Research on a Novel Combination System on the Basis of Deep Learning and Swarm Intelligence Optimization Algorithm for Wind Speed Forecasting

XIAOHUI HE[1], YING NIE[2], HENGLIANG GUO[1], AND JIANZHOU WANG[2]

[1]School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China
[2]School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

Corresponding author: Ying Nie (nieyingaa@163.com)

ABSTRACT Wind speed forecasting takes a significant place in electric system owing to the fact that it has significant influence on operation efficiency and economic benefits. Aimming at improving forecast performance, a substantial number of wind speed prediction models have been proposed. However, these models have disregarded the limits of individual prediction models and the necessity of data preprocessing, resulting in poor prediction accuracy. In this study, a novel forecasting system is proposed consisting of three modules: data preprocessing module, individual forecasting module and weight optimization module, which effectively achieve better forecasting ability. For data preprocessing and individual forecasting module, more regular sequences are obtained by decomposition technology, and association features are extracted by deep learning algorithm in prediction module. In the weight optimized module, the combination method base on the multi-objective optimization algorithm and nonnegative constraint theory are used to improve the prediction effectiveness. The combination model successfully exceeds the limits of individual predicton models and comparatively improves prediction accuracy. The effectiveness of the developed combination system is evaluated by 10-min wind speed in Penglai, China. The experiment results indicate that proposed forecasting system is better than other traditional forecasting models on three real wind speed datasets indeed.

INDEX TERMS Wind speed forecasting, deep learning, multi-objective optimization algorithm, combination system.

NOMENCLATURE

MLP	Multi-layer perceptron model
LSTM	Long Short-Term Memory network
ARIMA	Autoregressive Integrated Moving Average
RNN	Recycle neural network
ANN	Artificial neural network
SVM	Support vector machine
GRNN	Generalized Regression Neural Network
RBF	Radial Basis Function network
WNN	Wavelet neural network
BP	Back Propagation Neural Network
ARMA	Auto-Regressive Moving Average method
FARIMA	Fractional ARIMA
ES	Exponential smoothing
EEMD	Ensemble Empirical Mode Decomposition
MOP	Multi-objective optimization problem
MOBA	Multi-objective bat algorithm
AE	Average forecast error
MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean square error
MAPE	Mean absolute percentage error
I_{MAE}	The percentage error of MAE
I_{MSE}	The percentage error of MSE
I_{MAPE}	The percentage error of MSE
MOPSO	Multi-objective particle swarm optimization algorithm

BFGS	Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm
EEMD-LSTM	EEMD preprocessing data for LSTM prediction
EEMD-MLP	EEMD preprocessing data for MLP prediction
EEMD-ARIMA	EEMD preprocessing data for ARIMA prediction

I. INTRODUCTION

With the improving attention to clean energy, the utilization rate of resources is raising day by day. At the same time, the resource lacking has also become a critical problem to be solved. As widely distributed, pollution-free renewable energy, wind energy is the theme of new energy resources analysis and development. Wind energy is a kind of fastest growing renewables, and is considered as an alternative to traditional fuel-fired electricity generation. Prediction of wind speed is the soul of wind energy system and takes a vital place in the supervision of wind farms. The precisely wind speed forecast is also significant for improving wind energy utilization and stable electric system operation. In contrast, incorrect wind speed prediction can lead to unfavorable decisions and wind power systems can be caused huge economic losses. Recently, the electric power produced by wind power has been increasing greatly. Wind power is advance in the aspects of its reliability, good ability and low price, and the utilization of wind energy helps to reduce air pollution, which is the largest environment task for most regions and countries [1].

Wind speed forecasting brings the decision-making challenge to electric system running due to the prediction has its own error [2]. Conventional approaches of wind speed prediction pay close attention to the potential features of previous data and the effects of numerical weather on wind speed [3]. In recent decades, some artificial intelligence forecasting models have been developed for wind speed prediction due to the rapid development of artificial technology, such as artificial neural network (ANN) [4], [5], fuzzy logic method [6] and SVM [7]. Short term wind speed prediction methods could be separated into 4 classes [8]: (a) physical approaches; (b) statistical approaches; (c) artificial intelligence methods; and (d) hybrid model.

The physical model is the numerical weather forecast, which mainly uses the detailed information of the lower atmosphere to analyze, mine and forecast [9]. This model is based on the basic information of the wind turbine provided by the numerical weather forecast system, parameterizes the physical phenomena according to the primary conditions and the nonlinear partial differential equation system, and then obtains a series of different meteorological parameters [10]. For example, Wilgan *et al.* [11] established a comprehensive neutral atmosphere model with high spatial-temporal resolution for prediction. The combined model will be according to the NWP model, with high spatial resolution and vertical message about different meteorological parameters. Statistical model is according to a large number of historical data to study the prediction model, without considering the impact of many meteorological factors. Milligan *et al.* [12], [13] used ARMA model to forecast wind power of wind farms in the United States. Maatallah *et al.* [14] combined Hammerstein model and AR model, and thus proposed a new wind speed forecasting model, which finally achieved better prediction effectiveness. Nevertheless, on the basis of the hypothesis that there is a linear model between time series, statistical methods can't successfully capture the nonlinear structure [15].

In addition, with the rapid development and wide application of artificial intelligence algorithm, a lot of scholars already successfully used artificial intelligence method to carry on wind speed forecasting. JursaR [16] proposed a short-term wind power predicting model according to particle swarm optimization (PSO). Amjady *et al.* [17] established a prediction model on the basis of the ridge neural network (RNN) as a prediction engine to effectively predict wind power. An advanced prediction model put forward for very short-term wind power prediction in [18] is combined the adaptive Bayesian learning and approximated Gaussian process. In [19], Zhang *et al.* used the radial basis function (RBF) network as well as the swarm intelligence optimization algorithm to forecast the wind speed in the interval, and finally achieve high prediction accuracy. As the artificial intelligence model has strong nonlinear prediction ability, it is generally better than the time series model [20].

Reviewing the previous literature, it is believed that the above prediction methods own certain of intrinsic shortcomings. The shortcomings of these approaches are summed up bellow:

(1) Physical model is easy to model and cost less. However, compared with the statistical model, the physical model has more requirements on data. Models based on air pressure, terrain and temperature are usually used for wind speed prediction of long-term situation of various weather variables [21].

(2) Statistical methods include ARMA [22], ARIMA [23], fractional ARIMA (FARIMA) [24] and so on, whose major restriction is the pre-assumed linear form of the model. And exponential smoothing (ES) [25] is established according to the relationship between the variables, using mathematical statistics method to represent the latent correlation of historical samples in wind speed prediction. Nevertheless, the elegant statistical model according to the linear structure between time series does not grasp the nonlinear mode of wind speed time series well. Only in some specific cases can statistical models achieve higher accuracy [26].

(3) Different from other methods, artificial intelligence algorithm can effectively find the implicit non-linear relationship between historical data, which has been widely studied and applied in solving complex relationship and accurate prediction. However, there are still some shortcomings in artificial intelligence methods, such as local optimization, over adaptation and relatively low convergence rate.

(4) The hybrid model integrates two or more methods, such as data preprocessing and prediction model, and integrates the advantages of each method. With the disregard of the limits of individual prediction models, the forecasting power of the hybrid model seems to need improvement.

Therefore, in view of the shortcomings of the above prediction methods, several combination wind speed prediction models are proposed. Bates and Granger proposed a combination prediction theory with good results in 1969 [27]. Since the 1970s, the research of combinatorial forecasting model has been widely concerned [28], [29]. In the aspect of wind speed forecast, Retrospected and sorted the combination wind speed prediction models, Xiao *et al.* proposed two combined wind speed prediction models which have preferable prediction abilities: the first is no negative constraint theory (NNCT) combined model, the second is meta-heuristic algorithm combined model [27]. A powerful combination prediction model was developed by Wang *et al.* in 2017, which includes GRNN, RBF, WNN and BP. The model combines these single prediction short-term wind speed prediction models applying meta-heuristic optimization method MOBA. The results show the established combination model is superior to individual model in the aspect of accuracy and reliability of forecasting.

Generally speaking, the goal of combinatorial prediction mechanism is to search the best weight by diminishing the total of prediction error squares of training sets provided by each model. Nevertheless, quite a few conventional wind speed prediction models ignore the importance of accuracy and stability, and these two are equally significant in the effective prediction of wind speed. Ideally, when the weights of member model in the combination model are obtained by optimization algorithm, the excellent accuracy and stability are obtained at the same time. Multiple objective problems (MOPs) aim at the simultaneous optimization of multiple conflicting objectives, which has aroused extensive research interest. For a nontrivial optimized question, a range of solutions exist when optimizing each objective concurrently, rather than a single solution, and this is different from single objective optimization.

However, wind speeds are often highly nonlinear, irregular, and nonstationary, but upward moving in secular trend. Many of the traditional predictive models ignore the both the accuracy and stability, which is significance to the wind speed prediction. **The leading dedications and innovations of this article are summed up as:**

(1) *An effective data preprocessing method is selected as the step before the model forecast.* The preprocessing approach is used to diminish the objectionable effects of highly frequency noise and collect the major features of the data to achieve higher prediction accuracy, according to the decomposition and integration technique,.

(2) *The deep recurrent neural network is successfully adopted in this study to build the combined model.* Using the deep recurrent neural network to predict the wind speed has higher accuracy than the classifier without considering the time dependence of the encoding in the wind speed time series signal.

(3) *A more effective combination weight optimization method, multi-objective optimization algorithm, is adopted to calculate the weights of member model in the combination system.* A group of non-dominated solutions are usually produced by the MOP, which are called Pareto optimal solutions. Every goal corresponds to the point on the Pareto frontier, and can be enhanced only by reducing one or more other goal.

(4) *According to the three hybrid artificial neural networks, the new combination model developed successfully to enhance the prediction effectiveness of wind speed.* The developed combination model utilizes the advantages of the member model and conquers the drawbacks of the traditional individual model with low accuracy and instability effectively.

The process of established combination prediction system architecture is shown as:

I. In the first stage, in consideration of the randomness and non-determinacy of the original wind speed series, a nonparametric, data driven and adaptive time series preprocessing technology is used for wind speed forecast to collect the major characters of the original wind speed sequences.

II. Three individual models including the MLP, LSTM and ARIMA are look upon as prediction methods to forecast the preprocessed wind speed sequences respectively.

III. According to the prediction results of MLP, LSTM and ARIMA, a combined model is established, and the **Fig.1** shows the structure of the combined model.

IV. MOPSO is applied to obtain the optimization of the weights of member models in the combined model. The proposed MOPSO architecture is also displayed in **Fig.1**.

V. Through meta-heuristic algorithm discussing and prediction effectiveness evaluation of the prediction system, the forecast ability of the model and the forecasting availability is further verified.

The other part of this study is designed as: In Section 2, EEMD and MOPSO, as well as the establishment of the combination model on the basis of MLP, LSTM and ARIMA model and the theory of the combination prediction model are introduced. The numerical experiment results of the forecasting system are presented in Section 3. The discussion on the meta-heuristic algorithm and forecasting system are shown in Section 4. Ultimately, the conclusions are presented in Section 5.

