IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 20, 2020, accepted March 27, 2020, date of publication April 1, 2020, date of current version April 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984851

Multi-Step Short-Term Wind Power Prediction Based on Three-level Decomposition and Improved Grey Wolf Optimization

YUCHAO HAN AND XIANGQIAN TONG
School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China
Corresponding author: Xiangqian Tong (lstong@mail.xaut.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51677151.

ABSTRACT Wind power prediction is of great importance in enhancing wind energy penetration. This paper proposes a novel wind power prediction method which combining three-level decomposition with optimized prediction method. In the decomposition part, the Wavelet Packet Decomposition (WPD) is introduced as the first level decomposition, then the obtained sub-series are further decomposed by Variable Mode Decomposition (VMD). At last, Singular Spectrum Analysis (SSA) is carried out for each Intrinsic Mode Function (IMF), and the dominant component and residual components are separated as the input of the prediction. In the prediction part, Kernel Extreme Learning Machine (KELM) is adopted to complete the multi-steps wind power prediction. In this paper, an Improved Grey Wolf Optimization (IGWO) algorithm with redesign of the hierarchy and architecture is proposed, which especially suitable for optimizing wind power prediction. Finally, ten different models are compared, and the results show that the proposed method in this paper can extract the trend information of wind power greatly and has achieved excellent accuracy in short-term wind power prediction.

INDEX TERMS Wind power prediction, three-level decomposition, improved grey wolf algorithm, multi-step prediction, kernel extreme learning machine.

NOMENCLATURE

ABC	Artificial bee colony algorithm
ARMA	Auto-Regressive and Moving Average model
ARIMA	Autoregressive Integrated Moving Average model
BSA	Backtracking Search Algorithm
EMD	Empirical Mode Decomposition
EEMD	Ensemble EMD
ELM	Extreme Learning Machine
FNN	Fuzzy-Neural Network
GWO-SCA	Grey Wolf Optimized Sine Cosine Algorithm
GPR	Gaussian Process Regression
GMDH	Group Method Data Handling
IMF	Intrinsic Mode Function
IGWO	Improved Grey Wolf Optimization
KELM	Kernel Extreme Learning Machine
LSSVM	Least Square Support Vector Machine
LSTM	Long Short-Term Memory
NWP	Numerical Weather Prediction
NN	Neural Networks
PSR	Phase Space Reconstruction
SSA	Singular Spectrum Analysis
GWO	Grey Wolf Optimization
SVM	Support Vector Machine
TDCNN	Two-Dimensional Convolution Neural Network
VMD	Variable Mode Decomposition
WPD	Wavelet Packet Decomposition

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

I. INTRODUCTION

High intermittency and large capacity of wind power set up a new challenge for the power system. Storage system and wind power prediction are two ways to fix the randomness of the wind power and reduce the impact to the grid [1]. Due to its high cost of storage system, the construction is relatively slow. The wind power prediction can significantly alleviate this problem with a small investment. To improve the penetration rate of wind power while reducing the reserve capacity, the more accurate wind power prediction is needed. The purpose of wind power prediction varies according to the length of prediction time. Short-term wind power prediction which lasts from minutes to hours mainly contributes to the optimization of capacity of spinning reserve. Medium-term wind power prediction ranging from one day to several days is mainly used for wind farm maintenance and scheduling plan. The existing wind power prediction methods can be divided into three main types: Physical models [2], Statistical models [3] and Hybrid models [4]. Physical models mainly focused on using mathematical model to describe the terrain and obtain more accurate Numerical Weather Prediction (NWP). In the medium-term and long-term prediction, the physical model has incomparable advantages on accuracy. But when it comes to short-term wind power prediction, large calculation consumption makes the physical model not competent for high-precision prediction. And the statistical model such as Auto-Regressive and Moving Average model (ARMA) [5], Autoregressive Integrated Moving Average model (ARIMA) [6], Support Vector Machine (SVM) [7], Least Square Support Vector Machine (LSSVM) [8], Extreme Learning Machine (ELM) [9]–[11] are adopted because of its fast calculation speed and high accuracy. The hybrid model not only uses the statistical model but also integrates the physical model, especially using NWP. So, the hybrid model is also unsuitable for short-term wind power prediction.

In the field of statistical models, there are two ways to predict wind power. One is predicting wind power from the wind turbine power data directly [12], or fitting the power curve with wind speed, the main influencing factor, can be used for indirect prediction [13], [14]. Regardless of input data type, the original time series is not suitable for prediction. The signal decomposition methods are employed to find the change law in obtained sub-series. WPD divides the time-frequency plane more carefully, and its resolution to the high-frequency part of the signal is higher than that of the wavelet decomposition [15]. However, the number of decomposition layers and the type of wavelet basis function influence the decomposition effect, the Empirical Mode Decomposition (EMD) method [16] overcomes this shortcoming and decomposes the input signal into several different IMF components. These IMF components reflect the characteristics of different frequency bands in the original signal. In order to overcome the mode aliasing in EMD algorithm, Improved EMD, Ensemble EMD (EEMD) and VMD methods [17]–[19] are proposed. In [20], VMD method is used to decompose the wind power time series. By extracting the feature of IMFs, wonderful prediction accuracy is achieved. In [21], the author using WPD as the first decomposition method, after that EEMD method is used for each sub-series. Numerical experiment has proofed that single decompose is not quite enough for accurate wind power prediction, and the two-level decomposition of contrast has good effect. But with the increasing numbers of the integrated decomposition methods, the sub-series numbers explosive growth which brings irreconcilable contradiction for short-term wind power prediction.

After decomposing the wind power data, the efficiency and accuracy of the predict method determine the precise of the wind power prediction. Machine learning method has been widely applied in wind power prediction for its splendid performance in fitting nonlinear time series. SVM, Gaussian Process Regression (GPR) and Neural Networks (NN) are representative methods in prediction and classification [22]–[24]. With the emergence of artificial intelligence, a large number of combined methods have emerged. The traditional machine learning has drawbacks like local minima, overtraining and high computation cost especially in chaotic time series like wind power time series. Because of these shortcomings, traditional methods are complicated in handling the non-stationary wind power time series for its lack of adaptation to the changing environment. Combined methods can solve this problem with appropriate combinations. To remedy the chaotic behaviour of wind power time series, Liu *et al.* [21] adopted two-level decomposition into wind power time series, the WPD-EMD method shows enormous superiority upon the conventional single-level decomposition method. Fu *et al.* [25] employed Phase Space Reconstruction (PSR) and SSA to eliminate the influence of the chaotic sequence and looking for the real law of wind speed sequence. Safari *et al.* [26] used maximum Lyapunov exponent to decide whether to further decompose the IMF components with SSA. Since ELM comes out, the computing ability and satisfactory calculation results make this method being applied to solve many problems. But it also has disadvantages such as unstable, many works have been carried out to solve this problem. In [27], the Improved Extreme Learning machine method was proposed to satisfy the need for short-term wind power prediction. In [28], an optimized EM-ELM algorithm was adopted to solve the problem of parameter selection. In [29] KELM combined kernel function and ELM greatly improves the learning speed of the forward neural network and avoided suck into local maximal optimal at the same time.

Since the predict method always has several input parameters, the enumeration method can hardly find the best solution through parameter combinations. Many excellent works focused on the optimization of the prediction parameters. Li *et al.* [23] used dynamic adaptive learning factor and differential evolution strategy to improve the traditional dragonfly algorithm. The improved dragonfly algorithm showed its advantages in improving the prediction accuracy. Fu *et al.* [30] used a hybrid Grey Wolf Optimized Sine Cosine Algorithm (GWO-SCA) to optimize the input parameters of ELM, the combined method showed superior advantages than the compared method in wind speed forecasting. Tian [31] adopted Improved GWO algorithm to improve the performance of LSSVM. This method performs well in dynamic liquid level forecasting of beam pump. Abedinia *et al.* [32] using improved PSO algorithm to optimize the weight of Two-Dimensional Convolution Neural Network (TDCNN) and by using combinatory approach this work obtained good results. Amjady and Abedinia [33] adopted Kriging Interpolation Method into wind power prediction. By using evolutionary algorithm to optimize the basic setting of proposed method enhanced good effect. Naik *et al.* [34] using vaporization and precipitation based water cycle algorithm to optimize the combined prediction model. Many other algorithms like Backtracking Search Algorithm (BSA) [35], Grey Wolf Optimization (GWO) algorithm [36], Particle Swarm Optimization (PSO) [37] Artificial bee colony algorithm (ABC) [36] have been used in wind power or wind speed prediction. Due to the limitation of calculation time, there is less research working on three-level decomposition and parameter optimization in wind power prediction.

