Energy Conversion and Management 156 (2018) 525–541

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm

Hui Liu, Zhu Duan, Feng-ze Han, Yan-fei Li*

Institute of Artificial Intelligence and Robotics, Key Laboratory for Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, Hunan, China

ARTICLE INFO

Keywords:
Big multi-step wind speed forecasting
Wavelet decomposition
Variational mode decomposition
Sample entropy
Modified adaBoost.RT
Wavelet filter

ABSTRACT

Wind power is one of the most promising powers. Wind speed forecasting can eliminate the harmful effect caused by the intermittent and fluctuation of wind power, and big multi-step forecasting can provide more time for the power grid to be adjusted. To achieve the high-precision big multi-step forecasting, a novel hybrid model named as the WD-SampEn-VMD-MadaBoost-BFGS-WF is proposed in the study, which consisting of three main modeling steps including the secondary decomposition, the ensemble method and the error correction. The detail of the proposed model is given as follows: (a) wind speed series are decomposed by the WD (*Wavelet Decomposition*) to obtain wind speed subseries. The SampEn (*Sample Entropy*) algorithm is used to estimate the unpredictability of these wind speed subseries. The most unpredictable subseries will be decomposed secondarily by the VMD (*Variational Mode Decomposition*); (b) the subseries are proceeded by the MAdaBoost (*Modified AdaBoost.RT*) with the BFGS (*Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton Back Propagation*) neuron network to obtain forecasting subseries; (c) all of the forecasting subseries will be combined with the original subseries to form the combined wind speed series, which will be further proceeded by the WF (*Wavelet Filter*) to obtain the corrected forecasting series from the point of the frequency domain; (d) the corrected forecasting series are reconstructed to get the final forecasting series. To validate the effectiveness of the proposed model, several forecasting cases are provided in the study. The result indicates that the proposed model has satisfactory forecasting performance in the big multi-step extremely strong simulating wind speed forecasting.

1. Introduction

Energy storage and pollution has hindered the usage of petroleum, natural gas, etc. Comparing to those conventional energy, the wind power is renewable, which catches the world's attention. Recently, the wind power increases rapidly [1]. The worldwide wind power reached 486.661 MW at the end of 2016, with 54.846 MW added in 2016, which represents a growth rate of 11.8%. The electricity, which is generated by the wind turbines installed worldwide by the end of 2016, is around 5% of the world's electricity demand [2].

In the process of using the wind power, accurate and stable forecasting of wind speed is expected, which can achieve the timely and effective management of the power grid and help eliminate the adverse impact in the growing wind energy scenario as well as increasing the revenues from the power market with bidding strategies optimized [3]. However, because of the inherent randomness and volatility of wind speed, it is a big challenge to propose an effective and practical method to forecasting the wind speed accurately in the big multiple steps. The proposed wind speed forecasting methods can be classified as four main types as: (a) statistical methods, (b) physical methods, (c) intelligent methods and (d) hybrid methods [4]. The physical methods include the forecasting models as spatial correlation, NWP (*Numerical Weather*

Abbreviations: NWP, numerical weather prediction; CS, Cuckoo search; FS, fuzzy system; WRF, weather research and forecasting; KF, Kalman filter; ARIMA, auto-regressive integrated moving average; ARCH, autoregressive conditional heteroskedasticity; ANN, artificial neural networks; SVM, support vector machine; CRO, Coral Reefs optimization algorithm; ELM, extreme learning machine; MFNN, multi-layer feed-forward neural network; SPSA, simultaneous perturbation stochastic approximation; HM, Hammerstein Model; AR, auto-regressive; AdaBoost, adaptive boosting; MLP, multilayer perceptron; DNN-MRT, deep neural network based meta regression and transfer learning; WD, wavelet decomposition; FEEMD, fast ensemble empirical mode decomposition; EMD, empirical mode decomposition; WPD, wavelet packet decomposition; SSA, singular spectrum analysis; BFGS, Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton Back Propagation; LSSVM, least square support vector machine; PSOGSA, partial swarm optimization combined with gravitational search algorithm; FCM, fuzzy C-means; EEMD, ensemble empirical mode decomposition; SampEn, sample entropy; VMD, variational mode decomposition; MAdaBoost, Modified AdaBoost.RT; WF, wavelet filter; MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean squared error; CWT, continuous wavelet transform; DWT, discrete wavelet transform; LMD, local mean decomposition; ADMM, alternate direction method of multipliers

* Corresponding author.
E-mail address: yanfeili@csu.edu.cn (Y.-f. Li).

https://doi.org/10.1016/j.enconman.2017.11.049
Received 15 October 2017; Accepted 18 November 2017
Available online 28 November 2017
0196-8904/

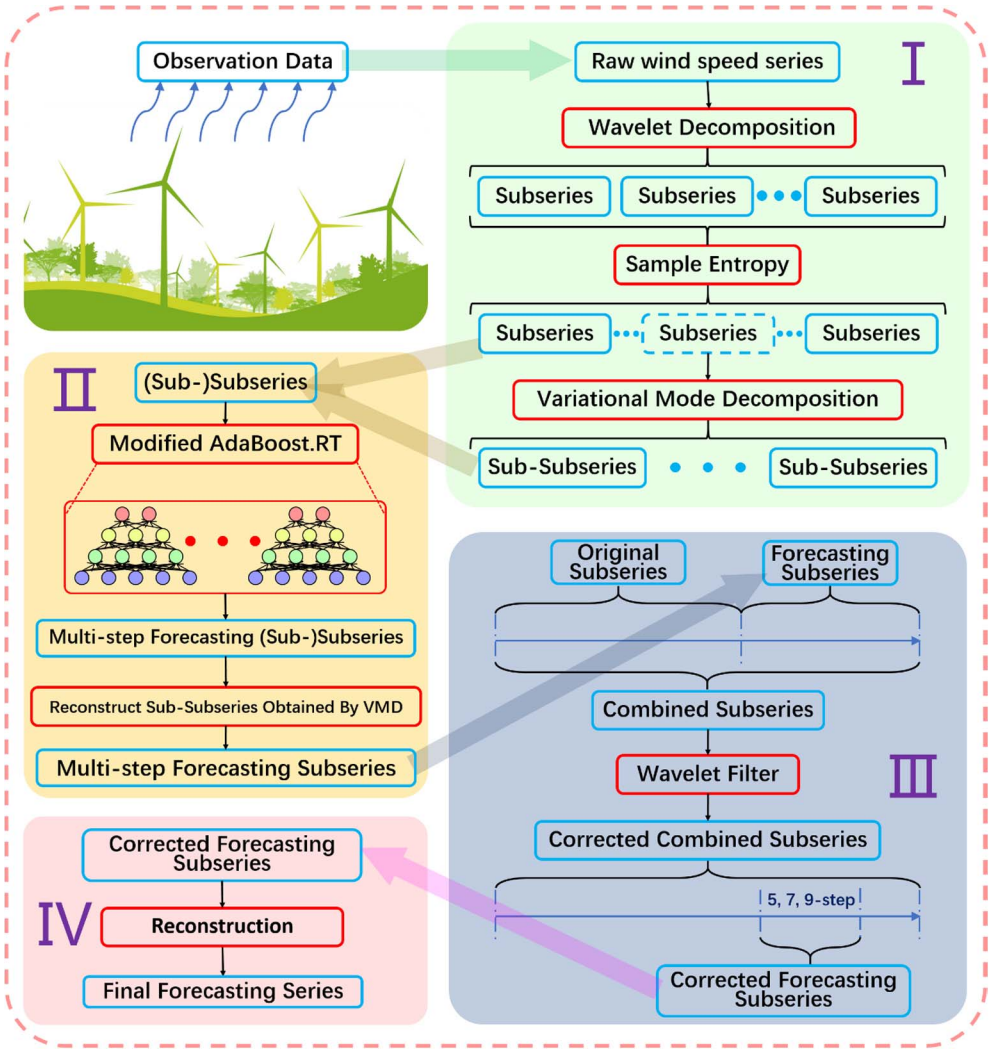

Fig. 1. The framework of the proposed forecasting model.

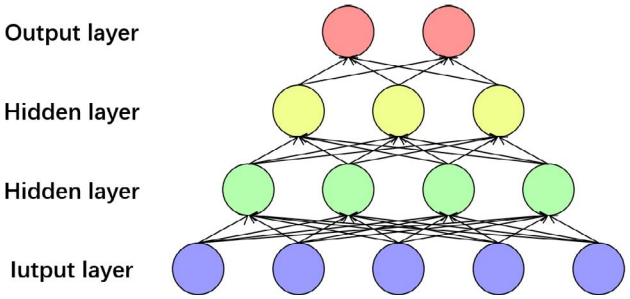

Fig. 2. The structure of the adopted BFGS neural network.

Prediction), etc [5]. Zhao et al. [6] put forward a CS-FS-WRF-E model consisting of the CS (*Cuckoo Search*) algorithm, the FS (*Fuzzy System*) and the WRF (*Weather Research and Forecasting*). Cassola and Burlando [7] proposed a new wind speed and wind power forecasting approach using the KF (*Kalman Filter*) to process the generations of a NWP model. The KF is aimed to minimize the predicting errors. In general, the physical forecasting models always have worse accuracy in the short-term forecasting than the statistical methods [4]. Besides, the real-time performance of the physical forecasting models is also an important point for them to be applied in the real-time forecasting systems [8]. In terms of the statistical methods, the most widely used model is ARIMA (*Auto-Regressive Integrated Moving Average*) model, Masseran [9] introduced a forecasting model composed of the ARIMA and the ARCH (*Autoregressive Conditional Heteroskedasticity*), which is proved that the proposed model had better performance than single ARIMA model. In addition, the statistical methods also can be used to improve the accuracy performance of other forecasting models. Wang et al. [10] proposed residual modification models to improve the precision of seasonal ARIMA, and the performance of the modification models appeared to be more workable than that of the single seasonal ARIMA. *Liu* et al. [11] used ARIMA model to discover the statistical rule of wind speed series. At the same time, the built ARIMA model was also adopted to determine the initial parameter of the Kalman filtering and the numbers of inputting neurons of the ANN (*Artificial Neural Networks*) forecasting models. Although the statistical methods are simple and effective, they have potential to be further promoted in the future. Because the intelligent methods are smart and robust, they are widely adopted as one of the basic foresting frameworks to forecast the non-stationary wind speed. There are numerous algorithms in the wind speed intelligent modeling and forecasting, such as ANN, SVM (*Support Vector Machine*), etc. Li and Shi [12] investigated three types of typical neural networks, namely, the adaptive linear element, the back propagation and the radial basis function, to investigate the wind speed forecasting performance based on the neural networks. Salcedo-Sanz et al. [13] presented a novel approach for the short-term wind speed prediction based on a CRO (*Coral Reefs Optimization Algorithm*) and an ELM (*Extreme Learning Machine*). Hong et al. [14] proposed a new method of wind power and speed forecasting using a MFNN (*Multi-layer Feed-forward Neural Network*) trained by the SPSA (*Simultaneous*

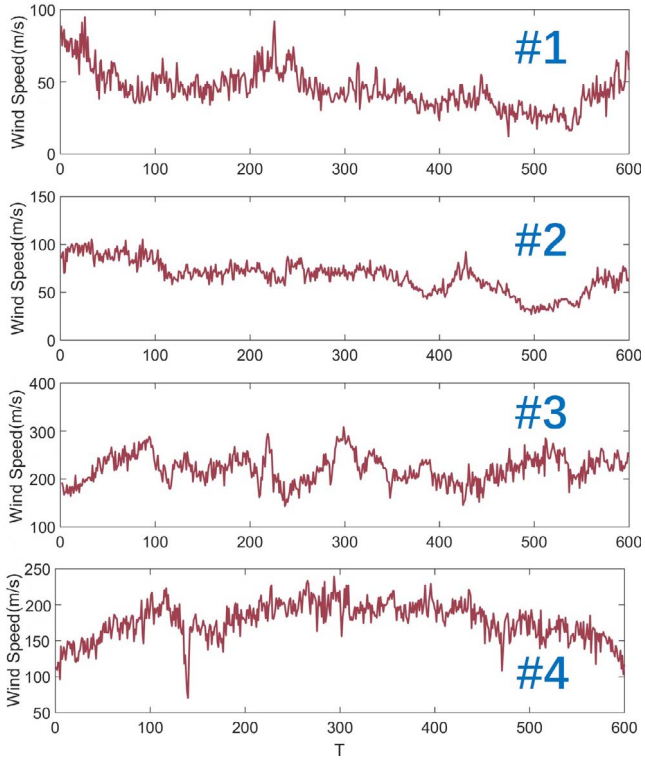

Fig. 3. The experimental modeling wind speed time series.