II. COMBINED FORECASTING SYSTEM DESIGN

For the sake of acquire precise and steady forecasting results at the same time, the system developed in this article is divided into three stages: data preprocessing stage, single model forecasting stage, combined forecasting with multi-objective optimization stage. MOPSO is adopted to calculate the weighting coefficient of each model to establish a combination model. **Fig.1** displays the flowchart of the established combination system.
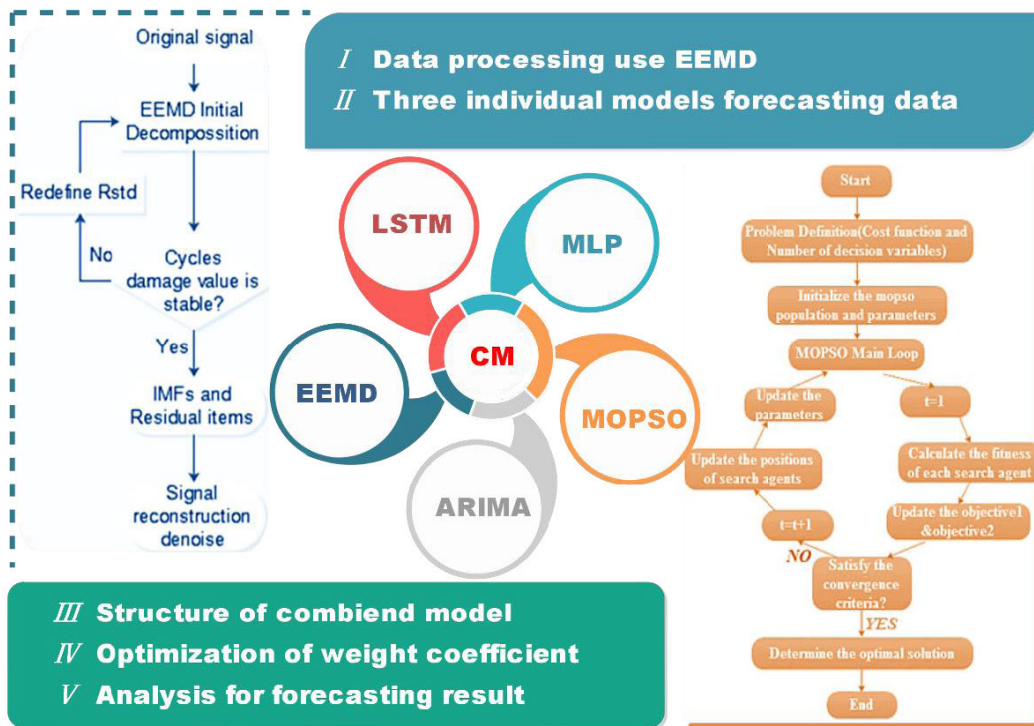
FIGURE 1. The structure of the developed combination model. On the left is the specific step of EEMD preprocessing, and on the right is the process of MOPSO optimizing weight. The middle part shows the model used in the model.

A. ENSEMBLE EMPIRICAL MODE DECOMPOSITION (EEMD)

Firstly, the empirical mode decomposition (EMD) put forward by Huang *et al.* is reviewed. Then decompose the signals, in order to handle the nonlinear and non-stationary data. The EMD can disintegrate the complex initial signal into intrinsic mode functions (IMF) and residuals. In the EMD method [31]:

Definition 1: The data $X_{(t)}$ is disintegrated into IMFs, c_j where r_n is the residue of data $X_{(t)}$, and $X_{(t)}$ is defined as follows [32]:

$$X_{(t)} = \sum_{j=1}^{n} c_j + r_n \tag{1}$$

As a result, a series of IMFs c_j and a residual r_n are obtained. And the IMFs $c_1, c_2, \cdots, c_n$ contain the bands of different frequency.

Definition 2: The EEMD produces a set of datasets which add different white noise $\omega_{(t)}^i$ to the initial data firstly. In this case, the *i*-th observation $X_{(t)}^i$ will be set as

$$X_{(t)}^i = X_{(t)} + \omega_{(t)}^i \tag{2}$$

Next, these new datasets are conducted using EMD. Last, the set average of IMFs obtained by different decomposition is calculated. For a time series $X_{(t)}$, the steps of EEMD are [33]:

A new sequence $Y_{(t)}$ is generated by adding white noise to the original sequence $X_{(t)}$. Determine all the local maximum

and minimum of the time series $Y_{(t)}$. The upper envelopes $Y_{(t)}$ and lower envelopes $e_l(t)$ of $Y_{(t)}$ are generated.

Calculate the average $m_{(t)}$ from upper and lower envelope.

$$m_{(t)} = \frac{e_u(t) + e_l(t)}{2} \tag{3}$$

Pick up the different value between the data $Y_{(t)}$ and $m_{(t)}$ as the first component $h_{(t)}$

$$Y_{(t)} - m_{(t)} = h_{(t)} \tag{4}$$

The residue r_i is regarded as a new series and the above steps are repeated to obtain the whole r_j and a residue c_n. By summarizing all IMFs and residues:

$$Y_{(t)} = \sum_{j=1}^{n} c_j + r_n \tag{5}$$

B. MULTI-LAYER PERCEPTRON (MLP)

MLP is a feedforward method mapping the input set to output set, and its weight is an adjustable model parameter [34].

MLP is composed of multi-layer sigmoid processed neurons or elements, which reach interaction through weighted connecting. Considering the structure shown in **Fig. 2**, the neuron outputs in each layer except the input layer are shown below:

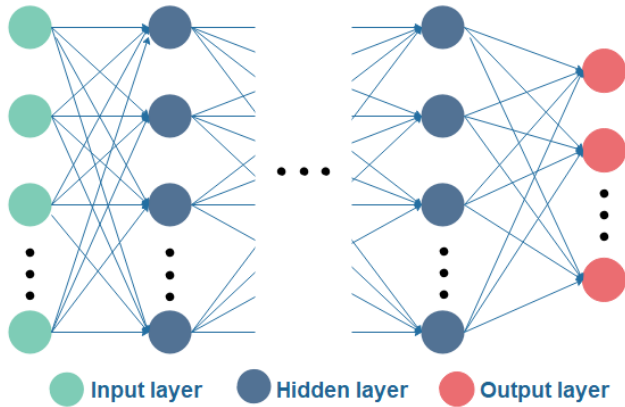$$y_j^{h+1} = \frac{1}{1 + e^{-\sum_i y_i^h \omega_{ji}^h}} \tag{6}$$

FIGURE 2. The construction of MLP.

Among them, $\boldsymbol{\omega}_{ji}^{h}$ is connecting weight between the *i*-th neuron in the *h*-th layer and the *j*-th neuron of the $(h+1)$-th layer, and $\boldsymbol{y}_{i}^{h}$ is the condition of the *i*-th neuron in the previous *h*-th layer. For input layer nodes, it can be obtained that $y_{j}^{0} = x_{j}^{0}$, and this is the *j*-th component of the input vector.

For a given network weight vector *w*, the minimum mean square error in the output vector is decided as

$$E(\omega) = \frac{1}{2}\sum_{j,c}(y_{j,c}^{H}(w) - d_{j,c})^{2} \tag{7}$$

where $\boldsymbol{d}_{j,c}$ is the expected condition provided by the *teacher* and $\boldsymbol{y}_{j,c}^{H}(w)$ is the condition received by the output node *j* in the *H* layer of the input-output situation *c*. One way to minimize *E* is to use gradient descent method to update each weight repeatedly starting from any group of weights.

$$\Delta\omega_{ji}^{h}(t) = -\varepsilon\frac{\partial E}{\partial\omega_{ji}} + \alpha\Delta\omega_{ji}^{h}(t-1) \tag{8}$$

Among them, ε is the positive constant controls the descent and $0 \le \alpha \le 1$ is the momentum coefficient. Besides, *t* is the current iteration numbers in procedure. The error *E* in Eq.(8) can be reduced to minimum after the multiple times of scan of the training set.

The main purpose of the training is to simulate the procedure of data generation [35]. Excellent generalized ability can make well prediction for the experiment data [36]. For the sake of assessing the performance of the model, the available data must be separated into three subsets: training set, testing set and validation set [37], [38]. The train set can be applied to calculate network weight value and bias values. In the train stage, testing set is applied to ensure the generalized ability of the training model, while validation set is applied to test the generalized ability of the training model. The error function is applied to assess the prediction ability of the model. By measuring the range from the network prediction to the target, the range from the target to the prediction is provided. **Fig.2** displays the construction of MLP.
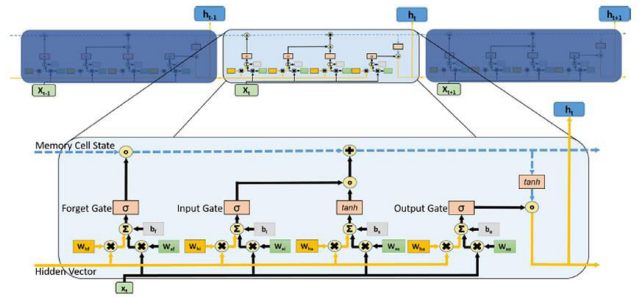

FIGURE 3. The structure of LSTM.

C. LONG SHORT-TERM MEMORY (LSTM)

LSTM network is a special recycle neural network (RNN) [39], and its inner structure is shown in **Fig.3** LSTM provides precise control over the algorithms embedded in memory and dislodged from the hidden layer memory. This is obtained by combining three gates, and they dominate the inflow and outflow of memory units: input, forgetting and output gate [43].

Definition 2: The condition of a memory cell at time *t* are described and the connection from time $t-1$ to time $t+1$ are repeated. **Fig.3** shows that two vectors remain unchanged in time $t-1$: the hidden vector $\boldsymbol{h}_{t-1}$ and the storage unit state $\boldsymbol{s}_{t-1}$. Eq. (2) shows that the forgetting gate $\boldsymbol{f}_{t}$ decides what can be removed from the memory unit state. That is to say, it compels the storage unit to forget the unimportant stuffs based on error back propagation:

$$\boldsymbol{f}_{t} = \sigma\left(\boldsymbol{w}_{xf}\boldsymbol{x}_{t} + \boldsymbol{w}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{f}\right) \tag{9}$$

Among them, the weight matrix $\boldsymbol{w}_{xf}$ is from input *x* to forgetting gate $\boldsymbol{f}_{t}$, the weight matrix $\boldsymbol{w}_{hf}$ is from the prior hidden vector $\boldsymbol{h}_{t-1}$ to forgetting gate $\boldsymbol{f}_{t}$, $\boldsymbol{x}_{t}$ represents the input at *t* time, and $\boldsymbol{b}_{f}$ is forgetting gate deviation. The $\boldsymbol{i}_{t}$ is the input gate which decides the amount of every element of the candidate renovate vector to the relevant storage unit element in time *t* according to the recursive connecting between the hidden vector $\boldsymbol{h}_{t-1}$ and the successive input $\boldsymbol{x}_{t}$ at time *t*.

Definition 3: The gate outputs from the complete connection tanh layer and scales the candidate update vector:

$$\boldsymbol{i}_{t} = \sigma(\boldsymbol{w}_{xi}\boldsymbol{x}_{t} + \boldsymbol{w}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{i}) \tag{10}$$

$$\hat{\boldsymbol{S}}_{t} = \tanh(\boldsymbol{w}_{xs}\boldsymbol{x}_{t} + \boldsymbol{w}_{hs}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{s}) \tag{11}$$

The delicate balance needed to maintain the state of memory cells in long series is achieved by discarding old message and merging new message. And provide new message by joining a part of each value given by the input gate product of in the candidate renovate:

$$\boldsymbol{s}_{t} = \boldsymbol{i}_{t} * \hat{\boldsymbol{s}}_{t} + \boldsymbol{f}_{t} * \boldsymbol{s}_{t-1} \tag{12}$$

Finally, for the designative time context, the output gate determines the content of the output from the storage unit state to the hidden vector to minimize errors.

