Applied Energy 228 (2018) 1783–1800

Contents lists available at ScienceDirect

Applied Energy

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

A novel ensemble model of different mother wavelets for wind speed multi-step forecasting

Hui Liu*, Zhu Duan, Yanfei Li, Haibo Lu

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

HIGHLIGHTS

- A novel ensemble model is proposed by combining hybrid forecasting models with different mother wavelets.
- AdaBoost.MRT and wavelet packet decomposition are used to improve the forecasting performance of the base predictors.
- Multi-objective grey wolf optimizer is adopted to obtain the optimal values of the coefficients in the base predictors.
- Forecasting performance of the base predictors with different mother wavelets are compared completely.
- The capacity of single mother wavelet for the wind speed forecasting is upgraded.

ARTICLE INFO

Keywords:
Wind speed forecasting
Wavelet packet decomposition
Mother wavelet
Vanishing moment
AdaBoost.MRT
Outlier-robust extreme learning machine

ABSTRACT

Accurate wind speed forecasting is essential for smart wind power conversion and integration. In the study, a novel ensemble model, using four novel hybrid models as base predictors to obtain high prediction accuracy, is proposed for the multi-step wind speed forecasting. The hybrid base predictors consist of the Wavelet Packet Decomposition (WPD), the Multi-Objective Grey Wolf Optimizer (MOGWO), the Adaptive Boosting.MRT (AdaBoost.MRT) and the Outlier-Robust Extreme Learning Machine (ORELM). The proposed ensemble model is named as the MOGWO-WPD -AdaBoost.MRT-ORELM model. The accuracy and diversity of the base predictors have significant positive influences on the performance of the proposed ensemble model. To guarantee the diversity of the base predictors, one of the most important hyper-parameters in the WPD computation (i.e., the mother wavelet) for every base predictor is investigated. In addition, the MOGWO is used to assemble the base predictors. By combining various models with different hyper-parameters, the ensemble structure can be used to improve the forecasting performance of the hybrid model with single hyper-parameter. To investigate the performance of the proposed forecasting architecture, four sets of experiments were conducted in the study. The results show that: (a) the proposed ensemble model has good convergence and forecasting performance; (b) the forecasting accuracy of the base predictor increases as the vanishing moment increases; and (c) the proposed ensemble model outperforms other benchmark models significantly.

1. Introduction

The usage of wind energy has been increasing rapidly in recent years [1]. Accurate wind speed forecasting is essential for protecting the safety of the power conversion and integration. Due to the inherent intermittency and randomness of the wind speed data, it is difficult to make an accurate forecasting of the wind speed.

To realize accurate wind speed predicting, a number of forecasting models are proposed, which include physical methods, statistical methods, intelligence methods and hybrid methods. The physical methods use meteorological data to predict the wind speed [2]. The NWP (Numerical Weather Prediction) is one of the most successful models in the physical methods [3]. The physical methods have a feature that they always can perform well on the long-term forecasting but need significant time [4]. The statistical methods are simpler than the physical methods. The most popular statistical methods consist of Auto-Regressive Integrated Moving Average (ARIMA) model, Auto-Regressive Conditional Heteroskedasticity (ARCH), random walking model, etc. El-Fouly et al. [5] used the relationships between the current wind speed and the corresponding one-year and two-year old wind

* Corresponding author.
E-mail address: csuliuhui@csu.edu.cn (H. Liu).

https://doi.org/10.1016/j.apenergy.2018.07.050
Received 10 February 2018; Received in revised form 10 July 2018; Accepted 13 July 2018
Available online 20 July 2018
0306-2619/