Table 1
The statistical parameters of the simulated wind speed data.

Series	Mean	Standard Deriation	Min	Max
#1	44.66	13.62	95	12
#2	66.48	16.65	105	27
#3	211.12	29.79	308	143
#4	177.08	26.06	239	70

Table 2
The SampEn results of the subseries after the WD decomposition.

Subseries	#1	#2	#3	#4	#5	#6
SampEn	2.4445	2.7807	2.7055	2.7228	2.7633	**2.8012**

Table 3
The SampEn results of the subseries #6 after the VMD decomposition.

Sub-Subseries	#1	#2	#3
SampEn	2.7401	2.7560	2.7280

Perturbation Stochastic Approximation) algorithm to forecast different time-scales series, whose inputs were selected by the fuzzy. Ait Maatallah et al. [15] put forward an artificial intelligent wind speed forecasting model with the HM (*Hammerstein Model*) and the AR (*Auto-Regressive*) approach. The results proved that this intelligent model outperformed the conventional method like the ARIMA model and the ANN model. The hybrid methods are one kind of methods that they combine the forecasting abilities of different forecasting methods to get better performance than the single forecasting methods. In the hybrid methods, there are several common ensemble algorithms which are popularly used in the wind speed forecasting, such as the AdaBoost (*Adaptive Boosting*) [16], the stacked generalization algorithms [17,18]. Liu et al. [16] proposed four different hybrid methods for the high-precision multi-step wind speed predictions based on the AdaBoost algorithm and the MLP (*Multilayer Perceptron*) neural networks. Qureshi et al. [17] proposed a DNN-MRT (*Deep Neural Network based Meta Regression and Transfer Learning*) scheme based on the stacked generalization structure, in which the deep auto-encoders acted as a base-regressor, whereas the deep belief network worked as a meta-regressor. Zameer et al. [18] put forward an effective short-term wind power prediction methodology combining several artificial neural networks and genetic programming. It was proved that the wind speed decomposition was one of the most effective processing algorithms in the wind speed predictions [19]. For instance, the WD (*Wavelet Decomposition*) [20,21], FEEMD (*Fast Ensemble Empirical Mode Decomposition*) [22], EMD (*Empirical Mode Decomposition*) [23], WPD (*Wavelet Packet Decomposition*) [24], SSA (*Singular Spectrum Analysis*) [5] is widely used. Besides these single processing methods, the combined signal processing methods are always more effective. Liu et al. [25] demonstrated a secondary decomposition algorithm consisting of the WPD and the FEEMD. In the experiments, it was proved that the WPD-FEEMD-BFGS model outperformed both of the WPD-BFGS model and the standard BFGS (*Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton Back Propagation*) neural networks considerably. Beside the upper kind of hybrid methods, the error correction is also one of the useful measures to promote the forecasting performance of the hybrid methods. Wang et al. [26] presented a novel hybrid intelligent forecasting model based on the phase space reconstruction and the LSSVM (*Least Square Support Vector Machine*) optimized by the PSOGSA (*Partial Swarm Optimization Combined With Gravitational Search Algorithm*). The Markov model was used to correct the forecasting errors with the state ranges determined by the FCM (*Fuzzy C-Means*) model. Jiang et al. [27] proposed a method that decomposed the original wind speed series into a group of wind speed subseries by the EEMD (*Ensemble Empirical Mode Decomposition*). The proposed method could reduce the disturbance of illusive components by two feature selection methods, and corrected the errors by the combination form of the least squares support vector machine and the generalized auto-regressive conditionally heteroscedastic model. Different to the other published works, in this study the big multi-step wind speed forecasting is focused. Due to the complexity of mapping relationships, the big multi-step forecasting is more difficult and complicated than the small or short multi-step forecasting. To archive the big multi-step wind speed forecasting, the hybrid modeling strategy is adopted. In addition, the recursive computation as given in [28] is executed to realize the big multi-step forecasting.

In the study, a new big multi-step wind speed hybrid forecasting structure is proposed, which consists of a secondary decomposition algorithm (*Wavelet Decomposition + Sample Entropy + Variational Mode Decomposition*), an ensemble method (*Modified AdaBoost.RT*) and error correction method (*Wavelet filter*). More details of the proposed hybrid computation are given as follows:

(a) The original wind speed series are decomposed by the WD to get a group of wind speed subseries. The SampEn (*Sample Entropy*) is employed to estimate the unpredictability of each wind speed subseries. The most unpredictable subseries among them will be further decomposed by the VMD (*Variational Mode Decomposition*) to obtain the secondary sub-subseries.
(b) The MAdaBoost (*Modified AdaBoost.RT*) is used as an ensemble model which could enhance the forecasting ability of the individual forecasting methods, while the BFGS neuron network is selected as a base regressor. All of the subseries obtained by the executed secondary decomposition method will be processed by the MAdaBoost with the BFGS neuron network method to get the multi-step forecasting wind speed subseries. The multi-step forecasting sub-subseries will be reconstructed into the multi-step forecasting subseries.
(c) All of the forecasting subseries will be combined with the original wind speed subseries which are in front of the multi-step

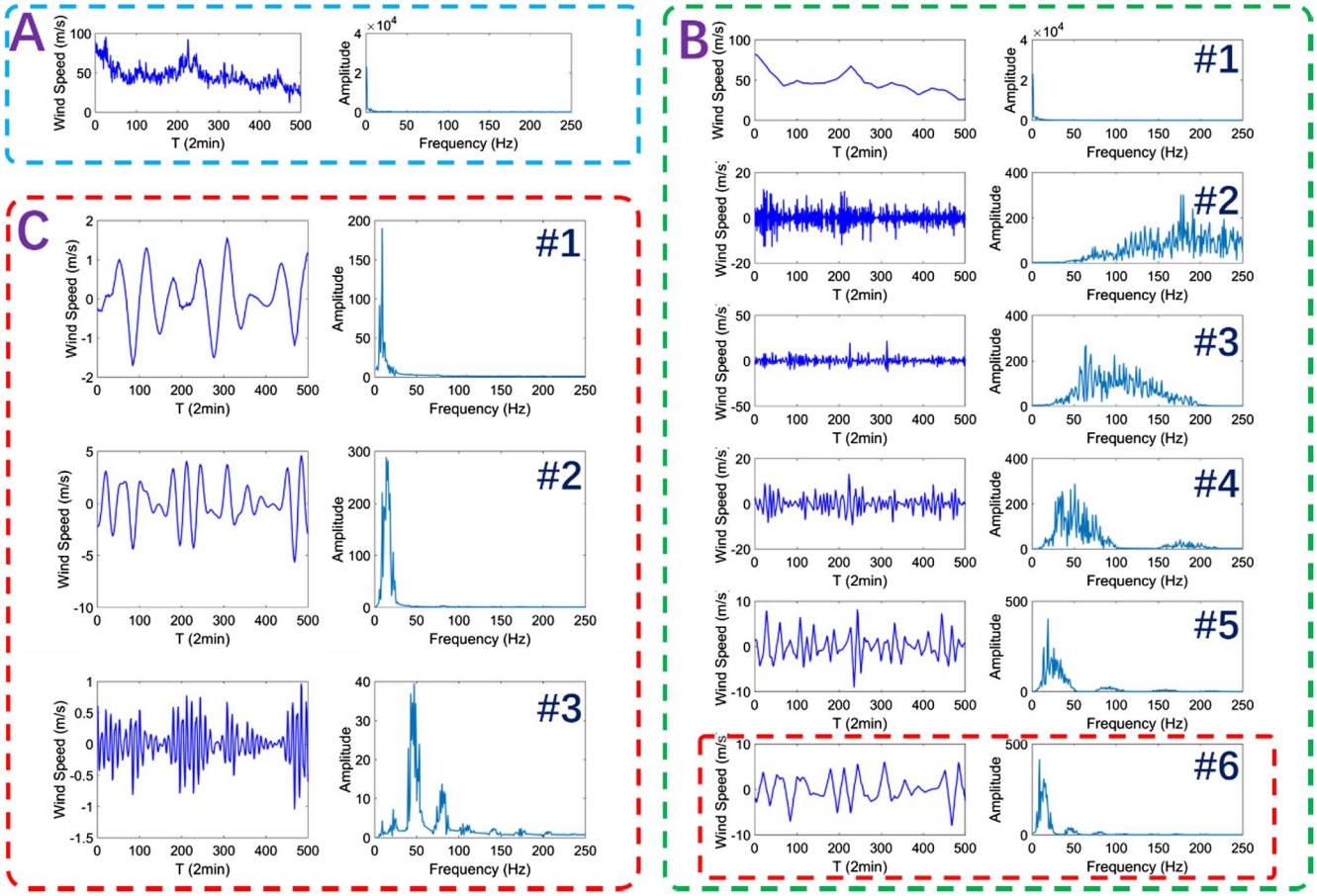

Fig. 4. The process results of the executed secondary decomposition.

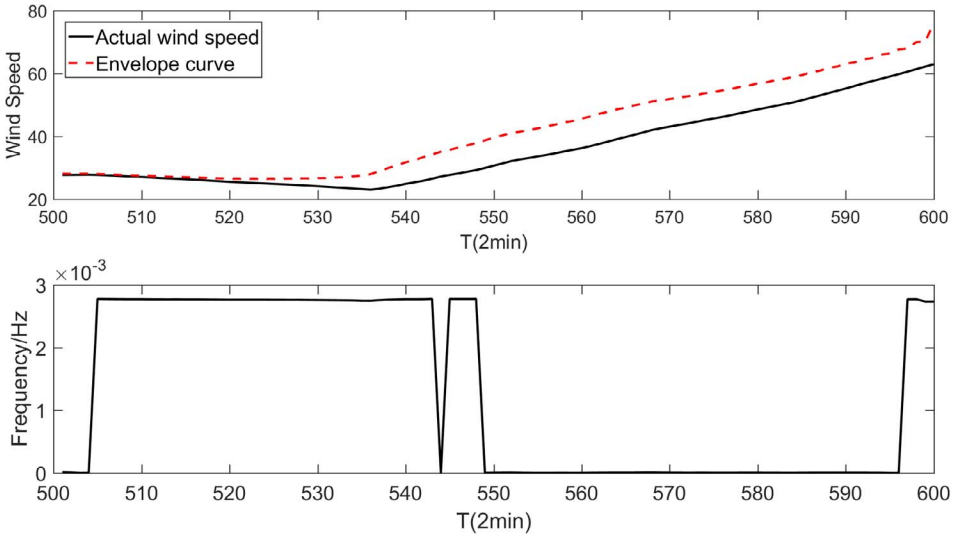

Fig. 5. The time-domain and instantaneous frequency plots of the actual subseries #1 of the series #1.

forecasting subseries to create a new combined wind speed subseries. The WF (*Wavelet filter*) is employed to correct all forecasting errors, which smooth the extra frequency components of the combined subseries to make the frequency spectrum of combined subseries is close to the frequency spectrum of the actual subseries. Thereafter the corrected forecasting subseries are obtained in the corresponding forecasting steps.

(d) The corrected multi-step forecasting subseries are reconstructed into the final multi-step forecasting series to finalize the whole hybrid forecasting computation.