Definition 4: The output gate o_t and hidden vector h_t can be calculated as:

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \tag{13}$$

$$h_t = o_t * \tanh(s_t) \tag{14}$$

Every gate uses the S-shaped function σ in the direction of the element to proportional change each gate vector element to the value in the range of [0, 1]. The gate function is realized by selecting the value vector in the range of [0, 1], and multiple with another vector, so as to specify the value that the second vector passing the gating account for the whole vector; otherwise, determine which parts are blocked [40]–[42].

D. AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

ARIMA model, proposed by Box and Jenkins [44], is an important prediction model.

Definition 5: ARIMA model is defined as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_q \varepsilon_{t-q} \tag{15}$$

Among them, $y_i(i = 1, 2, \ldots, t)$ is the actual value, ϕ_i and θ_i are the coefficients, $\varepsilon_i(i = 1, 2, \ldots, t)$ is the random error at time t, and p, q are interrelated numbers, usually called autoregressive polynomials and moving average polynomials [45]. This model considers the objective sequence as a random sequence, and applies the certain model to approach the sequence. The model is divided into three stages: model identifying, parameter estimating and diagnosis test. Then the model is applied to predict the sequence.

E. MULTI-OBJECTIVE PARTICLE SWARM (MOPSO)

Multi-objective optimization algorithm (MOPSO) has been widely used in a series of optimization problems [46], [47], including inventory control problems [48], [49].

Step 1. n individuals are randomly opted from the population;

Step 2. Non-dominated position of each individual is obtained;

Step 3. The crowd distance of rank equaled solution is calculated;

Step 4. Finally, the solution with the smallest rank is selected. In addition, the individual with the largest crowd distance should be chosen in the case that more than one individual has the lowest level.

Similar to NSGA-II, the step1, step2 and step3 of the algorithm MOPSO are initialization, fast non-dominated sequencing and congestion distance. The fourth step is to update the speed and position of particles (chromosomes) through formula (13) and formula (14):

$$v^i_{k+1,d} = w * v^i_{k,d} + c_1 * r_1 * \left(pbest^i_{k,d} - x^i_{k,d}\right)$$
$$+ c_2 * r_2 * \left(gbest^i_{k,d} - x^i_{k,d}\right) \tag{16}$$

$$x^i_{k+1,d} = x^i_{k,d} + v^i_{k+1,d} \tag{17}$$

where w is inertia weight, and coefficients c_1 and c_2 are the required acceleration constants of **pbest** and **gbest**. r_1 and r_2 are random uniform numbers in [0, 1]. By studying the parameters of inertia weight, a linear decreasing inertia weight is introduced into the initial particle swarm optimization algorithm, which largely enhances the performance of the algorithm [50]. Moreover, the linear distribution of inertia weight is shown below [51]:

$$w = w_{max} - \frac{w_{max} - w_{win}}{NOG} iteration \tag{18}$$

where NOG is the maximal iteration numbers and *iteration* is the current iteration number s. Eq.(15) gives the updating method of inertia weight, which considered that w_{max} and w_{min} are primary weight and final weight. In this study, we use the parameters $w_{max} = 0.9$ and $w_{min} = 0.4$, which Naka *et al.* [50] as well as Kennedy *et al.* [51] have been investigated in their work. Specially, the arranges and crowd distance of new chromosomes are determined firstly. The pseudo code of MOPSO is described in Appendix.

F. COMBINED FORECASTING THEORY

The combination prediction model is a kind of prediction approach that selects the optimal weight to combine each of the prediction models according to the forecasting results of different forecasting models. The main purpose of this method is to diminish the disadvantage that single prediction models may meet and enhance the prediction ability. The combined of single model could take full advantage of wind speed message due to the different prediction models own different forecasting abilities. Each prediction model possesses its special features, reflecting different aspects of the prediction object. This kind of feature does not represent the whole feature of the prediction object, but cannot be ignored. Better prediction performance can be obtained by properly combining different individual artificial networks. The framework of the composite model is as follows:

Definition 6: The conventional prediction combined theory tries to get the optimum weights of the combination model on the basis of minimizing SSE:

$$\min J = L^T E L = \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} l_i l_j e_{it} e_{jt} \begin{cases} R^T L = 1 \\ L \geq 0 \end{cases} \tag{19}$$

where $L = (l_1, l_2, \cdots, l_m)^T$ is the weight vector, $R = (1, 1, \ldots, 1)^T$ is a column vector whose entire elements are 1. $E_{ij} = e_i^T e_j$, for $e_i = (e_{i1}, e_{i2}, \cdots e_{iN})$ and $E = (E_{ij})_{m \times m}$ named the error matrix.

The weight coefficients has no limit in the interval [0, 1] in Eq. (19). The final results indicates that when the weight vector is obtained the value in the interval [−2, 2], the combined model can get ideal results. This work supplies a weight determination approach that is evaluated by numerical simulation rather than theories testify.

G. COMBINATION SYSTEM ESTABLISHMENT

According to the mythologies discussed above, the established system is mainly established by three modules, including data preprocessing module, individual forecasting module and weight optimization module.

One: Data preprocessing module.

The SSA preprocessing technique is selected in the established model to obtain a reconstructed sequence by refining and identifying the period and vibration parts of the original signal. Through this method, a time series with less noise signal and random volatility can receive to apply in the following forecasting steps.

Two: Individual forecasting module.

Three individual forecasting models—MLP, LSTM and ARIMA—are selected to carry on wind speed forecasting, respectively. And three forecasting results are obtained by this operation. There are two neural networks for nonlinear prediction and one linear prediction, and the models of both the linear and nonlinear prediction is excellent in predicting wind speed.

Three: Weight optimization module.

For the sake of obtaining the weight coefficients of each model, a kind of decision-making weight method on the basic of MOPSO algorithm and nonnegative constraint theory is proposed to gain the optimal consequences. Say concretely, the last three days of the training set are retained to obtain the weights of selected models. It is remarkable that when the algorithm reaches the maximum iteration number or the minimum fitness function value, it will stop. According to the weight coefficients of each model, the prediction consequences of each model are combined for getting the final wind speed prediction results.

III. NUMERICAL EXPERIMENTATION

In this section, the study area of three datasets and the data structure are introduced firstly. Then the model parameter selection and evaluation metrics are displayed. The last is the experimental results for three datasets.

A. STUDY SITE INTRODUCTION AND DATA ANALYSIS

As exhibited in **Fig.4**, Penglai is situated in the northeast of Shandong, China. Due to its geographical feature of along the coast and monsoon climate, it has rich and huge wind energy resources. It is reported that, the power generation capacity of Penglai electric grid has more than 200 million kilowatts in June 2017, ranking in the forefront of Shandong electric system. In this paper, number seven wind turbine of wind farm of Penglai is selected as the experiment object. The data point is located in the coastal hilly area, with a measurement height of 100m. The average sampling period is ten minute, and the rate of scanning is 144 times every day.

For the sake of assess the practicality, effectiveness and universality of the established new combination system, three datasets are selected from three stations in Penglai for the numerical simulation. Each dataset contains 3600 data points

of ten-minute wind speed data, and these observations are separated into training set and testing set. Currently, there is no clear theory explain the method to choose the quantity of train samples and test samples. Too few samples cannot train the neural network well, while too many samples easily make the network over fit. In fact, two thirds of the data in the dataset is usually used as training data, and one third of the data is retained as testing data [52]. In this article, the ratio between training and testing is 5:1. In other words, training samples selects the initial 3000 data points and the testing samples selects the rest of 600 data points. The statistical measurement results of the datasets of wind speed data sampling, including minimum, maximum, mean value and standard deviation, are shown in **Fig.4**.

B. PARAMETER OPTION OF EXPERIMENT AND FORECASTING

The three datasets A, B and C of wind speed data are applied to evaluate the effectiveness of the new combination system. Three prediction models, EEMD-LSTM, EEMD-MLP and EEMD-ARIMA, are used for comparison, which are the member model of the combined model. National energy administration issued the energy industry standard NB/T31046-2011 and formulated the wind energy measurement rules in 2011. The wind speed time provided by wind farm must be more than 10 minutes, the maximum error predicted by wind energy curve on the next day should be no more than 0.25, and RMSE (root mean square error) must be no more than 0.2.

1) OPERATING ENVIRONMENT

The operating environment of the ARIMA method and MOPSO algorithm is: 3.20 GHz CPU, 8.00 GB RAM, Windows 7, and MATLAB R2016a. The operating environment of the MLP and LSTM is: 3.08GHz CPU, 4GB RAM, Windows 10, Anaconda 3, Tensorflow 1.2 and Keras 2.0. Moreover, the parameter settings in the model are specified below.

2) PARAMETER SELECTION

Through many trials, the model parameters are seted as follows:

(a) For LSTM, the network is established by applying Keras. The dimensions of each layer are: input layer is 4, the first hidden layer is 50, the second hidden layer is 100, and the output layer is 1. Setting some parameters as Activation='relu', loss='mean squared error', optimizer='nadam', epochs=400, batch size=16, verbose=2.

(b) For MLP, the network is established by applying Keras. The dimensions of each layer are: input layer is 4, the first hidden layer is 12, the second hidden layer is 8, and the output layer is 1. Setting some parameters as 'activation ='relu', loss='mean squared error', optimizer='nadam', epochs=400, batch size=16, verbose=2.
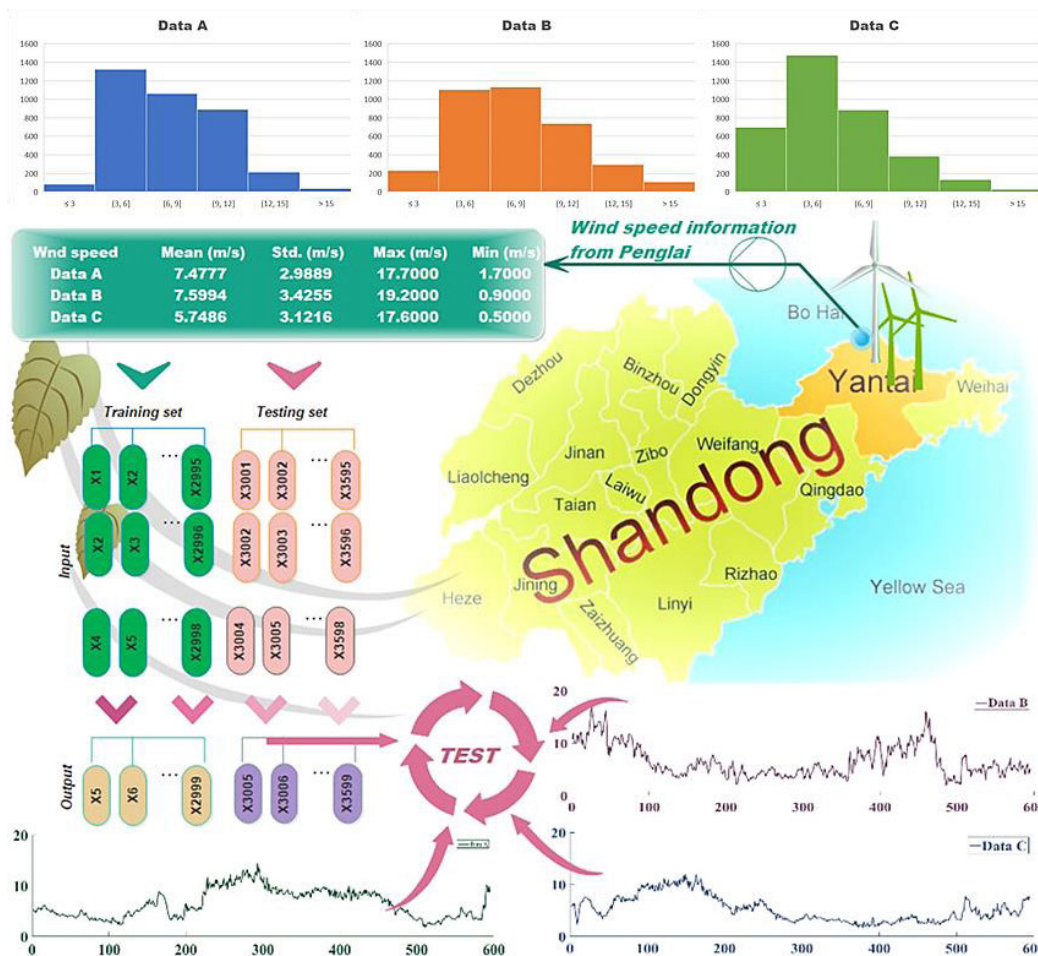
FIGURE 4. The selected sites of the Penglai and the data structure of selected datasets. The upper part of the figure shows the data distribution and statistical characteristics of the three datasets. On the left is the input-output structure of the wind speed prediction model, which is called rolling forecasting method. Below is the original time series of wind speed.