From the above discussion, the main purpose of this paper is to apply the decomposition method and optimization method synthetically in order to eliminate the randomness of wind power time series and hence improve the accuracy of short-term wind power prediction. The innovations of this study are explained as: (a) three-level decomposition algorithm is presented based on WPD VMD and SSA; (b) by redesigned the framework of GWO algorithm, previous best solution can be inherited, and the new hierarchy enhanced the algorithm the global optimization; (c) we combined the three-level decomposition with IGWO optimized KELM to obtain stable prediction accuracy. This paper is organized as follows: In section II, we present the structure of proposed method. Section III introduces the detailed three-level decomposition method. The improved grey wolf optimized KELM is introduced in section IV. The section V uses real wind power time series to verify the effectiveness of the proposed method. Section VI concludes this paper.

II. PROPOSED COMBINATION STRUCTURE

In this section, we give a brief introduction to the structure of proposed nonlinear combination model. By adopting the decomposition method, we can get several components that contain the basic law of wind power change. A suitable combination of decomposition methods can provide lower global optimum for optimization. IGWO-KELM method provides an efficient way to extend the trend of the sub-components. By combining the decomposition method and the predict method, we propose the WPD-VMD-SSA-IGWO-KELM model. To be specific, we use three-level decomposition to extract the detail information of the wind power series. Then the obtained dominant component and residual component of each IMF will be used to predict the future wind power. Redesign the hierarchy and inherit the previous optimization results, so that the IGWO can obtain the global optimal parameters. In this study, the multi-step short-term wind power prediction is required to be implemented every 10 minutes and the prediction length is 4 hours which contained 24 wind power data points. The training data set is established with 10-day wind power data. The parameters to be optimized in the prediction include the lag length of training set, the parameters of RBF kernel function and the regularization coefficient.

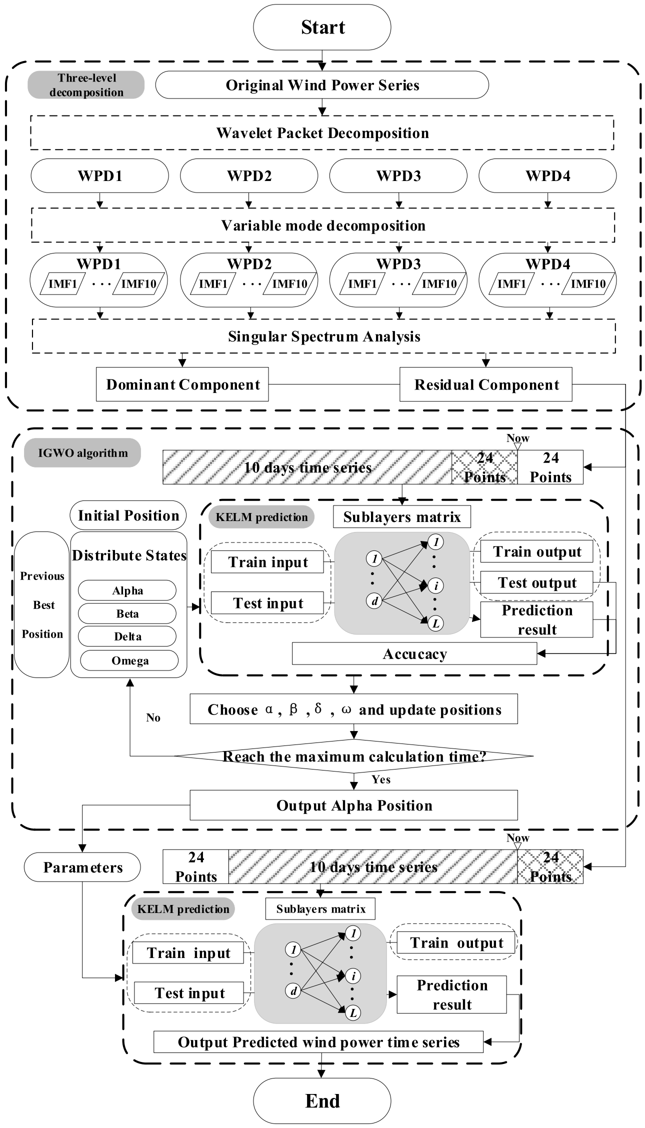


FIGURE 1. The structure of the proposed model.

As shown in figure 1, the detailed structure of proposed model can be described as:

Step 1: Collect 10 days wind power data and clear invalid points;

Step 2: Decompose the wind power time series with the WPD method, and obtained 4 sub-series;

Step 3: Use VMD as the secondary decomposition method. Further decomposed the four sub-series into 10 IMFs components respectively. Total of 40 sub-series are obtained after this step;

Step 4: Input the obtained IMFs to the SSA algorithm and extract the dominant component and residual component respectively. Add up to 80 sub-series are formed up as the input of prediction;

Step 5: Each sublayer is divided into train input, train output and test input according to the lag length. Use KELM

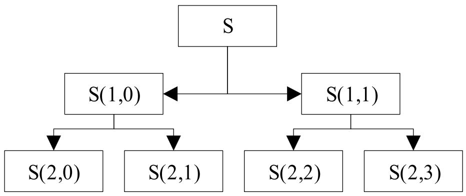


FIGURE 2. Schematic diagram of wavelet packet decomposition.

with initial parameters to predict the corresponding 24 data points. Sum up the prediction result of all the sub-series, and calculate the fitness under initial parameters;

Step 6: Use IGWO algorithm to optimize the parameters. The final Alpha position is used as the parameters of the real-time prediction. Move the training set back 24 data points as a new training set of KELM;

Step 7: Move test set back 24 data points. And use the optimized parameter to predict the wind power in next 24 time-points.

III. THREE-LEVEL DECOMPOSITION METHOD

More decomposed sub-series mean larger chances to find the trend of wind power change. We propose the three-level decomposition structure which consists of three different signal processing methods that good at the nonstationary signal decomposition. The proposed three-level method not only overcome the insufficient decomposition of WD, but also eliminate the remaining chaotic components in the VMD. This section will provide a brief introduction to WPD, VMD and SSA.

A. WPD

WPD method is a significant improvement from Wavelet Decomposition (WD), which only consists of the appropriate coefficients. WPD is superior to WD in that it divides the frequency band into several levels. And WPD further decomposes the high-frequency part which is not subdivided in the multi-resolution analysis and can adaptively select the corresponding frequency band according to the characteristics of the analysis signal. According to [38], we are using 'db4' wavelet to decompose the original signal into two levels.

As shown in figure 2, the WPD is obtained by decomposing the high-frequency part of WD, and the original signal is mapped into 2^j wavelet packet subspaces, by completing the map process a binary tree is obtained. The wavelet transform can be described as:

$$WT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \psi^{*}(\frac{t-\tau}{a}) \mathrm{d}t \tag{1}$$

where $*$ is the complex conjugate, a is the scale coefficient and τ is the translation coefficient. ψ is the mother wavelet function. By using WPD, four sub-series are shown in figure 3.

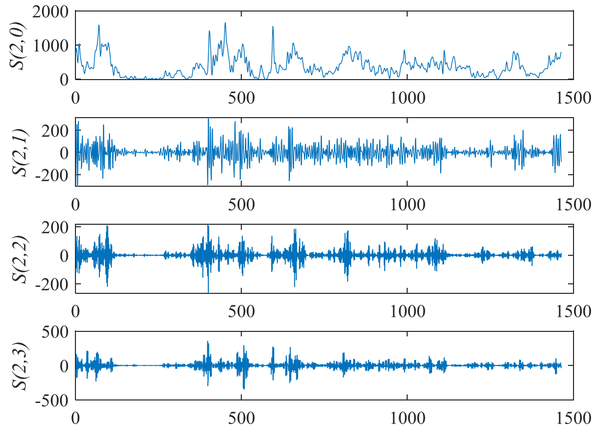


FIGURE 3. Decomposition result of WPD.

B. VMD

Empirical mode decomposition is well-known for its excellent performance in decomposing nonlinear and nonstationary signals. Several IMFs separated from different base frequencies can be extracted from the time series. But when features of time series become close to each other will cause model mixing phenomenon. For the purpose of solving this loophole of EMD, VMD is proposed by Dragomiretskiy and Zosso [18].

