Short-term wind power prediction optimized by multi-objective dragonfly algorithm based on variational mode decomposition

Yilin Zhou[a], Jianzhou Wang[a, *], Haiyan Lu[b], Weigang Zhao[c]

[a]*School of Statistics, Dongbei University of Finance and Economics, Dalian, China*

[b]*School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia*

[c]*School of Management and Economics, Beijing Institute of Technology, Beijing, China*

Abstract

Short-term wind power prediction has a considerable effect on improving the productivity of wind energy systems and increasing economic benefits. In recently years, various wind velocity predictive models have been designed to raise the prediction effect. However, numerous predictive systems are limited by single type, and many ordinary predictive systems ignore the advantage of optimized parameters and the significance of data preparation, which bring about the lower predictive precision. To fill this gap, in this article, a novel predictive system is come up, which is on the basis of data denoising strategy, statistical predictive systems, artificial intelligence forecasting system and multi-objective optimization strategy. After using the data denoising strategy for denoising, the reconstructed data is used for the forecasting of different sub-systems, to obtain stable forecasting results, multi-objective dragonfly algorithm is used to estimate the weight coefficient of sub-systems. To evaluate the availability of the designed predictive system, five wind velocity datasets from different wind farms are used for the purpose of a case research. According four experiments and four analyses, it can be concluded that the designed combined system has a well predictive effect in short-term wind speed prediction. And it is in favor of grid regulation and operation.

Key words: Wind speed prediction; Variational mode decomposition; Multi-objective dragonfly algorithm; Hybrid models;

* Corresponding author.

Email address: jianzhouwdufe@163.com (J-Z. Wang)

1. Introduction

Due to routine energy resources such as petroleum and coal, have brought severe problems of environment, the research of renewable energy sources has aroused widespread concern from countries the world over [1]. After extensive research, wind energy has been recognized as a green, sustainable and emission-free energy resource, so it has been supported by national funding[2]. And in recent years, countries around the world have been positively goosing the growth of wind energy[3]. On the basis of the Global Wind Power Report 2019, the new installed capacity exceeded the 60GW milestone for the second time in history, with a year-on-year increase of 19%. In 2019, the installed capacity of onshore market was 54.2 GW, a growth of 17% compared with 2018[4], and the circumstance of other major markets can be seen in **Fig.1**. Nevertheless, in the wake of the increase in the equipped capacitor, the efficient use of wind energy source has become a problem that must be solved. In wind farms, since wind energy is highly random and intermittent, the dependability of model prediction and the quality of wind power will decline[5]. Accordingly, to guarantees the safety operation and rationally regulation of the electrical power systems, it is essential to correctly forecast wind speed[6].

Motivated by the need for enhance the forecasting accuracy, various researches in view of wind velocity prediction have been carried out over the past several years [7]. Through these researches, wind speed prediction technology has been relatively mature. And there are mainly two prediction modes for data from different sources. The first mode is mainly used for numerical forecasts where forecasting periods of over than four hours[8]. Nevertheless, the forecasting system is more costly and complex. The second mode uses some wind power mathematical models to direct forecast wind power based on historical wind velocity series, wind direction and temperature[9]. And it is noteworthy that the principal body of this research is using historical wind velocity sequences to make short-term prediction.

Since the wind speed series depend on the climate system, there is noise in the sequences, which bring about large errors in the prediction results. Therefore, before predicting wind speed, some preprocessing algorithms must be performed on the wind speed sequence data[10,11]. In [12] **WD** as an traditional data denoising, it can be used to identify the chaotic and repeatable features of frequencies with the signal, and Li LL etc.[13] use this method to decompose the original wind

power data. But in general, it's difficult to choose an appropriate wavelet base. Although **PCA** can decrease the dataset's dimension and the complicacy of the wind velocity series[14], this method may cause data loss and it is not easy to perfectly reserve the valid information of the raw set. Amjady N etc.[15] proposed the use of **EMD** method to decompose the volatile wind power time sequences, and got relatively good results. And Abedinia O etc.[16] designed a novel strategy based on **EMD** to decompose the wind power measurement. By adding the white noise sequence to the primitive sequences, **EEMD** technique further processes the data set to solve the complex mixed mode problem[17]. However, it may cause residual noise and increase the time required for processing time sequences. The **VMD** is proposed on the basis of **EMD**, which is a totally non recursive and quasi-orthogonal decomposition method[18]. And this technique is widely used to decompose the raw wind velocity sequences, which can resolve the modal aliasing problem effective and increase the wind speed forecasting reliability [19]. Therefore, in this research, **VMD** is selected as the data denoising technique of the wind power time sequences.

After the data preprocessing, the quality of the data has been improved to a certain extent. Then, there are four main methods to deal with time series data: physics[20], statistics[21,22], artificial intelligence[23,24], and hybrid[25–27].

(1) Physical models usually need to use current meteorological and geographic data, such as temperature, velocity, density etc.[28]. Generally speaking, for this model, effective prediction results depend on physical information rather than wind speed time series. Therefore, we cannot easily obtain accurate prediction results through this model, and prediction requires various financial resources and time[29].

(2) Statistical models, include traditional **ARIMA**[30,31], exponential smoothing[32] and **KF**[33,34], can deal with linear features well, but they often cannot give very good results with the nonlinearity. Furthermore, under the input series are normal, the statistical models can produce good predict outcomes [35]. Nevertheless, these models are not often apply to the short-term wind velocity forecasts [36], since most of data in reality are non-linear, noisy, unstable and volatile, and the features in the original time series are always uncontrollable, which makes most modeling unable to obtain complete and effective information.

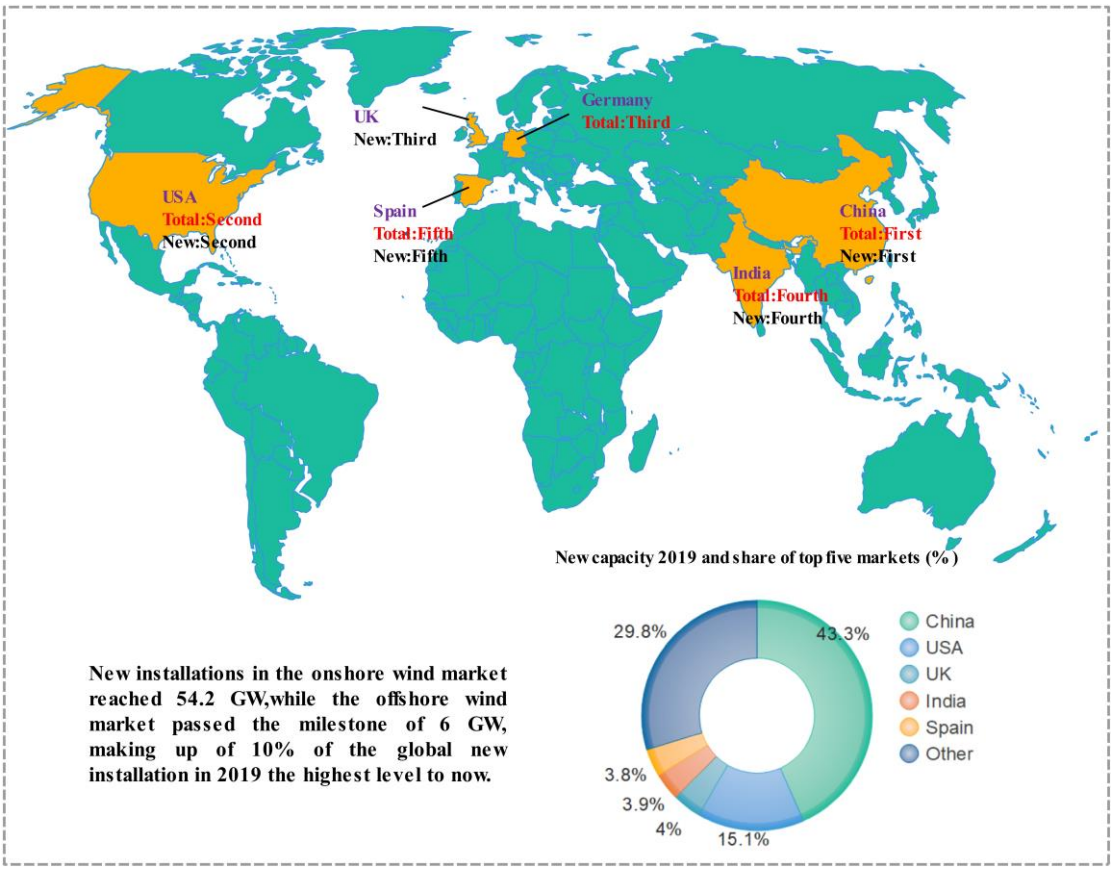


Fig. 1. New and total capacity of installed wind power in six major markets in 2019.

(3) In recent year, the emergence of artificial intelligence algorithms [37]has promoted the development of short-term wind speed prediction. Artificial intelligence algorithms include **SVM**[38], **ANN**[39,40] and **TCN**[41] can provide greater prediction precision than physical and statistical models. But, it cannot be ignored that the performance of its prediction mostly depends on the training set, which also makes it easy to fall into situations like over-fitting and local optimum [42].

(4) Since each of the above models has its inherent shortcomings, researchers began to study the effect of combined models on prediction. In[43], a new combined system in the light of neural network and chaotic shark smell optimization was designed to handle the intermittency and variability of the wind power. Particularly, the hybrid techniques combine different single algorithms to get better prediction performance, and increase the precision, effectiveness and constancy of wind velocity forecasts[44]. Therefore, the application of the hybrid models in wind speed forecast has much enhanced in the past few years. And the forecast accuracy and effect are also satisfactory[45–47]. Furthermore, **Table 1** summarizes the pros and cons of the above four types of models.

Therefore, in this research, a hybrid predictive model is designed, combining multiple forecasting models. The test and predictive results display that the designed model not only effectively enhances the prediction precision and stability, but also has a good prediction effect for data from multiple sites.

The contribution of this article can be summarized as:

1) **This research has designed a special model in the light of the data denoising method and the multi-objective optimization algorithm.** Which to enhance the efficiency and accuracy of prediction. Furthermore, the data from five sites were chosen to validate the performance of designed predictive system.

2) **A unique data processing technique is selected, which can significantly decrease the noisy information in the time sequences data and decrease or reject the influences of high-frequency noise.** Moreover, the reconstructed data is used for model prediction, which improves the performance of forecasting.

3) **In this research, five systems are chosen as sub-systems of the designed system. And the multi-objective dragonfly algorithm is chosen to estimate the weights of the sub-systems.** Comparing with the ordinary model, the predictive combined model significantly increases the prediction precision and prediction constancy.

4) **In this research, four compact and rational indicators are selected as the evaluation metrics for evaluating whether the predictive result of the designed model is good.** Which including root mean square error, mean square error, mean absolute percentage error and mean absolute error.

5) **To research the predictive stability and availability of the designed model, the Diebold-Mariano test, Sensitivity analysis and Forecasting effectiveness test are selected to compare the difference and degree of stability between the designed system and other models.** Moreover, the experimental conclusions indicate that in this research the predictive model developed is greater than other models in terms of stability and effectiveness.

Table 1
Overview of four types of models.

Model	Ref.	Variable	Result	Advantage	Disadvantage
Physical models					
NWP	[20]	Weather data	Using the weather data from NWP model can improve performance, especially in the hottest regions where temperature is most affected by power consumption.	The physical model shows preferable performance in terms of long-term prediction and have broad application prospects.	Physical model requires extensive financial resources and time.
Statistical models					
ARIMA	[21]	Wind power	The designed predictive system has a certain efficiency on the forecast of wind velocity sequences data.	Under the input series are normal, the statistical model such as traditional **ARIMA**, exponential smoothing and Kalman filter can produce good prediction outcomes. And statistical methods can deal with linear features well.	In short-term wind energy forecasts, especially multi-step forecasts, the statistical model often performs poorly. And they often cannot give very good results when nonlinear problems are involved.
	[31]	Time sequence	The designed model is a suitable strategy for analyzing statistics.		
Kalman filter	[22]	Wind power	The proposed system can supply more accurate estimation results than other comparison systems.		
	[33]	Wind power	The results show that the model forecasting skill have been obviously enhanced, especially in the aspect of wind power forecasts.		
STES	[32]	Daily volatility	Experimental result shows that compared to the multiple GARCH system and the fixed parameter exponential smoothing, the proposed strategy performs better.		
Artificial intelligence models					
ANN/SMEs	[23]	Wind energy	The **SMES** controlled by **ANN** improves the temporary state constancy of the wind farms connected to the multi machine power systems.	The artificial intelligence models include **SVM** and artificial neural networks can provide higher prediction accuracy than statistical models and physical models.	The performance of artificial intelligence model prediction mostly depends on the training set, which also makes it easy to fall into situations such as local optimal, and over-fitting.
BP	[39]	Wind power	Neural network wind power predictive model on the basis of tabu search algorithm can enhance the forecasting precision.		
	[40]	Stock price	The prediction result of the stock price trend predictive system based on BP model is greater than other comparison model.		
Jaya-SVM	[38]	Wind speed	The designed system can produce the best results than other eight ordinary models and has accurate wind power prediction capabilities.		
Hybrid forecasting models					
DWT/LSSVM/GARCH	[45]	Wind speed	In the part of precision and constancy, the designed model has approved performance.	The hybrid system can enhance the accuracy, effectiveness and constancy of wind velocity prediction. And it is more adaptable to the predicting situations than the single model.	In most studies, a single data preprocessing and target algorithm are needed to improve the stability of prediction.
CEEMDAN/LSTM	[46]	Wind speed	The proposed system performs greater than other considered systems without double decomposition.		
MLP-WOA	[47]	Wind power	As for the MLP model, the WOA optimization strategy can enhance the predictive precision of it.		

The paper is structured as below. The principle of each method in the combined model is introduces in the second section, and the third section introduces the general formation of the model which is proposed. The fourth part introduces the wind speed sequences data used in the experiment in detail, and not only designs four control experiments to conclude the forecasting precision of the predictive model, but details the compare outcomes of the designed model and other commonly used models. Moreover, in the fifth section, four test technologies are chosen to discuss the precision validity and constancy of the designed predictive system and compare with other commonly used systems. And lastly, the sixth section is the conclusion part.