(c) For MOPSO, the fitness function is MAPE and RMSE. Number of decision variables is 3, the lower bound of variables is −2, the upper bound of variables is 2, the maximum iterations are 1000 and the population size is 200.

C. EVALUATION METRICS

For the sake of effectively testify the prediction ability of the combination prediction model with changing weights, the model is usually evaluated based on evaluation indicators. In this study, eight performance evaluation criteria are adopted to estimate the prediction effectiveness of the combined model, which contain AE (average error) [53], MAE (mean absolute error) [54], MAPE (mean absolute percentage error) [55], MSE (mean square error) [56] and RMSE (root mean square error) [57] to evaluate the average error between the forecasted and actual value, and the I_{MAE}, I_{MSE}, I_{MAPE} represent the percentage error of the benchmark wind speed forecasting models in this study compared to the developed combination prediction model. If the value of these indexes is the smaller, the developed model is considered as owning the

excellent prediction performance. See **Table 1** for the details of the mentioned metrics.

D. EXPERIMENT I: THE PROPOSED MODEL COMPARED WITH INDIVIDUAL MODELS

In the trial, using the dataset A, B and C (10-min wind speed data), while each dataset selects 3000 samples as the training set, the combination model is trained to predict 600 samples in the future rested. The time series of the predicted value is shown in **Fig.5** and **Fig.6**.

1) EXPERIMENTAL RESULTS DESCRIPTION

The experiment is set as a comparison the forecasting performance between proposed model and three single models, including EEMD-MLP, EEMD-LSTM and EEMD-ARIMA. The results are shown in the **Table 2** and **Table 3**, as well as the **Fig.5** and **Fig.6**. From the table and figure obtained, the information of experiment results can be shown as:

(1) **Table 2** describes the forecasting results statistical properties of the four models. This statistical description

TABLE 1. Definition of the performance metrics.

Metric	Definition	Equation
AE	The average forecast error of *n* times forecast results	$\mathbf{AE} = \dfrac{1}{N}\sum\limits_{n=1}^{N}\left(y_n - y_n\right)$
MAE	The average absolute forecast error of *n* times forecast results	$\mathbf{MAE} = \dfrac{1}{N}\sum\limits_{n=1}^{N}\left|y_n - y_n\right|$
MSE	The average of the prediction error squares	$\mathbf{MSE} = \dfrac{1}{N}\sum\limits_{n=1}^{N}\left(y_n - y_n\right)^2$
RMSE	The root average of the prediction error squares	$\mathbf{RMSE} = \sqrt{\dfrac{1}{N}\sum\limits_{n=1}^{N}\left(y_n - y_n\right)^2}$
MAPE	The average of absolute error	$\mathbf{MAPE} = \dfrac{1}{N}\sum\limits_{n=1}^{N}\dfrac{\left|y_n - y_n\right|}{y_n}\times 100\%$
$\mathbf{I_{MAE}}$	The percentage error of MAE	$\mathbf{I_{MAE}} = \left|\dfrac{MAE_a - MAE_b}{MAE_a}\right|\times 100\%$
$\mathbf{I_{MSE}}$	The percentage error of MSE	$\mathbf{I_{MSE}} = \left|\dfrac{MSE_a - MSE_b}{MSE_a}\right|\times 100\%$
$\mathbf{I_{MAPE}}$	The percentage error of MSE	$\mathbf{I_{MAPE}} = \left|\dfrac{MAPE_a - MAPE_b}{MAPE_a}\right|\times 100\%$

Among the formula, y_i and $\hat{y}_i$ represent the true value and predicted value of wind speed respectively. And *N* represents the testing set number.
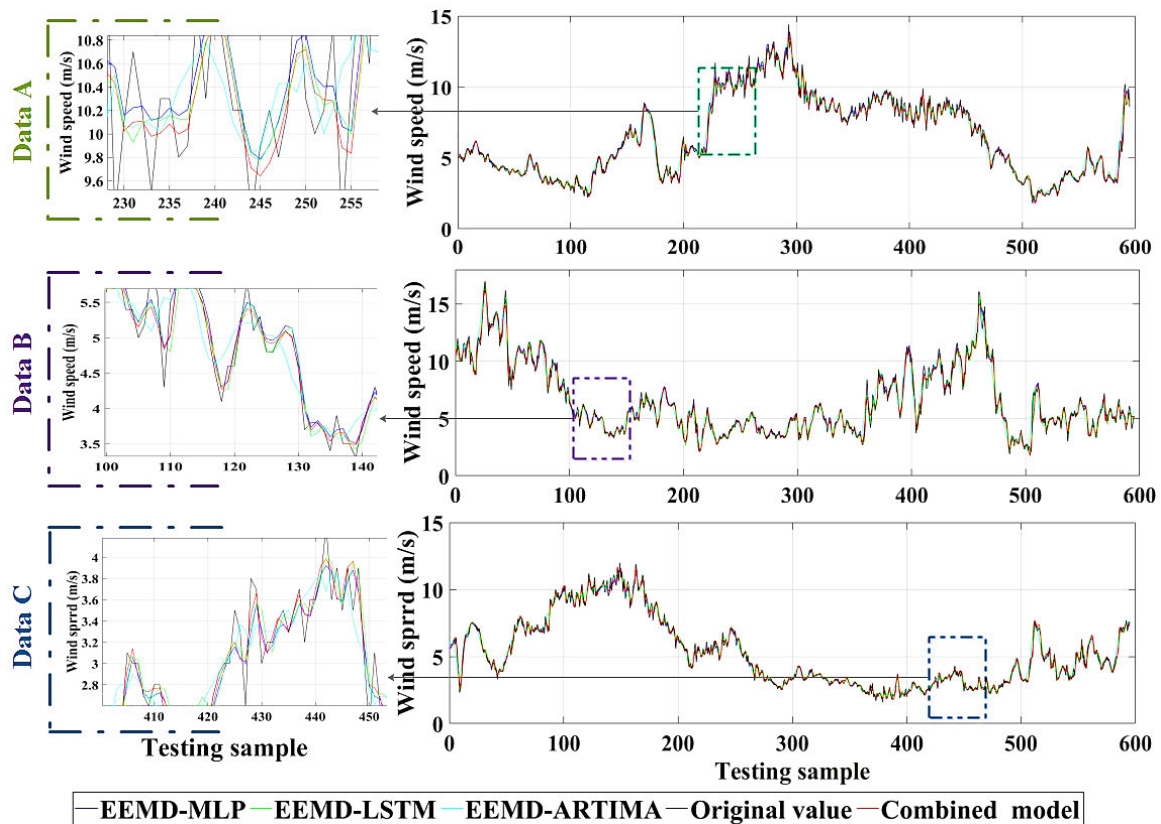

FIGURE 5. The results of every forecasting model on three datasets. On the right is the entire wind speed time series, including original data and forecast data. On the left is the partial enlarged diagram of time series, which can clearly and intuitively see the fitting situation of each model for wind speed data.

can show the prediction state of each model more intuitively. **Table 3** shows the prediction error of EEMD-MLP, EEMD-LSTM, EEMD-ARIMA and combined model for

dataset *A, B* and *C*, which can reflect the prediction ability of the model. From the prediction results, the developed model is superior to all others in all evaluation indexes. For example,
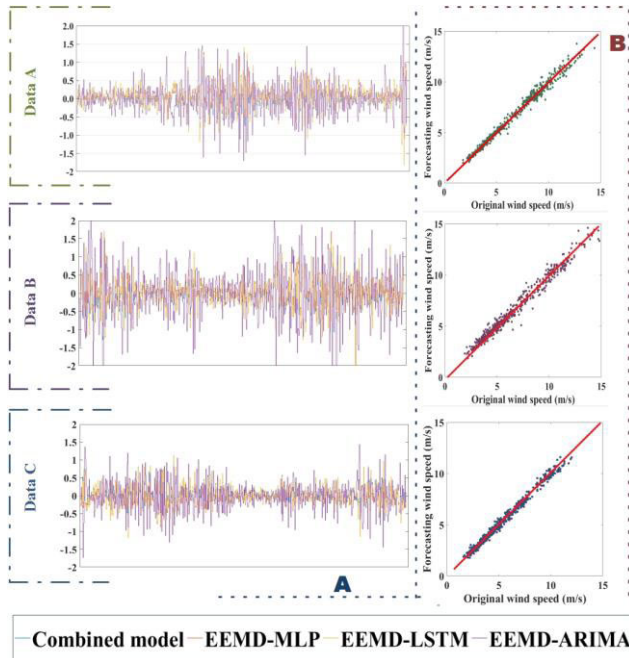
FIGURE 6. Comparison of the forecasting error for three datasets. On the left is the display of the prediction error of each model on the time scale. The figure on the right shows how discrete the forecast data is to the real data.

TABLE 2. Statistical properties for each method forecasting result.

Data	Model	Mean (m/s)	Std. (m/s)	Max (m/s)	Min (m/s)
A	EEMD-LSTM	6.4984	2.7436	13.4465	2.0753
	EEMD-MLP	6.5360	2.7889	13.8692	2.0094
	EEMD-ARIMA	6.4723	2.7506	13.1578	2.0089
	Combined model	**6.4510**	**2.7588**	**13.7644**	**1.9550**
B	EEMD-LSTM	6.5333	3.0006	16.5282	1.8135
	EEMD-MLP	6.5934	2.9894	16.5431	1.7485
	EEMD-ARIMA	6.5493	2.9077	15.4549	2.1557
	Combined model	**6.5071**	**2.9386**	**16.2185**	**1.8469**
C	EEMD-LSTM	5.2965	2.5187	11.5761	1.7809
	EEMD-MLP	5.2136	2.4836	11.4204	1.7923
	EEMD-ARIMA	5.2663	2.4940	11.3624	1.9046
	Combined model	**5.2645**	**2.5165**	**11.6111**	**1.7786**

Specially, the mean value and the standard deviation (Std.) of the forecasting data reflect the basic information of the data, which is calculated as

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i \,,\ \text{Std.} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{\hat{y}} - \hat{y}_i)^2}$$

for *dataset A*, the MAE, MSE, RMSE and MAPE (%) of the combined model are 0.2259, 0.0994, 0.3153 and 3.5271%, respectively, while the errors of the other models are higher than them. For three comparison models, EEMD-MLP is the best model, and its error index MAE, MSE and RMSE are all 0.01 higher than the combined model, and the MAPE is also 0.212 higher. Other datasets can reach the same conclusion. The worst performance is EEMD-ARIMA. Therefore, the combined model is considered to improve the prediction accuracy.