The innovation and scientific contribution of the proposed hybrid forecasting method are further explained as follows: (a) the secondary composition-ensemble method-error correction structure is investigated. The secondary composition in the proposed hybrid structure can enhance the composition ability, decrease the mode mixing problem and make the wind speed subseries smoother to be predicted. At the same time, to avoid the significant increasing of the computation demand caused by the adopted secondary decomposition, the SampEn algorithm is utilized to select the most unpredictable subseries to be executed the secondary decomposition; (b) the Modified AdaBoost.RT

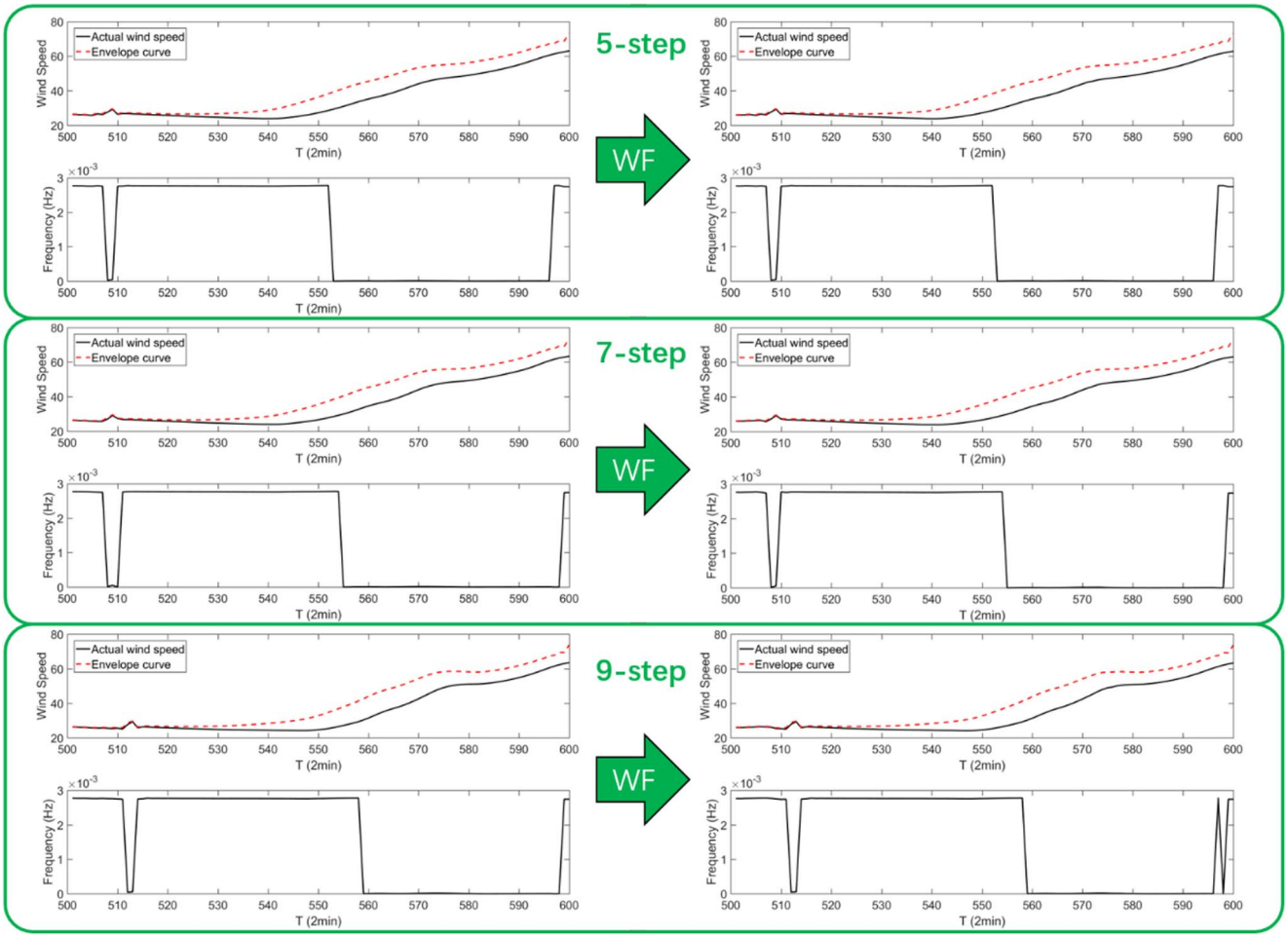

Fig. 6. The effecting phenomenon of the WF computation for the time-domain and instantaneous frequency plots of the subseries #1 of the series #1.

Table 4
The instantaneous frequency cross-correlation coefficient analysis of the subseries #1 of the series #1.

	Before WF			After WF		
Forecasting Step	5-step	7-step	9-step	5-step	7-step	9-step
Cross-correlation Coefficient	0.894	0.842	0.824	0.894	0.854	0.836

algorithm is used as an ensemble method to enhance the base regressor. Although the Adaboost method has been proposed for different signal processing engineering applications for a few decades, the investigation of the AdaBoost algorithm is insufficient in the wind speed forecasting; (c) the Wavelet filter is employed as an error correcting method; (d) to demonstrate the efficiency of the proposed hybrid forecasting method, several forecasting cases are provided in the study. In the given cases, different forecasting algorithms are used, which include the WD-SampEn-VMD-BFGS, the MAdaBoost-BFGS, the BFGS-WF and the single BFGS. In addition, the different error estimating standards are executed to find the real performance of the proposed hybrid method, which includes the MAE (*Mean Absolute Error*), the MAPE (*Mean Absolute Percentage Error*), the RMSE (*Root Mean Squared Error*) and the Pearson test.

This paper is organized as follows: (a) the framework of the hybrid forecasting model in this study is provided in Section 2; (b) the algorithms involved in the proposed model are explained in Section 3; (c) model analysis and four wind speed forecasting case study are presented in Section 4; and (d) Section 5 concludes the paper.

2. Framework of the proposed hybrid model

The framework of the proposed hybrid model in the study is demonstrated in Fig. 1. As shown in Fig.1, the proposed study can be summarized as follows:

- In the processing I, the original wind speed series is decomposed by the WD, and the SampEn is used to evaluate the unpredictability of each wind speed subseries. These subseries having the high entropy characters will be further decomposed by the VMD to generate the wind speed sub-subseries.
- In the processing II, the MAdaBoost algorithm with the BFGS neuron network is proposed to forecast all of these (sub-)subseries obtained in the upper processing I. The obtained multi-step forecasting wind speed sub-subseries are reconstructed to generate the multi-step forecasting subseries.
- The multi-step forecasting subseries are combined with the original subseries to obtain the combined wind speed subseries. The WF deprives the extra frequency components of the combined wind speed subseries. The corresponding forecasting step data in the corrected combined subseries are taken as the corrected forecasting subseries.
- In the processing IV, all the corrected forecasting subseries are reconstructed to obtain the final forecasting series.

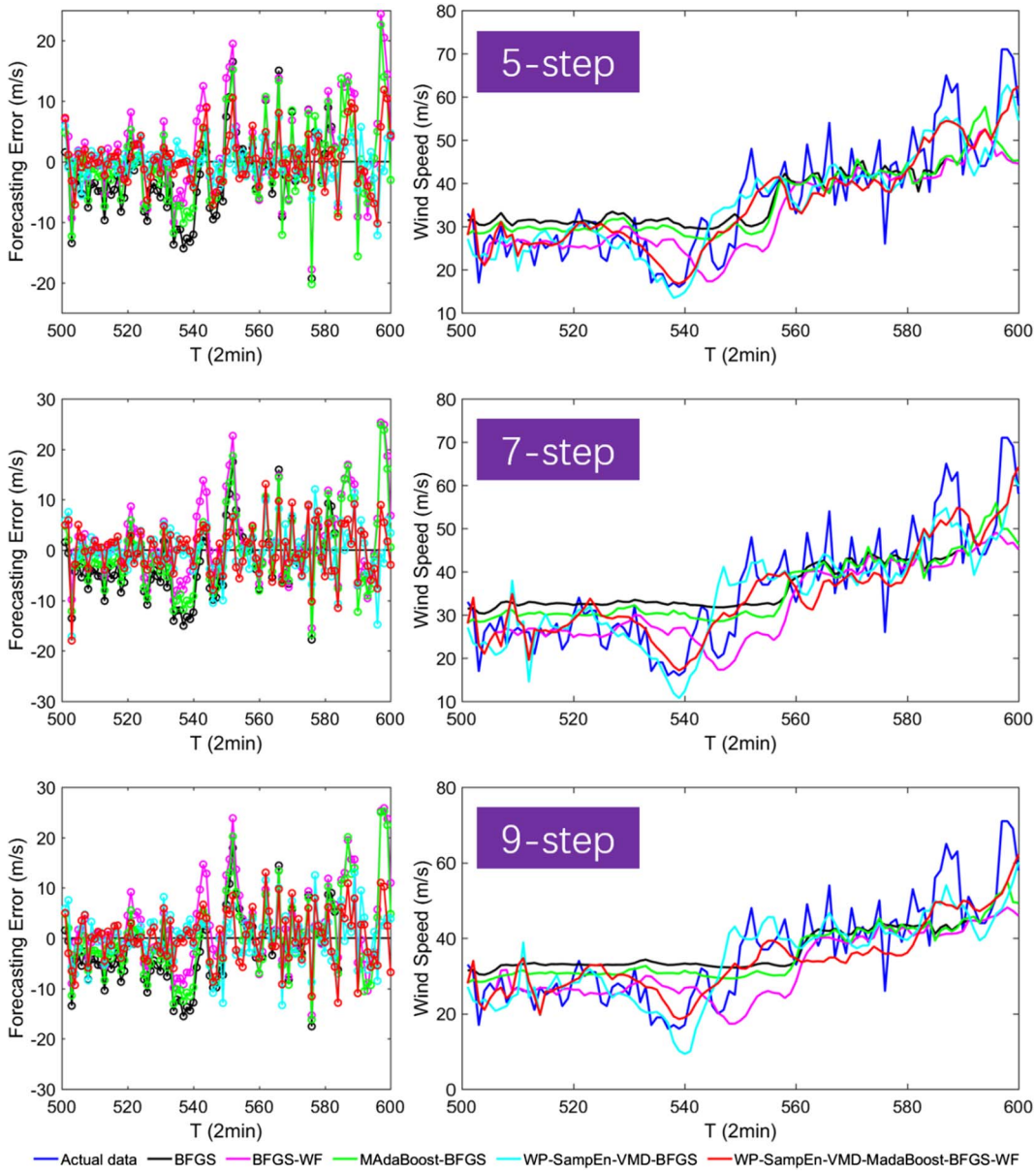

Fig. 7. The forecasting results of the series #1.

3. Methodology

3.1. Time series decomposition method

3.1.1. Wavelet decomposition

The WD is a mathematic method, which has been generally used to process the non-stationary signals. The WD decomposes the original signals into different frequencies. In this study, the WD is used to analyze the original wind speed series into different subseries and to improve the ANN forecasting potential performance.

1) **Continuous Wavelet Transform**

The CWT (*Continuous Wavelet Transform*) can decompound a signal $f(t)$ with a mother wavelet function $\psi(t)$. it can be defined as below [29].

$$CWT_f(a,b) = \langle f(t),\psi(t)\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)dt \tag{3.1}$$

Where $*$ denotes the complex conjugate, a is a scale coefficient and b is a translation coefficient.

2) **Discrete Wavelet Transform**

The DWT (*Discrete Wavelet Transform*) is a digital counterpart of the CWT. It can offer necessary information and reduce the time of calculation. It can be defined as below.

$$\begin{cases} a = 2^j \\ b = k \times 2^j \end{cases} \tag{3.2}$$

The scale and translation coefficient is j and k, respectively.

3) **Mallat algorithm**

The Mallat algorithm is used to process the multi-solution DWT computation. The function of DWT works as a series of filters. The process of Wavelet decomposition of a signal can be regarded as dividing the signal into a hierarchy of step-by-step approximations. The signal is decompounded into a high pass and low pass filter. The signal A means the low-frequency information. The signal D represents the

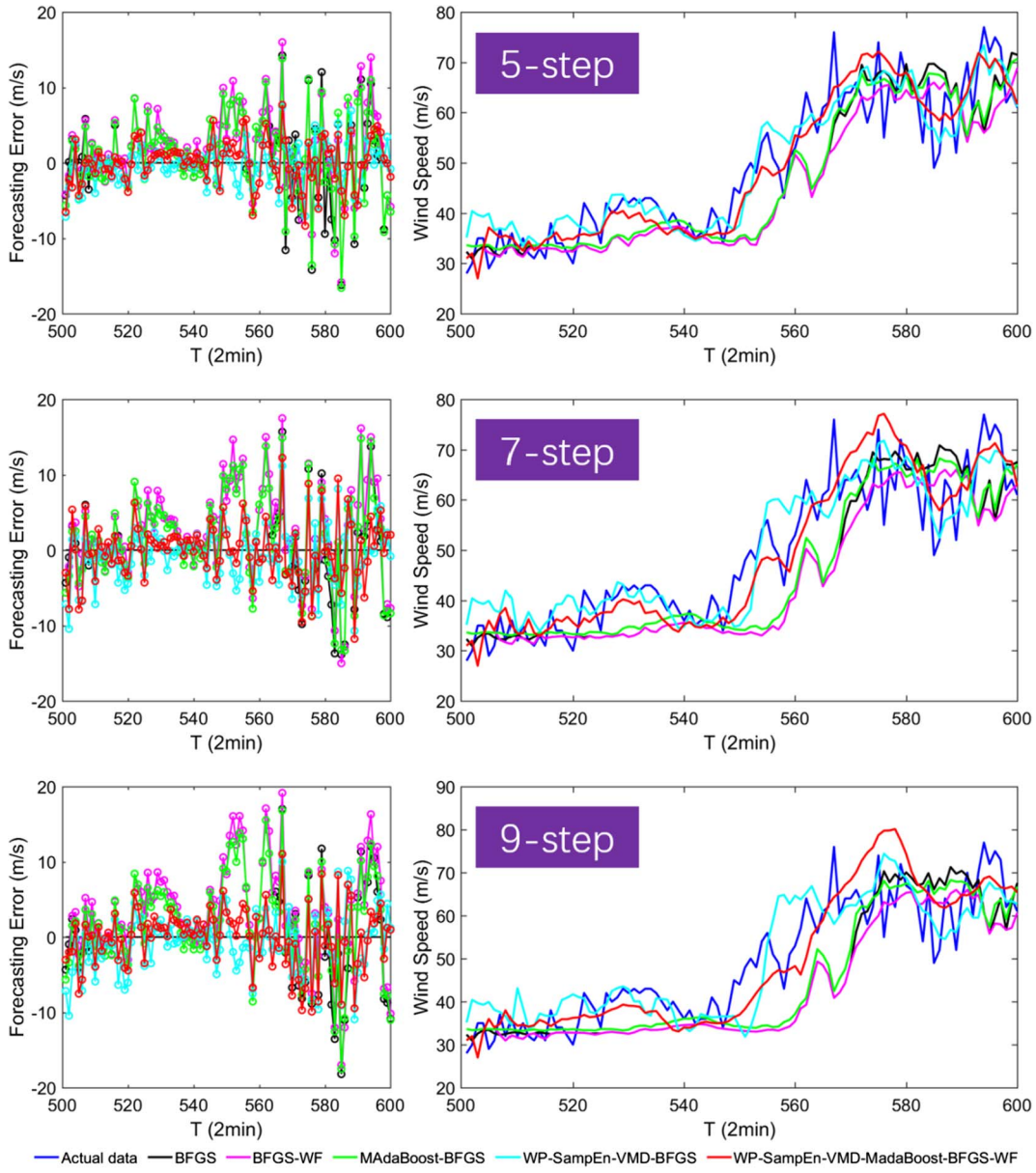

Fig. 8. The forecasting results of the series #2.

high-frequency approximation component. In the study, the number of decomposition layer is chosen as 5 and the wavelet function is selected as sym3.

