMA	0.3847	0.1081	10.59[a]	0.5360	0.0828	9.48[a]	0.7893	0.0935	8.34[a]
DE-BP	0.3795	0.0683	10.60[a]	0.5407	0.0547	8.98[a]	0.7826	0.0667	8.53[a]
WDO-BP	0.3812	0.0446	10.63[a]	0.5186	0.0597	9.26[a]	0.7807	0.0719	8.39[a]
MWDO-BP	0.3663	0.0611	10.45[a]	0.5149	0.0527	9.57[a]	0.7725	0.0712	8.49[a]
MWDO-SSA-BP	0.1647	0.0201	5.75[a]	0.2093	0.0231	5.85[a]	0.2731	0.0262	3.67[a]
MWDO-WT-BP	0.2904	0.0287	8.64[a]	0.3790	0.0189	8.68[a]	0.5521	0.0388	8.16[a]
MWDO-CEEMD-BP	**0.0896**	**0.0143**	–	**0.1149**	**0.0142**	–	**0.1953**	**0.0238**	–

[a] Is the 1% significance level.

decreased by 40.80%, 39.97% and 26.30%, respectively, when compared with the results of BP, MWDO-BP and MWDO-SSA-BP. In other words, both CEEMD and MWDO play a vital role in contributing to the forecasting accuracy from the aspect of de-noising and optimization, respectively. The proposed model has good nonlinear forecasting ability, and it can also be applied in many other fields.

Acknowledgment

This work was supported by the National Social Science Foundation of China [Grant564 No.13&ZD171].

References

[1] National Energy Administration. The eleventh five-year plan for electric power development (2016-2020). http://www.nea.gov.cn/news/jwzdt.htm.
[2] Meng Anbo, Ge Jiafei, Yin Hao, Chen Sizhe. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers Manag 2016;114:75–88.
[3] Zhao Weigang, Wei Yi-Ming, Su Zhongyue. One day ahead wind speed forecasting: a resampling-based approach. Appl Energy 2016;178:886–901.
[4] Xiao Ling, Wang Jianzhou, Dong Yao, Wu Jie. Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 2015;44:271–88.
[5] Lei M, Jiang C, Liu H, Zhang Y. A review on forecasting wind data and wind output. Renew Sustain Energy Rev 2009;13:915–20.
[6] Zhao Jing, Guo Zhen-Hai, Su Zhong-Yue, Zhao Zhi-Yuan, Xiao Xia, Liu Feng. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. Appl Energy 2016;162:808–26.
[7] Lynch Conor, OMahony Michael J, Scully Ted. Simplified method to derive the Kalman filter covariance matrices to predict wind speeds from a NWP model. Energy Procedia 2014;62:676–85.
[8] Khashei M, Bijari M, Raissi Ardali GA. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). Neurocomputing 2009;72:956–67.
[9] Maatallah OA, Achuthan A, Janoyan K, Marzocca P. Recursive wind speed forecasting based on Hammerstein auto-regressive model. Appl Energy 2015;145:191–7.
[10] Liu Xiuli, Moreno Blanca, García Ana Salome. A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors. Energy 2016;115:1042–54.
[11] Bahrami Saadat, Hooshmand Rahmat-Allah, Parastegari Moein. Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. Energy 2014;72:434–42.
[12] Kavasseri Rajesh G, Seetharaman Krithika. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[13] Erdem Ergin, Shi Jing. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[14] Dowell Jethro, Pinson Pierre. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. IEEE Trans Smart Grid 2016;7.2:763–70.
[15] Raza Muhammad Qamar, Khosravi Abbas. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev 2015;50:1352–72.
[16] Okumus Inci, Dinler Ali. Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Convers Manag 2016;123:362–71.
[17] Ata Rasit. Artificial neural networks applications in wind energy systems: a review. Renew Sustain Energy Rev 2015;49:534–62.
[18] Wang Deyun, Luo Hongyuan, Grunder Olivier, Lina Yanbing, Guo Haixiang. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl Energy 2017;190:390–407.
[19] Ren Chao, An Ning, Wang Jianzhou, Li Lian, Hu Bin, Shang Duo. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl Base Syst 2014;56:226–39.
[20] Liu Hui, Tian Hong-qi, Liang Xifeng, Li Yanfei. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94.
[21] Wang Jianzhou, Qin Shanshan, Zhou Qingping, Jiang Haiyan. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. Renew Energy 2015;76:91–101.
[22] Zhang Wenyu, Wang Jujie, Wang Jianzhou, Zhao Zengbao, Tian Meng. Short-term wind speed forecasting based on a hybrid model. Appl Soft Comput 2013;13:3225–33.
[23] Kong Xiaobing, Liu Xiangjie, Shi Ruifeng, Lee Kwang Y. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[24] Zhou Junyi, Shi Jing, Li Gong. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manag 2011;52:1990–8.
[25] Liu Da, Niu Dongxiao, Wang Hui, Fan Leilei. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62:592–7.
[26] Jiang Ping, Ma Xuejiao. A hybrid forecasting approach applied in the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms. Appl Math Model 2016;000:1–19.
[27] Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. Renew

Energy 2016;85:790–809.
[28] Dong Qingli, Sun Yuhuan, Li Peizhi. A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: a case study of wind farms in China. Renew Energy 2017;102:241–57.
[29] Wang Shouxiang, Zhang Na, Wu Lei, Wang Yamin. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew Energy 2016;94:629–36.
[30] Aghajani Afshin, Kazemzadeh Rasool, Ebrahimi Afshin. A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. Energy Convers Manag 2016;121:232–40.
[31] Huang N, Shen Z, Long S, Wu M, Shih H, Zheng Q. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc. R. Soc. Lond. A 1998;454:903–95.
[32] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 2009;01:1–41.
[33] Delgado-Arredondo Paulo Antonio, Morinigo-Sotelo Daniel, Osornio-Rios Roque Alfredo, Avina-Cervantes Juan Gabriel, Rostro-Gonzalez Horacio, Romero-Troncoso Rene de Jesus. Methodology for fault detection in induction motors via sound and vibration signals. Mech Syst Signal Process 2017;83:568–89.
[34] Yeh R, Shieh JS, Huang NE. Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. Adv Adapt Data Anal 2010;2:135–56.
[35] Noorollahi Younes, Ali Jokar Mohammad, Kalhor Ahmad. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. Energy Convers Manag 2016;115:17–25.
[36] Bello Cruz JY, Ferreira OP, Németh SZ, Prudente LF. A semi-smooth Newton method for projection equations and linear complementarity problems with respect to the second order cone. Lin Algebra Appl 2017;513:160–81.
[37] Byrd RH, Liu DC, Nocedal J. On the behavior of Broyden's class of quasi-Newton methods. SIAM J Optim 1992;2:533–57.
[38] Bazaraa Mokhtar S, Sherali Hanif D, Shetty Chitharanjan M. Nonlinear programming: theory and algorithms. John Wiley & Sons; 2013.
[39] Broyden Charles George. The convergence of a class of double-rank minimization algorithms 1. general considerations. IMA J Appl Math 1970;6:76–90.
[40] Bayraktar Z, Komurcu M, Werner DH. Wind driven optimization (WDO): a novel nature-inspired optimization algorithm and its application to electromagnetics. In: 2010 IEEE antennas and propagation society international symposium (APSURSI). IEEE; 2010. p. 1–4.
[41] Diebold FX, Mariano R. Comparing predictive accuracy. J Bus Econ Stat 1995;13:253–63.