The abbreviations and corresponding explanations of the models and algorithms involved in this article are displayed in **Table 2**.

2. Methodology

In this part, the theoretical strategies of the designed predictive system are introduced, which include the **VMD** strategy and **MODA**.

2.1. Variational Mode Decomposition

The **VMD** proposed on the basis of **EMD** is not only a new type of complex signal decomposition method, but also an self-adaptive signal processing technique[48]. Generally speaking, under the present count of patterns decomposed K, the primitive signals $f(t)$ is break down into K pattern functions and a center frequency, namely $\boldsymbol{U}_k(t)$ and $\omega_k(k=1,2,...,K)$ in turn. Furthermore, the variational mode decomposition technique break real signals down into a specified count of bandwidth limited patterns and minimizes the sum of evaluated bandwidth for patterns.[49,50]

Solving the variational problem is the essence of **VMD** strategy. And the main steps are as below.

1) The Hilbert transform to be used for computing the analytical signals of each mode function, then to get the unilateral spectrum, and the mathematical modeling of the Hilbert transform is $\left[\boldsymbol{\delta}(t)+j/\pi t\right]*\boldsymbol{U}_k(t)$. Where $\{\boldsymbol{U}_k\}=\{u_1,u_2,\cdots,u_K\}$ represents the set of k patterns decomposed, $\boldsymbol{\delta}(t)$ represents the impulse function and $*$ represents convolution.

Table 2
Symbol abbreviation and corresponding explanation.

Nomenclature			
Abbreviate			
\MOALO	Multi-objective ant lion optimization	**\MOGWO**	Multi-objective grey wolf optimization
\EEMD	Ensemble empirical mode decomposition	**\EMD**	Empirical mode decomposition
\MODA	Multi-objective dragonfly algorithm	**\WD**	Wavelet denoising
\BPNN	Back propagation neural network	**\VMD**	Variational mode decomposition
\ELMAN	Elman neural network	**\TCN**	Temporal convolutional networks
\LS-SVM	Least square support vector machine	**\LSTM**	Long short-term memory
\ELM	Extreme learning machine	**\CNN**	Convolutional neural network
\GRNN	General regression neural network	**\MAPE**	Mean absolute percent error
\ARIMA	Autoregressive integrated moving average	**\MAE**	Mean absolute error
\ANN	Artificial neural network	**\RMSE**	Root mean square error
\PCA	Principal component analysis	**\MSE**	Mean square error
\KF	Kalman filter	**\FE**	Forecasting effectiveness
\NWP	Numerical weather predictor	**\DM**	Diebold-Mariano
\STES	Smooth transition exponential smoothing	**\SMEs**	Superconducting magnetic energy storage system
Function			
$\boldsymbol{U}_k(\cdot)$	Mode function	$\boldsymbol{\omega}_k$	Center frequency
$\boldsymbol{\delta}(\cdot)$	Impulse function	$\boldsymbol{Le'vy}(\cdot)$	The levy fight method
$\hat{\lambda}^n(\cdot)$	Augmented Lagrangian function	$\boldsymbol{S}_i$	The separation function of *ith* individual
$\boldsymbol{A}_i$	The alignment function of *ith* individual	$\boldsymbol{C}_i$	The cohesion function of *ith* individual
$\boldsymbol{F}_i$	The attraction function of *ith* individual	$\boldsymbol{E}_i$	The Distraction function of *ith* individual
$\tilde{\boldsymbol{A}}_n$	Forecasting accuracy	ξ_n	Forecasting error
Variables			
τ	Convergence threshold	$\boldsymbol{X}^+$	The position of food source
$\boldsymbol{X}^-$	The position of enemy	*ub*	The upper bound of weight
lb	The lower bound of weight	*t*	The number of current iterations
$\boldsymbol{MAPE}_i^j$	**MAPE** value obtained by the *ith* model at Site *j*	$\overrightarrow{\boldsymbol{MAPE}}$	The value matrix of **MAPE**
$\overline{\overline{\mathbf{DM}}}$	The maximum value of **DM**	$\underline{\underline{\mathbf{DM}}}$	The minimum value of **DM**

2) Obtain the baseband signal by mixing the pattern function with the unilateral spectrum, the center frequency and the exponential signal. The mathematical modeling of that is $\left[\left(\boldsymbol{\delta}_t + j/\pi t\right) * \boldsymbol{U}_k(t)\right] e^{-j\omega_k t}$ .Here, $\{\omega_k\} = \{\omega_1, \omega_2, \cdots, \omega_K\}$ expresses the central frequency sets of the patterns decomposed.

3) For each pattern, computer the bandwidth of them, and the obtained model of variational constraint is as below.

$$\begin{cases} \min\limits_{\{\boldsymbol{u}_k\},\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \left\| \wp_t \left[\left(\boldsymbol{\delta}_t + \dfrac{j}{\pi t} \right) * \boldsymbol{U}_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum\limits_{k=1}^{K} \boldsymbol{U}_k = f(t) \end{cases} \tag{1}$$

4) For transform the restricted question into an unrestricted question, we introduce he Lagrangian function and secondary penalty term, and the augmented Lagrangian expression is obtained as:

$$\begin{aligned} L\left(\{\boldsymbol{U}_k\},\{\omega_k\},\lambda\right) = \alpha \sum_{k=1}^{K} \left\| \wp_t \left[\left(\boldsymbol{\delta}_t + \frac{j}{\pi t} \right) * \boldsymbol{U}_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f(t) - \left[\boldsymbol{U}_k(t)\right] \right\|_2^2 + \left[\lambda(t), f(t) - \sum_{k=1}^{K} \boldsymbol{U}_k(t) \right] \end{aligned} \tag{2}$$

In Eq. (2), λ and α indicate the augmented Lagrangian function and the secondary penalty term respectively.

5) Let $n = 1$ and start the loop, as $k = 1:K$, bring up to date $\{\boldsymbol{U}_k\}$, $\{\omega_k\}$ and λ .

i. As $\omega \geq 0$, iteratively bring up to date the $\boldsymbol{U}_k$ values by using Eq. (3).

$$\hat{\boldsymbol{U}}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{\boldsymbol{U}}_i(\omega) + \dfrac{\hat{\boldsymbol{U}}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

ii. Iteratively bring up to date the values of ω_k , the specific mathematical expression is:

$$\omega_k^{n+1} = \frac{\int_0^{\infty} \omega \left| \hat{\boldsymbol{U}}_k(\omega) \right|^2 d\omega}{\int_0^{\infty} \left| \hat{\boldsymbol{U}}_k(\omega) \right|^2 d\omega} \tag{4}$$

iii. Iteratively bring up to date the λ values by using Eq. (5).

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \tau\left(\hat{f}(\omega) - \sum_{k=1}^{K}\hat{\boldsymbol{U}}_k^{n+1}(\omega)\right) \tag{5}$$

Where τ represents a convergence threshold.

6) Set convergence threshold ε, then stop the iteration if the following conditions are met:

$$\frac{\sum_{k=1}^{K}\left\|\hat{\boldsymbol{U}}_k^{n+1} - \hat{\boldsymbol{U}}_k^{n}\right\|_2^2}{\left\|\hat{\boldsymbol{U}}_k^{n}\right\|_2^2} < \varepsilon \tag{6}$$

2.2. Multi-Objective Dragonfly Algorithm

In recent years, heuristic algorithms based on natural group behavior have been widely used. Which including the biologically inspired computing[51] and Swarm Intelligence [52] have attracted the attention of many researchers[53]. So in this research, we use a special natural heuristic optimized algorithm, the multi-objective dragonfly algorithm.[54]

The mathematic model of **MODA** is expressed in the sub-section that follow[55]. In the following equations, the positions of current and *jth* nearby individual are represented by $\boldsymbol{X}$ and $\boldsymbol{X}_j$ in turn. And $\boldsymbol{N}$ represents the number of neighboring individuals.

Separation refers to the method that guarantees keeping the search individual away from each other nearby individuals. This behavior is calculated by using $\boldsymbol{S}_i = -\sum_{j=1}^{N}\boldsymbol{X} - \boldsymbol{X}_j$.

Alignment shows how an individual's speed matches that of other neighboring individuals. The alignment is estimated by using $\boldsymbol{A}_i = (1/\boldsymbol{N})\sum_{j=1}^{N}\boldsymbol{V}_j$. Here, $\boldsymbol{V}_j$ expresses the *jth* nearby individual velocity.

Cohesion expresses the individual tend to approach the nearby centroid. This behavior is calculated by using $\boldsymbol{C}_i = (1/\boldsymbol{N})\sum_{j=1}^{N}\boldsymbol{X}_j - \boldsymbol{X}$.

Attraction refers to the individuals tend to approach the source of food. The attraction is estimated by using $\boldsymbol{F}_i = \boldsymbol{X}^{+} - \boldsymbol{X}$. Here, $\boldsymbol{X}^{+}$ represents the food source's position.

Distraction indicates the tendency which an individual is far away from an enemy in nature. And this behavior is described by using $\boldsymbol{E}_i = \boldsymbol{X}^{-} + \boldsymbol{X}$ .Here, $\boldsymbol{X}^{-}$ expresses the location of enemy.

The step vector is expressed as the dragonflies' movement direction and the mathematical model of it is as below: $\Delta\boldsymbol{X}_{t+1} = \left(\boldsymbol{a}\boldsymbol{A}_i + \boldsymbol{s}\boldsymbol{S}_i + \boldsymbol{c}\boldsymbol{C}_i + \boldsymbol{f}\boldsymbol{F}_i + \boldsymbol{e}\boldsymbol{E}_i\right) + \boldsymbol{w}\Delta\boldsymbol{X}_t$. Here, $\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{f}, \boldsymbol{s}$ and $\boldsymbol{e}$ represent

the weights of the cohesion C_i, the alignment A_i, the attraction F_i, the separation S_i and the distraction E_i of the ith individual in turn. In addition, w is the inertia weight.

According to the calculated step vector, the mathematical model for updating the position of an individual is calculated by using $X_{t+1} = X_t + \Delta X_{t+1}$.

In order to further optimize the algorithm, when there is no nearby solution of neighboring similar individuals, the individual's position is updated by using the $Le'vy$ method: $X_{t+1} = X_t + Le'vy(d) \times X_{t+1}$. Here, the position vector's dimension is represented by d, and the $Le'vy$ is estimated as below.

$$Le'vy(x) = 0.01 * \frac{r_1 * \delta}{|r_2|^{1/\beta}} \tag{7}$$

Here, r_1, r_2 express two random number, β is a constant. In addition, σ is represented by Eq. (8).

$$\sigma = \left(\frac{\Gamma(1+\beta) * \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) * \beta * 2\left(\frac{\beta-1}{2}\right)} \right)^{\frac{1}{\beta}}, \Gamma(x) = (x-1)! \tag{8}$$

In this algorithm, Seyedali Mirjalili use the non-inferior solutions obtained in the archive set storage optimization process in the swarm intelligence algorithm, then the roulette strategy and the adaptive grid strategy are used to choose the position of the enemy and the location of the food source from the archive set, the best result is finally obtained.

Algorithm: **MODA.**

Parameters:

N_d—Dragonfly number.
I_n—Iteration number.
A_s—Archive size.
X_i, X_j—the positions of ith and jth grasshoppers.

1: /* Install the parameters of MODA */
2: Initialize the grasshopper position $X_i, N_d, I_m, A_s, ub, lb$;
3: Initialize external archive W;
4: $iter = 1$;
5: /* Calculate the *inertia weight* */
6: **WHILE** $(iter \le I_m)$ **DO**
7: $r = (1/4)(ub - lb) + (ub - lb) * 2 * t / I_m$;
8: $w = 0.9 - 0.7 * iter / I_m$;

9: /* Initialize relative weights include f,a,c,s,e */
10: **FOR** (all agents) **DO**
11: /* Normalize distance between grasshoppers in [1,4] */
12: Update f, a,c,s and e;
13: **IF** (*neighboring outcome* > 1) **THEN**
14: /*Update the location of the current search entity */
15: $\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \Delta \boldsymbol{X}_{t+1}$;
16: Update W;
17: **ELSE**
18: $\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \boldsymbol{Le'vy}(d) \times \boldsymbol{X}_{t+1}$;
19: $\Delta \boldsymbol{X} = \mathbf{0}$;
20: Add the new solution to W;
21: **END IF**
22: **END FOR**
23: $iter = iter + 1$;
24: **END WHILE**
25: $\hat{z}(i) = W(1) * \hat{Y}_1(i) + W(2) * \hat{Y}_2(i) + W(3) * \hat{Y}_3(i) + W(4) * \hat{Y}_4(i) + W(5) * \hat{Y}_5(i)$;
26: **RETURN** $\hat{z}$

3. Framework of the designed system

Combined models, which is a weighted and combined combination of other single models, and it is widely applied in terms of forecasting[56]. Nevertheless, the result obtained by the conventional weighting method may differ greatly from the actual value[57]. To improve the forecast effect, a novel predictive system is designed in this article. The main flow of this designed system and specific sequence feature are both demonstrated in **Fig. 2**, and the more concrete processes are described as follows.

Stage 1 Data decomposition

Since the raw wind velocity sequences are usually random and unstable, the use of the raw data modeling to predict will cause large errors. The **VMD** technique is chosen to decrease the volatility of the wind velocity sequences, the original series are decomposed into several sub-sequences, after the data reconstruction, the final series are used for the future prediction. On the basis of the numerical results of the trial-and-error method, we can conclude that when the main parameter values of **VMD** strategy are $K_d = 7$, $\alpha_d = 2000$ and $\tau = 0.001$, the reconstructed sequence has the optimal effect of prediction.

Stage 2 Sub-systems selection

Since no single system can adapt well to all types of data sequences, in this paper, we combine several single systems to fit as many data types as possible. And according to the reference and the accuracy of prediction, five single predictive systems with better prediction

effects are selected as the sub-systems of designed combined system, namely **BPNN, ELM, LS-SVM,** Elman neural network (**ELMAN**) and **GRNN** respectively, and the main parameter values of those five systems and other major forecasting system are shown in **Table 3**.