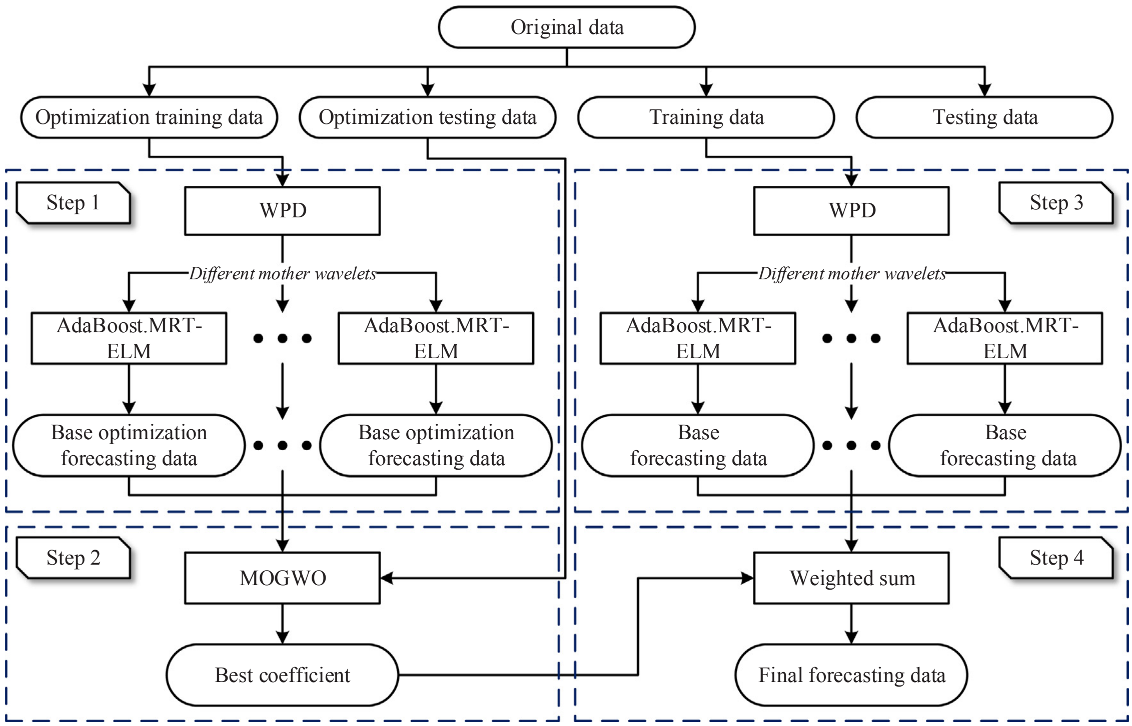

Fig. 1. The framework of the proposed wind speed forecasting model.

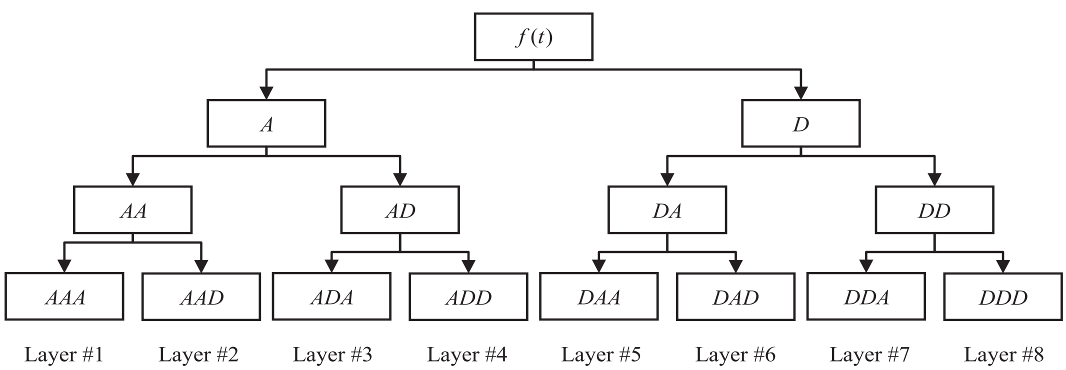

Fig. 2. The three-layer binary tree of the WPD.

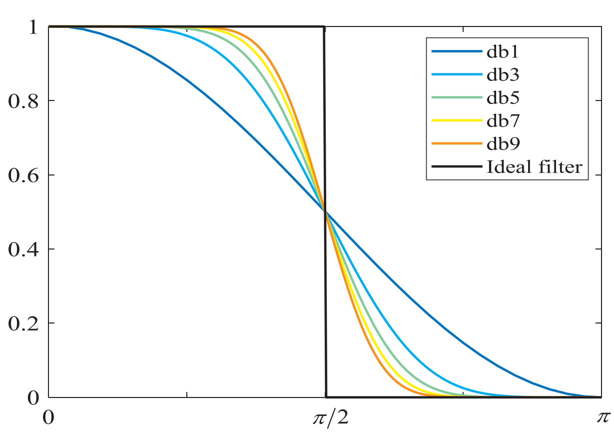

Fig. 3. The magnitude-frequency characteristics of the wavelet low-pass filters with different vanishing moments and the ideal filter.

speed to propose two forecasting models. The case studies verified the effectiveness of the proposed models. The proposed models were further proved in reference [6]. Kiplangat et al. [7] employed the fractional-ARIMA (f-ARIMA) to forecast the wind speed. Their experimental results indicated that the f-ARIMA method was more accurate than both of the ARIMA and the persistence methods. Masseran et al. [8] combined the ARIMA with the ARCH to generate a new hybrid ARIMA-ARCH model. The results presented in the study showed that the ARIMA-ARCH had better performance than the ARIMA model. Wang et al. [9] applied the Particle Swarm Optimization (PSO) to correct the seasonal ARIMA model. The results showed that the proposed models outperformed both of the PSO and the ARIMA model. Zuluaga et al. [10] tried to use three different methods to make the Kalman Filter (KF) model more robust. Liu et al. [11] combined the ARIMA with the KF in the wind speed multi-step prediction. The ARIMA was used to determine the initial parameter of the KF. The case studies showed the proposed model outperformed the single ARIMA model.