3.1.2. Sample Entropy

Entropy is originally a thermodynamic concept, which is used to describe the degree of disorder of a thermodynamic system. SampEn, which is proposed by Richman and Moorman [30], has been widely used in statistical fields, such as the electroencephalography [31] and the battery health monitoring [32]. The SampEn is a modification of approximate entropy, which has two advantages as: the data length independence and the relatively trouble-free implementation [33].

The SampEn has three parameters as: the time series length N, the run length m and the tolerance window r. Assuming there is a time series of size N: $\mathbf{x} = \{x(1),...,x(N)\}$, the concrete step in Sample Entropy can be given as follows [30].

(1) Reconstruct $\mathbf{x}$ into matrix form, the reconstructed matrix is $\mathbf{X} = \{X_m(1),...,X_m(N-m+1)\}$, where $X_m(i) = \{x(i),x(i+1),...,x(i+m-1)\}$, for $1 \leqslant i \leqslant N-m+1$.

(2) Define the distance of $X_m(i)$ and $X_m(j)$, $d[X_m(i),X_m(j)]$, as the absolute maximum difference between their scalar components:

$$d[X(i),X(j)] = \max_{k=0\sim m-1} |x(i+k)-x(j+k)|. \tag{3.3}$$

(3) For given $X_m(i)$, count the number of elements in a collection $\{j|d[X_m(i),X_m(j)] < r, j \neq i, 1 \leqslant j \leqslant N-m\}$ as B_i. Define the ratio of $N_m(i)$ and total quantity of distance as below:

$$B_i^m(r) = \frac{B_i}{N-m+1}. \tag{3.4}$$

(4) For all i, compute the average of $B_i^m(r)$, written as $B^m(r)$, which means the probability that two sequences will match for m points:

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m(r). \tag{3.5}$$

(5) Increase m into $m = m+1$, and calculate $B^{m+1}(r)$ as step (1–4).

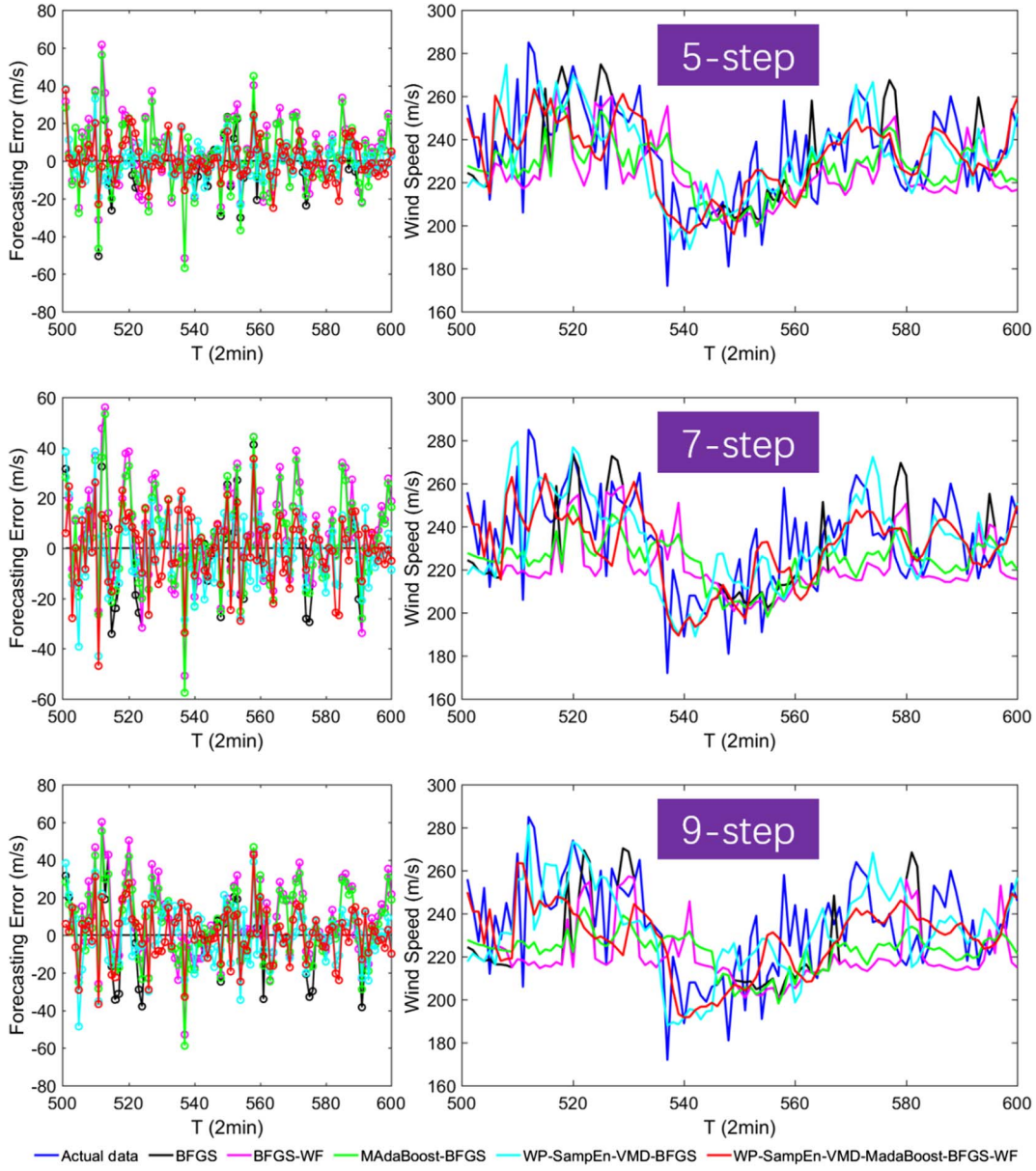

Fig. 9. The forecasting results of the series #3.

(6) Theoretically, the SampEn of **x** is defined as below:

$$SampEn(m,r) = \lim_{N\to\infty}\left\{-ln\left[\frac{B^{m+1}(r)}{B^{m}(r)}\right]\right\}. \tag{3.6}$$

In fact, time series length cannot be infinite. when N is finite, the SampEn can be estimate by the statistic as:

$$SampEn(\mathrm{N},m,r) = -\ln\left[\frac{B^{m+1}(r)}{B^{m}(r)}\right]. \tag{3.7}$$

The Parameter of the SampEn can affect the result. If the value of m is small and the value of r is high, the estimate of the SampEn can be accurate. But if the value of r is too high, some details could lost, and the SampEn tends to 0 for all processes [34]. When tuning parameter, the time series length N is determined once input time series is determine, m and r can be tuned according to [35], In this paper, $m = 1$, and r is 0.1 times the standard deviation of the original time series.

3.1.3. Variational mode decomposition

The VMD is a kind of state-of-art decomposition algorithm proposed by Dragomiretskiy and Zosso [36], which can decompose a multi-component signal into a number of quasi-orthogonal intrinsic mode functions by the non-recursively way [37]. Comparing with the LMD (*Local Mean Decomposition*) and the EMD, the VMD avoid error caused during calculate recursively and ending effect.

Given f is an input signal, $\{u_k\} = \{u_1,...,u_K\}$ and $\{\omega_k\} = \{\omega_1,...,\omega_K\}$ are all modes and their center frequencies. The theory of VMD is demonstrated as follows [36].

(1) **Construction of variational problem**

1) The analytic signal of each mode is obtained through Hilbert transform to get unilateral spectrum:

$$\left(\sigma(t) + \frac{j}{\pi t}\right) * \mathrm{u}_k(t) \tag{3.8}$$

2) The analytic signal is modulated by corresponding center frequencies:

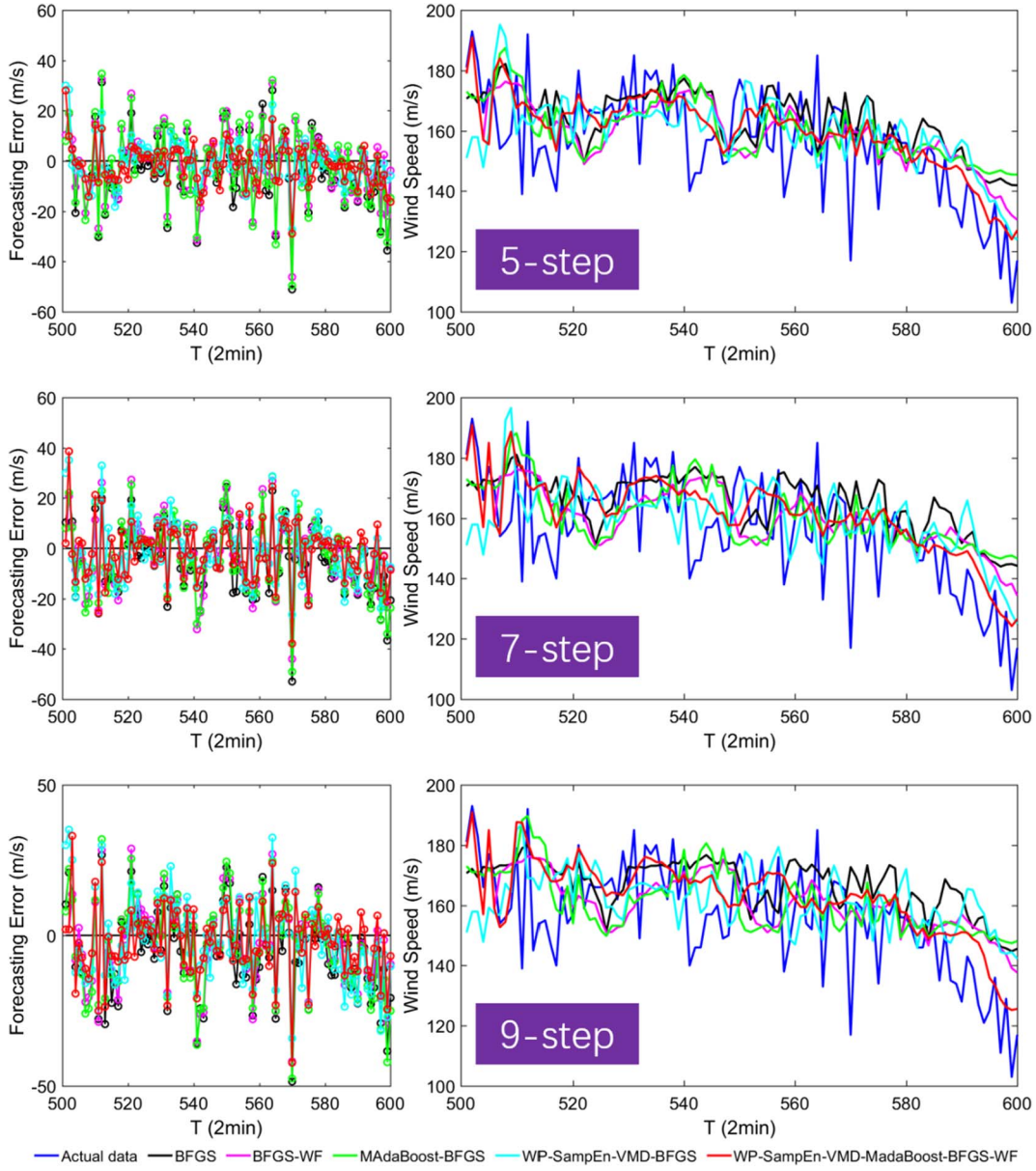

Fig. 10. The forecasting results of the series #4.