First, Hilbert transformation is used to obtain the associated analytic signal, shifting the frequency spectrum by mixing with estimated center frequency, so that the variational problem can be described as:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t [(\delta(t) + \frac{\mathrm{j}}{m}) \cdot u_k(t)] e^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ &\text{s.t.} \sum_k u_k = f \quad k = 1, 2, \ldots K \end{aligned} \tag{2}$$

where $\{u_k\} = \{u_1, u_2, \ldots u_k\}$ is set of IMFs, and $\{\omega_k\} = \{\omega_1, \omega_2, \ldots \omega_k\}$ is set of the corresponding center frequency of IMFs. k is the assumed number of decomposed IMFs.

For the sake of solving this problem without constrained, the quadratic penalty term and Lagrangian multipliers are applied:

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \lambda) \\ &= \alpha \sum_k \left\| \partial_t [(\delta(t) + \frac{\mathrm{j}}{m}) \cdot u_k(t)] e^{-\mathrm{j}\omega_k t} \right\|_2^2 \\ &\quad + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{3}$$

So, finding the saddle point of the (3) will solve the original minimization problem:

$$u_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) - \sum_{i \neq k} \widehat{u}_i(\omega) + \frac{\widehat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left|\widehat{u}_k(\omega)\right|^2 \mathrm{d}\omega}{\int_0^\infty \left|\widehat{u}_k(\omega)\right|^2 \mathrm{d}\omega} \tag{5}$$

And the dual ascent for all $\omega \geq 0$

$$\hat{\lambda}^{n+1} = \hat{\lambda}^n + \tau(\hat{f}(\omega) - \sum_k u_k^{n+1}(\omega)) \tag{6}$$

Repeating (4)-(6) until reaching the iteration limit or convergence error less than:

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 < \varepsilon \tag{7}$$

Decomposing the sublayer S (2, 0) into 10 IMFs, the result is shown in figure 4.

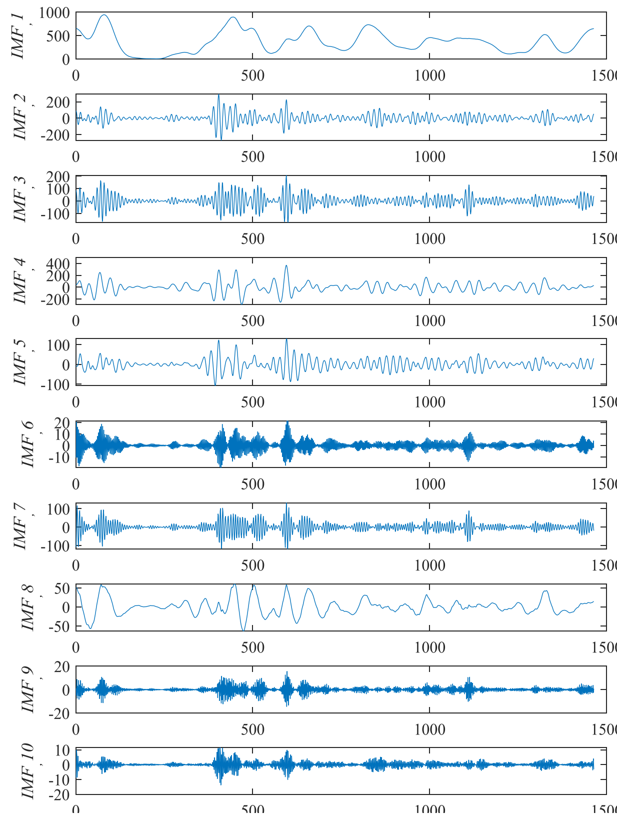


FIGURE 4. Decomposition of VMD.

C. SSA

The wind power time series comprise of many kinds of intrinsic components. The over decomposition of VMD will also cause period oscillations. SSA method is particularly suitable for dealing nonlinear part of the wind power time series. The technological process of SSA can be regarded as three parts:

First, in order to find the trend of the input component, time series need to be mapped into multi-dimension space.

$$X = \begin{Bmatrix} x_1 & x_2 & \dots & x_p \\ x_2 & x_3 & \dots & x_{p+1} \\ \dots & \dots & \dots & \dots \\ x_l & x_{l+1} & \dots & x_{p+l-1} \end{Bmatrix} \tag{8}$$

In (8), l is the ensemble dimension, and the matrix X represents l-dimensional space.

Then the singular value decomposition (SVD) method is used to calculate the eigenvalues and corresponding eigenvalue matrix. In this part, the i-th triple eigenvalues (σ_i, U_i, V_i) are obtained by decomposing the matrix XX^T with SVD. So the matrix X can be expressed as:

$$X = X_1 + X_2 + \ldots + X_l X_i = \sigma_i U_i V_i^T \tag{9}$$

In (9) σ_i is the singular value of the i-th eigenvalue vector, and U_i, V_i is the corresponding vector of the singular value.

The second part of SSA is grouping, several subsets can be obtained in this procedure. For the trend of the time series, we can separate the dominant singular values to gathering a trend subset:

$$X_t = X_{t1} + X_{t2} + \ldots + X_{tr} X_{tr} = \sigma_i U_i V_i^T \tag{10}$$

At last, diagonal averaging can convert the subsets extracted above to time series. The refactored time series can be described as:

$$y_m = \begin{cases} \frac{1}{m} \sum_{i=1}^{m} x_{i,m-i+1}^* & 1 \leq m < l \\ \frac{1}{l} \sum_{i=1}^{l} x_{i,m-i+1}^* & l \leq m \leq p \\ \frac{1}{N-m+1} \sum_{i=m-p+1}^{T-p} x_{i,m-i+1}^* & p < m < N \end{cases} \tag{11}$$

As shown in figure 5, the largest eigenvalue corresponding eigenvalue matrix can be used for extracting the main trend of the time series, and the smallest one corresponding to the non-information time series which can be removed from the next procedure. Two components named dominant component and residual component respectively can be obtained in SSA.

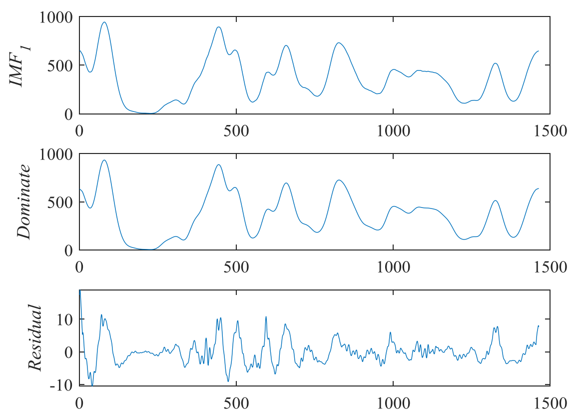


FIGURE 5. Decomposition result of IMF1 with SSA.

IV. IMPROVED GREY WOLF OPTIMIZIED KELM

In this section, we propose the IGWO-KELM algorithm. By optimized with IGWO algorithm, KELM can easily enhance its performance.

A. KERNEL EXTREME LEARNING MACHINE

ELM is typically a single-layer-feedforward neural network. By using only one hidden layer, the ELM has fast convergence and satisfactory generalization performance. But because of the random initialization, the ELM algorithm behaves unstable in prediction accuracy. To solve this problem Huang *et al.* [29] create KELM by the inspiration of the SVM. Compared with the traditional ELM, the KELM has less adjustable parameters and faster convergence speed. With the advantages of stable, the optimization of input parameters can be easily extended to next data point. Fast computing speed enables a large number of decomposition layers to be completed within a time limit.

The output function of ELM for generalized single hidden layer feedforward neural network is:

$$f(x) = h(x)\beta = H\beta \tag{12}$$

In (12) x is input, β is the output weight coefficient, H is the input mapping matrix between input and hidden layers. The output weight coefficient matrix can be obtained by:

$$\beta = H^T\left(\frac{\mathrm{I}}{C} + HH^T\right)^{-1}T \tag{13}$$

where:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_n^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{n1} & \cdots & t_{nm} \end{bmatrix} \tag{14}$$

The output of ELM can be concluded as:

$$f(x) = h(x)H^T\left(\frac{I}{C} + HH^T\right)^{-1}T \tag{15}$$

In KELM, almost all the nonlinear piecewise continuous functions can be used as the function between input and hidden layers. So, the KELM can be obtained by:

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \cdots \\ K(x, x_n) \end{bmatrix} H^T\left(\frac{I}{C} + \Omega_{KELM}\right)^{-1}T$$
$$\Omega_{KELM} = HH^T : \Omega_{KELMi,j} = h(x_i)\cdot h(x_j) = K(x_i, x_j) \tag{16}$$

In (16), Ω_{KELM} is the kernel matrix which can be obtained by radial basis function, linear kernel function or polynomial kernel function.