Although statistical methods have good performance in the short-term wind speed forecasting, their capacity of fitting nonlinear time-series data can be still promoted. Agrawal et al. [12] proposed a novel model, namely Artificial Neural Network based Yearly Auto-Regressive (ANNYAR), to forecast the wind speed. The results of their cases indicated the proposed model had good forecasting performance. Agrawal et al. [13] completed a wind speed comparing experiment to

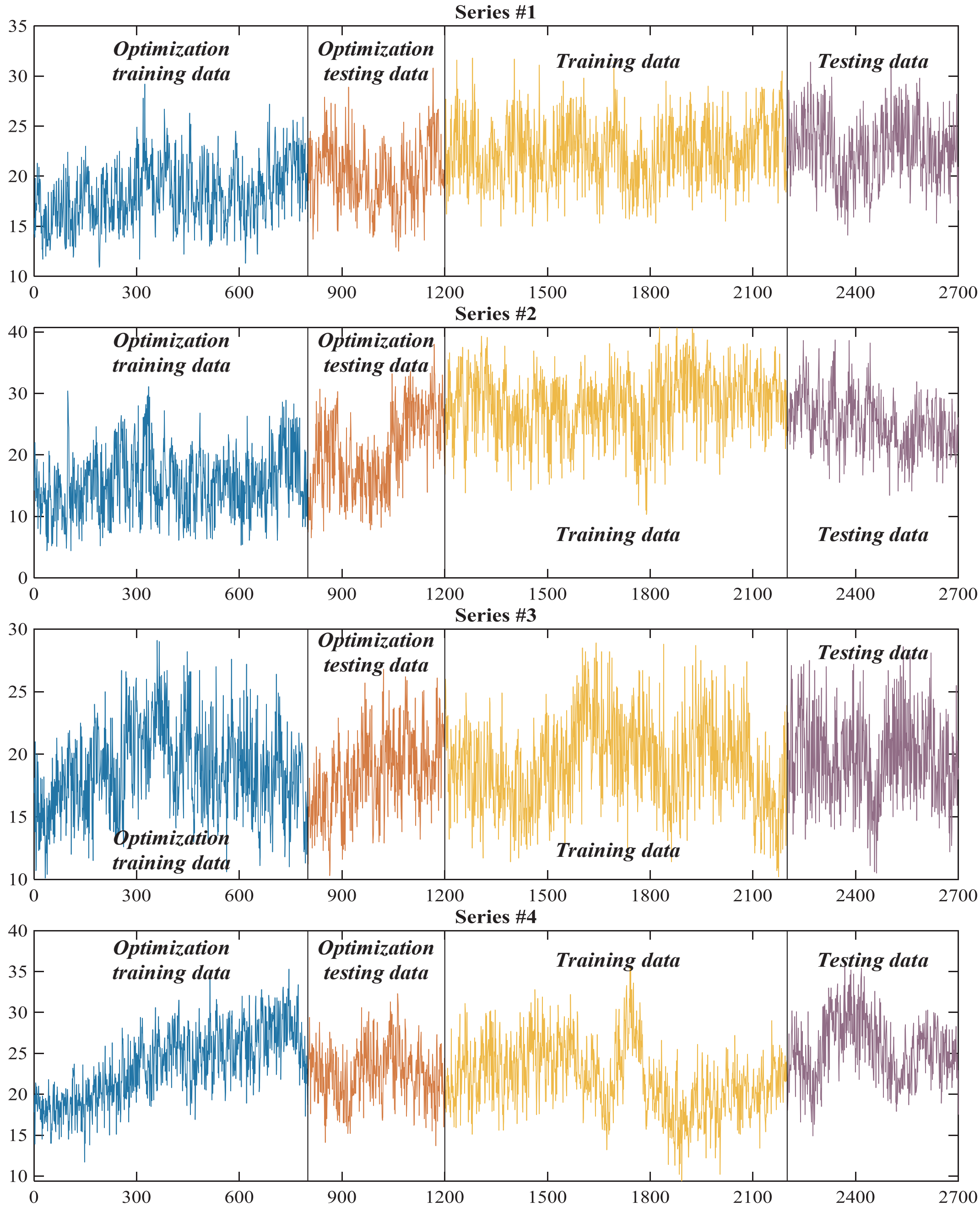

Fig. 4. The original wind speed time series of the latter forecasting experiments.