$$\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mathrm{u}_k(t)\right]e^{-j\omega_k t} \tag{3.9}$$

3) Calculating the L^2 -norm of gradient of those modulated signal, and estimating the band width of each modes, the variational problem is shown as below.

$$\{u_k,\omega_k\}=\operatorname{argmin}\left\{\sum_{k=1}^{K}\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mathrm{u}_k(t)\right]e^{-j\omega_k t}\right\|^2\right\}$$
$$\text{s.t.}\sum_{k=1}^{K}u_k=f \tag{3.10}$$

(2) **The solution of the variational problem**

1) Introducing the quadratic penalty α and the Lagrangian multipliers $\lambda(t)$ to turn the primal constrained problem into dual unconstrained problems, where the α can keep reconstruction accuracy, and the $\lambda(t)$ can ensure duel problem is be equivalent to the primal problem. Dual problem is shown as follow.

$$L\{\{u_k\},\{\omega_k\},\lambda\}=\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mathrm{u}_k(t)\right]e^{-j\omega_k t}\right\|^2+$$
$$\|f(t)-\sum_{k=1}^{K}u_k(\mathrm{t})\|_2^2+\left\langle\lambda(t),\mathrm{f}(t)-\sum_{k=1}^{K}u_k(t)\right\rangle \tag{3.11}$$

2) The mentioned variation problem is solved by the ADMM (*Alternate Direction Method of Multipliers*), which can find the saddle point of the Lagrangian expression by updating u_k^{n+1}, ω_k^{n+1} and λ^{n+1}. u_k^{n+1} can be updated using the equation as below:

$$u_k^{n+1}=\underset{u_k\in X}{\operatorname{argmin}}\left\{\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mathrm{u}_k(t)\right]e^{-j\omega_k t}\right\|^2\right.$$
$$\left.+\left\|f(t)-\sum_{k=1}^{K}u_k(t)+\frac{\lambda(t)}{2}\right\|_2^2\right\} \tag{3.12}$$

Where the ω_k is an identical parameter with ω_k^{n+1}, which can be solved in the spectral domain by the following Parseval Fourier

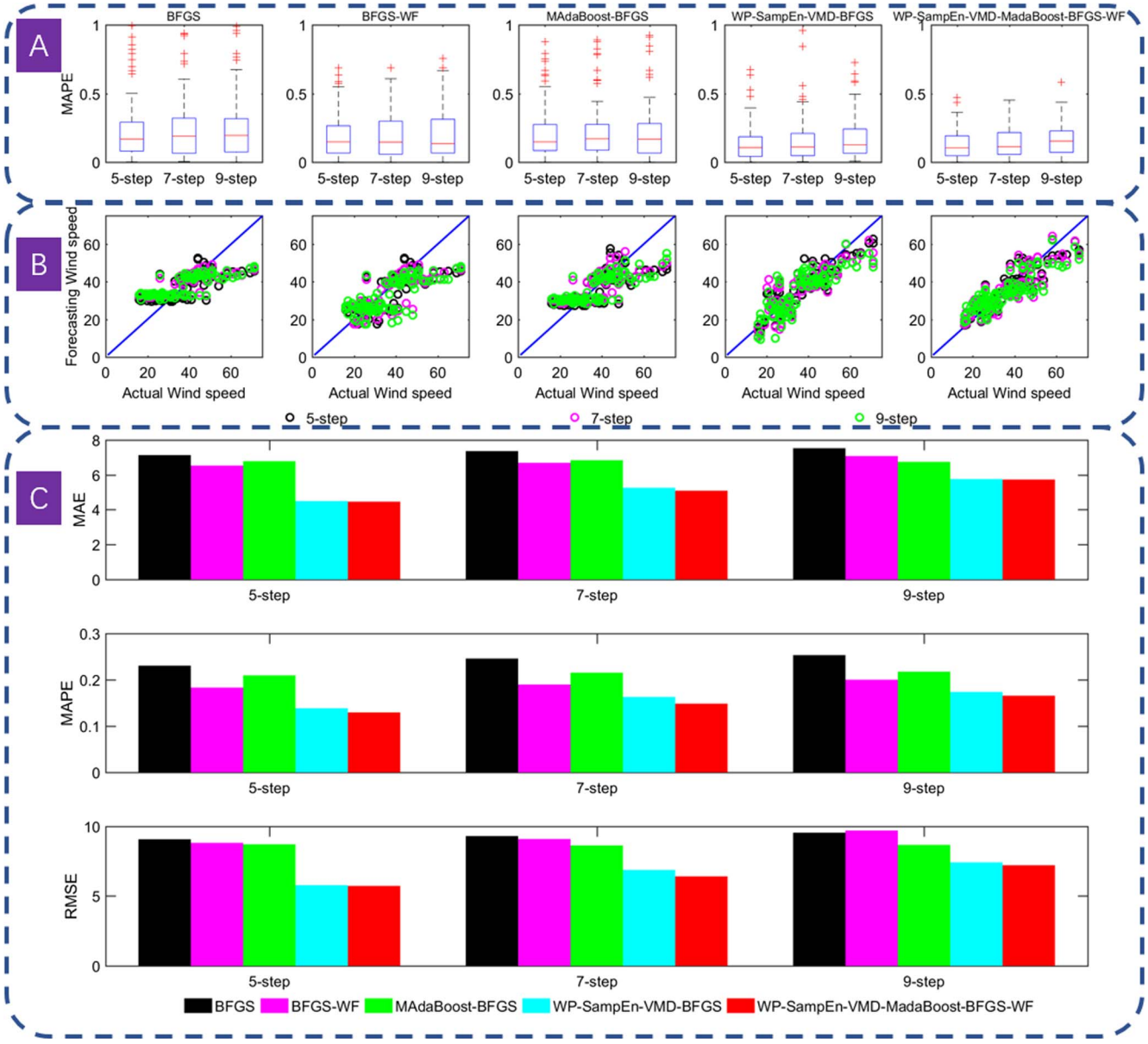

Fig. 11. The error evaluation analysis of the series #1.

isometry as:

$$\hat{u}_k^{n+1}(\omega) = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \alpha \| j\omega [(1 + \operatorname{sgn}(\omega + \omega_k))\hat{u}_k(\omega + \omega_k)] \|_2^2 + \left\| \hat{f}(t) - \sum_{k=1}^{K} \hat{u}_k(\omega) + \frac{\lambda(\omega)}{2} \right\|_2^2 \right\} \tag{3.13}$$

Then the ω is changed by $\omega - \omega_k$.

$$\hat{u}_k^{n+1}(\omega) = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \alpha \| j(\omega - \omega_k) [(1 + \operatorname{sgn}(\omega))\hat{u}_k(\omega)] \|_2^2 + \left\| \hat{f}(t) - \sum_{k=1}^{K} \hat{u}_k(\omega) + \frac{\lambda(\omega)}{2} \right\|_2^2 \right\} \tag{3.14}$$

Those problems can be changed into half-space integrals over the non-negative frequencies:

$$\hat{u}_k^{n+1}(\omega) = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \int_0^{\infty} 4\alpha(\omega - \omega_k)^2 |\hat{u}_k(\omega)|^2 + 2 \left| \hat{f}(t) - \sum_{k=1}^{K} \hat{u}_k(\omega) + \frac{\lambda(\omega)}{2} \right|^2 d\omega \right\} \tag{3.15}$$

The solution of this quadratic optimization problem can be defined as:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3.16}$$

This solution is the same with the Wiener filtering of the current residual $\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega)$. The $\hat{u}_k^{n+1}(\omega)$ is a solution parameter in spectral domain, the $u_k^{n+1}(t)$ is the real part inverse Fourier transform form of the $\hat{u}_k^{n+1}(\omega)$. In a similar way, the center frequencies can be updated as:

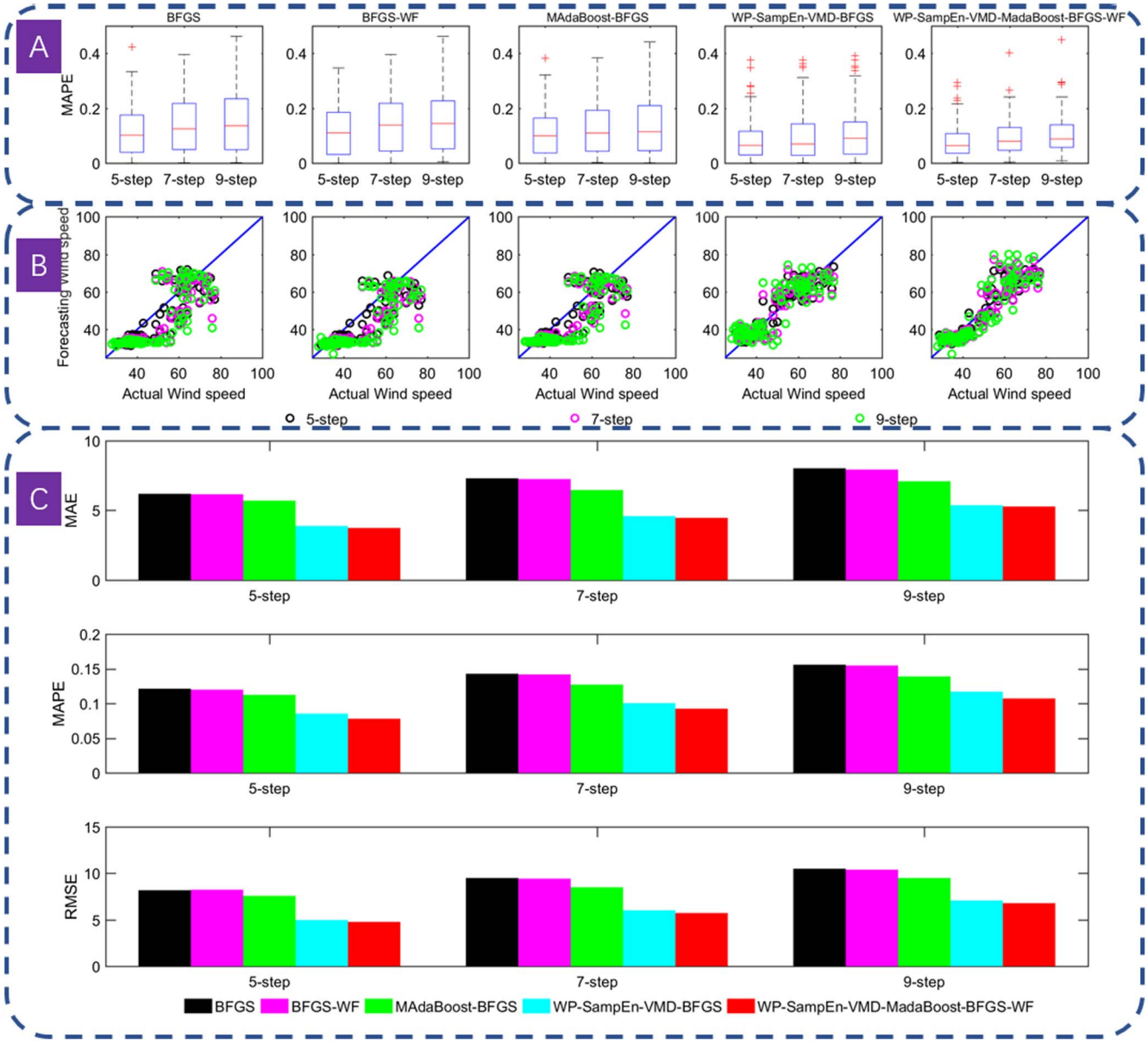

Fig. 12. The error evaluation analysis of the series #2.

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{3.17}$$

The ω_k^{n+1} can be regarded as the center of gravity of the corresponding mode's power spectrum.

According to the ADMM algorithm, the $\hat{\lambda}^{n+1}(\omega)$ can be updated as:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_{k=1}^{K} \hat{u}_k^{n+1}(\omega)) \tag{3.18}$$

Finally, the VMD algorithm can be obtained as shown in Algorithm 1.

Algorithm 1 (*Variational Mode Decomposition*).

Input:

- Input signal f.
- Initial modes $\{u_k^1\} = \{u_1^1,...,u_K^1\}$.
- Initial center frequencies $\{\omega_k^1\} = \{\omega_1^1,...,\omega_K^1\}$.
- Initial Lagrangian multipliers λ^1.
- Balancing parameter α.