B. IMPROVED GREY WOLF OPTIMIZATION

Optimization algorithm can make the prediction method better by adjusting the parameters. In this section, the framework of GWO is redesigned and IGWO algorithm is proposed. The traditional GWO using four groups of wolves searching for their goal to optimize the problem. As shown in the figure 6, the dominant wolf at the top position is Alpha. It has priority over determine the time and place of hunting. The next level is Beta who plays the assistant of the leader. The third part of the wolves are Delta, who are the roles of reconnaissance, sentry, and nurse. At the bottom of the society are Omega wolves.

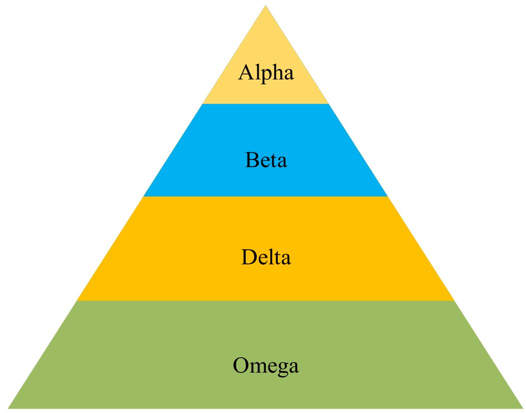


FIGURE 6. The hierarchy of the GWO algorithm.

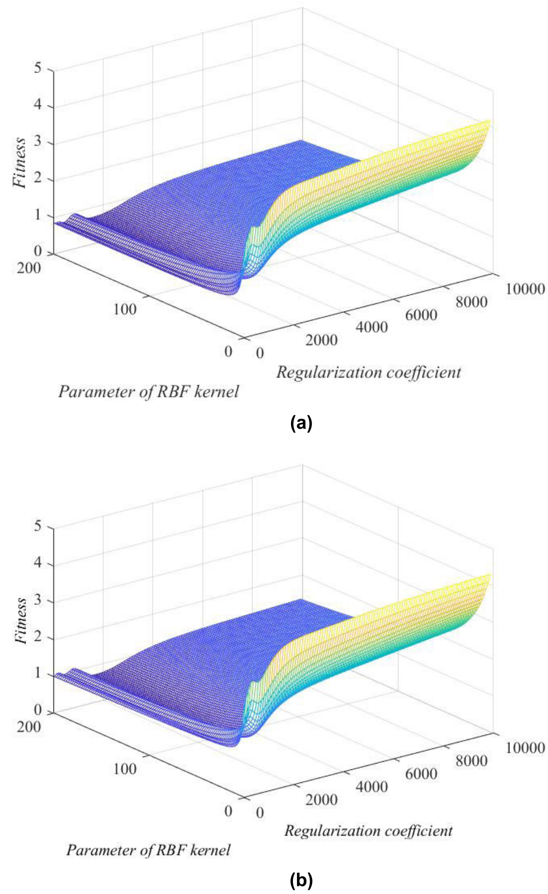


FIGURE 7. Fitness surface of different input parameters.

Although the traditional method almost suitable for all the condition, the wolf group easy to fall into local optimum cause of the Alpha Beta and Delta only has one wolf.

As shown in figure 7, the fitness surface of time t is shown in (a), and the next time scale is shown in (b). It is obvious that the best solution can be most likely founded near the previous best solution. So, it is of significance that the optimization algorithm inherits optimization results.

Inspired by Cai *et al.* [39], to avoid falling into local optimum, the hierarchy of the grey wolf algorithm is redesigned.

TABLE 1. Result of testing composition functions.

Test function	GA		FOA		PSO		GWO		IGWO	
	Ave.	Std.	Ave.	Std.	Ave.	Std.	Ave.	Std.	Ave.	Std.
TF1	1.34E-06	1.42E-06	2.76E-08	1.06E+01	2.27E+01	1.91E-116	1.09E-115	1.47E-128	**1.41E-127**	1.34E-06
TF2	3.93E-05	3.83E-03	3.24E-05	1.28E+00	1.39E+00	9.17E-67	2.06E-66	2.12E-76	**8.17E-76**	3.93E-05
TF3	1.25E-05	5.17E-05	1.09E-06	4.18E+02	3.60E+02	1.60E-51	1.01E-50	1.64E-59	**1.21E-58**	1.25E-05
TF4	8.35E-02	7.00E-04	2.91E-05	4.88	2.28	5.85E-37	**2.37E-36**	4.18E-32	1.45E-31	8.35E-02
TF5	2.06	8.60	2.25	2.87	624	6.48	7.03	8.15	**1.61**	1.06
TF6	8.49E-07	2.50	3.64E-05	16.9	39.5	0.01	4.95E-02	5.87E-09	**3.15E-09**	8.49E-07
TF7	1.26E-02	2.64E-03	2.09E-03	6.33E-02	4.42E-02	2.84E-04	2.03E-04	1.16E-04	**1.02E-04**	1.26E-02

In the new hierarchy, the Beta wolves have more population and the Delta wolves become the main part of the wolves. The Omega wolves are no longer being the useless part of the wolf group, they will search randomly in case the first three kinds of wolf fall into local optimum.

Based on the new hierarchy above, the Improved Grey Wolf Optimization (IGWO) is proposed. Beta wolves now become several wolves searching around the alpha wolf, the movement of Beta wolves can be described as:

$$Beta = Alpha + r^{*}D \tag{17}$$

In (17) *Beta* is the position matrix of Beta wolves, and the *Alpha* is the Alpha wolf's position. r is the random value from $[-1\ 1]$. D is the distance matrix between the Alpha wolf and the best Delta wolf:

$$D_j = |Alpha - Delta_{best}| \tag{18}$$

In the new hierarchy, if Beta wolves have a better solution than Alpha wolf, then the leader position will be given to the Beta wolf who has the best score. The Delta wolves will take the responsibility of searching the area close to Alpha wolf, which can be formulated as:

$$Delta = Alpha - r_1 \left(2 - \frac{2t}{T_{\max}}\right) |C \cdot Alpha - Delta| \tag{19}$$

In (19), the *Delta* is the position matrix of Delta wolf, C is the oscillation factor between [0 2], r_1 is the random values between [0 1], t is the iteration and $T_{\max}$ is the maximum iteration. The complete procedure of the IGWO algorithm proposed in this paper is shown in figure 8. In order to evaluate the effectiveness of proposed IGWO algorithm, seven benchmark functions were used for evaluation.

The comparative model includes GA, FOA [40], IPSO, GWO [41] and proposed IGWO. For fair comparison, the population number is set to 30 and max iteration is 1000. The benchmark functions including:

$$f_1(x) = \sum_{i=1}^{n} x_i^2 \tag{20}$$

$$f_2(x) = \sum_{i=1}^{n} |x_i| + \prod_{i=1}^{n} |x_i| \tag{21}$$

$$f_3(x) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j\right)^2 \tag{22}$$

$$f_4(x) = \max_i \{|x_i|, 1 \le i \le n\} \tag{23}$$

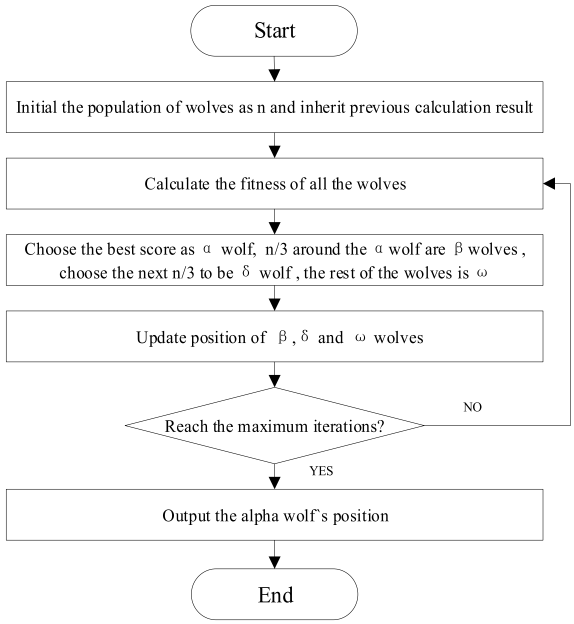


FIGURE 8. The procedure of IGWO method.

$$f_5(x) = \sum_{i=1}^{n-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right] \tag{24}$$

$$f_6(x) = \sum_{i=1}^{n} (|x_i + 0.5|)^2 \tag{25}$$

$$f_7(x) = \sum_{i=1}^{n} ix^4 + random[0, 1] \tag{26}$$

The table 1 listed the detail results of comparison under different benchmark functions. The dimension of benchmark function set to 10. Maximum iteration times restricted to 1000. The population of all the contrast models is set to 30. All of the benchmark functions have an optimized solution which is zero. So, the result of contrast model close to zero means better optimization result. The result shows proposed IGWO algorithm is the best in most benchmark functions. The mutation of GA is set to 0.1. Cross possibility is 0.4. The parameters of other models are the same as the citation.