TABLE 3. Wind speed forecasting errors of different models.

Data	Model	AE	MAE	MSE	RMSE	MAPE (%)
A	EEMD-LSTM	0.0336	0.2599	0.1305	0.3612	4.2318
	EEMD-MLP	0.0711	0.2347	0.1045	0.3232	3.7391
	EEMD-ARIMA	0.0074	0.3386	0.2182	0.4671	5.5545
	Combined model	**-0.0139**	**0.2259**	**0.0994**	**0.3153**	**3.5271**
B	EEMD-LSTM	-0.0171	0.3422	0.2118	0.4602	5.7187
	EEMD-MLP	0.0430	0.3046	0.1737	0.4167	5.1124
	EEMD-ARIMA	-0.0012	0.5427	0.5407	0.7353	9.2347
	Combined model	**-0.0433**	**0.3014**	**0.1688**	**0.4109**	**5.0176**
C	EEMD-LSTM	0.0297	0.2318	0.0950	0.3082	4.8838
	EEMD-MLP	-0.0531	0.2046	0.0717	0.2678	4.2168
	EEMD-ARIMA	-0.0004	0.3068	0.1712	0.4138	6.4343
	Combined model	**-0.0023**	**0.1920**	**0.0621**	**0.2493**	**4.0142**

EEMD-LSTM model, EEMD-MLP model and EEMD-ARIMA model are prediction models using the wind speed data processed by EEMD. These models are the member models in the combination mechanism.

(2) The **Fig.5** shows the results of the advance prediction from EEMD-MLP, EEMD-LSTM, EEMD-ARIMA and the combined model, respectively. Obviously, the error of the combined model to the observation value is smaller compared with the single model. The detail diagram on the left is the partial amplification of the EEMD-MLP, EEMD-LSTM, EEMD-ARIMA and the combined model. As shown in the figure that the red line (the combined model forecasting result) is closet on the black line (the original value forecasting result). It is worth noting that when the observed wind speed value encounters difficulties in the slope, the high accuracy prediction result is carried out. In this case, the prediction effect is deteriorated.

(3) **Fig.6** displays the actual plots and the predictions wind speed plots of the combination model and other models of *dataset A, B* and *C*. The conclusion is that the combination model obtains accurate predictions in most cases. The sequence of prediction errors is shown in part A, and these errors always fluctuate around 0. In addition, it can be seen from part B that the combined model results in a strong relation between the observed wind speed and the predict wind speed. In the same time, the prediction result for *dataset B* is divergence than the other two datasets, and it can also be confirmed it in **Table 3**.

Remark: According to the results, the developed model has the best performance and keeps the original curve form better for prediction. Moreover, the proposed model is the best according to the five error indexes. In general, the comprehensive model utilizes the advantages of the other three modes, reduces the redundant information and improves the local optimal value. The prediction accuracy is improved obviously, and the prediction error fluctuation is small.

2) EXPERIMENTAL RESULTS ANALYSIS

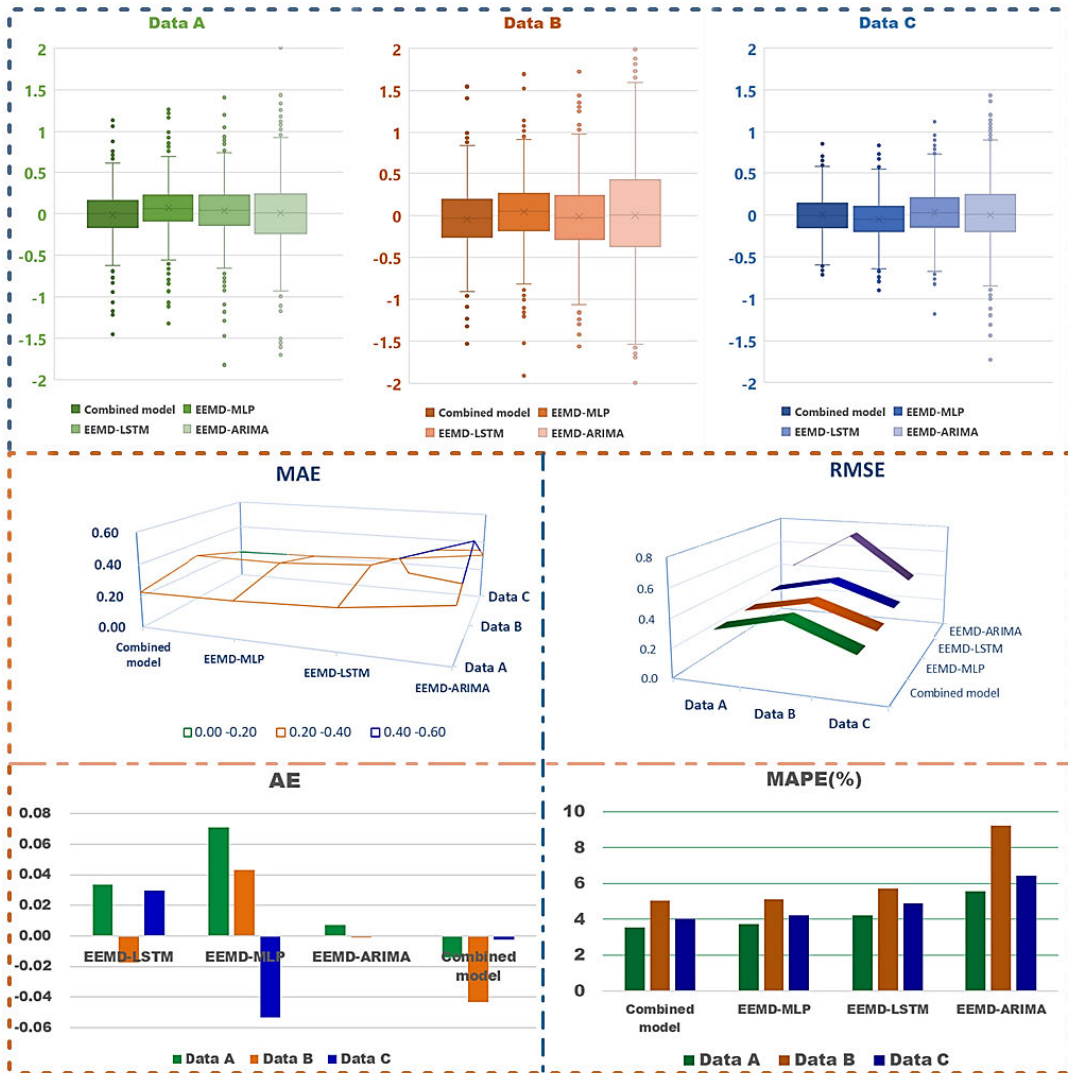Through comparing the combination model with three benchmark models, the evalution metrics are calculated and

FIGURE 7. Comparison of the prediction error for three datasets.

TABLE 4. The improvement rate of MAE, MSE and MAPE by the combination model.

Data	Model	I_{MAE} (%)	I_{MSE} (%)	I_{MAPE} (%)
	EEMD-MLP	3.7495	4.8804	5.6698
A	EEMD-LSTM	13.0820	23.8314	16.6525
	EEMD-ARIMA	33.2841	54.4455	36.5001
	EEMD-MLP	1.0506	2.8210	1.8543
B	EEMD-LSTM	11.9229	20.3022	12.2598
	EEMD-ARIMA	44.4629	68.7812	45.6658
	EEMD-MLP	6.1584	13.3891	4.8046
C	EEMD-LSTM	17.1700	34.6316	17.8058
	EEMD-ARIMA	37.4185	63.7266	37.6125

shown in **Table 4** and **Fig.7**. Through the analysis, the conclusion can be obtained as follows:

(1) From **Table 4**, it can be clearly seen that based on the three standards of I_{MAE} (%), I_{MSE} (%) and I_{MAPE} (%),

the proposed combination model is more accuracy than other models discussed in this work. The results present that the model could get more messages from the wind speed and get the prediction results more precisely. More concretely, taking the prediction results of *dataset C* as an example, compared with the three benchmark models, the improvement of the combined model improved MAPE by 4.8046%, 17.8058%, 37.6125%. Also, MAE and RMSE increased by 6.1584%, 17.1700%, 37.4185%, and 13.3891%, 34.6316%, 63.7266%, respectively. Comparing with other models, the prediction ability of combined model is largely improved, which shows that the combined model does make use of the other three models and improve the prediction accuracy effectively.

(2) **Fig.7** includes five parts, where the top three box-plots show the errors of EEMD-MLP, EEMD-LSTM, EEMD-ARIMA and combined model in the three datasets. The results show that the error of the proposed model is the minimum, which means the predicting ability of the

TABLE 5. Wind speed forecasting errors of combined model using different preprocessing method.

Data	Model	AE	MAE	MSE	RMSE	MAPE (%)
	EMD-Combined	0.0253	0.3267	0.1871	0.4326	4.4647
	SSA-Combined	-0.0791	0.4470	0.4122	0.6420	6.1316
A	WDD-Combined	0.0638	0.4585	0.3969	0.6300	6.2443
	EEMD-Combined	**-0.0139**	**0.2259**	**0.0994**	**0.3153**	**3.5271**
	EMD-Combined	-0.0551	0.4518	0.3941	0.6278	6.1931
	SSA-Combined	0.0805	0.4457	0.3206	0.5662	7.3411
B	WDD-Combined	-0.0811	0.5409	0.5017	0.7083	9.6467
	EEMD-Combined	**-0.0433**	**0.3014**	**0.1688**	**0.4109**	**5.0176**
	EMD-Combined	0.0381	0.3375	0.2023	0.4498	5.6482
	SSA-Combined	-0.0790	0.4678	0.3752	0.6125	7.3667
C	WDD-Combined	0.0823	0.5450	0.5461	0.7390	8.5275
	EEMD-Combined	**-0.0023**	**0.1920**	**0.0621**	**0.2493**	**4.0142**

EMD-combined model, **SSA-combined** model and **WDD-combined** model are compared models using the wind speed data processed by different preprocessing methods. And **EEMD-combined** model is the proposed model in this study. These models are the member models in the combination mechanism. The comparison of these models is to highlight the advantages of EEMD preprocessing method.

TABLE 6. The improvement rate of the combined models using different preprocessing methods.

Data	Model	I_{MAE} (%)	I_{MSE} (%)	I_{MAPE} (%)
	EMD-Combined	30.8540%	46.8733%	21.0003%
A	SSA-Combined	49.4631%	75.8855%	42.4767%
	WDD-Combined	50.7306%	74.9559%	43.5149%
	EMD-Combined	33.2891%	57.1682%	18.9808%
B	SSA-Combined	32.3760%	47.3487%	31.6506%
	WDD-Combined	44.2781%	66.3544%	47.9864%
	EMD-Combined	43.1111%	69.3030%	28.9296%
C	SSA-Combined	58.9568%	83.4488%	45.5088%
	WDD-Combined	64.7706%	88.6285%	52.9264%

developed model is excellent compared with the other three models. Each point errors of the combination model are not only closely related to the zero axis, but also have a very small dispersion level. Part MAE presents that from left to right the surface chart is raising along the X axis, which means that for the same dataset, the combined method produces the smallest MAE, and the EEMD-ARIMA yields the largest MAE value. The other two parts, RMSE and MAPE, can also prove that the combination model is better than the individual model in each aspect.

Remark: Compared with the other three models, the conclusion can be drawn that the combination model owns the highest precision of prediction results. The comparative experiment of EEMD-MLP, EEMD-LSTM and EEMD-ARIMA shows that the prediction ability of EEMD-MLP is better than the other three methods, and the prediction performance of EEMD-ARIMA is the worst. No matter dataset A, B and C, EEMD-MLP always perform best than others.