Table 1
The statistical parameters of all series.

Series	Site	Year	Mean	Standard derivation	Min	Max	Skewness	Kurtosis
#1	#1	2015	20.8884	3.6560	10.9000	31.8000	0.0602	2.6007
#2	#1	2017	22.7686	7.5827	4.4000	40.7000	-0.1807	2.2406
#3	#2	2015	19.0307	3.4702	10.1000	29.1000	0.1767	2.6394
#4	#2	2017	23.1244	4.2304	9.4000	36.2000	0.1070	2.6830

find the best forecasting algorithm for different multi-step results. Hong et al. [14] utilized the Simultaneous Perturbation Stochastic Approximation (SPSA) to optimize the Multi-layer Feed-forward Neural Network (MFNN). As demonstrated in the cases, the proposed model had better forecasting performance than the classical ARIMA model. Li et al. [15] compared three kinds of ANN models, which included the adaptive linear element, the back propagation and the radial basis function. The results indicated that the adopted ANN models have no significant difference in the wind speed forecasting results. Salcedo-Sanz et al. [16] employed the Extreme Learning machine (ELM) to predict wind speed and used the Coral Reefs Optimization (CRO) to optimize the forecasting performance. The results indicated the proposed hybrid forecasting model had excellent performance in the wind speed forecasting.

In the field of wind speed forecasting, it has been recognized that

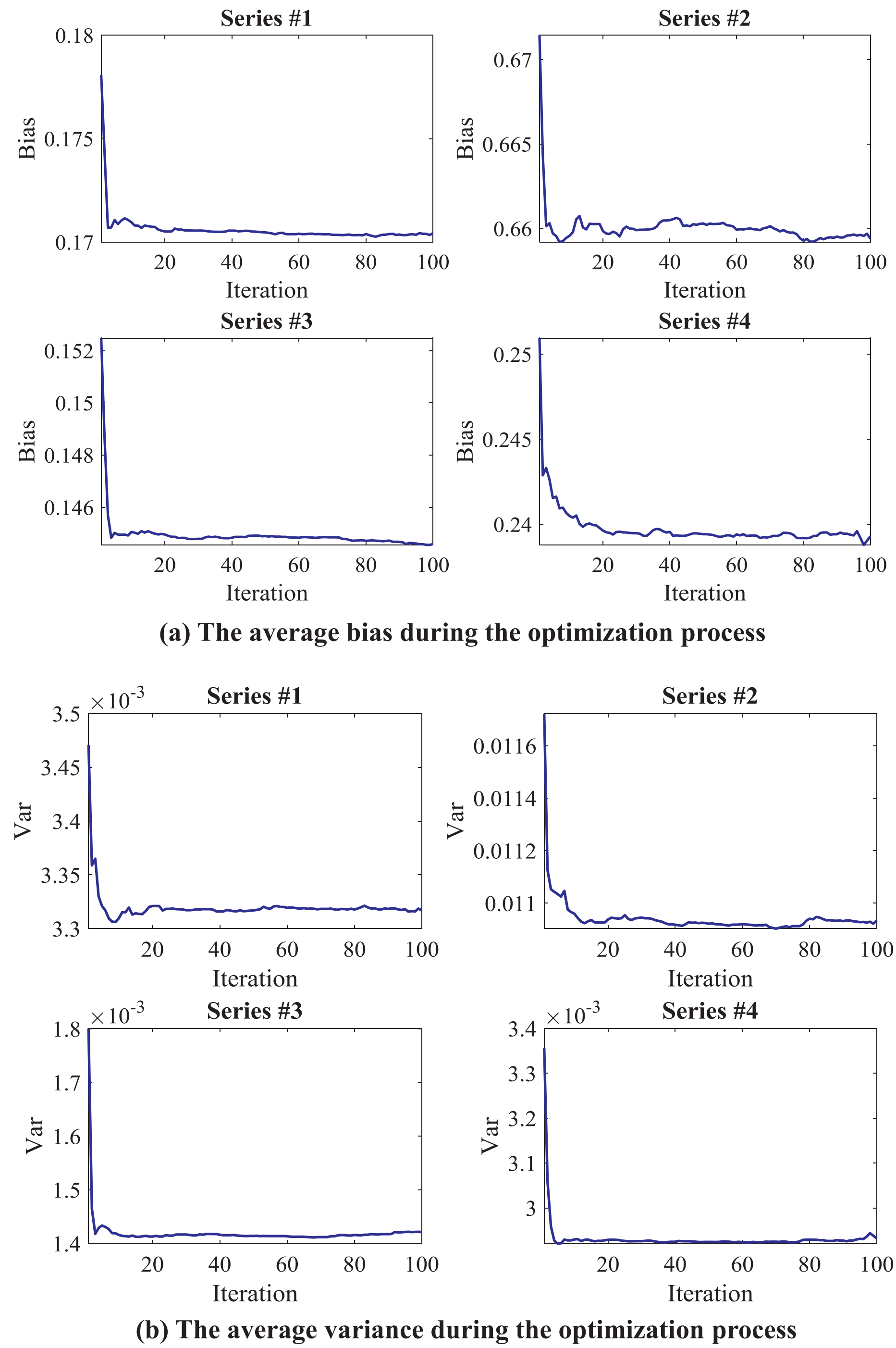

Fig. 5. The average objective functions during the optimization process.