V. CASE STUDIED

A. DATASET

The La Haute Borne wind farm is located in the northwest of France. The dataset has open-sourced on the internet and can

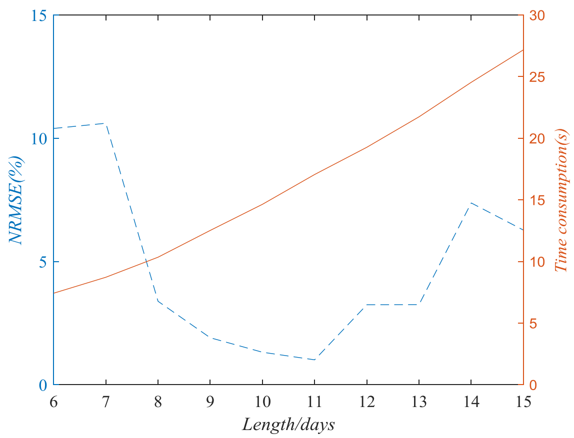


FIGURE 9. Computation time and NRMSE of different train set length.

be download from the website. The single machine capacity is 2050 kW, and there are two wind power generators. In this study we used the 2017 dataset, the time scale of the dataset is 10min. The dataset is divided into four subsets by seasons. In each set, we selected 1484 points for wind power prediction. The prediction is taking place by means of rolling calculation. Experiment was taken by using Spring set to measure the accuracy and time consumption with the length of input dataset. As shown in figure 9, with the increase of the length of the input dataset the accuracy improved significantly. When further increase the value of the train set, the accuracy is not improved as it supposed to be. Unfortunately, the overlong training set will cause the calculation to exceed the prediction time limit. Comprehensive consideration of the complexity and the accuracy, we choose ten days data as the input dataset.

B. EXPERIMENTAL DESCRIPTION

The proposed combined method was implemented by MATLAB. The computer used in this study has i7-7700 CPU (3.6GHz) and 16GB RAM, and parallel computing was used in the prediction process. There are many indicators to describe the performance of the prediction, such as R-Square(R^2), Index of agreement(IA), Relative RMSE(RRMSE), Diebold Mariano TEST, Normalized Root Mean Square Error (NRMSE), Normalized Mean Absolute Error (NMAE), Mean Absolute Percentage Error (MAPE) [42]–[45]:

$$E_{R-\text{square}} = 1 - \frac{\sum_{i=1}^{N} w_i \left(P_i - \overline{Y}_i\right)^2}{\sum_{i=1}^{N} w_i \left(Y_i - \overline{Y}_i\right)^2} \tag{27}$$

$$E_{NRMSE} = \frac{1}{P}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - Y_i)^2} \times 100\% \tag{28}$$

$$E_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - Y_i}{P_i}\right| \times 100\% \tag{29}$$

$$E_{NMAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - Y_i}{P}\right| \times 100\% \tag{30}$$

$$E_{IA} = 1 - \frac{\sum_{i=1}^{N}(P_i - Y_i)^2}{\sum_{i=1}^{N}\left(\left|P_i - \overline{Y}\right| + \left|Y_i + \overline{Y}\right|\right)^2} \tag{31}$$

The time complexity of proposed method consists of five parts: WPD, VMD, SSA, KELM, IGWO. The length of the input time series is L. So, the time complexity of WPD is O(M * L * logL), M is the decomposition level of WPD. The time complexity of VMD is O(K * s * L^3), s is the iteration times when reach specified precision and K is the decomposition level. The time complexity of SSA is O(L^3). The time complexity of KELM is O(L^3). The time complexity of IGWO is O(g * N * O(L^3)), N is the population number, g is the maximum iteration times. The time complexity of the proposed method can be reduced to O(L^3) by ignore the constants and lower order term. It can be observed that the proposed method doesn‘t increase the magnitude of time complexity. The space complexity of the method is O(L). It can be concluded that the time complexity of proposed combined method has not increase too much.

In order to show the superiority of the method proposed in this paper, two groups of different methods are set up for comparison. In order to verify the effectiveness of propose three-level decomposition method, model 1 to model 5 using different decomposition level. Model 1 is the proposed method. The decomposition level of WPD is set to 2. In order to keep the fitness surface stable, we choose the compromise scheme by select 10 as the VMD decompose level. With the proposed three-level decomposition method, fixed decomposition level of VMD will accelerate the prediction process when there are more layers in optimal scheme, and the effect of over decomposition can be reduced when the optimal number of decomposition layers is less than the fixed number of decomposition layers. Model 3 and model 5 use WPD and VMD respectively, and model 4 use the original wind power time series directly.

In order to verify the superiority of the proposed prediction method, different prediction models are set as control group. Model 6 is using the same decomposition method and BPNN which the input weight is optimized by genetic algorithm. The mutation of GA is set to 0.1, cross possibility is 0.4. Model 7 using PSO optimized LSSVM to predict the wind power. wc_1 and c_2 in PSO is set to 0.4 1 and 1 respectively. Since the three-level decomposition method has too many sub-series that PSO-LSSVM can’t finish the prediction in 10 min. So, the two-level decomposition method is used in model 7. Model 8 using WPD as the input of Group Method Data Handling(GMDH) neural networks. The maximum number of neurons in a Layer is 60 and have 6 layers total. Model 9 adopted the traditional ARIMA method as the prediction method. Akaike Information Criterion and Bayesian Information Criterion are used to verify the order of AR and MA.

TABLE 2. Short name for the comparative model.

Models	Short name
WPD-VMD-SSA-IGWO-KELM	Model1
WPD-VMD-IGWO-KELM	Model2
WPD-IGWO-KELM	Model3
Original -IGWO-KELM	Model4
VMD-IGWO-KELM	Model5
WPD-VMD-SSA-GABPNN	Model6
WPD-VMD -PSO-LSSVM	Model7
WPD-GMDH	Model8
VMD-ARIMA	Model9
LTSM	Model10

TABLE 3. Statistic indices of datasets.

Datasets	Statistic indices			
	Mean(m/s)	Max.(m/s)	Min.(m/s)	Std. (m/s)
Spring	5.4483	12.19	0.37	2.2180
Summer	5.9774	11.98	0.16	1.6347
Fall	5.4195	9.52	0.01	1.8752
winter	6.0406	15.63	0.01	2.9771

Ten decomposition layers decomposed by VMD are set as the input of ARIMA. Model 10 using Long Short-Term Memory (LTSM) to predict the wind power time series directly.

The population number of IGWO and PSO algorithm is set to 20. The maximum iteration times is limited by the calculation time. The prediction step is set to 24, which is 4 hours at 10 min interval.

The fitness function is used in this paper to reflect the accuracy of the prediction in optimization. A smaller value means better solution for wind power prediction. The fitness of the IGWO is described as:

$$Fitness = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - P_{pre,i}\right)^2 \tag{32}$$

In traditional analysis, the researchers always put algorithms with different time complexity into comparison. But with the increase of time complexity, the timeliness of sending predict results to the power companies cannot be guaranteed. And the low time consumption method has always been put into disadvantages positions by getting results early. In this study, we put the comparison of different methods into the same time zone which is ten minutes. When the deadline for calculation is up, the final optimization result will be outputted immediately.

Four datasets with different characteristic is illustrated in table 3. The statistical information of four datasets from four seasons are illustrated in Table 2. The basic statistic indices including mean values, maximum values, minimum values and standard deviation of the main influence factor are shown by wind speed in Table 2. Train dataset is established by using ten days wind power time series. After one prediction finished, the training set and test set moved forward one data point for the next rolling prediction.