E. EXPERIMENT II: THE EEMD COMPARED WITH OTHER PREPROCESSING METHODS

This experiment designs the comparison based on different preprocessing methods. The preprocessing methods include EMD, SSA and WDD, as well as EEMD applied in the proposed model. These models are built to emphasize the importance of selecting an excellent data processing technology. Compared with the above methods, the superiority of the model proposed in this study is further illustrated. The prediction error of the experiment is shown in **Table 5** and **Table 6**. The details of the experiment are as follows:

(1) For **Table 5**, the proposed model has the best MAE, MSE, RMSE and MAPE (%) respectively. Secondly, EEMD-combined model, EMD-combined model, SSA-combined model and WDD-combined model are the

models with high to low prediction accuracy, for the MAPE value is gradually decreasing, which shows that EEMD is the most suitable preprocessing method for this data. In addition, there are still a few SSA and WDD do not follow this level, but EEMD is always the best preprocessing method. The MAPE of EEMD-combined model is about 3%-5%, which is the best in three datasets. The results show that EEMD-combined forecasting model is the best one at present.

(2) It can be clearly seen from **Table 6** that for the three standards of I_{MAE}, I_{MSE} and I_{MAPE}, the effect of EEMD-combined model is significantly improved compared with other preprocessing methods. Specifically, taking *dataset A* as an example, EEMD improved MAPE by 21.0003%, 42.4767% and 43.5149% compared with the other three preprocessing methods. MAE and RMSE increased by around 30%-50% respectively. For the other two datasets, the prediction accuracy of the proposed model is significantly improved, which shows that EEMD-combined method is indeed more suitable for processing wind speed data than other preprocessing methods, and the signal decomposition is more complete, which can effectively increase the prediction ability of the predict model.

Remarks: It can be seen from this experiment that EEMD-combined model has the highest prediction accuracy and the best MAPE value, and significantly improved MAE, MSE and MAPE. In addition, on the basis of different preprocessing methods, it shows that the proposed model is superior to other models, thus verifying the effectiveness of the model.

IV. ANALYSIS AND DISCUSSION

A great amount of statistical and machine learning models can be adopted to forecast time series, including economic growth, currency inflation, and wind speed time series. This section is absorbed in the efficient characteristic of the developed model in terms of meta-heuristic algorithm performance and computational ability of the system.

A. META-HEURISTIC ALGORITHM DISCUSSING

In order to testify the performance multi-objective optimization algorithm, this section will discuss the fitness functions and convergence of the MOPSO algorithm in detail.

TABLE 7. The definition of MOPSO parameters.

Cost Function	Object 1&2
Number of Decision Variables	3
Bound of Variables	[-2,2]
Maximum Number of Iterations	1000
Population Size	200
Repository Size	100
Inertia Weight	0.5
Mutation Rate	0.1

1) FITNESS FUNCTIONS

The accuracy and stability of prediction are two commonly selected assessment standard to assess the performance of prediction model. Therefore, it is not enough to consider only one aspect of prediction results, whether accuracy or stability. In the combination prediction model, the goal of weight coefficient optimization should be both good accuracy and stability.

The framework of bias-variance error [58] is used to evaluate both the prediction accuracy and stability of the prediction model. Among this architecture, the accuracy and stability of the value prediction model are respectively reflected by the value of $\left|Bias(\hat{y})\right|$ and $Variance(\hat{y})$. According to the bias-variance framework, the fitness function contains accuracy and stability of the optimization algorithm is established. The difference between the original and the predict value is considered as the error caused by bias. The variation of prediction results is considered as the error caused by variance.

The $y_i - \hat{y}_i$ is regarded as the difference value between the original value y_i and the predict value $\hat{y}_i$. The predicted expected value is calculated as $E(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i$, the expected value of the observed value is expressed as $E(y) = \frac{1}{N}\sum_{i=1}^{N}y_i$, where N is the data point numbers to be compared. The bias-variance architecture is decomposed below:

$$E(y - \hat{y})^2 = E\left[\hat{y} - E(\hat{y}) + E(\hat{y}) - y\right]^2$$
$$= E\left[\hat{y} - E(\hat{y})\right]^2 + E\left[E(\hat{y}) - y\right]^2$$
$$= Bias(\hat{y})^2 + Variance(\hat{y}) \quad (20)$$

Therefore, the fitness function target of the combined model is the setting that minimizing the accuracy and stability of the prediction, and it is obtained as:

$$\min imam \begin{cases} Objective\ 1 = \left|Bias(\hat{y})\right| \\ Objective\ 2 = \left|Variance(\hat{y})\right| \end{cases} \quad (21)$$

where a smaller $\left|Bias(\hat{y})\right|$ proved that the prediction model has high prediction accuracy. In the same way, a smaller $Variance(\hat{y})$ represents more stable. Specially, the parameter definition of object 1 and 2 is shown in **Table 7**.

2) CONVERGENCE

From the **Fig.8**, we can see that as the iteration numbers increases, the fitness rate drops rapidly. EEMD-MLP and
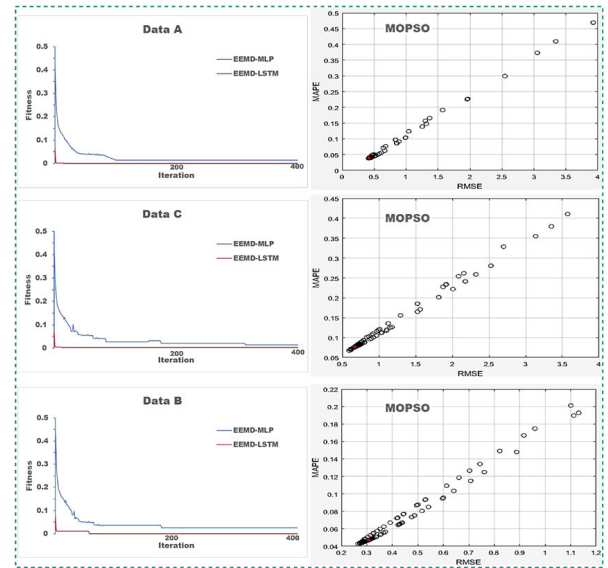

FIGURE 8. Convergence of different models under three sets of data.

TABLE 8. MOPSO operating status.

Data	RMSE	MAPE	CPU time
Data A	0.3982	0.0311	24.2 min
Data B	0.5726	0.0613	26.7 min
Data C	0.2621	0.0408	24.5 min

Specially, the calculation formula of RMSE and MAPE has been given above, and now it is restated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_n - y_n\right)^2}, \quad MAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{\left|y_n - y_n\right|}{y_n}\times100\%$$

EEMD-LSTM reached convergence in 400 iterations, and deep neural network convergence performance was superior to other comparison models. In the initial stage, the fitness of EEMD-MLP is about 0.45, indicating that the performance of the model using random initial parameters is poor. However, as the iteration numbers increases, the fitness rate decreases rapidly, indicating that the neural network finds better parameters in a short time. After 200 times, there was no significant change in fitness, indicating that the model obtained the best parameters. The fitness function value of EEMD-LSTM is very small in the initial state, and it converges to the best area in the first 10 times. It can be seen that the deep neural network is highly efficient in data feature mining during the learning process.

In addition, the global optimization ability of MOPSO in the iterative process can be seen from the right side of **Fig.8**. The final global optimal solution is the point marked in red in the figure by defining the two objective functions as the fitness function from precision and skewness. And use these two aspects to control the final convergence position. During the experiment, the overall optimization process is greater than 20 minutes (see **Table 8**), but for the three groups of dataset, the first 5 minutes locked the convergence interval, and the latter iteration moved almost within the non-inferior
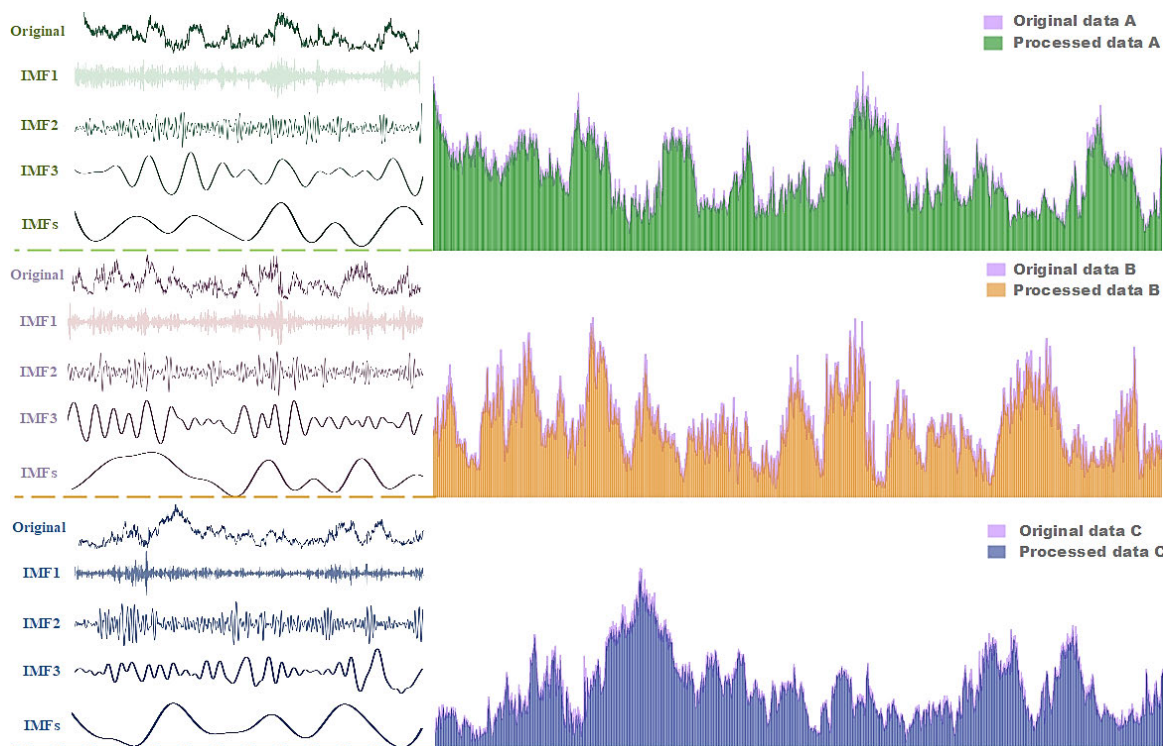
FIGURE 9. The result of data preprocessing. On the left is the sequence of components decomposed by EEMD, and on the right is the time series of original data and preprocessed data.

solution set. The black circle in the figure represents the historical trajectories for each iteration, most of which are concentrated in the convergence set in the lower left corner. From the **Table 8**, we can see that the dataset B has a higher MAPE than 0.05 when using multi-objective optimization, and RMSE is greater than 0.5, which is larger than the values of the other two datasets. This indicates that the nonlinearity of dataset B is stronger and the forecasting trend is difficult to grasp. Inference can also be confirmed from the experimental results. The proposed combination method has a better search for the convergence performance of each hidden neuron optimal threshold in a highly nonlinear prediction than other methods.

B. ISCUSS FORECASTING SYSTEM

This section conducts five profound discussions of the forecasting system we established, including data processing, train-to-verify ratios, Diebold-Mariano Test, forecasting effectiveness and bias-variance test.

1) DATA PROCESSING

EEMD is applied to preprocess initial time series prior to the forecast the wind speed by the models. Compared with the original data, it is obviously that processed data have accurate and stable display because of the removing of random perturbation. The theory of EEMD has been expounded in section III. **Fig.9** is the details of data preprocessing.