the hybrid models are always effective to obtain the high-accuracy forecasting results. In the proposed hybrid models, some ensemble methods and decomposition methods were used, which can be summarized as follows:

(a) The ensemble methods can improve the forecasting performance by combining several base predictors. There are three kinds of popular ensemble methods: the boosting method, the stacking generalization method and the heuristic algorithm method. The boosting method can train base predictors and obtain their best coefficients simultaneously. Liu et al. [17] combined the MAdaBoost (Modified AdaBoost.RT) with the ENN (Elman Neural Network). Xiao et al. [18] proposed a novel boosting algorithm called TW-FE-Adaboost (Time Vary Forecasting Effectiveness Adaptive boosting). The proposed boosting algorithm was upgraded from the traditional AdaBoost by considering the time-variation of wind speed data. The stacking generalization method combines all outputs of the base predictors using an additional model to obtain the nonlinear combination of base predictors. Qureshi et al. [19] employed a number of deep auto-encoders as the base predictors and at the same time the Deep Belief Network (DBN) model as a predictor in the proposed forecasting framework. The simulation results showed that the proposed hybrid model outperformed other involved models. The heuristic algorithm method optimized the coefficients of all base predictors. Wang et al. [20] used the MOBA (Multi-Objective Bat Algorithm) to obtain the optimal coefficients in various ANN

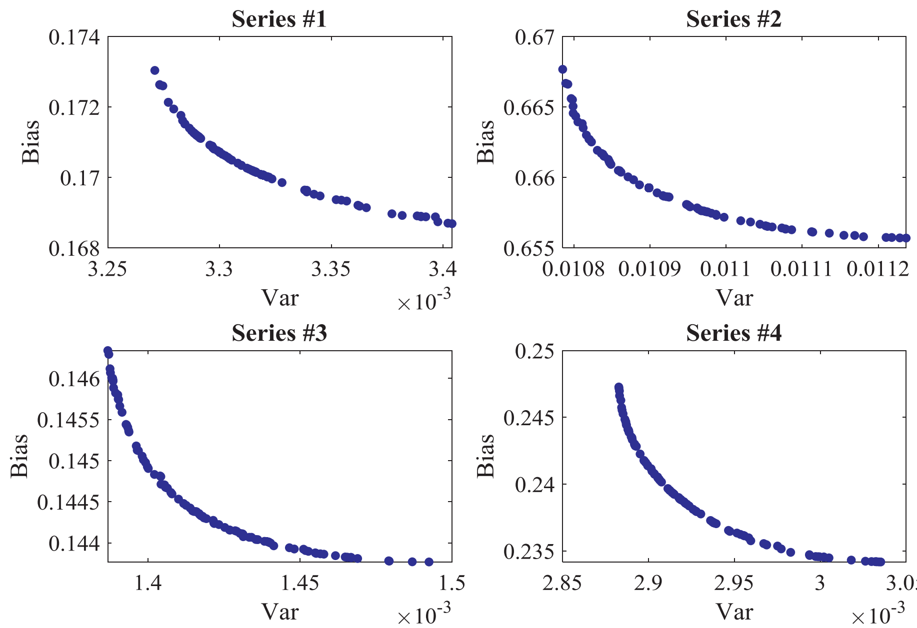

Fig. 6. The final Pareto font.