TABLE 4. Prediction performance of different models in the spring set.

Models	E_{R2}	E_{NRMSE}	E_{NMAE}	E_{MAPE}	E_{IA}
Model1	**0.9922**	**0.5071**	**0.3861**	**1.4424**	**0.99991**
Model2	0.9830	0.7836	0.5654	2.2296	0.99978
Model3	0.3335	4.2731	3.3270	12.473	0.99360
Model4	0.1544	5.0705	3.9224	14.530	0.99137
Model5	0.9084	1.7203	1.3347	5.2162	0.99889
Model6	0.9869	0.5802	0.4249	1.7859	0.99980
Model7	0.9786	0.8795	0.6563	2.5984	0.99974
Model8	0.4250	4.2829	3.4893	13.458	0.99261
Model9	0.8409	1.8018	1.4898	6.1330	0.99860
Model10	0.0473	4.8595	3.7063	13.588	0.99178

As shown in Table, the average wind speed of the Spring set is 5.4483 m/s, the maximum wind speed is 12.19 m/s, the lowest wind speed is 0.37 m/s, and the standard deviation is 2.218 m/s.

Input the above wind power data into the proposed prediction method and the comparative prediction method. Calculate the average value of each index after ten cycles of calculation. The prediction results are shown in Table 4.

By using C language in MATLAB, the proposed three-level decomposition method only takes 6.2746 seconds to obtain 80 sub-series. The prediction process costs 2.8231 seconds on average to finish prediction of 80 sub-series. Under the advantage of fast calculation, the IGWO algorithm can finish over 200 times calculation in the proposed method. Model 2 and model 3 have less calculation time than the proposed method, but the fast calculation is not equal to high accuracy. Since all of the indicators are averaged from rolling calculation, in some specific time span the other models may get better scores than model 1. Model 6 using genetic algorithm to overcome the shortcomings of BP neural network falling into local optimum. As for BPNN optimized by GA, the more layer means higher accuracy and longer calculation time. The train and the test time cost 494.45 seconds, so there is not enough time for optimization. Model 7 used LSSVM optimized by PSO to predict the future wind power curve. Because of a single calculation of the predict procedure cost 53.2 seconds, the PSO algorithm can only finish 11 times calculation, which is far from enough calculation times to reach the global best position.

As shown in Table 4, the best score of each index is shown in bold. From the index values in Table 4, the proposed method outperforms in all of the six methods. For instance, the NRMSE of the proposed method is 0.5071%, which is the best score upon the list. It shows 35.3%, 88.1%, 90.0%, 70.5% improvement than model 2 to model 5 respectively. The proposed three-level decomposition method shows great improvement than other decomposition level method. Model 2 behave best in the comparison models except the proposed model. Compared with model 6 and model 7 the proposed method is improved by 14.3% and 42.3% respectively. As for NMAE and MAPE, the proposed method improved by 9.1% and 19.2% respectively than the

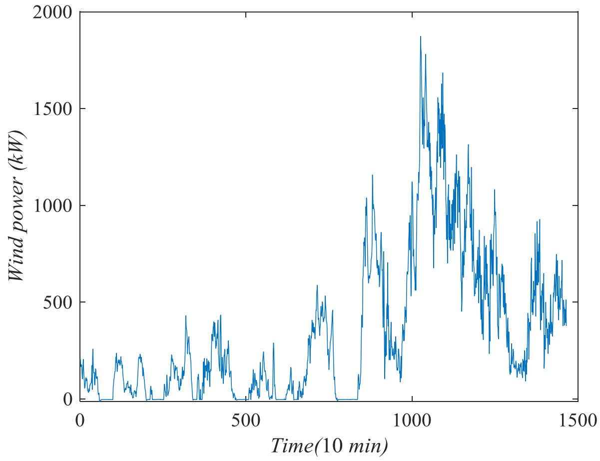


FIGURE 10. Wind speed of Spring set.

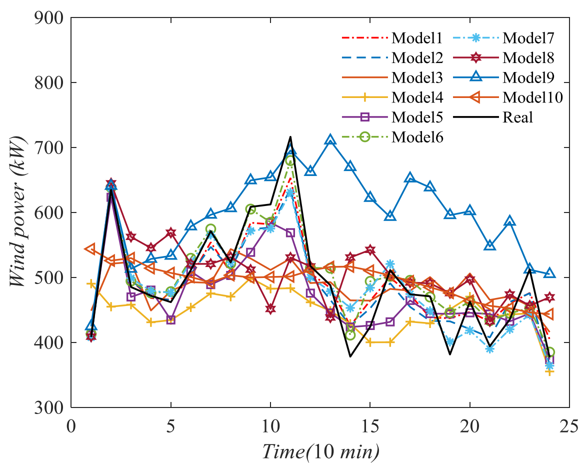


FIGURE 11. Prediction result of all six methods in Spring set.

second-place score. It can be concluded that the proposed method successfully eliminated the predict error, and the prediction accuracy is guaranteed to the greatest extent.

Figure 11 shows the result of the first prediction of the rolling calculation in the Spring set. The proposed method sticks tightly with the real wind power curve. Same trend as the former analysis, the proposed model outperforms all the contrast models.

The summer set has a higher average wind speed of 5.9774 m/s and a lower amplitude of variation 1.6347 m/s. The maximum and minimum values of wind speed are 11.98 m/s and 0.16 m/s respectively. As shown in figure 12, the summer set is relatively balanced. There is no obvious high-power zone and low power zone like spring set. With several wind power climbing phenomena, the wind power prediction may face huge difficulty in improving accuracy.

The performance under this dataset is shown in Table 5. In the NRMSE index, the proposed method shows 9.7%, 86%, 83.4%, 51.1% improvement than model 2 to model5 respectively. 23.9% and 37.1% improvement respectively than model 6 and 7. The MAPE index increased significantly,

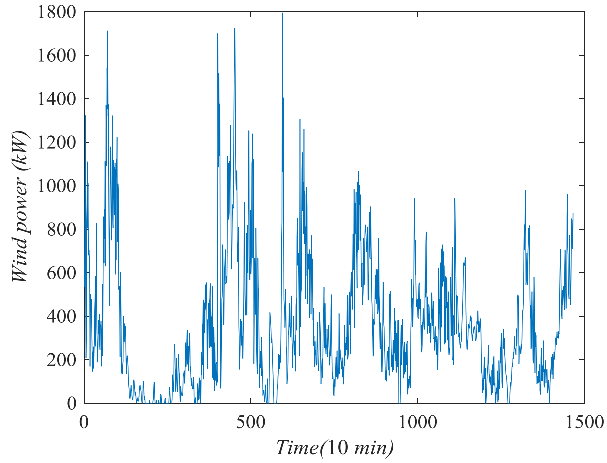


FIGURE 12. Wind speed in Summer set.

TABLE 5. Forecast performance of different models in the summer set.

Models	E_{R2}	E_{NRMSE}	E_{NMAE}	E_{MAPE}	E_{IA}
Model1	0.9903	**0.8404**	**0.6666**	**2.3087**	**0.99984**
Model2	**0.9911**	0.9309	0.7510	2.4124	0.99977
Model3	0.6217	6.0047	4.7439	15.591	0.99154
Model4	0.1545	5.0705	3.9224	14.530	0.99142
Model5	0.9085	1.7203	1.3347	5.2162	0.99892
Model6	0.9658	1.1047	0.8565	2.7719	0.99941
Model7	0.9838	1.3366	1.0245	3.4046	0.99956
Model8	0.32375	6.3640	5.2609	16.969	0.98948
Model9	0.84688	3.0829	2.4571	8.2612	0.99753
Model10	0.10846	7.8763	6.7107	21.877	0.98546

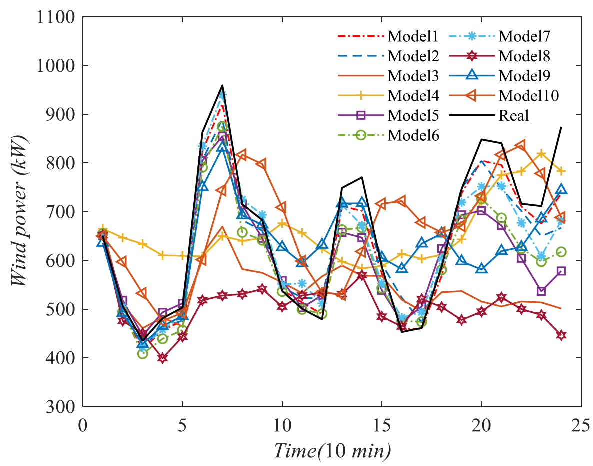


FIGURE 13. Prediction result of all six methods in Summer set.

because of the prediction section belongs to low power area. The proposed method shows 11.2% and 4.2% respectively improvement than the second place in NMAE and MAPE. The proposed method outperforms in all the index in summer set.