2) THE TRAIN-VERIFY RATIO

The train-verify ratio can indicate the extent to which the latest sequence is being used, or its impact on the forecast results. We have configured a number of sequence train-verify ratios to study the influence of train-verify ratios on the results. For the wind speed data of the datasets, the train verification ratio is respectively configured as 1:1, 2:1, 3:1, 4:1 and 5:1. A large ratio means more samples are put into training. On the other hand, a small ratio indicates a small number of samples are involved in training. The experimental results show that the better precision can be obtained by increasing the ratio. It is beacuse appling updated data can enhance the accuracy of training. Nevertheless, that doesn't mean the train-verify ratio can be infinitely expanded in practical application, because in the case of few verification samples, the train verification ratio lacks reliability. Therefore, we suggest to choose a higher ratio when considering the number of samples.

3) DIEBOLD-MARIANO TEST

For the sake of further assessing and discussing the effectiveness of the developed combination prediction model, this section discusses the Diebold-Mariano (DM) test for prediction availability and evaluation of prediction performance [58].

The DM test is used to verify the difference of the prediction effectiveness between the established model and other

TABLE 9. The definition of MOPSO parameters.

Data	Model	Combined model	EEMD-MLP	EEMD-LSTM
	EEMD-MLP	1.6352※※		
A	EEMD-LSTM	6.2949※	4.5527※	
	EEMD-ARIMA	7.1629※	7.0757※	5.1289※
	EEMD-MLP	0.6297※※※		
B	EEMD-LSTM	3.2710※	3.9079※	
	EEMD-ARIMA	8.3663※	7.4773※	8.6176※
	EEMD-MLP	3.8890※		
C	EEMD-LSTM	6.7316※	5.8325※	
	EEMD-ARIMA	9.2991※	8.8990※	6.3036※
	EEMD-MLP	2.0513※		
MEAN	EEMD-LSTM	5.4325※	4.7644※	
	EEMD-ARIMA	8.2761※	7.8174※	6.6834※

TABLE 10. The definition of MOPSO parameters.

Data / Model	Data A		Data B		Data C	
	1-order	2-order	1-order	2-order	1-order	2-order
Combined model	**0.9647**	**0.9350**	**0.9500**	**0.9058**	**0.9596**	**0.9248**
EEMD-MLP	0.9626	0.9313	0.9489	0.9028	0.9578	0.9231
EEMD-LSTM	0.9577	0.9195	0.9428	0.8942	0.9512	0.9094
EEMD-ARIMA	0.9445	0.8934	0.9083	0.8212	0.9357	0.8767

comparative models. Considering the significance level α, zero hypothesis H_0 indicates the predictive effectiveness of the established model and the comparason model are not significantly different. The meaning of H_1 is contrast with H_0. The relevant formulas is shown as:

$$H_0 : E[F(e^1)] = E[F(e^2)] \& H_1 : E[F(e_t^1)] \neq E[F(e_t^2)] \tag{22}$$

In the fomula, F represents the loss function of prediction error. And e_t^1, e_t^2 are the error sequence predicted by selected model. In addition, the statistics of DM test can be defined in the following ways:

$$DM = \frac{\sum_{i=1}^{n}(L(err_i^1) - L(err_i^2))}{\sqrt{S^2/n}} s^2 \tag{23}$$

In which, S^2 is the estimate of the variance of $d_i = L(err_i^1) - L(err_i^2)$. Assuming a certain significance level α, the obtained value DM is in comparison with that of $z_{\alpha/2}$. Once DM statistics exceed the interval $[-z_{\alpha/2}, z_{\alpha/2}]$, H_0 can be rejected. This shows the predictive performance of the establishd model and that of the comparative model are significantly different, which means that H_1 will be accepted.

In this part, we use the DM test to testify the validity of the proposed model. The comparison are set between all of the following models and the proposed model, that is, EEMD-LSTM and EEMD-ARIMA compared with EEMD-MLP, EEMD-ARIMA compared with EEMD-LSTM. On the basis of the fundamental principle of DM test, no significant difference between the two models of the forecasting effectiveness is the main idea of the zero hypothesis, while there exists significant difference in the forecasting effectiveness of the two models is the main idea of the alternative hypothesis. **Table 9** shows the average DM test results for the three datasets.

As can be seen from **Table 9**, in the multi-step prediction, at the 1% significance level, the established combination model is different from others. In addition, at 5% significance level, for the comparison result of models with different preprocessing methods, the minimum value of $|DM|$ is 0.629747, which can makes the zero hypothesis be rejected. Moreover, for some traditional individual models, including

EEMD-MLP, EEMD-LSTM and EEMD-ARIMA, all the values are far greater than $Z_{\alpha/2}(Z_{0.005} = 2.58, Z_{0.025} = 1.96)$. Hence, at the 1% significance level, the difference compared the combined model with the individual model is significant. Hence, the established combination model is remarkably superior to other comparative models.

4) FORECASTING EFFECTIVENESS
This part is mainly used to verify the prediction efficiency of the developed model. The forecast effectiveness of the model can be obtained not only by the average prediction error, but also by the mean variance of prediction accuracy [59]. The next section gives the main idea of prediction validity.

The k-th order prediction efficiency unit is obtained by $m^k = \sum_{i=1}^{n} Q_i A_i^k$, where Q_i is the discrete probability distribution and A_i is the prediction accuracy at time i, and $\sum_{i=1}^{n} Q_i = 1, Q_i > 0$. An exception is that Q_i is defined as $Q_i = \frac{1}{n}, i = 1, 2, \cdots, n$, when the previous information of Q_i can't be known under certain circumstance.

Afterwards the k-th order prediction efficiency is judged by $H(m^1, m^2, \cdots, m^k)$, among which H is a continuous function, and it has a series of unit. The 1-order prediction efficiency is obtained as $H(m^1) = m^1$ in the case of $H(x) = x$ be the continuous function of one variable. Then if $H(x, y) = x(1 - \sqrt{y - x^2})$ represents a continuous function of two variables, $H(m^1, m^2) = m^1 \left(1 - \sqrt{m^2 - (m^1)^2}\right)$ is called the 2-order prediction efficiency.

In this section, the prediction effectiveness is applied to assess and compare the prediction accuracy of the developed combination model and other comparison models. If the prediction effective value of the model is large, the prediction ability of the model is relatively effective. The mean value of the three datasets is presented in **Table 10**.

The specific details in **Table 10** shows, on the one hand, the predicted effective results of the combined model is always higher than the predicted effective value of other models, regardless of 1-order or 2-order. Hence, the proposed combination model is obviously more effective than the whole other comparative models. On the other hand, the results show that the prediction validity value of EEMD-MLP model is slightly lower than that of other comparison models, which is level at the second effective model compared with the other comparison models.

Algorithm 1 MOPSO

Input:
    $\boldsymbol{y}_t^{(0)} = (\boldsymbol{y}^{(0)}(1), \boldsymbol{y}^{(0)}(2), \cdots, \boldsymbol{y}^{(0)}(q))$-the training samples
    $\boldsymbol{y}_f^{(0)} = (\boldsymbol{y}^{(0)}(q + 1), \boldsymbol{y}^{(0)}(q + 2), \cdots, \boldsymbol{y}^{(0)}(q + l))$-the testing samples
Output:
    $\hat{\boldsymbol{y}}_f^{(0)} = (\boldsymbol{y}^{(0)}(q + 1), \boldsymbol{y}^{(0)}(q + 2), \cdots, \boldsymbol{y}^{(0)}(q + l))$-the forecasting data
Parameters:
    t - the current iteration number
    d - the number of dimension
    j - currently evolving particles

1   /*Initialize the population P and archive set A. */
2   /* Calculating the density information of particles in the archive set. */
3   /*Calculation of particle set A^j better than $P_{j,t}$ in A_t. */
4   **FOR** $k = 1$ **TO** $|A_t|$
5    $A^j = A^j + \{A_{k,t}|A_{k,t} \prec P_{j,t}, A_{k,t} \in A_t\}$
6   /*Calculating the particle set GJ with the lowest density in AJ.*/
7    $G^j = \min\{Density(A_k), k = 1, 2, \ldots, |A^j|, A_k \in A^j\}$
8   /*Update particle position and velocity in the population.*/
9   **IF** $|G^j| > 1$ **THEN** $g_{j,t} = Rand\{G^j\}$
10   **FOR** $i = 1$ **TO** d
11   /*Calculate position range $[\underline{x}_{j,t}^i, \bar{x}_{j,t}^i]$ and speed range $[\underline{v}_{j,t}^i, \bar{v}_{j,t}^i]$*/
12    $v_{j,t+1}^i = \chi(w v_{j,t}^i + c_1 R_1(p_{j,t}^i - x_{j,t}^i) + c_2 R_2(g_{j,t}^i - x_{j,t}^i))$
13    $x_{j,t+1}^i = x_{j,t}^i + v_{j,t+1}^i$
14    $x_{j,t+1}^i$ cross-border handling
15   /*Update archive set.*/
16   **IF** $A_t = \Phi$ **THEN**
17   **FOR** k=1 **TO** $|P_{t+1}|$
18    $A_{t+1} = A_{t+1} + \{P_{k,t+1}|P_{k,t+1} \prec P_{i,t+1} or P_{k,t+1} \prec\succ P_{i,t+1}, i = 1, 2, \ldots, |P_{t+1}|, i \neq k\}$
19   **ELSE**
20   **FOR** k=1 **TO** $|P_{t+1}|$
21    $A_{t+1} = A_{t+1} + \{P_{k,t+1}|P_{k,t+1} \prec A_{i,t} or P_{k,t+1} \prec\succ A_{i,t}, i = 1, 2, \ldots, |P_{t+1}|, i \neq k\}$
22   **END IF**
23   /*Output particle information in archive set.*/

---

5) FORECASTING EFFECTIVENESS

The bias-variance architecture is selected to assess the effectiveness of the combination model and other single models. The validity of prediction model is a comprehensive standard to measure the accuracy and stability of forecasting model. When evaluating the validity of the forecasting model, the accuracy and stability of the forecasting model are very significant. No matter accuracy or stability, it is not enough to consider only one standard. The average difference between the observation value y_n and the prediction value $\hat{y}_n$ on all the observation and prediction data is the composition of the absolute value of $Bias(\hat{y})$. This indicates that if the absolute value of $Bias(\hat{y})$ is small, the predict ability of the prediction model is more accurate. In terms of variance, the smaller the $Var(\hat{y})$, the more stable the prediction results of the prediction model. Therefore, the bias values in **Table 11** indicate that the bias absolute value of other models is greater than the absolute deviation value of the combined model, indicating that the combination model show more accuracy than other models. The values of variance analysis show that the

TABLE 11. The definition of MOPSO parameters.

Bias-Variance	Combined model	EEMD-MLP	EEMD-LSTM	EEMD-ARIMA
Bias	0.0036	0.0588	0.0456	0.1714
Variance	0.1123	0.1162	0.1455	0.3100

combination model has the most stabilization performance. The results show the combined model is the most accurate and stable model in wind speed forecasting, and its prediction effect is obviously superior to that of a single model.

V. CONCLUSION

With the increasing demand of renewable energy for pollution free energy, the requirement for renewable energy utilization and management into the electric system is also increasing. It should be noted that in the field of prediction, accuracy and stability should be equally important. Therefore, it is imperative to develop a technology which could obtain satisfactory accuracy and stability simultaneously. Nevertheless, as a

result of the randomness and intermittence of wind speed, it is hard to get both the accuracy and stability by a single model. In order to conquer this tricky issue, this paper develops a combined model on the basis of EEMD-MLP, EEMD-LSTM and EMD-ARIMA model, and uses the MOPSO algorithm to calculate the weight of the combined model of wind speed prediction. Specially, in each EEMD-ANN model, the EEMD is selected to process the wind speed sequence to improve the prediction performance.