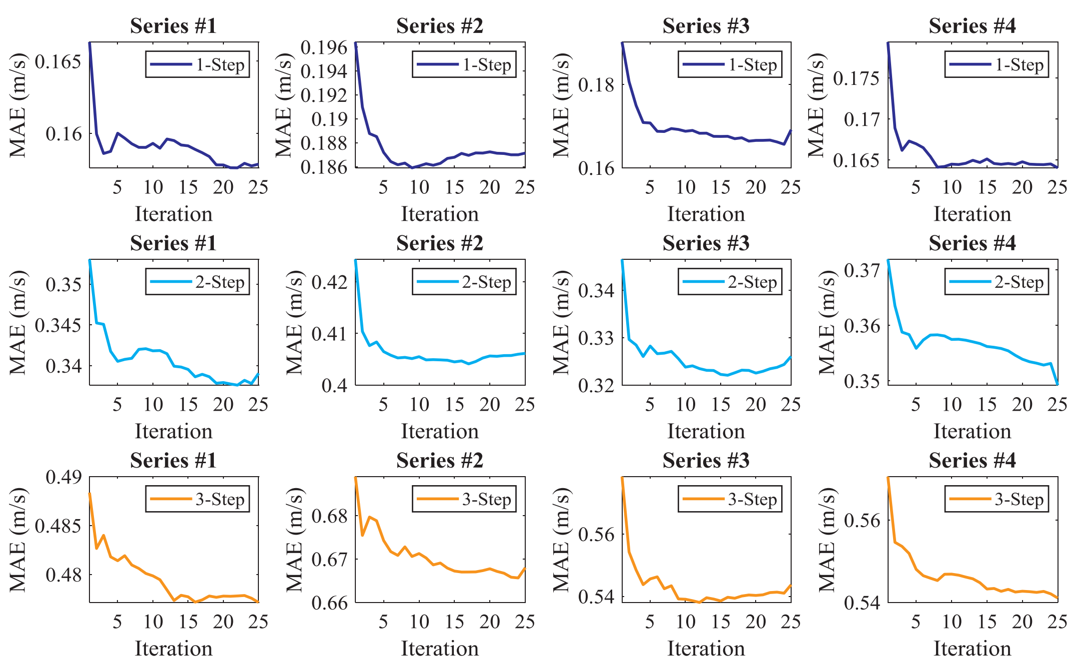

Fig. 7. The MAE during the training process of the AdaBoost.MRT-ORELM.

models. Qu et al. [21] proposed a novel multi-objective optimization algorithm, namely the MOBSFPA (Multi-Objective Flower Pollination Algorithm with Bat Search Algorithm), to combine different ANNs to reach the high-precision forecasting. The results presented in references [20] and [21] verified the good performance of the heuristic algorithm methods in the wind speed forecasting.

(b) The decomposition method could change the original wind speed series into different frequency signal bands. Liu et al. [22] proposed a novel model based on the EMD (Empirical Mode Decomposition) and the ELM. The experimental results in the study indicated that the proposed model outperformed the single ELM significantly. Wang et al. [23] proposed a hybrid model using the EMD and the ENN. The results showed that the hybrid EMD-ENN had better performance than the single ENN. Lei et al. [24] utilized the Wavelet Decomposition (WD) to decompose the original wind speed into several subseries. The Auto-regressive Moving Average (ARMA) model was used to forecasting each subseries. The results verified the effectiveness of the WD algorithm in wind speed forecasting. Li et al. [25] proposed a hybrid decomposition model, consisting of the Variational Mode Decomposition (VMD), the Gram-Schmidt orthogonal and the ELM. The parameters of the involved three algorithms were optimized by the Gravitational Search Algorithm (GSA). The results indicated the proposed synchronous optimization model outperformed other benchmark models significantly. To further improve the performance of the decomposition methods, the secondary decomposition method was proposed. The proposed secondary methods can be divided into two kinds. In the first kind, the original wind speed layers are estimated by a number of error indicators before being executed the secondary

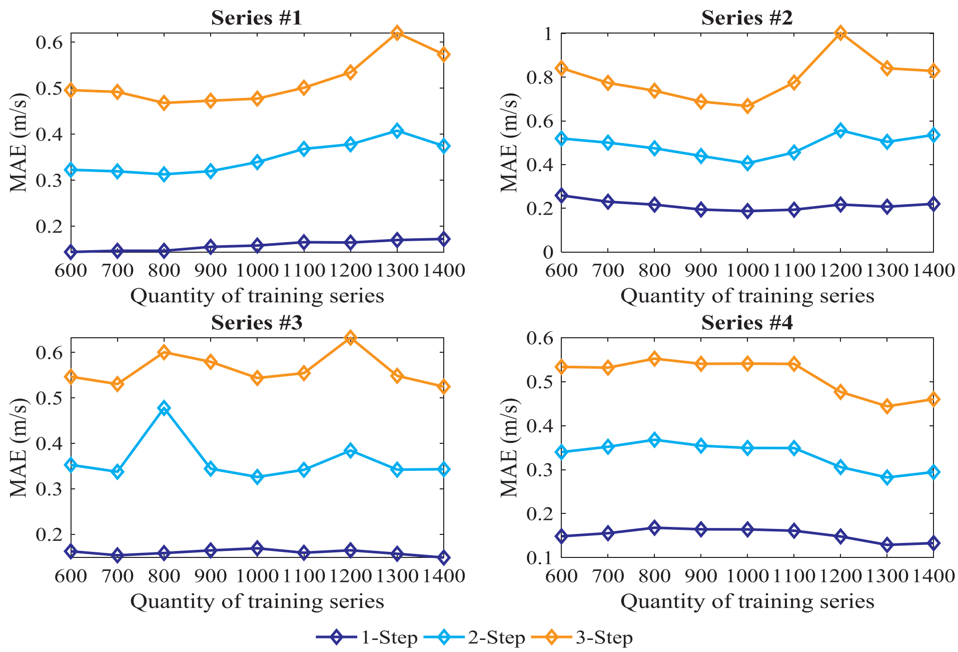

Fig. 8. The influences of training data quantity on MAE.