Stage 3 Optimization Algorithm

To additional increase the predictive efficiency, the weight of each sub-system is determined by the **MODA**. After data preprocessing and sub-systems forecasting, we can obtain a value matrix $[\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \hat{Y}_4, \hat{Y}_5]$, where $\hat{Y}_j$ indicates the forecasting sequence obtained by jth single system. Then through the **MODA** method, the weights coefficient W can be determined. And for this method, the iteration number is $\boldsymbol{I_n} = 200$, the dragonfly number is $\boldsymbol{N_d} = 40$, and the single weight upper and lower bound are $ub = 2$ and $lb = -2$ in turn. In the light of the prediction outcomes of sub-systems and weights coefficient obtained, we can gain the final result.

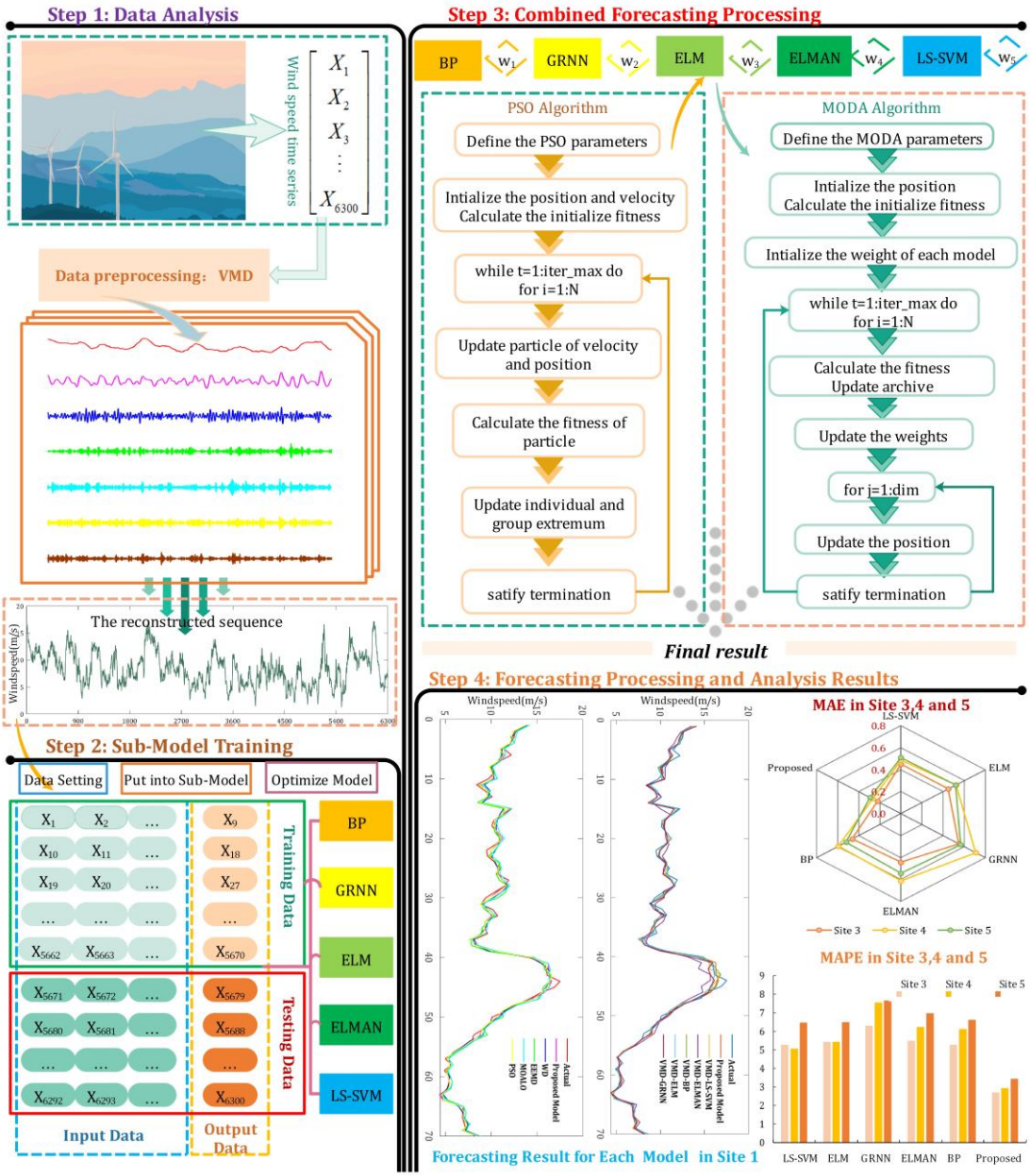


Fig. 2. Structure of the designed system

4. Analysis and results of experiment

To test and prove the predictive performance of the designed predictive system, in this part, we design four experiments, which compare the availability of designed system with **VMD**-based single predictive systems, traditional single predictive systems, combined system on the basis of different denoising strategy, and other popular predictive systems respectively. Moreover, the more concrete procedures are described as follows.

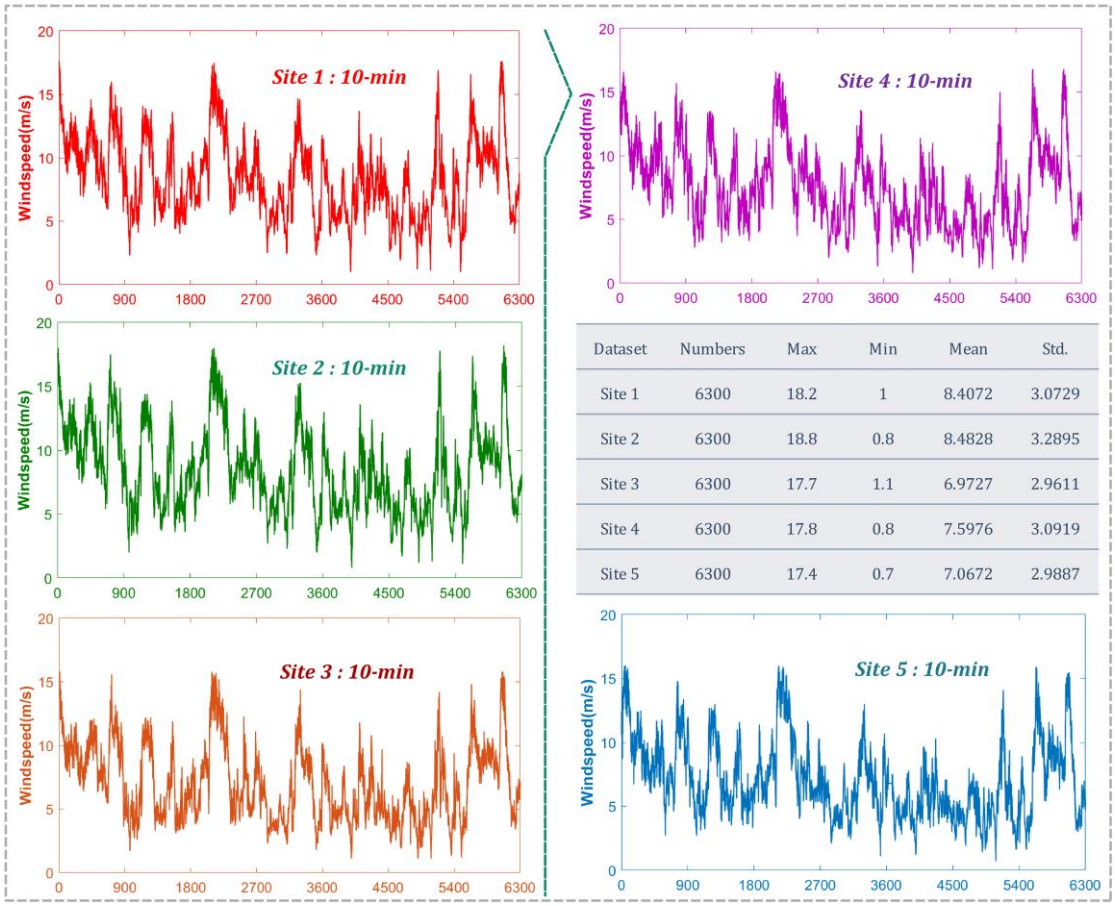
Table 3
Parameters explanations and details of main systems.

System	Symbol	Explanation	Value	Reason
GRNN	m_i	Number of input layer nodes	8	Number of feature inputs
	d_h	Desire spread	-	Iterative optimization
	m_o	Number of output layer nodes	1	Time series regression
ELM、BPNN、ELMAN	m_i	Number of input layer nodes	8	Number of feature inputs
	m_o	Number of output layer nodes	5	Trial-and-error manner
	m_h	Number of hidden layer nodes	1	Time series regression
LS-SVM	c_l	Penalty parameter	2	Trial-and-error manner
	g_l	Kernel function parameter	1	Trial-and-error manner
WD	D_l	Decomposition Layer Number	5	Trial-and-error manner
EEMD	N_{std}	Standard deviation of Gaussian white noise	0.1	Common value [0.01,0.05,01, ⋯]
	N_e	Number of times to add noise	100	Common value [50,100,⋯]
VMD	K_d	Count of modes decomposed	7	Trial-and-error manner
	α_d	Penalty factor	2000	Common value [500,1000,2000, ⋯]
LSTM	e_τ	Epochs of training	250	Trial-and-error manner
CNN	K_s	Kernel size of the convolutional layer	40	Trial-and-error manner
	K_n	Number of kernels in the convolutional layer	3	Trial-and-error manner
	m_h	Number of hidden layer nodes	[384, 384]	Trial-and-error manner
MODA	A_n	Archive size	500	Common value [100, 200, 300, 400, 500]
	I_m	Iteration number	200	Common value [50, 100, 150, 200, 250]
	N_d	Dragonfly Number	40	Common value [20, 40, 60, 80, 100]
MOGWO、MOALO	A_n	Archive size	500	Common value [100, 200, 300, 400, 500]
	I_n	Iteration number	200	Common value [50, 100, 150, 200, 250]
	N_d	Dragonfly Number	40	Common value [20, 40, 60, 80, 100]

4.1. Information of datasets

In this research, five different wind generators in a certain region in eastern China are selected. Therefore, five groups of wind power sequences are obtained, and each group includes 6300 observations. We divide each group of samples into two parts, the training sample and verifying sample. Among them, the 1st to 5670th data constitute the training sample, and the 5671th to 6300th data constitute the verifying sample. The major numeric characteristics of the chosen series are displayed in **Fig.3** and **Table 4**.

To particular, this paper uses historical wind velocity series as the only input for wind velocity prediction. Therefore, the feature selection of the data mainly includes the processing of missing values and outliers. For observations where the wind speed or power in the data is negative, considering it will cause unnecessary errors in the prediction, we use the nearest neighbor mean interpolation method to reduce the error.



Dataset	Numbers	Max	Min	Mean	Std.
Site 1	6300	18.2	1	8.4072	3.0729
Site 2	6300	18.8	0.8	8.4828	3.2895
Site 3	6300	17.7	1.1	6.9727	2.9611
Site 4	6300	17.8	0.8	7.5976	3.0919
Site 5	6300	17.4	0.7	7.0672	2.9887

Fig. 3. Information of the original data.

Table 4
Main numerical characteristics of five wind speed series.

Period		**Site 1**			**Site 2**			**Site 3**			**Site 4**			**Site 5**		
Sample		**All**	**Train**	**Verify**	**All**	**Train**	**Verify**	**All**	**Train**	**Verify**	**All**	**Train**	**Verify**	**All**	**Train**	**Verify**
Numbers		6300	5670	630	6300	5670	630	6300	5670	630	6300	5670	630	6300	5670	630
statistical indicator (*m/s*)	**Max**	18.2	18.1	18.2	18.8	18.7	18.8	17.7	17.7	17.1	17.8	17.8	17.3	17.4	17.4	17
	Min	1	1	4	0.8	0.8	4.3	1.1	1.1	3.4	0.8	0.8	3.3	0.7	0.7	3
	Mean	8.4072	8.2288	10.0124	8.4828	8.3367	9.7979	6.9727	6.7709	8.7887	7.5976	7.4153	9.2376	7.0672	6.9033	8.5422
	Std.	3.0729	3.028	3.0104	3.2895	3.2908	2.9731	2.9611	2.9081	2.8136	3.0919	3.0401	3.0729	2.9887	2.9428	2.9962

Note: The number of samples in all samples, training samples and verifying samples are represented by **All, Train, Verify** in turn. There are 6300 observations in each of the five data sets, and the ratio of the verifying sample to the training sample is 1:9. In this table, the calculation formula of **Mean** is $\mathbf{mean} = (1/N)\sum_{i=1}^{N} \boldsymbol{Y}_i$, N expresses the sample capacity. And the calculation formula of **Std.** is $\mathbf{std.} = (1/(N-1))\sum_{i=1}^{N} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})^2$.

4.2. Evaluation criteria of forecast results

To quantify and analyze the forecast performance of the designed system, it is indispensable to use some evaluation indices to test the precision of different model predictions. In this research, four indicators are used as evaluation indicators, namely the mean square error, the root mean square error, the mean absolute percent error and the mean absolute error. In particular, the smaller the value of the above assessment metrics, the more accurate the forecast. And the mathematical formulae of main metrics are displayed in **Table 5**.

Table 5
Details of four assessment indicators.

Indicators	Definition	Equation of indicators
MSE	Mean square of error	$\boldsymbol{MSE} = (1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
RMSE	Root mean square error	$\boldsymbol{RMSE} = (1/N)\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
MAPE	Mean absolute percent error（%）	$\boldsymbol{MAPE} = (1/N)\sum_{i=1}^{N}(\lvert y_i - \hat{y}_i \rvert / y_i) \times 100\%$
MAE	Mean absolute error	$\boldsymbol{MAE} = (1/N)\sum_{i=1}^{N}\lvert y_i - \hat{y}_i \rvert$

Note: In **Table 5**, N represents the sample size, y_i represents the actual result of ith sample and $\hat{y}_i$ is the forecast result of ith sample.