In view of the prediction performance, in dataset A, the mean value of MAPE of the EEMD-MLP, EEMD-LSTM and EMD-ARIMA and combined model are 4.2318%, 3.7391%, 5.5545% and 3.5271%, respectively. In dataset B, the mean value of MAPE for these models is 5.7187%, 5.1124%, 9.2347% and 5.0176%, respectively. In dataset C, the mean value of MAPE of above models are 4.8838%, 4.2168%, 6.4343% and 4.0142%, respectively. Moreover, the error fluctuation of each prediction point of the wind speed combination model is the minimum, which shows that the combination model could enhance the accuracy and stability of the prediction. In this paper, one of the deep neural networks LSTM is adopted to wind speed prediction, and the LSTM is the recycle neural network (RNN) optimized algorithm. Furthermore, to enhance the accuracy and stability of wind speed predicting, a deep multi-layer perceptron is given. They all take a good express in forecasting accuracy and stability.

An example based on power grid shows that improving the prediction accuracy and stability is of great significance for wind power grid connection. The combination model has higher precision and steady performance, which can be used in electric system dispatching and has a wide range of economic and social benefits. For instance, timely adjust the scheduling plan, reduce the system reserve capacity, ensure the power quality, reduce environmental pollution, etc. For the future development direction, the combination prediction model proposed in this study can be applied to stock index prediction, traffic flow prediction, power load prediction and product sales prediction, as well as other forecasting fields. At the same time, flexibility can obtained according to the required accuracy and stability.

APPENDIX
The pseudo code of MOPSO.

CONFLICT OF INTEREST
The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] W. Yang, J. Wang, H. Lu, T. Niu, and P. Du, "Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: A case study in China," *J. Cleaner Prod.*, vol. 222, pp. 942–959, Jun. 2019.

[2] J. Wang, T. Niu, H. Lu, W. Yang, and P. Du, "A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 337–349, Jan. 2020.

[3] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," *Appl. Energy*, vol. 238, pp. 368–383, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.063.

[4] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks," *Appl. Energy*, vol. 107, pp. 191–208, Jul. 2013.

[5] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, Jul. 2010.

[6] Y.-Y. Hong, H.-L. Chang, and C.-S. Chiu, "Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs," *Energy*, vol. 35, no. 9, pp. 3870–3876, Sep. 2010.

[7] X. Zhang, J. Wang, and K. Zhang, "Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm," *Electr. Power Syst. Res.*, vol. 146, pp. 270–285, May 2017, doi: 10.1016/j.epsr.2017.01.035.

[8] T. Niu, J. Wang, H. Lu, and P. Du, "Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: Application to offshore wind speed," *Energy Convers. Manage.*, vol. 156, pp. 597–617, Jan. 2018, doi: 10.1016/j.enconman.2017.11.071.

[9] P. Jiang, R. Li, and H. Li, "Multi-objective algorithm for the design of prediction intervals for wind power forecasting model," *Appl. Math. Model.*, vol. 67, pp. 101–122, Mar. 2019, doi: 10.1016/j.apm.2018.10.019.

[10] L. Lazić, G. Pejanović, M. Živković, and L. Ilić, "Improved wind forecasts for wind power generation using the Eta model and MOS (Model Output Statistics) method," *Energy*, vol. 73, pp. 567–574, Aug. 2014.

[11] K. Wilgan, W. Rohm, and J. Bosy, "Multi-observation meteorological and GNSS data comparison with numerical weather prediction model," *Atmos. Res.*, vol. 156, pp. 29–42, Apr. 2015, doi: 10.1016/j.atmosres.2014.12.011.

[12] M. Milligan, M. Schwartz, and Y. Wan, "Statistical wind power forecasting models: Results for U.S. Wind farms," Nat. Renew. Energy Lab., Golden, CO, USA, 2003.

[13] M. Milligan, M. Schwartz, and Y. H. Wan, "Statistical wind power forecasting for U.S. Wind farms," *Bull. Amer. Meteorol. Soc.*, 2004.

[14] O. Ait Maatallah, A. Achuthan, K. Janoyan, and P. Marzocca, "Recursive wind speed forecasting based on hammerstein auto-regressive model," *Appl. Energy*, vol. 145, pp. 191–197, May 2015, doi: 10.1016/j.apenergy.2015.02.032.

[15] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, "A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 143, pp. 360–376, Jul. 2017, doi: 10.1016/j.enconman.2017.04.007.

[16] R. Jursa, "Variable selection for wind power prediction using particle swarm optimization," in *Proc. 9th Annu. Conf. Genetic Evol. Comput. (GECCO)*, 2007, pp. 2059–2065.

[17] N. Amjady, F. Keynia, and H. Zareipour, "Short-term wind power forecasting using ridgelet neural network," *Electr. Power Syst. Res.*, vol. 81, no. 12, pp. 2099–2107, Dec. 2011.

[18] R. Blonbou, "Very short-term wind power forecasting with neural networks and adaptive Bayesian learning," *Renew. Energy*, vol. 36, no. 3, pp. 1118–1124, Mar. 2011.

[19] C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, "Direct interval forecasting of wind speed using radial basis function neural networks in a multiobjective optimization framework," *Neurocomputing*, vol. 205, pp. 53–63, Sep. 2016, doi: 10.1016/j.neucom.2016.03.061.

[20] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renew. Sustain. Energy Rev.*, vol. 34, pp. 243–254, Jun. 2014, doi: 10.1016/j.rser.2014.03.033.

[21] C.-B. JL, J. Meng, E. Castillo-Guerra, and L. Chang, "A neural network approach to multi-step-ahead, short-term wind speed forecasting," in *Proc. 12th Int. Conf. Mach. Learn. Appl.*, Dec. 2013, pp. 243–248.

[22] J. L. Torres, A. García, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in navarre (Spain)," *Sol. Energy*, vol. 79, no. 1, pp. 65–77, Jul. 2005.

[23] X. Zhang, J. Wang, and Y. Gao, "A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM," *Energy Econ.*, vol. 81, pp. 899–913, Jun. 2019, doi: 10.1016/j.eneco.2019.05.026.

[24] 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.

[25] D. Yang, V. Sharma, Z. Ye, L. I. Lim, L. Zhao, and A. W. Aryaputera, "Forecasting of global horizontal irradiance by exponential smoothing, using decompositions," *Energy*, vol. 81, pp. 111–119, Mar. 2015.

[26] T. G. Barbounis and J. B. Theocharis, "A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation," *Neurocomputing*, vol. 70, nos. 7–9, pp. 1525–1542, Mar. 2007.

[27] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *J. Oper. Res. Soc.*, vol. 20, no. 4, pp. 451–468, 1969.

[28] J. Wang, L. Liang-dong, and W. Zuo-yi, "The status and development of the combination forecast method," *Forecast*, vol. 6, pp. 37–38, Aug. 1997.

[29] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Appl. Energy*, vol. 167, pp. 135–153, Apr. 2016.

[30] L. Xiao, J. Wang, R. Hou, and J. Wu, "A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting," *Energy*, vol. 82, pp. 524–549, Mar. 2015.

[31] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[32] 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.

[33] 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. A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[34] J. L. McClelland, D. Rumelhart, and The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Foundations. Psychological and Biological Models*, vols. 1–2. Cambridge, MA, USA: MIT Press, Jan. 1987.

[35] J. E. Dennis, Jr., and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia, PA, USA: SIAM, 1996.

[36] U. Schlink, S. Dorling, E. Pelikan, G. Nunnari, G. Cawley, H. Junninen, A. Greig, R. Foxall, K. Eben, T. Chatterton, J. Vondracek, M. Richter, M. Dostal, L. Bertucco, M. Kolehmainen, and M. Doyle, "A rigorous intercomparison of ground-level ozone predictions," *Atmos. Environ.*, vol. 37, no. 23, pp. 3237–3253, Jul. 2003.

[37] E. Agirre-Basurko, G. Ibarra-Berastegi, and I. Madariaga, "Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area," *Environ. Model. Softw.*, vol. 21, no. 4, pp. 430–446, Apr. 2006.

[38] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon Press, 1995.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Wein-Berger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[42] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[43] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[44] G. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. SanFrancisco, CA, USA: Holden-Day, 1970.

[45] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[46] M. Hu, J. D. Weir, and T. Wu, "An augmented multi-objective particle swarm optimizer for building cluster operation decisions," *Appl. Soft Comput.*, vol. 25, pp. 347–359, Dec. 2014.

[47] S. A. Torabi, N. Sahebjamnia, S. A. Mansouri, and M. A. Bajestani, "A particle swarm optimization for a fuzzy multi-objective unrelated parallel machines scheduling problem," *Appl. Soft Comput.*, vol. 13, no. 12, pp. 4750–4762, Dec. 2013.

[48] P. Fattahi, V. Hajipour, and A. Nobari, "A bi-objective continuous review inventory control model: Pareto-based meta-heuristic algorithms," *Appl. Soft Comput.*, vol. 32, pp. 211–223, Jul. 2015.

[49] S. H. R. Pasandideh, S. T. A. Niaki, and S. Sharafzadeh, "Optimizing a bi-objective multi-product EPQ model with defective items, rework and limited orders: NSGA-II and MOPSO algorithms," *J. Manuf. Syst.*, vol. 32, no. 4, pp. 764–770, Oct. 2013.

[50] S. Naka, T. Genji, T. Yura, and Y. Fukuyama, "Practical distribution state estimation using hybrid particle swarm optimization," in *Proc. IEEE Power Eng. Soc. Winter Meeting. Conf. Process.*, Jan. 2001, pp. 815–820.

[51] J. Kennedy, R. Eberhart, and Y. Shi, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.

[52] I. H. Witten and F. Eibe, *Data Mining Practical Machine Learning Tools and Techniques*, 2nd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[53] W. Yang, J. Wang, T. Niu, and P. Du, "A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting," *Appl. Energy*, vol. 235, pp. 1205–1225, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.034.

[54] Y. Hao and C. Tian, "The study and application of a novel hybrid system for air quality early-warning," *Appl. Soft Comput.*, vol. 74, pp. 729–746, Jan. 2019, doi: 10.1016/j.asoc.2018.09.005.

[55] J. Wang, P. Du, H. Lu, W. Yang, and T. Niu, "An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting," *Appl. Soft Comput.*, vol. 72, pp. 321–337, Nov. 2018, doi: 10.1016/j.asoc.2018.07.022.

[56] H. Li, J. Wang, R. Li, and H. Lu, "Novel analysis–forecast system based on multi-objective optimization for air quality index," *J. Cleaner Prod.*, vol. 208, pp. 1365–1383, Jan. 2019, doi: 10.1016/j.jclepro.2018.10.129.

[57] R. Li, Y. Dong, Z. Zhu, C. Li, and H. Yang, "A dynamic evaluation framework for ambient air pollution monitoring," *Appl. Math. Model.*, vol. 65, pp. 52–71, Jan. 2019, doi: 10.1016/j.apm.2018.07.052.

[58] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, "Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting," *Appl. Energy*, vol. 180, pp. 213–233, Oct. 2016.

[59] N. Srinivas and K. Deb, "Multi-objective optimization using nondominated sorting in genetic algorithms," *Evol. Comput.*, vol. 2, no. 3, pp. 221–248, Sep. 1994.

● ● ●