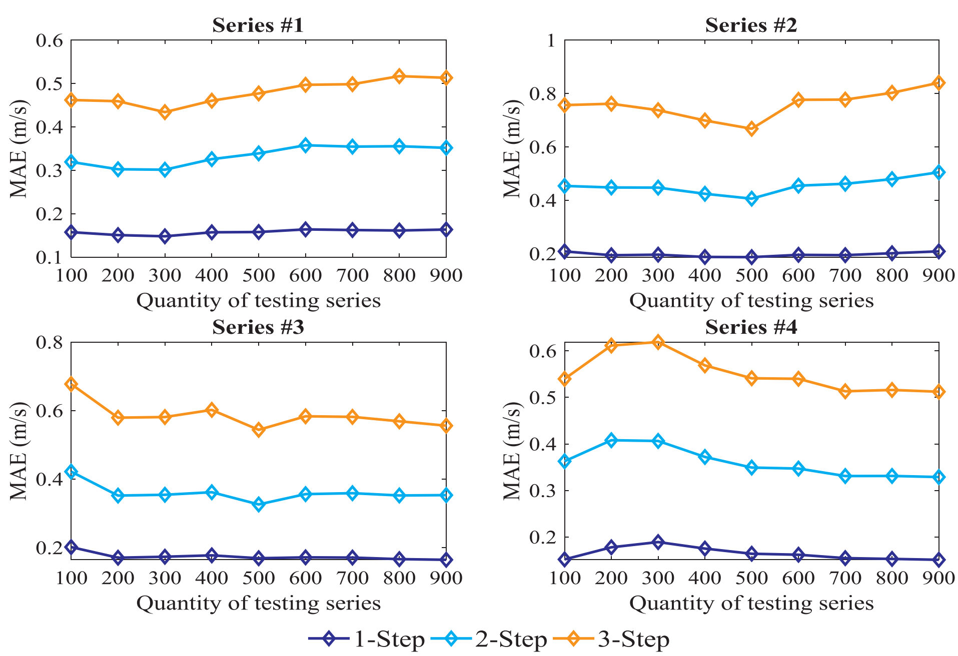

Fig. 9. The influences of testing data quantity on MAE.

decomposition. Liu et al. [17] decomposed the layer with minimum Sample Entropy (SampEn). Peng et al. [26] decomposed the Intrinsic Mode Functions (IMF) with the highest frequency. Both of these two methods were proved to be effective. In the second kind, a specified layer is expected to be decomposed without any performance estimation. Liu et al. [27] decomposed all detailed layers after the WPD computation. Yu et al. [28] extracted the trend of a IMF1 layer with the Singular Spectrum Analysis (SSA).

In the study, a new hybrid wind speed forecasting method is proposed based on the WPD, the AdaBoost.MRT [29] and the ORELM (Outlier-Robust Extreme Learning Machine) [30]. The proposed model is named as the WPD-AdaBoost.MRT-ORELM. Since the mother wavelet can affect the decomposition of the WPD as shown in reference [31], the proposed WPD-AdaBoost.MRT-ORELM method is studied with different mother wavelets. Additionally, in the proposed forecasting structure, a multi-objective optimization model named as the MOGWO [32] is applied to assemble all base predictors. Combining the upper mentioned algorithms, the novel ensemble model, named as the MOGWO-WPD-AdaBoost.MRT-ORELM, is finally proposed. The main contributions of the proposed model are explained in details as follows:

(a) A novel ensemble model is proposed. The base predictors of the ensemble model should be diverse and of high performance. The base predictors in the previous study are to combine different ANNs. In this study, the base predictors are selected to be the hybrid models with different hyper-parameters, which have been rarely studied. The hybrid models with different hyper-parameter have an

Table 2
The forecasting performance of the WPD-AdaBoost.MRT-ORELM model with different mother wavelets.