In figure 14, the wind resource in the autumn dataset is shown. The mean average wind speed in this dataset is 5.4195 m/s, which is at the same level as the spring set.

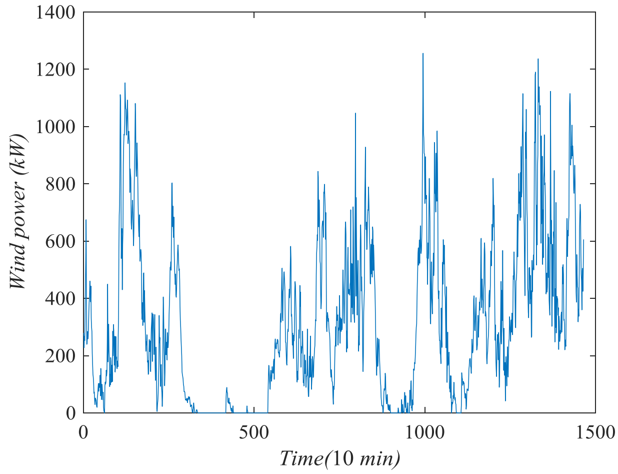


FIGURE 14. Wind speed in Autumn set.

TABLE 6. Prediction performance of different models in the autumn set.

Models	E_{R2}	E_{NRMSE}	E_{NMAE}	E_{MAPE}	E_{IA}
Model1	**0.9834**	**0.8881**	**0.6045**	**2.5055**	**0.99970**
Model2	0.9781	1.0129	0.6743	2.7627	0.99960
Model3	0.3081	6.2247	4.9368	19.772	0.98513
Model4	0.1545	5.0705	3.9224	14.531	0.99142
Model5	0.9085	1.7203	1.3347	5.2162	0.99893
Model6	0.9753	1.2648	0.9818	4.6390	0.99938
Model7	0.9690	1.3409	0.9663	4.1734	0.99930
Model8	0.3562	7.0981	6.2321	25.815	0.97983
Model9	0.9118	1.8454	1.5367	6.5921	0.99851
Model10	0.2367	5.6994	4.9032	21.290	0.98762

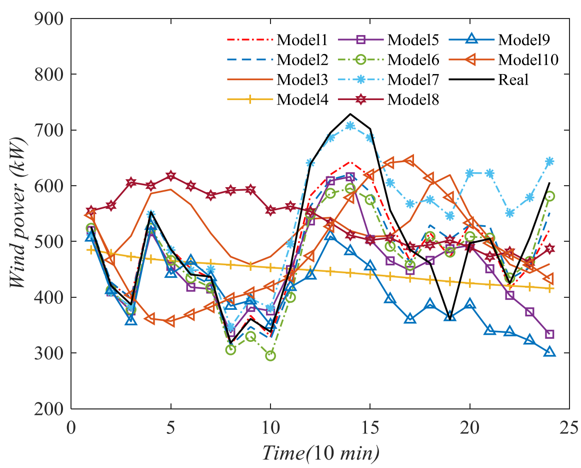


FIGURE 15. Prediction result of all six method in Autumn set.

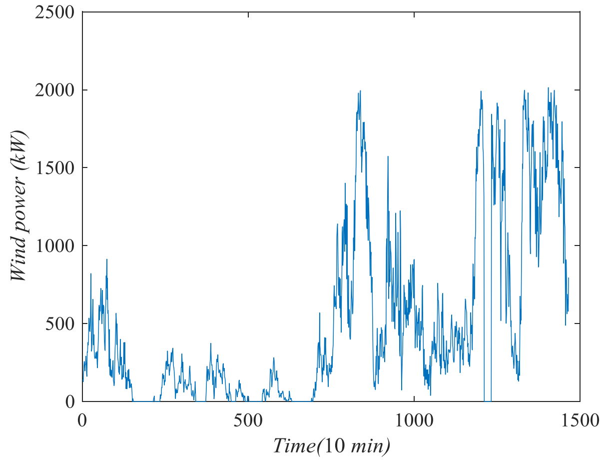


FIGURE 16. Wind speed in Winter set.

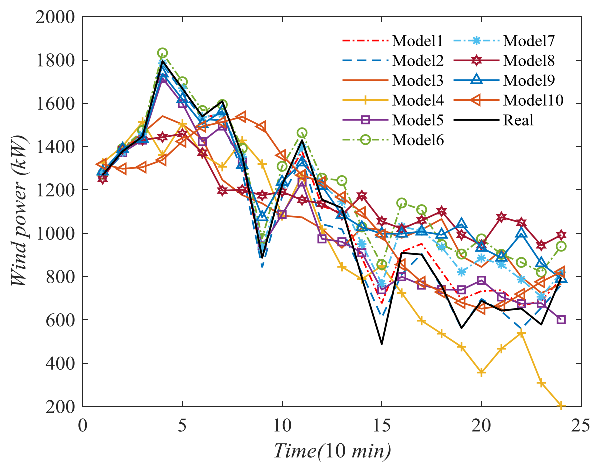


FIGURE 17. Prediction result of all six method in Winter set.

TABLE 7. Prediction performance of different models in the winter set.

Models	E_{R2}	E_{NRMSE}	E_{NMAE}	E_{MAPE}	E_{IA}
Model1	**0.9927**	**1.5603**	**1.0977**	**2.4445**	**0.99982**
Model2	0.9892	1.5909	1.1802	2.6749	0.99975
Model3	0.4721	14.205	11.123	22.392	0.98546
Model4	0.1545	5.0705	3.9224	14.5304	0.9914
Model5	0.9085	1.7203	1.3347	5.2162	0.9989
Model6	0.9897	1.6292	1.2341	2.7213	0.99982
Model7	0.9813	3.1742	2.4112	5.5118	0.99929
Model8	0.41198	13.13388	10.75107	25.65448	0.98530
Model9	0.69296	8.11187	5.77706	14.77517	0.99408
Model10	0.32404	12.49032	9.37548	20.59272	0.98940

The maximum wind speed stays low reaching 9.52 m/s, and the variance is 1.8752 m/s. It can be seen that the autumn set is a relatively stable and sustained breeze.

Among the ten individual combined models, the proposed method performs best in 4-hours short term wind power prediction. In the NRMSE index of the autumn set, the proposed method shows 12.3%, 85.7%, 82.5%, 48.8% improvement than model 2 to model5 respectively. The three-level decomposition method achieved the best result. In the index of NMAE and MAPE the proposed method shows 10.3% and 9.3% improvement than the second-place score.

Wind speed changes greatly in the winter dataset, the Maximum value of wind speed up to 15.63 m/s. The average wind speed is one level higher than the other three data sets up to 6.04 m/s. As shown in figure 16, the winter set experienced the change from low power area to high power area.

The proposed combined method still takes the first place upon all the indexes. In the difficult situation, the traditional method like model 8 or model 10 performs poorly. With the advance of proposed three-level decomposition method, the proposed model shows 1.9%, 89%, 69.2%, 9.3% improvement than model 2 to model 5 respectively in NRMSE.

As shown in the comparison, the proposed three-level decomposition method in this paper improved the prediction accuracy in all the dataset. And the proposed combined prediction structure achieved good result in all the prediction models.

VI. CONCLUSION

A novel combination of three-level decomposition method and optimized prediction method is proposed in this paper. In the decomposition part, we introduced the WPD as the first level decomposition, then the obtained sub-layer is decomposed by VMD. At last, SSA is carried out for each IMF, and the dominant component and other components are separated as the input of the prediction. In the stage of optimization and prediction, we propose an IGWO optimized KELM algorithm, which is especially suitable for wind power prediction.

In this paper, ten different combined methods are compared. Four sets of wind power time series are selected from La Haute Borne wind farm to evaluate the performance of the proposed method. The results show that the method proposed in this paper can extract the trend information of wind power greatly and has better accuracy in short-term wind power prediction. The proposed three-level decomposition method can significantly increase the accuracy in short term wind power prediction. Future work will be taken place in two aspects: (a) the suitable length of the train set is variable when wind power significantly changed, and (b) the error sequence can be used to correcting the prediction.