Output

- Output modes $u_k^{n+1}(t)$.

Algorithm

repeat

$n = n + 1$

for $k = 1: K$ **do**

Update $\hat{u}_k(\omega)$ for all $\omega \geqslant 0$ using (3.16)

Update $\hat{\omega}_k$ using (3.17)

end for

Update $\hat{\lambda}$ for all $\omega \geqslant 0$ using (3.18)

until convergence $\sum_{k=1}^{K} \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$

Obtain $u_k^{n+1}(t)$ by inverse Fourier transform of $\hat{u}_k^{n+1}(\omega)$.

return $u_k^{n+1}(t)$

3.2. Ensemble method

The ensemble method is a kind of algorithm that could ensemble

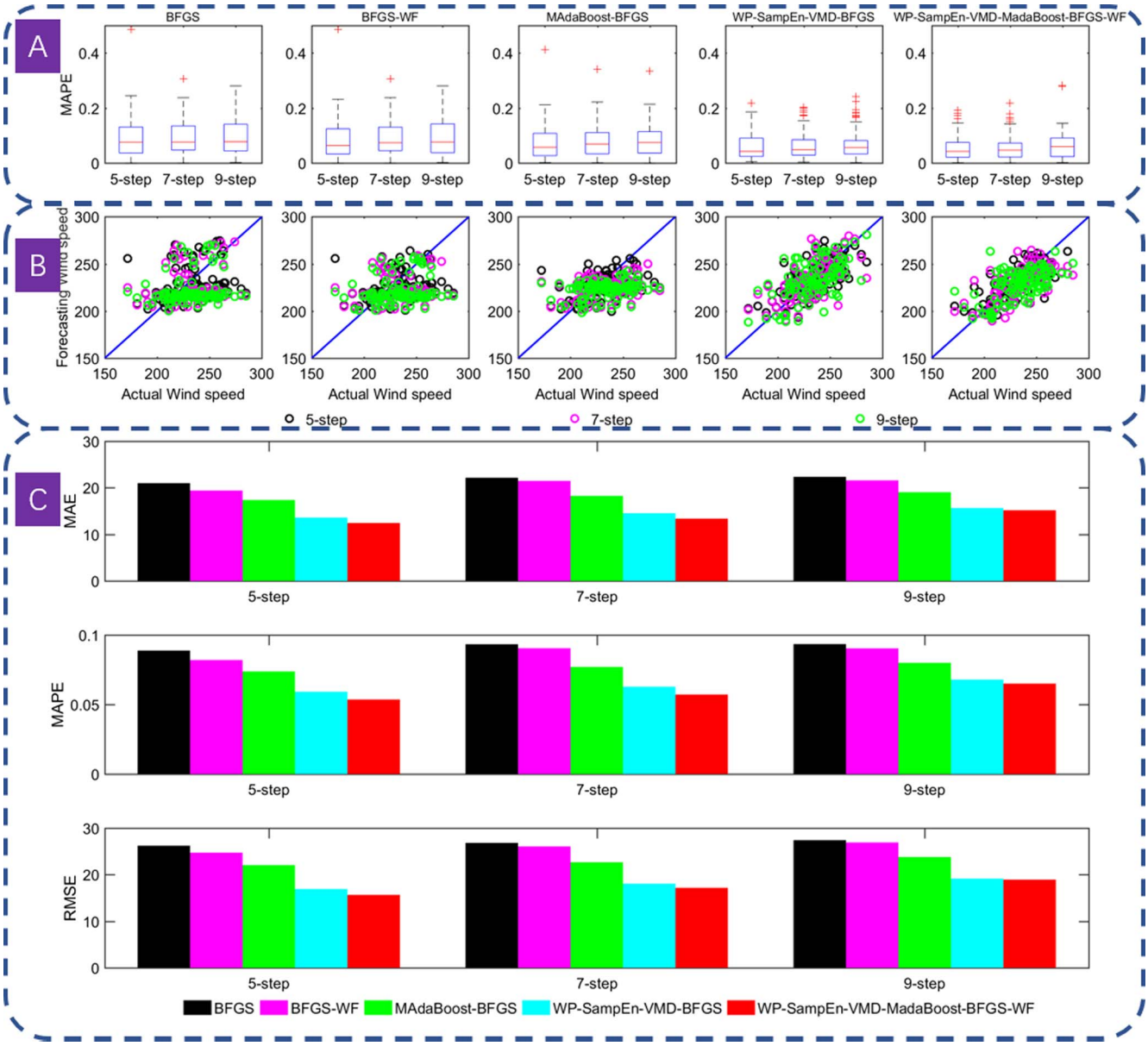

Fig. 13. The error evaluation analysis of the series #3.

several forecasting methods to accomplish better performance than individual forecasting method. The application of ensemble method can be divided into two kinds as: regression and classification. In the field of wind speed forecasting, the ensemble method is always used for the regression applications. One of the most famous ensemble methods in the regression applications is the AdaBoost. R (R stands for 'Regression'), which is firstly proposed by scientists Freund and Schapire [38]. After that, some important variates of the original AdaBoost algorithm are proposed. Such as, the AdaBoost.R was promoted as the AdaBoost.R2 [39]. AdaBoost.RT(RT stand for regression threshold) was proposed based on the AdaBoost.M1 [40]. By considering the accuracy of the AdaBoost.RT algorithm, a modified form named as the Modified AdaBoost.RT was presented by Kummer [41] by adjusting the thresholds adaptively according to the root mean square error (RMSE) of the last step iteration [42].

Due to the good performance of the Modified AdaBoost RT, it is selected as the basic ensemble framework in the study. The details of the Modified AdaBoost.RT are explained in Algorithm 2.

Algorithm 2 (*Modified AdaBoost.RT*).

Input:

- Training set $S = \{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_i,y_i),\ldots,(\mathbf{x}_N,y_N)\}$, where N is number of training sample.
- Weak learner algorithm WL_t.
- Distribution $D_1(i) = 1/N$ for all i.Positive integer T determine number of iterations.
- Threshold ϕ_t for distinguishing which forecasting error is correct or incorrect, initial threshold ϕ_1 determining the initial value of the iteration.
- Change rate r specifying the change rate of ϕ.

Output

- Output forecasting function f_{fin}.

Algorithm

Initialize weight vector $w_i^t = D_1(i) = 1/N$ for all i.

for $t = 1: T$ **do**

Call weak learner WL_t, providing it with distribution D_t.

Build the regression model with WL_t: $f_t(\mathbf{x}) \rightarrow y$

Compute absolute relative error for each training example as:

$$\mathrm{ARE}_t(i) = \left|\frac{f_t(\mathbf{x}_i) - y_i}{y_i}\right|.$$

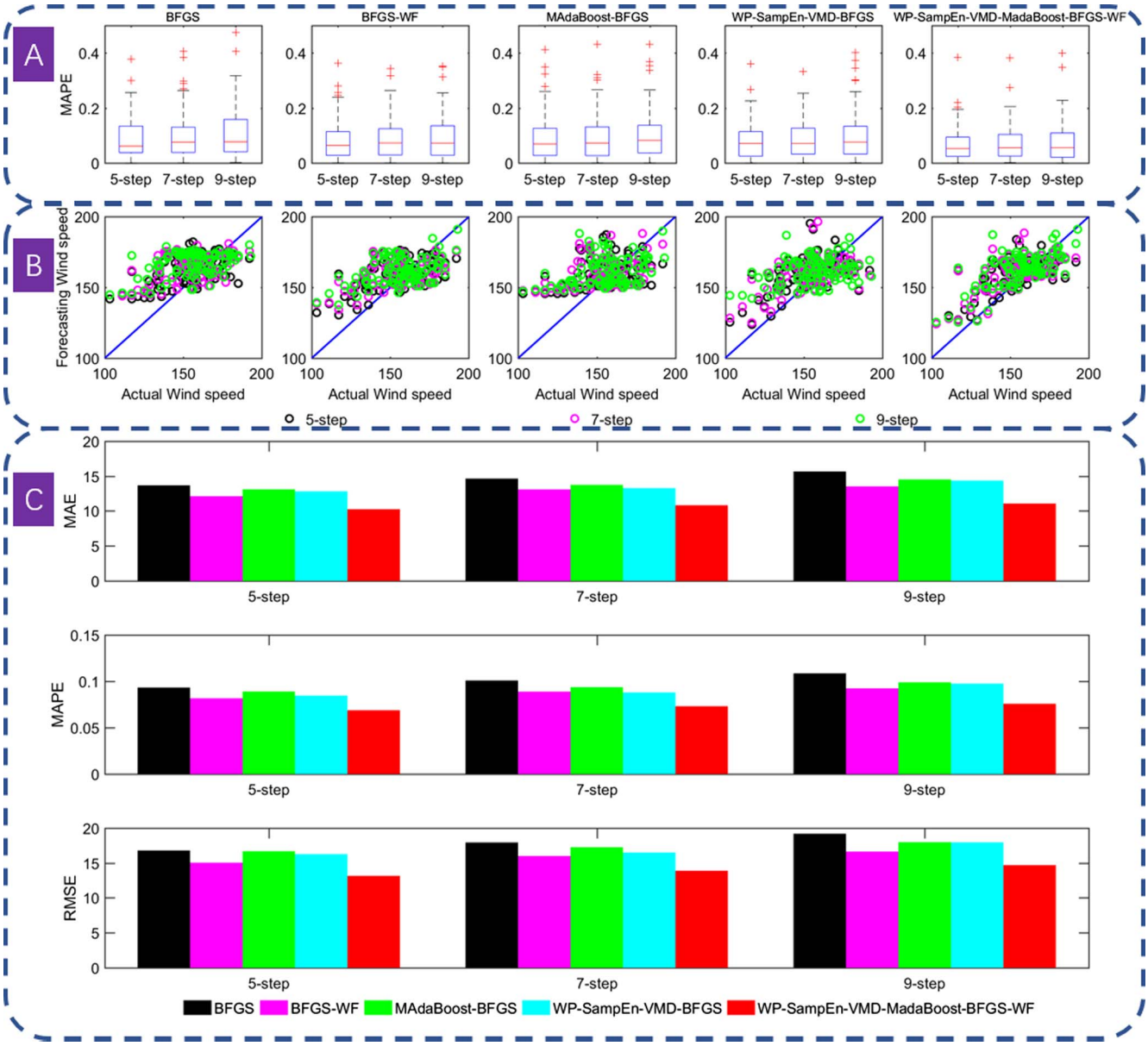

Fig. 14. The error evaluation analysis of the series #4.

Compute the error rate of $f_t(\mathbf{x})$: $\varepsilon_t = \sum_{i:\mathrm{ARE}_t(i) > \phi_t} D_t(i)$.

Update distribution $D_t(i)$ as $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \beta_t & if\ ARE_t(i) < \phi_t \\ 1 & otherwise \end{cases}$,

where $\beta_t = \varepsilon_t^2$, Z_t is normalization factor to keep $D_{t+1}(i)$ as distribution.

Update threshold ϕ_t as $\phi_{t+1}(i) = \phi_t(i) \times \begin{cases} (1-\lambda) & if\ e_t < e_{t-1} \\ (1+\lambda) & otherwise \end{cases}$,

where $\lambda = r \times \left|\frac{e_t - e_{t-1}}{e_t}\right|$, e is root mean square error(RMSE):

$e_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_t(\mathbf{x}_i) - y_i)^2}$.

Compute the weight of weak learner WL_t: $\alpha_t = -\log(\beta_t)$.

end for

Normalize weight $\boldsymbol{\alpha} = \{\alpha_1,...,\alpha_T\}$, such that $\sum \boldsymbol{\alpha} = 1$.

Output forecasting function $f_{fin}(x_i) = \sum_{t=1}^{T} \alpha_t \times f_t(x_i)$.

return f_{fin}

As shown in Algorithm 2, there are two important parameters as: the initial threshold ϕ_1 and the change rate r. In the study, the ϕ_1 is equal to 0.2 and the r equal is 0.5, respectively.

3.3. Forecasting method

BFGS neuron network is derived from the MLP, which uses the BFGS algorithm to solve the optimization problem in training networks [43]. The BFGS neural network used in this paper consists of three parts. There are input layer, hidden layer and output layer. In order to improve the non-linear mapping capability of the BFGS neuron network, a neural structure having two hidden layers is adopted. The detailed structure of the BFGS neural network is shown in Fig. 2.

The main parameters of the BFGS neuron network are the corresponding number of the neurons in the input layer these two hidden layers and the output layer. In this study, they are equal to six, ten, three and one, respectively.