4.3. Testing strategy

In this part, through Diebold-Mariano (**DM**) test, Forecasting effectiveness (**FE**) test, Model validity and stability test**,** and Sensitivity analysis, from the side of statistics, the practicality of the combined model is further verified[58].

4.3.1. Diebold-Mariano test

A hypothesis test such as the **DM** test is used to judge whether the prior hypothesis is right by computing the test statistics. In this research, for the prediction results of two models, we use this method to validate whether there is a significant difference, and the computational procedure of this test is shown below.

The error between the forecasting result and the true result is described as below.

$$\begin{cases} e_n^1 = y_n - \hat{y}_n^1 \\ e_n^2 = y_n - \hat{y}_n^2 \end{cases} \tag{9}$$

In Eq. (9), y_n represents the actual value of the sequence, $\hat{y}_n$ represents the forecasting result of the sequence and n is sequence size.

Loss function $F(e_n^i), i=1,2$ is defined as an index of predictive precision. For define the loss function, there are generally two ways, the calculation formula of absolute deviation error loss is $F(e_n^i)=\mid e_n^i \mid$, and the squared error loss is calculated by using $F(e_n^i)=(e_n^i)^2$.

Therefore, we can calculate the **DM** statistic as follows.

$$\mathbf{DM}=\frac{\frac{1}{T}\sum_{n=1}^{T}\left(F(e_n^1)-F(e_n^2)\right)}{\sqrt{S^2/T}} \tag{10}$$

In Eq. (10), the variance estimation of $F(e_n^1)-F(e_n^2)$ is represented by S^2.

In this test, we have established the priori hypothesis about testing. The null hypothesis is $\mathrm{H}_0: F(e_n^1)=F(e_n^2)$,and the alternative hypothesis is $\mathrm{H}_1: F(e_n^1)\neq F(e_n^2)$.

The null hypothesis indicates that between the forecasts of two systems have no significant differences, which means that the two models' predictive performance is identical. On the contrary, the alternative hypothesis indicates that the two models' forecast is significantly different. The **DM** statistics abide the standard normal distribution, and α represents the significance level. Therefore, $Z_{\alpha/2}$ represents the critical value. If $\mid \mathbf{DM} \mid > \mid Z_{\alpha/2} \mid$, then there is enough evidence that the two models are significant difference. And if $\mid \mathbf{DM} \mid < \mid Z_{\alpha/2} \mid$, the result is opposite.

4.3.2 Forecasting effectiveness (FE) test

In this paper, the mean squared deviation of the prediction accuracy and sum of squares of forecasting error are chosen as the metrics to measure the forecasting effectiveness. And the specific detail of this index is as below.

$\tilde{\boldsymbol{A}}_n$ is the prediction accuracy, and it can be calculated by using Eq. (11).

$$\tilde{\boldsymbol{A}}_n = 1 - \mid \xi_n \mid \tag{11}$$

$$\boldsymbol{\xi}_n=\begin{cases} -1, & (y_n-\hat{y}_n)/y_n < -1 \\ (y_n-\hat{y}_n)/y_n, & -1 \le (y_n-\hat{y}_n)/y_n < 1 \\ 1, & (y_n-\hat{y}_n)/y_n > 1 \end{cases} \tag{12}$$

In Eq. (12), $\boldsymbol{\xi}_n$ expresses the forecasting error of the *nth* predictive system. And the k-order **FE** ingredient is defined as $m_k=\sum_n Q_n \tilde{\boldsymbol{A}}_n^k$, $\sum_n Q_n = 1$. Where Q_n represents the discrete probability distribution at a certain time. Furthermore, **H** expresses continuous function of the **FE** ingredient, and $\mathbf{H}(m^1, m^2, \cdots, m^k)$ is named the k-order **FE**.

While there is a unary continuous function, namely $\mathbf{H}(a) = a$, the 1-order **FE** is represented as $\mathbf{H}(m^1) = m^1$.

4.3.3. Model validity and stability test

The effectuality and the stability of the model are reflected by the improvement rate and the performance difference of predictive indicators. In this research, the index selected is **MAPE**, and we calculate the index improvement rate by using Eq. (13).

$$\boldsymbol{P}_{\boldsymbol{MAPE}} = \left| \frac{MAPE_{compared} - MAPE_{proposed}}{MAPE_{compared}} \right| \times 100\% \tag{13}$$

Here, $MAPE_{compared}$ represents the **MAPE** value of the system for comparison, and $MAPE_{proposed}$ is the **MAPE** value of the designed predictive system.

4.3.4 Sensitivity analysis

In order to research the sensitivity of system state and output to changes in primary parameters of the model, we conduct the sensitivity analysis by changing a parameter of the designed system at a time, while keeping the remaining parameters fixed. And in this analysis, the standard deviations of **MAPE**, **MAE**, **RMSE** and **MSE** are selected to estimate the sensitivity.

When the parameter changes, the degree of the difference of error indicator $\boldsymbol{G}$ can be indicated by

$$\boldsymbol{S}(\boldsymbol{G}) = \frac{1}{\boldsymbol{m}} \sum_{i=1}^{\boldsymbol{m}} (\boldsymbol{G}_i - \bar{\boldsymbol{G}})^2 \tag{14}$$

where $\boldsymbol{G}_i$ expresses the *ith* forecasting time on the indicator, $\bar{\boldsymbol{G}}$ indicates the mean of all testing time, and $\boldsymbol{m}$ is the number of tests. It is clear that the sensitivity will increase as the evaluation indicators enhance.

4.4. Experiment and result analysis

4.4.1. Experiment 1: Comparison with other ordinary systems on the basis of VMD

In Experiment 1, to verify the prediction performance of the designed system, the designed predictive system is compared to the **VMD** ($\boldsymbol{K}_d = 7$, $\boldsymbol{\alpha}_d = 2000$)-based model, namely **VMD-ELM** ($\boldsymbol{m}_i = 8, \boldsymbol{m}_h = 5, \boldsymbol{m}_o = 1$), **VMD-ELMAN** ($\boldsymbol{m}_i = 8, \boldsymbol{m}_h = 5, \boldsymbol{m}_o = 1$), **VMD-LS-SVM** ($\boldsymbol{c}_l = 2, \boldsymbol{g}_l = 1$), **VMD-BPNN** ($\boldsymbol{m}_i = 8, \boldsymbol{m}_h = 5, \boldsymbol{m}_o = 1$) and **VMD-GRNN** ($\boldsymbol{d}_h = 0.1$) systems. The

specific values of each evaluation indicator of Site 1,2 and 3 are shown in **Table 6**, and the detailed description is described below.

(a) In this experiment, for Site 1, when the parameters of **MODA** are $A_s = 500$, $I_n = 200$ and $N_d = 40$, we can observe that the designed system gets the optimal predictive performance for the wind power forecast. With the values of assessment metrics are $MAPE^1_{moda} = 2.954\%$, $MAE^1_{moda} = 0.258$, $MSE^1_{moda} = 0.110$ and $RMSE^1_{moda} = 0.332$. Furthermore, the other single systems based on **VMD** ($K_d = 7$, $\alpha_d = 2000$) technique, from lowest to highest according to the predicted mean absolute percent error are **VMD-BP**, **VMD-ELM**, **VMD-LS-SVM**, **VMD-GRNN**, and **VMD-ELMAN** with the value matrix is $\overrightarrow{MAPE} = [2.772\%, 2.866\%, 3.066\%, 3.431\%, 4.833\%]$. This experimental result shows that for Site 1, the forecasting precision and effect of the designed system have a certain improvement compared with other **VMD**-based models.

(b) For Site 2, under the four error criteria, the forecasting performance of the **VMD** ($\alpha_d = 2000$, $K_d = 7$)-**MODA** ($A_s = 500$, $I_n = 200$, $N_d = 40$) system is optimal, and the **MAPE** value of this model is $MAPE^2_{moda} = 2.049\%$. And the values of other indicators are $MPE^2_{moda} = 0.202$, $MSE^2_{moda} = 0.084$ and $RMSE^2_{moda} = 0.289$ respectively. And among the other ordinary systems, it can be drawn that the **VMD-BPNN** ($m_o = 1$, $m_h = 5$, $m_i = 8$) system has the lower assessment metrics value, followed by **VMD-ELM**, **VMD-LS-SVM**, **VMD-GRNN**, and the predictive effect of **VMD-ELMAN** system is the worst. Take the value of **MAE** as an example, the metrics values of the five **VMD**-based systems are stable between 0.21 and 0.59, and the minimum gap with the designed model is about 8%, this gap is relatively notable. In addition, the forecast results of Site 2 and Site 3 are shown in **Fig. 4**.

(c) For Site 3 to Site 5, after processing the raw sequence with the data preprocessing method proposed, it can be seen that the predictive system on the basis of **VMD-MODA** still obtains the lowest values of assessment indicators, and the value matrix of **RMSE** is $\overrightarrow{RMSE}_{moda} = [0.280, 0.332, 0.378]$. Consequently, it can be drawn the conclusion that the designed predictive model is greater than the ordinary single model.

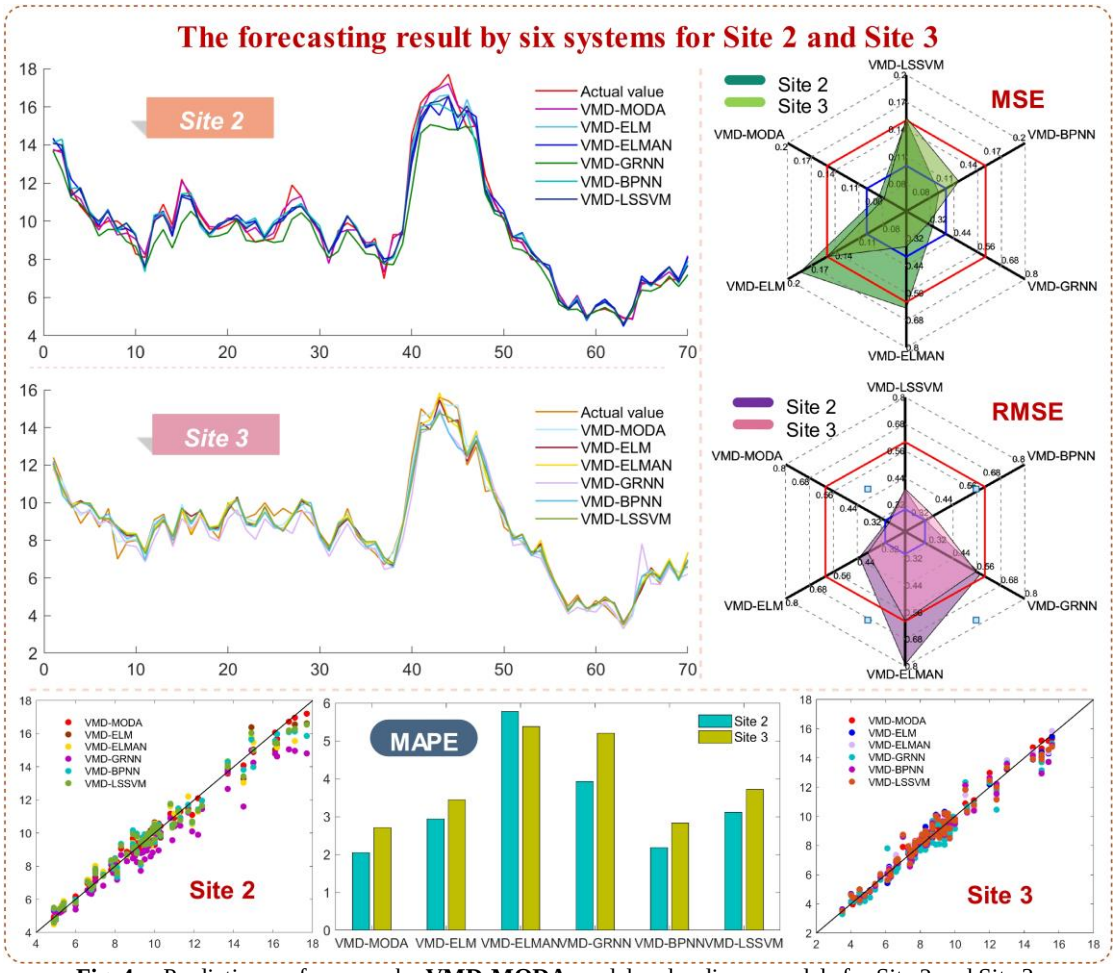


Fig. 4. Prediction performance by **VMD-MODA** model and ordinary models for Site 2 and Site 3.

Remark 1. According to Experiment 1, it can be concluded that the predictive system designed in this study has the smallest forecasting error and the best overall effect. The experiment results also demonstrate that the designed system is better than the ordinary system on the basis of **VMD** ($\alpha_d = 2000$, $K_d = 7$) technique. In addition, it is worth noting that **VMD-BPNN** ($m_h = 5$) and **VMD-ELM** ($m_h = 5$) have certain competitiveness in short-term wind power forecasting.

Table 6
Comparisons between the designed model with other **VMD**-based ordinary models.

Models	Site 1				Site 2				Site 3			
	MAPE (%)	**MAE**	**MSE**	**RMSE**	**MAPE (%)**	**MAE**	**MSE**	**RMSE**	**MAPE (%)**	**MAE**	**MSE**	**RMSE**
VMD-LS-SVM	3.066	0.311	0.161	0.401	3.117	0.301	0.150	0.388	3.715	0.310	0.153	0.391
VMD-ELM	2.866	0.294	0.159	0.399	2.945	0.298	0.184	0.429	3.444	0.295	0.150	0.388
VMD-GRNN	3.431	0.341	0.200	0.447	3.934	0.396	0.334	0.578	5.206	0.424	0.309	0.556
VMD-ELMAN	4.833	0.508	0.497	0.705	5.777	0.585	0.627	0.792	5.383	0.456	0.356	0.596
VMD-BPNN	2.772	0.281	0.130	0.361	2.176	0.218	0.091	0.302	2.835	0.243	0.116	0.341
VMD-MODA	**2.594**	**0.258**	**0.110**	**0.332**	**2.049**	**0.202**	**0.084**	**0.289**	**2.707**	**0.223**	**0.078**	**0.280**

Note: In this table, **VMD-MODA** represents our proposed forecasting model. And through the three datasets, we can notice that different predictive models get different results, which means that there is different between the comparison models. Moreover, as four assessment metrics, **MAPE** is calculated by using $\boldsymbol{MAPE} = (1/N)\sum_{i=1}^{N}(|y_i - \hat{y}_i|/y_i)\times 100\%$, **MAE** is estimated by using $\boldsymbol{MAE} = (1/N)\sum_{i=1}^{N}|y_i - \hat{y}_i|$, we computed **MSE** by using $\boldsymbol{MSE} = (1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, and the calculation formula of **RMSE** is $\boldsymbol{RMSE} = (1/N)\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$.