REFERENCES

[1] E. Oh and H. Wang, "Reinforcement-Learning-Based energy storage system operation strategies to manage wind power forecast uncertainty," *IEEE Access*, vol. 8, pp. 20965–20976, 2020.

[2] M. Lange, and U. Focken, *Physical Approach to Short-Term Wind Power Prediction*. 2006.

[3] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "ARIMA-based time series model of stochastic wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010.

[4] M. B. Ozkan and P. Karagoz, "Data mining-based upscaling approach for regional wind power forecasting: Regional statistical hybrid wind power forecast technique (RegionalSHWIP)," *IEEE Access*, vol. 7, pp. 171790–171800, 2019.

[5] E. Erdem and J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction," *Appl. Energy*, vol. 88, no. 4, pp. 1405–1414, Apr. 2011.

[6] H. Liu, H.-Q. Tian, and Y.-F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," *Appl. Energy*, vol. 98, pp. 415–424, Oct. 2012.

[7] Z. Tian, S. Li, Y. Wang, and X. Wang, "Wind power prediction method based on hybrid kernel function support vector machine," *Wind Eng.*, vol. 42, no. 3, pp. 252–264, Jun. 2018.

[8] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 52, no. 4, pp. 1990–1998, Apr. 2011.

[9] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[10] M.-R. Chen, G.-Q. Zeng, K.-D. Lu, and J. Weng, "A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6997–7010, Aug. 2019.

[11] C. Zhang, M. Ding, W. Wang, R. Bi, L. Miao, H. Yu, and L. Liu, "An improved ELM model based on CEEMD-LZC and manifold learning for short-term wind power prediction," *IEEE Access*, vol. 7, pp. 121472–121481, 2019.

[12] Y. Zhang, H. Sun, and Y. Guo, "Wind power prediction based on PSO-SVR and grey combination model," *IEEE Access*, vol. 7, pp. 136254–136267, 2019.

[13] Z. Shi, H. Liang, and V. Dinavahi, "Wavelet neural network based multiobjective interval prediction for short-term wind speed," *IEEE Access*, vol. 6, pp. 63352–63365, 2018.

[14] Y. Hao, L. Dong, X. Liao, J. Liang, L. Wang, and B. Wang, "A novel clustering algorithm based on mathematical morphology for wind power generation prediction," *Renew. Energy*, vol. 136, pp. 572–585, Jun. 2019.

[15] H. Liu, H.-Q. Tian, X.-F. Liang, and Y.-F. Li, "Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks," *Appl. Energy*, vol. 157, pp. 183–194, Nov. 2015.

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

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

[18] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[19] O. Abedinia, M. Lotfi, M. Bagheri, B. Sobhani, M. Shafie-khah, and J. P. S. Catalao, "Improved EMD-based complex prediction model for wind power forecasting," *IEEE Trans. Sustain. Energy*, early access, Feb. 28. 2020, doi: 10.1109/TSTE.2020.2976038.

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

[21] H. Liu, X. Mi, and Y. Li, "An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm," *Renew. Energy*, vol. 123, pp. 694–705, Aug. 2018.

[22] Y. Zhang, J. Le, X. Liao, F. Zheng, and Y. Li, "A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing," *Energy*, vol. 168, pp. 558–572, Feb. 2019.

[23] L.-L. Li, X. Zhao, M.-L. Tseng, and R. R. Tan, "Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm," *J. Cleaner Prod.*, vol. 242, Jan. 2020, Art. no. 118447.

[24] M. Sharifzadeh, A. Sikinioti-Lock, and N. Shah, "Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian process regression," *Renew. Sustain. Energy Rev.*, vol. 108, pp. 513–538, Jul. 2019.

[25] W. Fu, K. Wang, C. Li, and J. Tan, "Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM," *Energy Convers. Manage.*, vol. 187, pp. 356–377, May 2019.

[26] N. Safari, C. Y. Chung, and G. C. D. Price, "Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 590–601, Jan. 2018.

[27] Z. Tian, Y. Ren, and G. Wang, "Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine," *J. Electr. Eng. Technol.*, vol. 13, no. 5, pp. 1841–1851, Sep. 2018.

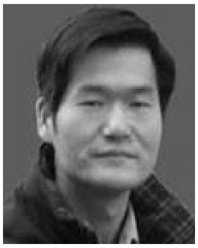
[28] Z. Tian, S. Li, Y. Wang, and X. Wang, "Network traffic prediction method based on improved ABC algorithm optimized EM-ELM," *The J. China Universities Posts Telecommun.*, vol. v.25, no. 3, pp. 37–48, 2018.

[29] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[30] T. Zhongda, "Kernel principal component analysis-based least squares support vector machine optimized by improved grey wolf optimization algorithm and application in dynamic liquid level forecasting of beam pump," *Trans. Inst. Meas. Control*, Nov. 2019, Art. no. 014233121988527.
[31] O. Abedinia, M. Bagheri, M. S. Naderi, and N. Ghadimi, "A new combinatory approach for wind power forecasting," *IEEE Syst. J.*, early access, Jan. 20, 2020, doi: 10.1109/JSYST.2019.2961172.
[32] N. Amjady and O. Abedinia, "Short term wind power prediction based on improved kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine," *Sustainability*, vol. 9, no. 11, p. 2104, 2017.
[33] J. Naik, S. Dash, P. K. Dash, and R. Bisoi, "Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network," *Renew. Energy*, vol. 118, pp. 180–212, Apr. 2018.
[34] S. Sun, J. Fu, and A. Li, "A compound wind power forecasting strategy based on clustering, two-stage decomposition, parameter optimization, and optimal combination of multiple machine learning approaches," *Energies*, vol. 12, no. 18, p. 3586, 2019.
[35] Z. Tian, G. Wang, S. Li, Y. Wang, and X. Wang, "Artificial bee colony algorithm–optimized error minimized extreme learning machine and its application in short-term wind speed prediction," *Wind Eng.*, vol. 43, no. 3, pp. 263–276, Jun. 2019.
[36] T. Ma, C. Wang, J. Wang, J. Cheng, and X. Chen, "Particle-swarm optimization of ensemble neural networks with negative correlation learning for forecasting short-term wind speed of wind farms in western China," *Inf. Sci.*, vol. 505, pp. 157–182, Dec. 2019.
[37] X.-W. Mi, H. Liu, and Y.-F. Li, "Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm," *Energy Convers. Manage.*, vol. 151, pp. 709–722, Nov. 2017.
[38] Z. Cai, J. Gu, J. Luo, Q. Zhang, H. Chen, Z. Pan, Y. Li, and C. Li, "Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112814.
[39] W.-T. Pan, "A new fruit fly optimization algorithm: Taking the financial distress model as an example," *Knowl.-Based Syst.*, vol. 26, pp. 69–74, Feb. 2012.
[40] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.
[41] T. Zhongda, L. Shujiang, W. Yanhong, and S. Yi, "A prediction method based on wavelet transform and multiple models fusion for chaotic time series," *Chaos, Solitons Fractals*, vol. 98, pp. 158–172, May 2017.
[42] Z. Tian, S. Li, and Y. Wang, "A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed," *Wind Energy*, vol. 23, no. 2, pp. 177–206, Feb. 2020.
[43] Z. Tian, Y. Ren, and G. Wang, "Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM," *Energy Sour. A, Recovery, Utilization, Environ. Effects*, vol. 41, no. 1, pp. 26–46, Jan. 2019.
[44] Z. Tian, G. Wang, Y. Ren, S. Li, and Y. Wang, "An adaptive online sequential extreme learning machine for short-term wind speed prediction based on improved artificial bee colony algorithm," *Neural Netw. World*, vol. 28, no. 3, pp. 191–212, 2018.

YUCHAO HAN was born in Shaanxi, China, in 1992. He received the bachelor's and master's degrees in electrical engineering from the Xi'an University of Technology, Xi'an, China, in 2015 and 2018, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include wind power prediction and machine learning.



XIANGQIAN TONG was born in Shaanxi, China, in 1961. He received the B.S. degree from the Shaanxi Institute of Technology, Hanzhong, China, in 1983, the M.S. degree from the Xi'an University of Technology, Xi'an, China, in 1989, and the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, in 2006.

In 1989, he joined the Xi'an University of Technology, where he has been a Professor and the Academic Leader of electrical engineering, since 2002. His research interests include the application of power electronics in power systems and control of power quality.