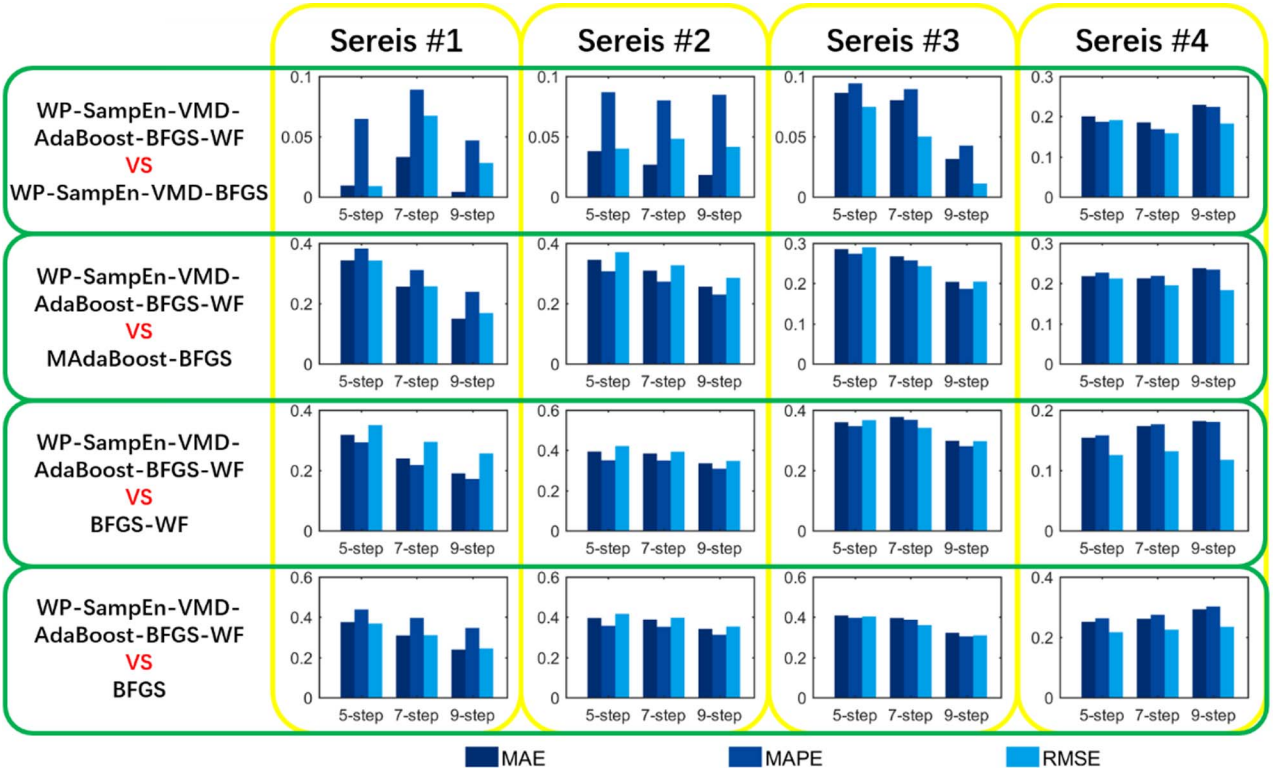

Fig. 15. The improving percentages of the involved forecasting models.

Table 5
The error evaluation results of the series forecasting.

Series	Model	MAE			MAPE			RMSE		
		5-step	7-step	9-step	5-step	7-step	9-step	5-step	7-step	9-step
#1	A	4.458	5.088	5.734	12.946	14.836	16.555	5.73	6.415	7.213
	B	4.5	5.262	5.757	13.841	16.282	17.366	5.781	6.878	7.422
	C	6.781	6.84	6.741	20.945	21.533	21.744	8.712	8.636	8.671
	D	6.532	6.691	7.08	18.314	18.974	19.996	8.819	9.094	9.702
	E	7.13	7.361	7.529	23.051	24.578	25.319	9.075	9.305	9.55
#2	A	3.739	4.467	5.277	7.794	9.249	10.713	4.77	5.725	6.785
	B	3.886	4.589	5.374	8.535	10.053	11.703	4.969	6.014	7.079
	C	5.704	6.46	7.09	11.237	12.711	13.902	7.571	8.497	9.493
	D	6.161	7.252	7.931	11.982	14.19	15.473	8.238	9.42	10.385
	E	6.19	7.301	8.019	12.121	14.274	15.589	8.173	9.497	10.49
#3	A	12.424	13.376	15.154	5.356	5.718	6.502	15.657	17.164	18.926
	B	13.594	14.543	15.646	5.912	6.279	6.791	16.919	18.068	19.137
	C	17.383	18.249	19.03	7.37	7.696	7.99	22.043	22.664	23.786
	D	19.399	21.479	21.577	8.196	9.045	9.031	24.714	26.044	26.9
	E	20.987	22.148	22.345	8.878	9.334	9.341	26.218	26.824	27.396
#4	A	10.251	10.825	11.075	6.878	7.317	7.573	13.156	13.894	14.687
	B	12.819	13.285	14.363	8.458	8.796	9.753	16.262	16.502	17.959
	C	13.107	13.745	14.521	8.894	9.363	9.891	16.696	17.258	17.978
	D	12.118	13.096	13.535	8.167	8.883	9.238	15.04	16.001	16.641
	E	13.683	14.646	15.659	9.33	10.088	10.854	16.797	17.931	19.186

Notes: Model A: WD-SampEn-VMD-MadaBoost-BFGS-WF; Model B: WD-SampEn-VMD-BFGS; Model C: MadaBoost-BFGS; Model D: BFGS-WF; Model E: BFGS.

3.4. Error correction method

A one-dimensional series can be defined as follows [44]:

$$s(i) = f(i) + \sigma e(i) \tag{3.19}$$

Where $f(i)$ is the real series, $e(i)$ is the noise, and $s(i)$ is the series with noise, σ is the noise intensity. A wavelet filter can be used to remove the noise $e(i)$ and find the real series.

In this study, an original subseries is obtained by the WD and has the fixed band. After the forecasting computation, the forecasting subseries have new frequency components, comparing to the corresponding actual subseries. The excrescent frequency component can be regarded as a forecasting error. The WF is adopted to correct the possible forecasting errors using the following 3 steps as given in [45]:

(1) The series $s(i)$ is decomposed by the wavelet. The coefficients of the wavelet are calculated. The wavelet and the level of decomposition N of the wavelet filter is the same with preceding WD to obtain the signals with identical frequency bands.
(2) The layer with the maximum energy remains the same, while the other layers is disposed by the appropriate thresholds.
(3) The filtered layers and layer of maximum energy is reconstructed to

Table 6
The improving percentages of the involved forecasting models.

Series	Comparison	ξ_{MAE}(%)			ξ_{MAPE}(%)			ξ_{RMSE}(%)		
		5-step	7-step	9-step	5-step	7-step	9-step	5-step	7-step	9-step
#1	A	0.934	3.307	0.41	6.462	8.879	4.672	0.886	6.725	2.809
	B	34.259	25.624	14.945	38.189	31.1	23.865	34.231	25.711	16.811
	C	31.758	23.962	19.014	29.31	21.808	17.209	35.03	29.454	25.651
	D	37.479	30.882	23.839	43.836	39.637	34.614	36.865	31.055	24.464
#2	A	3.793	2.67	1.807	8.678	7.999	8.464	4.003	4.81	4.141
	B	34.446	30.86	25.568	30.639	27.241	22.941	36.996	32.633	28.521
	C	39.314	38.404	33.465	34.953	34.821	30.768	42.092	39.228	34.658
	D	39.598	38.825	34.194	35.701	35.208	31.283	41.637	39.723	35.316
#3	A	8.608	8.024	3.144	9.409	8.929	4.255	7.46	5.005	1.101
	B	28.527	26.704	20.37	27.331	25.697	18.624	28.972	24.267	20.43
	C	35.954	37.727	29.77	34.652	36.781	28.003	36.648	34.096	29.644
	D	40.801	39.607	32.182	39.676	38.74	30.391	40.283	36.013	30.917
#4	A	20.031	18.522	22.894	18.676	16.821	22.355	19.1	15.8	18.22
	B	21.789	21.247	23.735	22.667	21.857	23.436	21.203	19.49	18.304
	C	15.405	17.343	18.175	15.78	17.633	18.026	12.524	13.163	11.741
	D	25.079	26.091	29.278	26.277	27.475	30.228	21.675	22.513	23.45

Notes: Comparison A: WD-SampEn-VMD-MadaBoost-BFGS-WF V.S. WD-SampEn-VMD-BFGS; Comparison B: WD-SampEn-VMD-MadaBoost-BFGS-WF V.S. MAdaBoost-BFGS; Comparison D: WD-SampEn-VMD-MadaBoost-BFGS-WF V.S. BFGS-WF; Comparison C: WD-SampEn-VMD-MadaBoost-BFGS-WF V.S. BFGS.

Table 7
The results of the Pearson's test.

Series	Step	Model				
		A	B	C	D	E
#1	5-step	0.903	0.9	0.764	0.776	0.773
	7-step	0.879	0.857	0.785	0.773	0.8
	9-step	0.86	0.834	0.797	0.737	0.788
#2	5-step	0.942	0.934	0.867	0.86	0.855
	7-step	0.924	0.902	0.842	0.833	0.82
	9-step	0.901	0.863	0.811	0.807	0.79
#3	5-step	0.702	0.661	0.356	0.241	0.176
	7-step	0.652	0.633	0.343	0.186	0.171
	9-step	0.577	0.591	0.279	0.167	0.161
#4	5-step	0.706	0.489	0.44	0.575	0.535
	7-step	0.69	0.467	0.396	0.512	0.501
	9-step	0.676	0.337	0.339	0.47	0.448

Notes: Model A: WD-SampEn-VMD-MadaBoost-BFGS-WF; Model B: WD-SampEn-VMD-BFGS; Model C: MadaBoost-BFGS; Model D: BFGS-WF; Model E: BFGS.

correct the wind speed subseries.

In the proposed wavelet filter there are two key computing steps. The first one is how to calculate the thresholds and the second one is how to use the thresholds to optimize the wavelet coefficients. By considering these two concerns, the following strategy is adopted as: the thresholds are obtained by averaging all coefficients; Based on the obtained thresholds, the layers can be filtered as follow. If the wavelet coefficients are less than the thresholds, the coefficients are retained the same, otherwise the coefficients are defined to be equal to the proposed thresholds.

4. Empirical study and contrast analysis

4.1. Series

In order to validate the performance of the proposed model, four forecasting cases are provided in the study. The experimental modeling wind speed series, which will be used in these four cases, are all generated to simulate the extremely strong wind speed data to check the potential forecasting capacity of the proposed forecasting models as displayed in Fig. 3. In all these forecasting experiments, the former 500 wind speed simulated data points are selected as the modeling datasets and the other corresponding 100 wind speed data points are chosen as the testing datasets at the same time. In this study, the extremely strong wind speed data and the big multi-step forecasting are both focused. The statistical parameters of these original wind speed datasets are given in Table 1.

4.2. Performance evaluation

4.2.1. Error evaluation

To evaluate the forecasting performance of the involved models fairly, a series of error evaluation indexes is used in the study, consisting of the MAE, the MAPE and the RMSE. All of these error evaluation indexes have been generally applied in the forecasting model estimation. The equations of these evaluation indexes are explained as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \tag{4.1}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{4.2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t|} \tag{4.3}$$

Where Y_t is the actual data, $\hat{Y}_t$ is the forecasting data, n is the number of forecasting data.

4.2.2. Pearson's test

Pearson's test is employed to evaluate the strength of association between the actual wind speed data and the forecasting wind speed data. The Pearson's test was proposed by scientist *Karl Pearson*. In the test, if the Pearson's correlation coefficient is equal to 0, it means that these two sets of data have no relationship. And if the Pearson's correlation coefficient is equal to 1, it means that these two sets of data have a linear relationship. Given the actual data as Y_t, and the forecasting data is $\hat{Y}_t$, the Pearson's correlation coefficient can be described as follows:

$$P = \frac{\sum_{t=1}^{T} (Y_t - Y_m)(\hat{Y}_t - \hat{Y}_m)}{\sqrt{\sum_{t=1}^{T} (Y_t - Y_m)^2 \times \sum_{t=1}^{T} (\hat{Y}_t - \hat{Y}_m)^2}} \tag{4.4}$$

Where Y_m and $\widehat{Y}_m$ are the means of the actual data and the forecasting data, respectively.

4.3. Analysis of proposed model

To explain the proposed forecasting model thoroughly, this section is presented to analyze how each part of the proposed model works in the hybrid combination.