4.4.2. Experiment 2: Comparison with other ordinary forecasting systems

In Experiment 2, the designed predictive model is compared to five classic predictive models, **ELM**, **ELMAN**, **BP**, **LS-SVM** and **GRNN**, to further validate the precision of the designed predictive system. Moreover, the specific values of the evaluation indicators of Site 3, 4 and 5 are shown in the **Table 7**, and the detailed description are given below.

(a) For Site 1, the predictive effect of **VMD** ($K_d = 7$, $\alpha_d = 2000$)-**MODA** ($A_s = 500$, $I_n = 200$, $N_d = 40$) model is the best among the comparison classic predictive model. And the results of assessment indexes are $MAE^1_{moda} = 0.258$, $RMSE^1_{moda} = 0.332$, $MSE^1_{moda} = 0.110$ and $MAPE^1_{moda} = 2.594\%$ respectively. By contract, the other ordinary models get the higher evaluation metrics values, for example, the specific **MSE** values are $MSE^1_{ls\text{-}svm} = 0.331$, $MSE^1_{elm} = 0.394$, $MSE^1_{grnn} = 0.396$, $MSE^1_{elman} = 0.391$ and $MSE^1_{bpnn} = 0.370$ in turn. The outcome can be obtained that for site 1, comparing with other five ordinary single models, the **VMD-MODA** based model is the satisfactory model in the field of forecasting precision and effect.

(b) For Site 2, in this experiment, it can be observed that the **VMD-MODA** system has the lowest metrics error of forecasting, as proven by four indicators values $MAPE^2_{moda} = 2.049\%$, $MAE^2_{moda} = 0.202$, $MSE^2_{moda} = 0.084$ and $RMSE^2_{moda} = 0.289$ respectively. Furthermore, when the parameter is $m_i = 8$, according the specific evaluation indicators (**MAPE**), the classic single predictive systems have the higher error, with the values are $MAPE^2_{ls\text{-}svm} = 4.280\%$, $MAPE^2_{elm} = 4.418\%$, $MAPE^2_{grnn} = 6.089\%$, $MAPE^2_{elman} = 5.786\%$ and $MAPE^2_{bpnn} = 4.495\%$ in turn. And for the rest three metrics, there are apparent differences between the designed system and other ordinary systems.

(c) For Site 3, 4 and 5, the metrics values of the designed system are generally lower than comparison predictive systems, with the values of **MAE** are $MAE^3_{moda} = 0.223$, $MAE^4_{moda} = 0.269$ and $MAE^5_{moda} = 0.292$. By contrast, for the other five systems, the **MAE** values are main between 0.4 and 0.8, and take the **BPNN** ($m_o = 1$, $m_h = 5$, $m_i = 8$) system as an example, the value matrix of error metrics is $\overrightarrow{MAE}_{bpnn} = [0.463, 0.595, 0.520]$. Moreover, the forecast results of Site 3, 4, and 5 are shown in **Fig. 5**.

Remark 2. According to Experiment 2, the result expresses that the values of evaluation metrics gathered from the **VMD-MODA** model are lower than collected from other ordinary model, it can be drawn the conclusion that the excellence of the designed model is more clearly, and the forecasting efficiency of it is significantly greater than other single predictive models, in short-term wind speed forecasting.

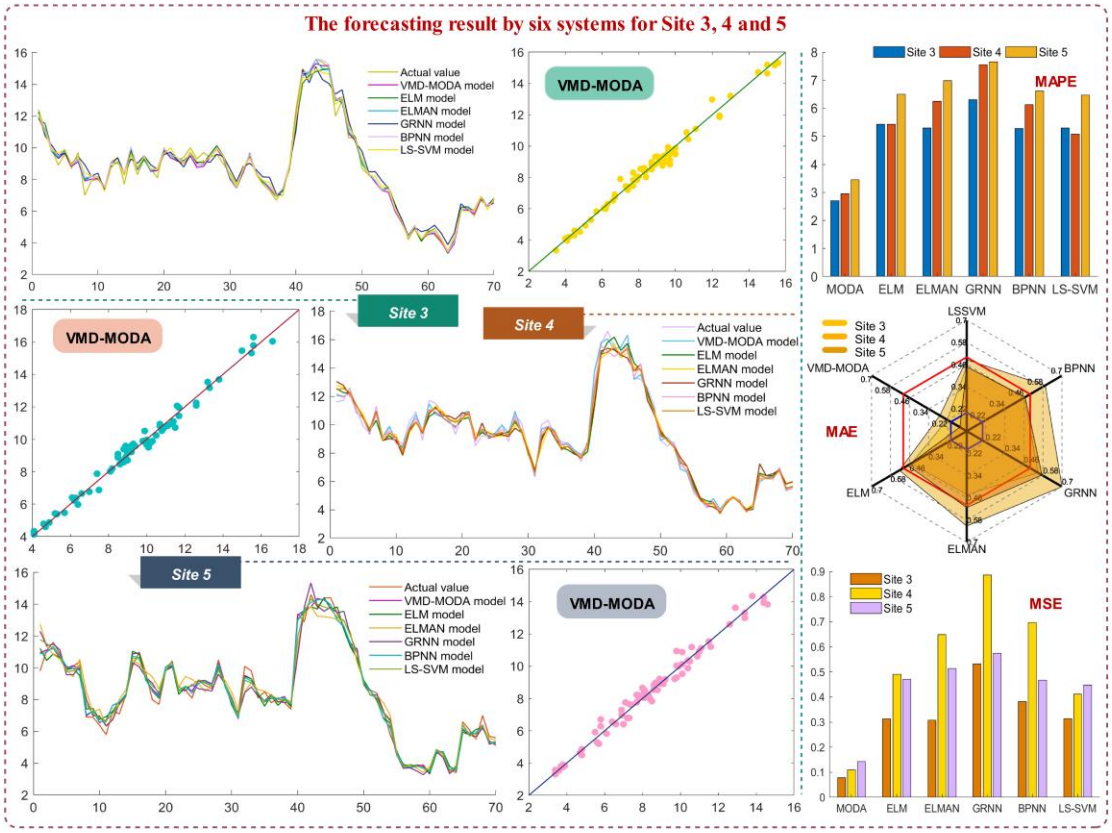


Fig. 5. The forecasting results by six systems for Site 3, 4 and 5.

4.4.3. Experiment 3: Comparison with hybrid systems on the basis of different data denoising techniques

In this part, we compare the predictive efficiency of the **MODA**-based combine system with different data denoising strategies, namely **VMD**, **EEMD** and **WD**. The assessment indicators are consistent with the above, which are still **MAPE**, **MAE**, **MSE** and **RMSE**. Furthermore, **Table 8** display the comparison results, and we can draw the conclusion that as below.

Table 7
Comparisons between the designed system and ordinary systems in site 3, 4 and 5.

Models	Site 3				Site 4				Site 5			
	MAPE (%)	**MAE**	**MSE**	**RMSE**	**MAPE (%)**	**MAE**	**MSE**	**RMSE**	**MAPE (%)**	**MAE**	**MSE**	**RMSE**
ELM	5.424	0.449	0.312	0.559	5.423	0.524	0.491	0.700	6.499	0.518	0.471	0.686
GRNN	6.314	0.550	0.531	0.729	7.552	0.711	0.886	0.941	7.659	0.566	0.574	0.758
ELMAN	5.295	0.443	0.308	0.555	6.247	0.613	0.649	0.806	6.988	0.544	0.513	0.716
BPNN	5.279	0.463	0.382	0.618	6.138	0.595	0.696	0.834	6.622	0.520	0.468	0.684
LS-SVM	5.291	0.449	0.314	0.561	5.074	0.486	0.411	0.641	6.468	0.510	0.448	0.669
VMD-MODA	**2.707**	**0.223**	**0.078**	**0.280**	**2.952**	**0.269**	**0.110**	**0.332**	**3.448**	**0.292**	**0.143**	**0.378**

Note: For three datasets, different predictive models get different results, which express that the prediction effect of the patterns are dissimilar. In this Table, **MAPE** is calculated by using $\boldsymbol{MAPE} = (1/N)\sum_{i=1}^{N}(|y_i - \hat{y}_i|/y_i)\times 100\%$, the calculation formula of **MAE** is $\boldsymbol{MAE} = (1/N)\sum_{i=1}^{N}|y_i - \hat{y}_i|$, we computed **MSE** by using $\boldsymbol{MSE} = (1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, and the calculation formula of **RMSE** is $\boldsymbol{RMSE} = (1/N)\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$.

(a) For Site 1, we can observe that the predictive model on the basis of **VMD** ($\alpha_d = 2000$, $K_d = 7$)-**MODA** obtains the smallest assessment indicators value. And the results of each assessment indicator are $MAE^1_{moda} = 0.258$, $MAPE^1_{moda} = 2.954\%$, $RMSE^1_{moda} = 0.332$ and $MSE^1_{moda} = 0.110$ respectively. Among the other two data denoising techniques in comparison, the combined system in the light of **WD** ($D_l = 5$) strategy has a lower evaluation values, for instance, the value of **MSE** is $MSE^1_{c\text{-}wd} = 0.307$, and the value of metrics predicted by the **EEMD** ($N_{std} = 0.1$, $N_e = 100$)-based combined system is $MSE^1_{c\text{-}eemd} = 0.318$.

(b) For Site 2, according to **Table 8**, our designed system still archives better predictive performance than the combined system on the basis of **WD** ($D_l = 5$) and **EEMD** ($N_{std} = 0.1$, $N_e = 100$) techniques, which the four assessment indicators values are $RMSE^2_{moda} = 0.289$, $MSE^2_{moda} = 0.084$, $MAE^2_{moda} = 0.202$, $MAPE^2_{moda} = 2.049\%$ in turn. Furthermore, opposite to Site 1, among other predictive systems, the forecasting effect of the **EEMD**-based combined system is slightly better than **WD**-based combined system, with the metrics values are $MAPE^2_{c\text{-}eemd} = 4.787\%$ and $MAPE^2_{c\text{-}wd} = 4.967\%$.

Table 8
Comparison the combined forecasting system on the basis of different data denoising methods.

Period	Models	Metrics			
		MAPE (%)	**MAE**	**MSE**	**RMSE**
Site 1	**C-EEMD**	4.298	0.424	0.318	0.564
	C-WD	3.725	0.359	0.307	0.455
	VMD-MODA	**2.594**	**0.258**	**0.110**	**0.332**
Site 2	**C-EEMD**	4.787	0.477	0.436	0.660
	C-WD	4.967	0.441	0.307	0.554
	VMD-MODA	**2.049**	**0.202**	**0.084**	**0.289**
Site 3	**C-EEMD**	5.018	0.430	0.311	0.558
	C-WD	4.86	0.383	0.248	0.498
	VMD-MODA	**2.707**	**0.223**	**0.078**	**0.280**
Site 4	**C-EEMD**	4.939	0.468	0.355	0.596
	C-WD	4.843	0.412	0.292	0.541
	VMD-MODA	**2.952**	**0.269**	**0.110**	**0.332**
Site 5	**C-EEMD**	6.175	0.501	0.428	0.654
	C-WD	5.594	0.416	0.300	0.547
	VMD-MODA	**3.448**	**0.292**	**0.143**	**0.378**

Note: C-EEMD, **C-WD** and **VMD-MODA** indicate the predictive system on the basis of **EEMD** technique, the predictive model based on **WD** technique and the combined model we proposed respectively. And for five datasets, different predictive systems get different results, which indicates that the predictive effect of the system are different. The mathematical formula of evaluation metrics can be defined as: $MAPE = (1/N)\sum_{i=1}^{N}(|y_i - \hat{y}_i|/y_i) \times 100\%$, $MAE = (1/N)\sum_{i=1}^{N}|y_i - \hat{y}_i|$, $MSE = (1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, and $RMSE = (1/N)\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$.

(c) For Site 3, 4 and 5, in general, the forecasting efficiency of our designed system is optimal, as can be observed from obviously smaller evaluation indicator values, take the **RMSE** as an example, the value matrix is $\overrightarrow{\overline{RMSE}}_{moda} = [0.280, 0.332, 0.378]$. And the metrics values of **EEMD**-based, and **WD**-based systems are $\overrightarrow{\overline{RMSE}}_{c\text{-}eemd} = [0.558, 0.596, 0.654]$ and $\overrightarrow{\overline{RMSE}}_{c\text{-}wd} = [0.498, 0.541, 0.547]$. It is noteworthy that the forecast effect of Site 5 in each system is worse than that of other sites, which indicates that the randomness of the data of this site fluctuates greatly, and it also make known that the advantage of the designed predictive system on the basis of **VMD** technique.

Remark 3. In this experiment, the prediction effect and accuracy of the designed predictive model is greater than those of the **WD** ($D_l = 5$)-based combined system and the **EEMD** ($N_{std} = 0.1$, $N_e = 100$)-based combined system. The experiment result not only illustrates the applicability of the designed model, but also displays the rationality of using **VMD** ($\alpha_d = 2000$, $K_d = 7$) strategy.