4.3.1. Secondary decomposition method

In the proposed model, the original wind speed time series are decomposed by two different decomposition methods. The wind speed dataset #4 is utilized as an experimental sample to analyze the process of the secondary decomposition method. First, the wind speed time series is decomposed by the WD. The SampEn results of each decomposed subseries are given in Table 2. From Table 2, it can be seen that the corresponding subseries having the highest SampEn value is the subseries #6. The subseries #6 is further decomposed by the VMD to obtain the sub-subseries, which SampEn results are shown in Table 3. From Table 2 and Table 3, it can be found that the SampEn results of the subseries #6 decrease through the VMD, which indicates the efficiency of the VMD decomposition in terms of entropy. The process of secondary decomposition is shown in Fig. 4, which contains both of the time-domain and frequency domain plots of each subseries.

4.3.2. Forecasting and error correction method

In the study, the forecasting subseries is obtained by executing the MAdaBoost, and the extra frequency components of the forecasting subseries are erased by completing the WF computation. Fig. 5 shows the time-domain and instantaneous frequency plots of the actual subseries #1, which is decomposed from the series #1. The comparing results between the subseries before and after WF are given in Fig. 6 and Table 4. Fig. 6 indicates the efficiency of the WF for the time-domain and instantaneous frequency plots of the different-step forecasting subseries. Table 4 displays the instantaneous frequency cross-correlation coefficient analysis of all the decomposed subseries of the series #1. From Table 4, it can be seen that the cross-correlation coefficients of subseries' instantaneous frequencies after the WF computation and the actual subseries' instantaneous frequencies are higher than the subseries' instantaneous frequencies before the WF computation and the actual subseries' instantaneous frequencies. This phenomenon exactly proves the efficiency of the WF computation, which has been adopted in the study.

4.4. Contrast analysis

The structure of the proposed model can be divided into three parts as: the secondary composition, the ensemble method and the error correction structure. To validate the real forecasting performance of the proposed structure according to the proposed model structures, several hybrid models, which includes the WD-SampEn -VMD-BFGS, the MAdaBoost-BFGS, the BFGS-WF and the BFGS, have been involved in the comparison of the forecasting performance.

4.4.1. Forecasting result

The forecasting results of the involved models are given in Figs. 7–10. From those figures, it can been found that: (a) the difficulty of the high-precision forecasting is up to the peak and valley points of the wind speed series, while the overall trends of the wind speed series can be forecasted accurately; (b) the performance of the single model is worse than the hybrid model. The front part of series #1, where has a downward trend, can provide a good contrast between the single model and the hybrid model. The BFGS's forecasting result series in the front part of the series #1 is almost straight and the WD-SampEn-VMD-MAdaBoost-BFGS-WF's forecasting result series can forecast the most trend wind speed points and parts of detailed wind speed points; and (c) with the increasing of the forecasting steps, the forecasting errors increase. This is mainly because of the forecasting error accumulation and the growing complexity in mapping the nonlinear relationships between the inputs and the outputs.

4.4.2. Error evaluation of models

The error evaluation results of the obtained wind speed forecasting results from Section 4.2.1 are shown in Figs. 11–15 and Tables 4–5.

1) The 'Part A' sections in Figs. 11–14 are the MAPE results of different forecasting models. Based on these results, it can be seen that: a) the MAPE has an increasing trend when the forecasting step increases. However, in some special cases, such as the MAdaboost-BFGS in the series #1, the 9-step's minimum MAPE value is less than those of the 5-step and the 7-step; b) among those forecasting series, the WD-SampEn-VMD-MAdaBoost-BFGS-WF achieves the minimum MAPE's mean with 95% confidence. However, there is also an expectation that MAPE's mean of the WD-SampEn-VMD-MAdaBoost-BFGS-WF is slightly higher than that of the WD-SampEn-VMD-BFGS in the 9-step forecasting results of the series #1; and c) the WD-SampEn-VMD-MAdaBoost-BFGS-WF has the smallest range of the variation, which indicates the stability of the proposed model.
2) The 'Part B' sections in Figs. 11–14 are the scatter diagrams between the actual wind speed and the corresponding forecasting wind speed in the 5, 7 and 9-step forecasting results. From the given curves, it can be seen that: a) comparing with the simple models, the proposed model has a big improvement; b) the scatter diagrams of the BFGS, the BFGS-WF and the MAdaBoost-BFGS are similar, while the WD-SampEn-VMD-BFGS and the WD-SampEn-VMD-MAdaBoost-BFGS-WF have the similar scatter diagrams. This phenomenon indicates that the secondary decomposition is able to improve the scatter diagram obviously.
3) The 'Part C' sections in Figs. 11–14 and Table 5 show the error evaluation. Based on the provided results, it can be analyzed that: a) the forecasting performance of the proposed model is the best, which indicates the efficiency of the combination of the secondary decomposition, the ensemble method and the error correction. The forecasting performance of the single BFGS has the worst result among those involved models; b) the WD-SampEn-VMD-BFGS have the best performance among three single hybrid models; c) As the number of steps increases, all models perform worse.
4) Fig. 15 and Table 6 show the improving percentage of models, the indexes ξ_{MAE}, ξ_{MAPE} and ξ_{RMSE} are the improving percentages of the MAE, the MAPE, the RMSE respectively. Based on their values, it can be seen that: a) all the improving percentages of error evaluation are positive, which means the proposed model can perform better than the other included models; and b) the improving percentages between the proposed model and the WD-SampEn-VMD-BFGS are insignificant, while these between the proposed model and the single BFGS are significant.

4.4.3. Pearson's test of models

The results of the Pearson's test are also shown in Table 7. From Table 7, it can be seen that the Pearson's test of the proposed model is higher than other models in all the series and forecasting steps.

5. Conclusion

In order to achieve higher accuracy of the big multi-step forecasting, in the study a novel hybrid model named as the WD-SampEn-VMD-MadaBoost-BFGS-WF is proposed, which consists of the secondary decomposition, the ensemble method and the error correction. To validate the efficiency and superiority of the proposed model, a number of forecasting experiments is also given in the study. The forecasting models included in the performance comparison consist of the proposed hybrid model, the WD-SampEn-VMD-BFGS model, the MAdaBoost-BFGS model, the BFGS-WF model and the BFGS model. The forecasting

experimental results show that the proposed hybrid model has the best forecasting performance among all the included models in the big multi-step wins speed prediction. In future, the further study will be to investigate the maxmum forecasting multi-steps using the proposed model.

Acknowledgements

This study is fully supported by the Shenghua Yu-ying Talents Program of the Central South University (Principle Investigator: Dr. Hui Liu), the innovation driven project of the Central South University (Project No. 502501002, Principle Investigator: Dr. Hui Liu), the engineering college 'double first-rate' supporting project of the Central South University and the Changsha Outstanding Innovative Youth Project (*Principle Investigator*: Dr. Hui Liu). This study is partly supported by the National Natural Science Foundation of China (Grant No. U1534210, U1334205).

Reference

[1] Wang J, Heng J, Xiao L, Wang C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 2017;125:591–613.
[2] WWEA. WORLD WIND MARKET HAS REACHED 486 GW FROM WHERE 54 GW HAS BEEN INSTALLED LAST YEAR. 2017.
[3] Croonenbroeck C, Stadtmann G. Minimizing asymmetric loss in medium-term wind power forecasting. Renew Energy 2015;81:197–208.
[4] 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.
[5] Dong Q, Sun Y, Li P. A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: a case study of wind farms in China. Renew Energy 2017;102:241–57.
[6] Zhao J, Guo Z-H, Su Z-Y, Zhao Z-Y, Xiao X, Liu F. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. Appl Energy 2016;162:808–26.
[7] Cassola F, Burlando M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. Appl Energy 2012;99:154–66.
[8] Pinson P, Chevallier C, Kariniotakis GN. Trading wind generation from short-term probabilistic forecasts of wind power. IEEE Trans Power Syst 2007;22:1148–56.
[9] Masseran N. Modeling the fluctuations of wind speed data by considering their mean and volatility effects. Renew Sustain Energy Rev 2016;54:777–84.
[10] Wang Y, Wang J, Zhao G, Dong Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. Energy Policy 2012;48:284–94.
[11] Liu H, Tian H-q, Li Y-f. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[12] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87:2313–20.
[13] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – Extreme learning machine approach. Energy Convers Manage 2014;87:10–8.
[14] Hong Y-Y, Chang H-L, Chiu C-S. Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. Energy. 2010;35:3870–6.
[15] Ait Maatallah O, Achuthan A, Janoyan K, Marzocca P. Recursive wind speed forecasting based on Hammerstein Auto-Regressive model. Appl Energy 2015;145:191–7.
[16] Liu H, Tian H-q, Li Y-f, Zhang L. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. Energy Convers Manage 2015;92:67–81.
[17] Qureshi AS, Khan A, Zameer A, Usman A. Wind power prediction using deep neural network based meta regression and transfer learning. Appl Soft Comput 2017;58:742–55.
[18] Zameer A, Arshad J, Khan A, Raja MAZ. Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. Energy Convers Manage 2017;134:361–72.
[19] Liu H, Tian H-q, Li Y-f. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Convers Manage 2015;100:16–22.
[20] Liu Y, Shi J, Yang Y, Lee W-J. Short-term wind-power prediction based on wavelet transform–support vector machine and statistic-characteristics analysis. IEEE Trans Ind Appl 2012;48:1136–41.
[21] Osório GJ, Matias JCO, Catalão JPS. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renew Energy 2015;75:301–7.
[22] Liu H, Tian H-q, Li Y-f. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. Energy Convers Manage 2015;89:1–11.
[23] Wang J, Zhang W, Li Y, Wang J, Dang Z. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl Soft Comput 2014;23:452–9.
[24] Liu H, Mi X, Li Y. Comparison of two new intelligent wind speed forecasting approaches based on Wavelet Packet Decomposition, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Artificial Neural Networks. Energy Convers Manage 2018;155:188–200.
[25] Liu H, Tian H-q, Liang X-f, Li Y-f. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94.
[26] Wang Y, Wang J, Wei X. A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: A case study of wind farms in northwest China. Energy 2015;91:556–72.
[27] Jiang Y, Huang G. Short-term wind speed prediction: hybrid of ensemble empirical mode decomposition, feature selection and error correction. Energy Convers Manage 2017;144:340–50.
[28] Wang J, Song Y, Liu F, Hou R. Analysis and application of forecasting models in wind power integration: a review of multi-step-ahead wind speed forecasting models. Renew Sustain Energy Rev 2016;60:960–81.
[29] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 1989;11:674–93.
[30] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol-Heart Circulatory Physiol 2000;278:H2039–49.
[31] Tsai PH, Lin C, Tsao J, Lin PF, Wang PC, Huang NE, et al. Empirical mode decomposition based detrended sample entropy in electroencephalography for Alzheimer's disease. J Neurosci Meth 2012;210:230–7.
[32] Widodo A, Shim M-C, Caesarendra W, Yang B-S. Intelligent prognostics for battery health monitoring based on sample entropy. Expert Syst Appl 2011;38:11763–9.
[33] Wiki. Sample entropy. 2017.
[34] Alcaraz R, Rieta JJ. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. Biomed Signal Process Control 2010;5:1–14.
[35] Pincus SM. Assessing serial irregularity and its implications for health. Ann NY Acad Sci 2001;954:245–67.
[36] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62:531–44.
[37] Wang Y, Markert R, Xiang J, Zheng W. Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. Mech Syst Sig Process 2015;60–61:243–51.
[38] Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. European conference on computational learning theory. Springer; 1995. p. 23–37.
[39] Drucker H. Improving regressors using boosting techniques. ICML1997. p. 107–15.
[40] Solomatine DP, Shrestha DL. AdaBoost. RT: a boosting algorithm for regression problems. Neural Networks, 2004 Proceedings 2004 IEEE International Joint Conference on. IEEE; 2004. p. 1163–8.
[41] Hui-Xin T, Zhi-Zhong M. An Ensemble ELM Based on Modified AdaBoost.RT Algorithm for Predicting the Temperature of Molten Steel in Ladle Furnace. IEEE Trans Autom Sci Eng 2010;7:73–80.
[42] Shrestha DL, Solomatine DP. Experiments with AdaBoost. RT, an improved boosting scheme for regression. Neural Comput 2006;18:1678–710.
[43] Xia J-H, Rusli, Kumta AS. Feedforward neural network trained by BFGS algorithm for modeling plasma etching of silicon carbide. IEEE Trans Plasma Sci 2010;38:142–8.
[44] Chang SG, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Process 2000;9:1532–46.
[45] Yu B, Lou L, Li S, Zhang Y, Qiu W, Wu X, et al. Prediction of protein structural class for low-similarity sequences using Chou's pseudo amino acid composition and wavelet denoising. J Mol Graph Model 2017;76:260–73.