4.4.4. Experiment 4: Compared to other popular forecasting systems

To preferably explore the forecast constancy of the designed model, for relatively popular predictive model are chosen to compare to the designed combined system, namely the predictive system based on multi-objective grey wolf algorithm (**MOGWO**), multi-objective ant lion optimization (**MOALO**), convolutional neural network (**CNN**) and long short-term memory network (**LSTM**). The comparison results are displayed in **Table 9**, and **Fig. 6** depicts the prediction conclusion in Experiment 3 and 4. Moreover, the specific description is shown below.

(a) Comparing with the combined systems in the light of **MOALO**, and **MOGWO**, when the parameter are ($A_s = 500$, $I_n = 200$, $N_d = 40$), it can be concluded from **Table 9** that for the data from Site 1 to 5, the forecast effect of the proposed predictive system is superior to the other two systems. About this proposed system, the values of **MAPE** for each site are $\overrightarrow{\overline{MAPE}}_{moda} = [2.954\%, 2.049\%, 2.707\%, 2.952\%, 3.448\%]$. For the predictive system on the basis of **MOGWO**, the **MAPE** values are $\overrightarrow{\overline{MAPE}}_{mogwo} = [3.791\%, 3.395\%, 3.212\%, 3.772\%, 4.634\%]$, and the assessment indicator values of hybrid system based on **MOALO** strategy are,

$\overline{MAPE}_{moalo} = [3.695\%, 5.007\%, 4.356\%, 4.194\%, 5.195\%]$. Compare those above sequences and other evaluation metrics values in **Table 9**, it can be drawn the conclusion that **MOGWO**-based system is more practical than **MOALO**-based system in this test.

Table 9
Comparison results of Experiment 4.

Period	Metrics	Models				
		MOGWO	**MOALO**	**LSTM**	**CNN**	**VMD-MODA**
Site 1	**MAPE (%)**	3.791	3.695	3.030	3.447	**2.594**
	MAE	0.333	0.350	0.308	0.368	**0.258**
	MSE	0.172	0.190	0.158	0.268	**0.110**
	RMSE	0.414	0.436	0.397	0.518	**0.332**
Site 2	**MAPE (%)**	3.395	5.007	2.777	2.710	**2.049**
	MAE	0.317	0.509	0.264	0.248	**0.202**
	MSE	0.165	0.425	0.120	0.113	**0.084**
	RMSE	0.407	0.652	0.347	0.335	**0.289**
Site 3	**MAPE (%)**	3.212	4.356	3.457	4.3965	**2.707**
	MAE	0.272	0.360	0.288	0.4015	**0.223**
	MSE	0.131	0.216	0.125	0.2841	**0.078**
	RMSE	0.361	0.465	0.354	0.533	**0.280**
Site 4	**MAPE (%)**	3.772	4.194	3.772	4.132	**2.952**
	MAE	0.360	0.402	0.360	0.377	**0.269**
	MSE	0.215	0.307	0.213	0.234	**0.110**
	RMSE	0.464	0.554	0.461	0.484	**0.332**
Site 5	**MAPE (%)**	4.634	5.195	4.636	4.521	**3.448**
	MAE	0.390	0.400	0.3772	0.378	**0.292**
	MSE	0.241	0.258	0.231	0.238	**0.143**
	RMSE	0.490	0.508	0.481	0.488	**0.378**

Note: This table displays the comparison results between the designed model with other popular predictive models, including the combined model based multi-objective optimization algorithms (**MOALO** and **MOGWO**), **LSTM** model and **CNN** model. In this table, **VMD-MODA** indicates the predictive system which we designed. The calculation formula of evaluation metrics can be defined as: $MAPE = (1/N)\sum_{i=1}^{N}(|y_i - \hat{y}_i|/y_i)\times 100\%$, $MAE = (1/N)\sum_{i=1}^{N}|y_i - \hat{y}_i|$, $MSE = (1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ and $RMSE = (1/N)\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$.

(b) For the comparison between the **VMD-MODA** system with **LSTM** and **CNN** system, we can observe that the designed system obtained the best accuracy of the prediction. And take Site 3 as an example, the main evaluation metrics values are $MAPE^{3}_{moda} = 2.707\%$, $MAE^{3}_{moda} = 0.223$, $MSE^{3}_{moda} = 0.078$, $RMSE^{3}_{moda} = 0.280$ respectively. Among the other two models, the **LSTM** model has the smaller metrics values, and the specific values are $MAPE^{3}_{lstm} = 3.457\%$, $MAE^{3}_{lstm} = 0.288$, $MSE^{3}_{lstm} = 0.125$, $RMSE^{3}_{lstm} = 0.354$. But at Site 2 and Site 5, the forecasting performances of **LSTM** ($e_\tau = 250$) model are lower than it of **CNN** ($K_s = 40$, $K_n = 3$, $m_h = [384, 384]$) model, it can be observed that there is relatively no difference between those two models.

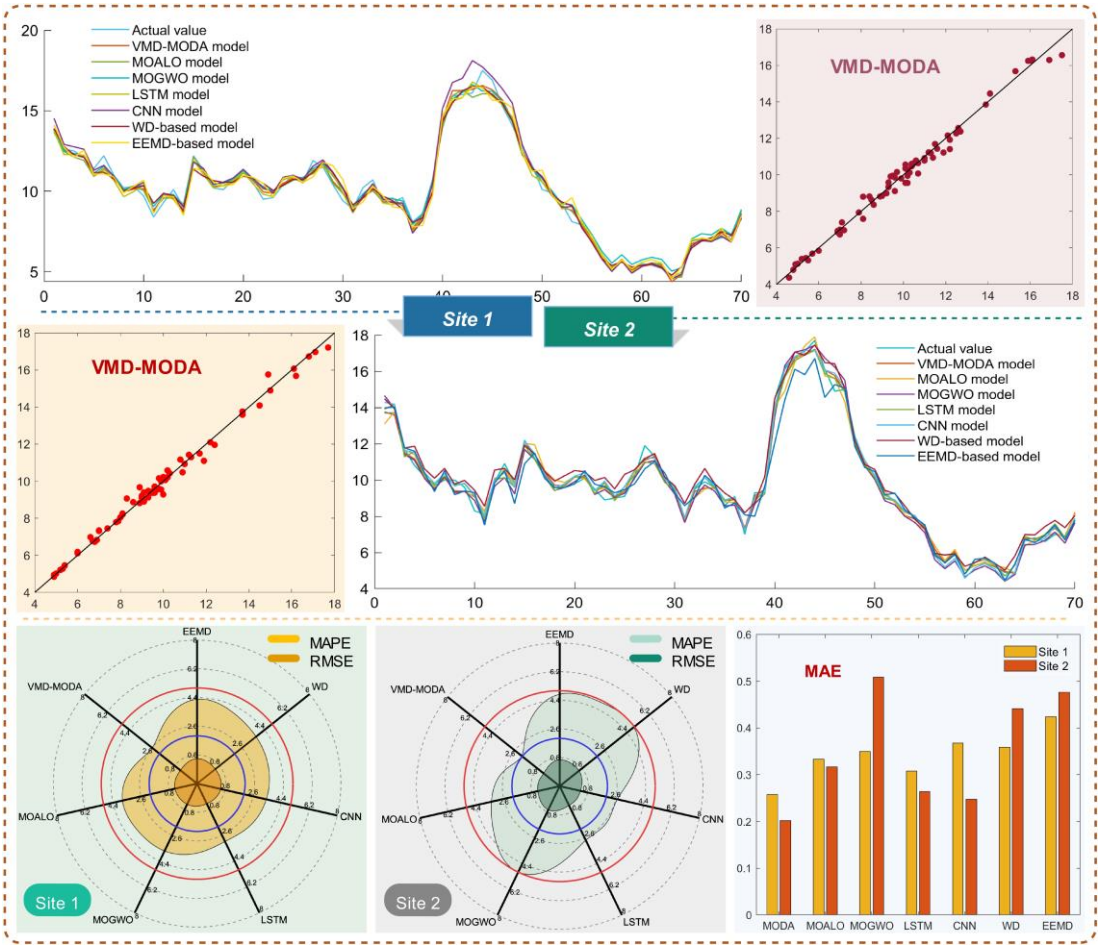


Fig. 6. The predicting outcomes of combined models for Site 1 and Site 2.

5. Discussion

In this section, four methods are selected to further analyze the comparison results of the four experiments, which includes the **DM** test, **FE** test, Improvement rate of the indicators, Sensitivity analysis. The result and specific description of each part are shown below.

5.1. Diebold–Mariano test and Forecasting effectiveness test

In order to examine the predictive performance of the designed predictive system further. In this part, the Diebold–Mariano test on account of the squared error loss function, and predictive validity test are chosen to compare the forecast differences between the designed combined model and comparison model. The results of **DM** test are listed in **Table 10**, and the specific description is as follows.

(a) In this test, the results of **DM** test between the designed predictive model on the basis of **WD** strategy and the predictive model in the light of **EEMD** strategy are obviously higher than 1.96, which expresses that between the two above models and the designed system, there are

significant differences. And it can also be seen that the superiority of choosing **VMD** as the data preprocessing method.

(b) Comparing with the ordinary system, it can be observed that the **DM** test values of all systems are much higher than 2.58, which is the critical value of the $\alpha = 0.01$ significance level. And for five sites, the **DM** maximum value of the **LS-SVM**, **ELMAN**, **BPNN**, the **ELM** model and the **GRNN** model are $\overline{\overline{\mathbf{DM}}}_{ls\text{-}svm} = 5.512$, $\overline{\overline{\mathbf{DM}}}_{elman} = 5.642$, $\overline{\overline{\mathbf{DM}}}_{bpnn} = 4.990$, $\overline{\overline{\mathbf{DM}}}_{elm} = 5.307$ and $\overline{\overline{\mathbf{DM}}}_{grnn} = 4.974$ respectively, which indicate that our designed system is significantly different from ordinary single system, in this comparison. we can also conclude that the designed predictive system is greater than the general predictive system.

(c) Comparing with the popular forecasting systems, including the combined predictive model on the basis of **MOALO** strategy, and **MOGWO** technique, the **LSTM** model and **CNN** model, most of **DM** values are higher than 1.96, and all value of **DM** is greater than 1.645, which means that at the significance level of 10%, H_0 could be rejected. Consequently, it can be considered that there are significant differences between the designed system and others.

Through the **DM** test, we can conclude that the designed predictive system not only has greater forecasting precision than others, but has significant differences in prediction precision level from other comparison models. It preferably validates the advantage of the designed predictive system in wind speed forecast.

Table 10
DM test of comparison models

Models	Period				
	Site 1	**Site 2**	**Site 3**	**Site 4**	**Site 5**
C-MOALO	3.051*	2.223	3.410*	2.191	3.291*
C-MOGWO	2.899*	3.044*	2.039	2.323	2.059
LSTM	2.690*	2.460	1.742	2.235	1.737
CNN	2.269	1.939	3.725*	3.084*	1.964
C-WD	4.253*	4.984*	3.926*	3.447*	2.921*
C-EEMD	4.293*	3.361*	4.132*	3.859*	3.418*
VMD-LS-SVM	2.620*	6.485*	2.206	2.521	1.993
VMD-ELMAN	3.488*	3.302*	1.695	3.164*	2.948*
VMD-BP	1.984	1.949	1.960	1.972	1.834
VMD-ELM	1.985	2.838*	2.662*	2.118	3.039*
VMD-GRNN	3.277*	2.748*	3.697*	3.040*	2.403
LS-SVM	5.192*	5.512*	4.303*	3.317*	3.268*
ELMAN	5.642*	4.873*	4.715*	3.839*	3.342*
BP	4.990*	4.119*	3.407*	3.032*	3.310*
ELM	4.675*	5.307*	4.291*	3.453*	3.263*
GRNN	4.974*	3.835*	4.249*	4.043*	3.515*

Note: This table displays the differences between the designed predictive model with other comparison models. **C-MOALO**, **C-MOGWO**, **C-WD** and **C-EEMD** represent the combined model based on **MOALO** strategy, and the **MOGWO** method, the predictive model in view of **WD** technique, and the **EEMD** technique.

And in this table, if the result is better than 2.58, That means H_0 could be rejected at the significance level of 1%, the symbol * indicates significance level $\alpha = 0.01$. Moreover, for the **DM** test, the null and the alternative hypotheses are expressed as below: $H_0 : F(e_n^1) = F(e_n^2)$, $H_1 : F(e_n^1) \neq F(e_n^2)$.

The next step is to evaluate the efficiency of the forecast to further validate the practicability of the designed predictive model. In this test, the effective result of the forecasting is closer to 1, the more accurate the prediction result is. And through the data in **Table 11**, it can be observed that our designed model obtains the highest value of assessment metrics, with the value matrix is $\vec{\mathbf{H}}_{moda} = [0.974, 0.980, 0.974, 0.971, 0.961]$. Consequently, we can conclude that the result of the designed combined model is greater than others, which also means that it is better than other models.

Table 11
Forecasting effectiveness of different systems.

Models	Period				
	Site 1	**Site 2**	**Site 3**	**Site 4**	**Site 5**
VMD-MODA	**0.974**	**0.980**	**0.974**	**0.971**	**0.961**
C-MOALO	0.963	0.975	0.956	0.958	0.948
C-MOGWO	0.962	0.966	0.968	0.962	0.954
LSTM	0.970	0.972	0.965	0.962	0.954
CNN	0.966	0.973	0.956	0.959	0.955
C-WD	0.963	0.95	0.951	0.952	0.944
C-EEMD	0.957	0.952	0.95	0.951	0.938
VMD-LS-SVM	0.969	0.891	0.963	0.961	0.952
VMD-ELMAN	0.952	0.942	0.973	0.949	0.935
VMD-BPNN	0.972	0.978	0.972	0.967	0.961
VMD-ELM	0.971	0.971	0.966	0.955	0.954
VMD-GRNN	0.966	0.961	0.948	0.949	0.941
LS-SVM	0.956	0.957	0.947	0.949	0.935
ELMAN	0.951	0.952	0.947	0.938	0.93
BP	0.953	0.955	0.947	0.939	0.934
ELM	0.952	0.956	0.946	0.946	0.935
GRNN	0.952	0.939	0.937	0.925	0.923

Note: C-MOALO, C-MOGWO, C-WD, C-EEMD, and **VMD-MODA** represent the combined model on the basis of **MOALO** strategy, and **MOGWO** method, the hybrid model in the light of **WD** technique, and the **EEMD** technique, and the combined model which we proposed respectively. And for this table, the value of test is closer to 1, the more effective the forecasting.

Remark 4. Based on the result of the **DM** test and Forecasting effectiveness test (see **Table 10** and **11**), the prediction performance of different systems is analyzed from the two sides of prediction error and prediction value. And we can see that the designed predictive system has higher predictive precision, lower error, and has significant differences compared with other models.

5.2. Model validity and stability test

To determine the prediction improvement efficiency of the designed model, we use **MAPE** value to calculate the improvement percentage of the designed model in this test, namely the

improvement rate, and the improvement rate is chosen to represent the forecasting improvement effectiveness of the combined system. Furthermore, four comparisons are performed with designed predictive system, the single systems based on **VMD** method, the traditional single predictive systems and other combined systems. **Table 12** is displaying the specific results, and the detailed data analysis is as follows.

(a) Comparing with other ordinary systems on the basis of **VMD** strategy, the forecast availability of the designed combined model is fairly improved. The average of improvement rate of predictive indicator is $\bar{\boldsymbol{P}}_{\boldsymbol{MAPE}} = 29.59\%$, the maximum and minimum of improvement rate of predictive indicator are $\overline{\overline{\boldsymbol{P}_{\boldsymbol{MAPE}}}} = 64.53\%$ and $\underline{\underline{\boldsymbol{P}_{\boldsymbol{MAPE}}}} = 4.50\%$ respectively.

(b) Comparing with traditional single predictive models, it can be observed that the designed system greatly improves the forecasting precision. Take Site 1 as an example, the maximum and minimum of $\boldsymbol{P}_{\boldsymbol{MAPE}}$ are $\overline{\overline{\boldsymbol{P}_{\boldsymbol{MAPE}}}} = 47.45\%$ and $\underline{\underline{\boldsymbol{P}_{\boldsymbol{MAPE}}}} = 41.13\%$ respectively, and the average of improvement rate of predictive indicator is $\bar{\boldsymbol{P}}_{\boldsymbol{MAPE}} = 44.95\%$. So, it can be clearly that our designed system has a satisfactory performance of wind speed forecasting.

(c) As contrasted with other predictive models based on different data denoising techniques, the proposed combined system has played a great improvement in forecasting precision. For example, comparing with the predictive model on the basis of **EEMD** method and the combined model based on **WD** strategy, the averages of improvement rate of predictive indicators are $\bar{\boldsymbol{P}}^{\boldsymbol{WD}}_{\boldsymbol{MAPE}} = 45.46\%$ and $\bar{\boldsymbol{P}}^{\boldsymbol{EEMD}}_{\boldsymbol{MAPE}} = 42.17\%$ in turn. And the minimum of $\boldsymbol{P}_{\boldsymbol{MAPE}}$ are $\underline{\underline{\boldsymbol{P}^{\boldsymbol{WD}}_{\boldsymbol{MAPE}}}} = 39.66\%$ and $\underline{\underline{\boldsymbol{P}^{\boldsymbol{EEMD}}_{\boldsymbol{MAPE}}}} = 30.38\%$ in turn. Consequently, the designed combined model can get better prediction effect.

(d) Compared with the combined model on the basis of **MOGWO** technique and the predictive model on the basis of **MOALO** strategy, the minimums of improvement rate of assessment indicators are $\underline{\underline{\boldsymbol{P}^{\boldsymbol{MOGWO}}_{\boldsymbol{MAPE}}}} = 15.73\%$ and $\underline{\underline{\boldsymbol{P}^{\boldsymbol{MOALO}}_{\boldsymbol{MAPE}}}} = 59.08\%$ in turn. Otherwise, as for the comparison between the designed predictive model with **LSTM** and **CNN** model, the improvement matrixes of evaluation index are $\overrightarrow{\boldsymbol{P}^{\boldsymbol{LSTM}}_{\boldsymbol{MAPE}}} = [14.39\%, 26.23\%, 21.70\%, 21.74\%, 25.60\%]$

and $\overrightarrow{P_{MAPE}^{CNN}} = [24.74\%, 24.38\%, 38.43\%, 28.56\%, 23.73\%]$ in turn, those values are all relatively significant. Consequently, we can conclude that the designed predictive system has a greater improvement than the comparison systems.

Table 12
The result of improvement rate of predictive indicator for each model.

Models	Period				
	Site 1	**Site 2**	**Site 3**	**Site 4**	**Site 5**
LSTM	14.39%	26.23%	21.70%	21.74%	25.63%
CNN	24.74%	24.38%	38.43%	28.56%	23.73%
ELM	45.63%	53.62%	50.09%	45.57%	46.95%
GRNN	46.01%	66.35%	57.13%	60.91%	54.99%
ELMAN	47.45%	64.59%	50.73%	52.74%	50.67%
BPNN	44.51%	54.42%	48.72%	51.91%	47.93%
LS-SVM	41.13%	52.12%	48.83%	41.82%	46.70%
C-WD	30.38%	58.75%	44.30%	39.05%	38.37%
C-EEMD	39.66%	57.19%	46.05%	40.23%	44.17%
C-MOALO	29.80%	59.08%	37.86%	29.61%	33.63%
C-MOGWO	31.57%	39.64%	15.73%	21.74%	25.60%
VMD-ELM	9.49%	30.42%	21.39%	34.80%	25.03%
VMD-GRNN	24.42%	47.91%	48.00%	42.24%	41.37%
VMD-ELMAN	46.34%	64.53%	49.71%	42.69%	46.64%
VMD-BPNN	6.44%	5.81%	4.50%	10.26%	7.69%
VMD-LS-SVM	15.42%	34.25%	27.13%	24.94%	28.46%

Note: In this table, **C-MOALO, C-MOGWO, C-WD** and **C-EEMD** represent the combined model based on **MOALO** strategy, and **MOGWO** method, the predictive model on the basis of **WD** technique and the combined model based on **EEMD** technique. For sixteen models, the improvement rate of predictive indicator is different of each other. It means that different model has different prediction. And in this table the improvement rate of the mean absolute percent error can be calculated by using $P_{MAPE} = |\frac{MAPE_{compared} - MAPE_{proposed}}{MAPE_{compared}}| \times 100\%$.

5.3. Sensitivity analysis

In this research, we mainly analyze the sensitivity of the designed predictive system, and there are three primary parameters in this system, namely the Iteration number, Archive size and Dragonfly number. And the matrix of Iteration number is set as $\vec{I}_n = [50, 100, 150, 200^{\bullet}, 250]$, the matrix of Dragonfly number is $\overrightarrow{N_d} = [50, 100, 150, 200^{\bullet}, 250]$, and matrix the Archive size is set as $\vec{A}_s = [100, 200, 300, 400, 500^{\bullet}]$. The symbol $\bullet$ expresses the satisfactory value of each parameter, which is determined by trial-and-error method. The assessment metrisc results are expressed in **Table 13**, and **Fig. 7** shows the comparison of three patterns at Site 5, and specific described is as follows.

Table 13
Sensitivity analysis results for the designed system.

Period	Parameters	S_{MAPE}	S_{MAE}	S_{MSE}	S_{RMSE}
Site 1	Iteration number	0.2806	0.0227	0.0153	0.0209
	Dragonfly number	0.1625	0.014	0.0084	0.0121
	Archive number	0.1385	0.0115	0.0068	0.0093
Site 2	Iteration number	0.6982	0.0573	0.0403	0.0563
	Dragonfly number	0.1737	0.0162	0.0119	0.0186
	Archive number	0.2481	0.0248	0.0169	0.0265
Site 3	Iteration number	0.2907	0.0284	0.0324	0.0453
	Dragonfly number	0.1832	0.0169	0.0108	0.0170
	Archive number	0.1033	0.0091	0.0096	0.0142
Site 4	Iteration number	0.3595	0.0236	0.0212	0.0286
	Dragonfly number	0.1078	0.007	0.0052	0.0077
	Archive number	0.2673	0.0177	0.0143	0.0201
Site 5	Iteration number	0.1029	0.0056	0.0062	0.0080
	Dragonfly number	0.0630	0.0048	0.0010	0.0020
	Archive number	0.1128	0.0107	0.0096	0.0126

Note: This table displays the sensitivity analysis results for the proposed system. Among the three main parameters, the Iteration number can be from 50 to 250 with a step size of 50, the Dragonfly number is set to 24, 40, 60, 80, 100, and the Archive number grow from 100 to 500 with the increase interval of 100. Moreover, in this table, the evaluation metrics can be computed by $S(G)=(1/n)\sum_{i=1}^{n}(G_i-\bar{G})^2$, where G_i indicates the *ith* forecasting time on the indicator, $\bar{G}$ express the mean of all testing time, and n is the number of testing times.

Based on the sensitivity analysis for the five sites, our combined model is most sensitive to the parameter of Iteration number. More specific, take Site 2 as an example, the evaluation metrics of this parameter are $S_{MAPE}=0.6982$, $S_{MAE}=0.0573$, $S_{MSE}=0.0403$ and $S_{RMSE}=0.0563$ respectively. Moreover, among the three primary parameters for the combined model, the Dragonfly number obtained the smallest value of the evaluation metrics. It means that the combined model has the lowest sensitivity and the highest stability to the changes of this parameter. In a word, although the sensitivity of the combined is difference between three primary parameters, the values of evaluation metrics in **Table 13** are not very large. Thence, it can be drawn the conclusion that the transformation of the main parameters has a low degree of influence on the prediction results, and the combined model is stable of the change of primary parameters.

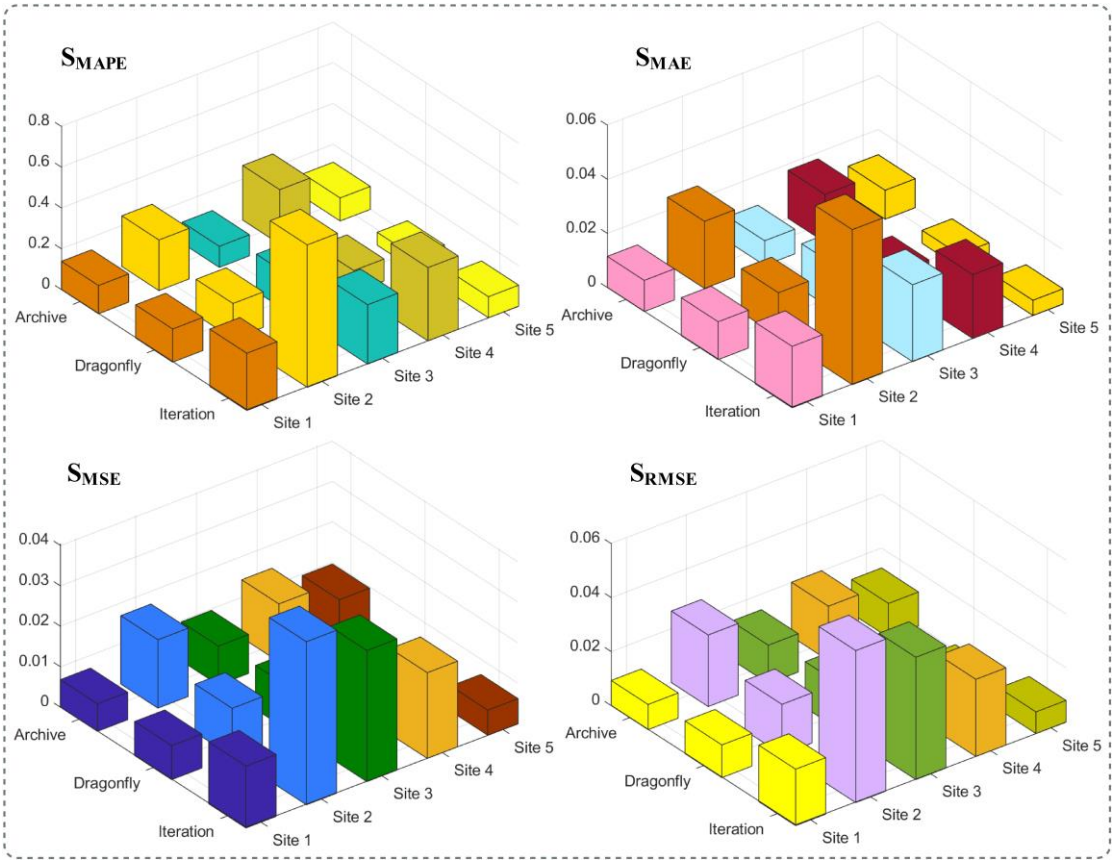


Fig. 7 Comparison results of three main parameters for the combined model

5.4 Future work

Since the combination proposed in this paper plays a role not be ignored in short-term wind speed forecast. It is recommended to apply an improved multi-objective optimization algorithm and more effective model combination as the future work of this article. Furthermore, in the study, the parameters of the data preprocessing method are obtained through trial and error. When the processed data is different, the parameters are possible not uniform. Therefore, the strategy of automatically optimizing parameters is also a part worthy of research.

6. Conclusions

Today, wind and solar energy are more cost-competitive than new coal or natural gas in two-thirds of the world. In the next ten years, building new wind and solar energy will be more cost-effective than operating existing coal or natural gas plants. Therefore, it is very essential and pressing to find an efficient and steady wind power prediction strategy. However, because the uncontrollability and variability of wind velocity, it is very laborious to predict velocity of wind with a general algorithm strategy. In this research, we combine the preponderance of variational mode decomposition, multi-objective dragonfly optimization algorithm and hybrid model, then a

new combined forecasting system is designed. Specifically, the variational mode decomposition is used for data decomposition and denoising, and the strong global search ability of the multi-objective dragonfly algorithm is chosen to certain the weights of the sub-systems. Moreover, the corresponding strategy is selected to optimizes the parameters of the sub-systems. In this research, for the five datasets from different wind farm, the final **MAPE** value matrix of the designed predictive system is $\overrightarrow{\overline{MAPE}}_{moda} = [2.954\%, 2.049\%, 2.707\%, 2.952\%, 3.448\%]$. In addition, this paper also proves the superiority and predictive stability of the combined system studied from the four aspects of Diebold-Mariano test, Forecasting effectiveness test, predictive index improvement rate and Sensitivity analysis. And verify that the designed system is optimal than other comparison systems.

On the whole, we got the following conclusions: 1) The proposed combined system has achieved great precision in short-term wind velocity forecast; 2) The combined effect of multi-objective dragonfly algorithm and variational mode decomposition technique is better than that of wavelet denoising strategy and ensemble empirical mode decomposition strategy; 3) 3) For the designed system, the promotion effect of the multi-objective dragonfly algorithm is superior to the promotion effect of the multi-objective ant lion optimization strategy, the multi-objective grey wolf optimization method and other popular predictive system.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 71601020, 71671029).

Reference

[1] Miao H, Dong D, Huang G, Hu K, Tian Q, Gong Y. Evaluation of Northern Hemisphere surface wind speed and wind power density in multiple reanalysis datasets. Energy 2020;200:117382. https://doi.org/10.1016/j.energy.2020.117382.

[2] Nazir MS, Ali N, Bilal M, Iqbal HMN. Potential environmental impacts of wind energy development: A global perspective. Curr Opin Environ Sci Heal 2020;13:85–90. https://doi.org/10.1016/j.coesh.2020.01.002.

[3] Shen X, Lyu S. Wind power development, government regulation structure, and vested interest groups: Analysis based on panel data of Province of China. Energy Policy 2019;128:487–94. https://doi.org/10.1016/j.enpol.2019.01.023.

[4] GWEC. Global Wind Report 2019. Glob Wind Energy Rep 2020:78.

[5] Xiang L, Li J, Hu A, Zhang Y. Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Convers Manag 2020;220:113098. https://doi.org/10.1016/j.enconman.2020.113098.

[6] Zhang J, Wei Y, Tan Z. An adaptive hybrid model for short term wind speed forecasting. Energy 2020;190. https://doi.org/10.1016/j.energy.2019.06.132.

[7] Liu H, Yang R, Duan Z. Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction. Energy Convers Manag 2020;217:112995. https://doi.org/10.1016/j.enconman.2020.112995.

[8] Wang J, Wang Y, Li Z, Li H, Yang H. A combined framework based on data preprocessing, neural networks and multi-tracker optimizer for wind speed prediction. Sustain Energy Technol Assessments 2020;40:100757. https://doi.org/10.1016/j.seta.2020.100757.

[9] Gangui Y, Yu L, Gang M, Yang C, Junhui L, Jigang L, et al. The Ultra-short Term Prediction of Wind Power Based on Chaotic Time Series. Energy Procedia 2012;17:1490–6. https://doi.org/10.1016/j.egypro.2012.02.271.

[10] Zhang Y, Pan G, Chen B, Han J, Zhao Y, Zhang C. Short-term wind speed prediction model based on GA-ANN improved by VMD. Renew Energy 2020;156:1373–88. https://doi.org/10.1016/j.renene.2019.12.047.

[11] Li C, Zhu Z, Yang H, Li R. An innovative hybrid system for wind speed forecasting based on fuzzy preprocessing scheme and multi-objective optimization. Energy 2019;174:1219–37. https://doi.org/10.1016/j.energy.2019.02.194.

[12] Jackson A, Turnbull B. Identification of particle-laden flow features from wavelet decomposition. Phys D Nonlinear Phenom 2017;361:12–27. https://doi.org/10.1016/j.physd.2017.09.009.

[13] Li LL, Chang YB, Tseng ML, Liu JQ, Lim MK. Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. J Clean Prod 2020;270:121817. https://doi.org/10.1016/j.jclepro.2020.121817.

[14] Gupta V, Mittal M. KNN and PCA classifier with Autoregressive modelling during different ECG signal interpretation. Procedia Comput Sci 2018;125:18–24. https://doi.org/10.1016/j.procs.2017.12.005.

[15] Amjady N, Abedinia O. Short term wind power prediction based on improved kriging interpolation, Empirical Mode Decomposition, and closed-loop forecasting engine. Sustain 2017;9. https://doi.org/10.3390/su9112104.

[16] Abedinia O, Lotfi M, Bagheri M, Sobhani B, Shafie-Khah M, Catalao JPS. Improved EMD-Based Complex Prediction Model for Wind Power Forecasting. IEEE Trans Sustain Energy 2020;11:2790–802. https://doi.org/10.1109/TSTE.2020.2976038.

[17] Wang S, Zhang N, Wu L, Wang Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew Energy 2016;94:629–36. https://doi.org/10.1016/j.renene.2016.03.103.

[18] Naik J, Dash S, Dash PK, Bisoi R. Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network. Renew Energy 2018;118:180–212. https://doi.org/10.1016/j.renene.2017.10.111.

[19] Li H, Liu T, Wu X, Chen Q. An optimized VMD method and its applications in bearing fault diagnosis. Meas J Int Meas Confed 2020;166:108185. https://doi.org/10.1016/j.measurement.2020.108185.

[20] De Felice M, Alessandri A, Ruti PM. Electricity demand forecasting over Italy: Potential benefits using numerical weather prediction models. Electr Power Syst Res 2013;104:71–9. https://doi.org/10.1016/j.epsr.2013.06.004.

[21] Song D, Yang J, Cai Z, Dong M, Su M, Wang Y. Wind estimation with a non-standard extended Kalman filter and its application on maximum power extraction for variable speed wind turbines. Appl Energy 2017;190:670–85. https://doi.org/10.1016/j.apenergy.2016.12.132.
[22] Hodge BM, Zeiler A, Brooks D, Blau G, Pekny J, Reklatis G. Improved Wind Power Forecasting with ARIMA Models. vol. 29. Elsevier B.V.; 2011. https://doi.org/10.1016/B978-0-444-54298-4.50136-7.
[23] Muyeen SM, Hasanien HM, Al-Durra A. Transient stability enhancement of wind farms connected to a multi-machine power system by using an adaptive ANN-controlled SMES. Energy Convers Manag 2014;78:412–20. https://doi.org/10.1016/j.enconman.2013.10.039.
[24] Çevik HH, Çunkaş M, Polat K. A new multistage short-term wind power forecast model using decomposition and artificial intelligence methods. Phys A Stat Mech Its Appl 2019;534:122177. https://doi.org/10.1016/j.physa.2019.122177.
[25] Wang G, Jia R, Liu J, Zhang H. A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning. Renew Energy 2020;145:2426–34. https://doi.org/10.1016/j.renene.2019.07.166.
[26] Wang C, Zhang H, Ma P. Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network. Appl Energy 2020;259:114139. https://doi.org/10.1016/j.apenergy.2019.114139.
[27] Wang R, Wang J, Xu Y. A novel combined model based on hybrid optimization algorithm for electrical load forecasting. Appl Soft Comput J 2019;82:105548. https://doi.org/10.1016/j.asoc.2019.105548.
[28] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13:915–20. https://doi.org/10.1016/j.rser.2008.02.002.
[29] Johnson PL, Negnevitsky M, Muttaqi KM. Short term wind power forecasting using Adaptive Neuro-Fuzzy Inference Systems. 2007 Australas Univ Power Eng Conf AUPEC 2007:1–4. https://doi.org/10.1109/AUPEC.2007.4548099.
[30] Alzyout MS, Alsmirat MA. Performance of design options of automated ARIMA model construction for dynamic vehicle GPS location prediction. Simul Model Pract Theory 2020;104:102148. https://doi.org/10.1016/j.simpat.2020.102148.
[31] Selvaraj JJ, Arunachalam V, Coronado-Franco KV, Romero-Orjuela LV, Ramírez-Yara YN. Time-series modeling of fishery landings in the Colombian Pacific Ocean using an ARIMA model. Reg Stud Mar Sci 2020;39:101477. https://doi.org/10.1016/j.rsma.2020.101477.
[32] Liu M, Taylor JW, Choo WC. Further empirical evidence on the forecasting of volatility with smooth transition exponential smoothing. Econ Model 2020;93:651–9. https://doi.org/10.1016/j.econmod.2020.02.021.
[33] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerodyn 2008;96:2348–62. https://doi.org/10.1016/j.jweia.2008.03.013.
[34] Zuluaga CD, Álvarez MA, Giraldo E. Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison. Appl Energy 2015;156:321–30. https://doi.org/10.1016/j.apenergy.2015.07.043.
[35] Ma X, Jin Y, Dong Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. Appl Soft Comput J 2017;54:296–312. https://doi.org/10.1016/j.asoc.2017.01.033.
[36] Hu J, Wang J, Zeng G. A hybrid forecasting approach applied to wind speed time series. Renew Energy 2013;60:185–94. https://doi.org/10.1016/j.renene.2013.05.012.
[37] Wang J, Niu X, Zhang L, Lv M. Point and interval prediction for non-ferrous metals based on a hybrid prediction framework. Resour Policy 2021;73:102222. https://doi.org/10.1016/j.resourpol.2021.102222.
[38] Liu M, Cao Z, Zhang J, Wang L, Huang C, Luo X. Short-term wind speed forecasting based on the Jaya-SVM model. Int J Electr Power Energy Syst 2020;121:106056. https://doi.org/10.1016/j.ijepes.2020.106056.
[39] Han S, Li J, Liu Y. Tabu search algorithm optimized ANN model for wind power prediction with NWP. Energy Procedia 2011;12:733–40. https://doi.org/10.1016/j.egypro.2011.10.099.
[40] Zhang D, Lou S. The application research of neural network and BP algorithm in stock price pattern classification and prediction. Futur Gener Comput Syst 2021;115:872–9. https://doi.org/10.1016/j.future.2020.10.009.

[41] Meka R, Alaeddini A, Bhaganagar K. A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables. Energy 2021;221:119759. https://doi.org/10.1016/j.energy.2021.119759.
[42] Iversen EB, Morales JM, Møller JK, Madsen H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. Int J Forecast 2016;32:981–90. https://doi.org/10.1016/j.ijforecast.2015.03.001.
[43] Abedinia O, Amjady N. Short-term wind power prediction based on Hybrid Neural Network and chaotic shark smell optimization. Int J Precis Eng Manuf - Green Technol 2015;2:245–54. https://doi.org/10.1007/s40684-015-0029-4.
[44] Wang Y, Wang J, Li Z. A novel hybrid air quality early-warning system based on phase-space reconstruction and multi-objective optimization: A case study in China. J Clean Prod 2020;260:121027. https://doi.org/10.1016/j.jclepro.2020.121027.
[45] Jiang Y, Huang G, Peng X, Li Y, Yang Q. A novel wind speed prediction method: Hybrid of correlation-aided DWT, LSSVM and GARCH. J Wind Eng Ind Aerodyn 2018;174:28–38. https://doi.org/10.1016/j.jweia.2017.12.019.
[46] Ma Z, Chen H, Wang J, Yang X, Yan R, Jia J, et al. Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. Energy Convers Manag 2020;205:112345. https://doi.org/10.1016/j.enconman.2019.112345.
[47] Samadianfard S, Hashemi S, Kargar K, Izadyar M, Mostafaeipour A, Mosavi A, et al. Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Reports 2020;6:1147–59. https://doi.org/10.1016/j.egyr.2020.05.001.
[48] Jiang P, Liu Z, Wang J, Zhang L. Decomposition-selection-ensemble forecasting system for energy futures price forecasting based on multi-objective version of chaos game optimization algorithm. Resour Policy 2021;73:102234. https://doi.org/10.1016/j.resourpol.2021.102234.
[49] Niu M, Hu Y, Sun S, Liu Y. A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. Appl Math Model 2018;57:163–78. https://doi.org/10.1016/j.apm.2018.01.014.
[50] Li J, Chen Y, Lu C. Application of an improved variational mode decomposition algorithm in leakage location detection of water supply pipeline. Meas J Int Meas Confed 2020:108587. https://doi.org/10.1016/j.measurement.2020.108587.
[51] Yarushev S, Averkin A. Time series analysis based on the biologically inspired modular approach. Procedia Comput Sci 2017;120:848–53. https://doi.org/10.1016/j.procs.2017.11.317.
[52] Figueiredo E, Macedo M, Siqueira HV, Santana CJ, Gokhale A, Bastos-Filho CJA. Swarm intelligence for clustering — A systematic review with new perspectives on data mining. Eng Appl Artif Intell 2019;82:313–29. https://doi.org/10.1016/j.engappai.2019.04.007.
[53] Wang J, Du P, Hao Y, Ma X, Niu T, Yang W. An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting. J Environ Manage 2020;255:109855. https://doi.org/10.1016/j.jenvman.2019.109855.
[54] Wang S, Wang J, Lu H, Zhao W. A novel combined model for wind speed prediction – Combination of Linear Model, Shallow Neural Networks, and Deep learning Approaches. Energy 2021;234:121275. https://doi.org/10.1016/j.energy.2021.121275.
[55] Hammouri AI, Mafarja M, Al-Betar MA, Awadallah MA, Abu-Doush I. An improved Dragonfly Algorithm for feature selection. Knowledge-Based Syst 2020;203:106131. https://doi.org/10.1016/j.knosys.2020.106131.
[56] Jiang P, Liu Z. Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting. Appl Soft Comput J 2019;82:105587. https://doi.org/10.1016/j.asoc.2019.105587.
[57] Wang J, Niu X, Liu Z, Zhang L. Analysis of the influence of international benchmark oil price on China's real exchange rate forecasting. Eng Appl Artif Intell 2020;94:103783. https://doi.org/10.1016/j.engappai.2020.103783.
[58] Liu Z, Jiang P, Wang J, Zhang L. Ensemble forecasting system for short-term wind speed forecasting based on optimal sub-model selection and multi-objective version of mayfly optimization algorithm. Expert Syst Appl 2021;177:114974. https://doi.org/10.1016/j.eswa.2021.114974.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Credit Author Statement

Yilin Zhou: Methodology, Software, Writing – Review &Editing.
Jianzhou Wang: Funding acquisition, Supervision, Data curation.
Haiyan Lu: Formal analysis, Validation.
Weigang Zhao: Funding acquisition, Visualization.