Series	Step	MAE (m/s)	MAPE (%)	RMSE (m/s)	MAE (m/s)	MAPE (%)	RMSE (m/s)
#1		db1			db3		
	1-step	0.8466	3.7765	1.5224	0.6091	2.7283	0.7651
	2-setp	1.3583	6.0662	2.1815	0.9322	4.2017	1.2407
	3-step	1.5925	7.1562	2.4236	1.2646	5.7088	1.6025
		db5			db7		
	1-step	0.3676	1.6468	0.4625	0.2662	1.1928	0.329
	2-setp	0.6309	2.8616	0.809	0.5327	2.3859	0.6721
	3-step	0.9089	4.1183	1.1638	0.735	3.2779	0.9124
		db9					
	1-step	0.205	0.9137	0.2577			
	2-setp	0.4488	2.0015	0.5612			
	3-step	0.6127	2.7236	0.7592			
#2		db1			db3		
	1-step	1.2532	4.9854	2.325	0.7398	2.9865	0.9575
	2-setp	1.8213	7.2079	2.9535	1.1231	4.5299	1.4921
	3-step	2.099	8.4113	3.1815	1.5606	6.3327	2.0119
		db5			db7		
	1-step	0.4667	1.8929	0.6031	0.3287	1.329	0.4143
	2-setp	0.8534	3.4201	1.0778	0.6405	2.585	0.8193
	3-step	0.4667	1.8929	0.6031	0.9304	3.7453	1.155
		db9					
	1-step	0.2572	1.042	0.3272			
	2-setp	0.5422	2.1931	0.684			
	3-step	0.8068	3.2231	1.0249			
#3		db1			db3		
	1-step	0.9635	4.9587	1.7922	0.6459	3.379	0.8354
	2-setp	1.4131	7.266	2.2601	1.0869	5.6997	1.4596
	3-step	1.695	8.6453	2.6133	1.3064	6.8639	1.6941
		db5			db7		
	1-step	0.4468	2.3531	0.5674	0.2985	1.5672	0.3799
	2-setp	0.7957	4.1489	1.0115	0.5672	2.9754	0.7218
	3-step	1.0302	5.3545	1.298	0.842	4.4049	1.0537
		db9					
	1-step	0.2365	1.2437	0.3064			
	2-setp	0.4442	2.3224	0.5668			
	3-step	0.6804	3.5284	0.8493			
#4		db1			db3		
	1-step	0.9084	3.5724	1.6329	0.5716	2.2837	0.7314
	2-setp	1.3455	5.3329	2.0482	0.9095	3.6268	1.223
	3-step	1.54	6.1257	2.2129	1.1437	4.5881	1.5086
		db5			db7		
	1-step	0.3463	1.3862	0.4509	0.2665	1.0698	0.3346
	2-setp	0.6641	2.6515	0.8323	0.5487	2.1782	0.7013
	3-step	0.8549	3.4384	1.0917	0.7565	3.0012	0.9533
		db9					
	1-step	0.2128	0.8467	0.2741			
	2-setp	0.4415	1.7576	0.5619			
	3-step	0.646	2.581	0.807			

important benefit that they can guarantee the diversity of the base predictors. To improve the performance of the base predictors, the proposed model uses a multi-objective optimizer MOGWO to combine all the base predictors.

(b) The decomposition and boosting algorithm is adopted to enhance the forecasting performance of the base predictor. The ORELM combined with the WPD and AdaBoost.MRT serves as the base predictor of the proposed model instead of the single ORELM. The WPD model can decompose the original series into more predictable subseries. The AdaBoost.MRT can combine several ORELMs to enhance the forecasting performance of the single ORELM.

(c) The impact of various mother wavelets in the proposed hybrid forecasting model is investigated. Although the WPD is generally used in the wind speed forecasting, the tuning and optimization of mother wavelets in the WPD for the wind speed forecasting has not been investigated before. The study will investigate how the mother wavelet affects the wind speed forecasting performance to find the optimal mother wavelet.

(d) Various forecasting models and wind speed series are included in the performance experiments to verify the superiority of the proposed model. The GWO-WPD -AdaBoost.MRT-ORELM model, the WPD-AdaBoost.MRT-ORELM-Best mother wavelet model, the MOGWO-WPD-ORELM model and other benchmark models are compared in the experiments. Four real wind speed series collected from different sites and at different time serve as simulation series to demonstrate the performance of the proposed model. At the same time, three mainstream indexes for the forecasting error estimation, including the MAE (Mean Absolute Error), the MAPE (Mean Absolute Percent Error) and the RMSE (Root Mean Square Error), are employed to estimate the forecasting performance of every used forecasting model.

2. Framework of the proposed hybrid model

The framework of the proposed model is shown in Fig. 1. All the wind speed data are divided into 4 data sets, including the set of optimization training data, the set of optimization testing data, the set of training data and the set of testing data. The framework of the proposed model can be summarized as follows:

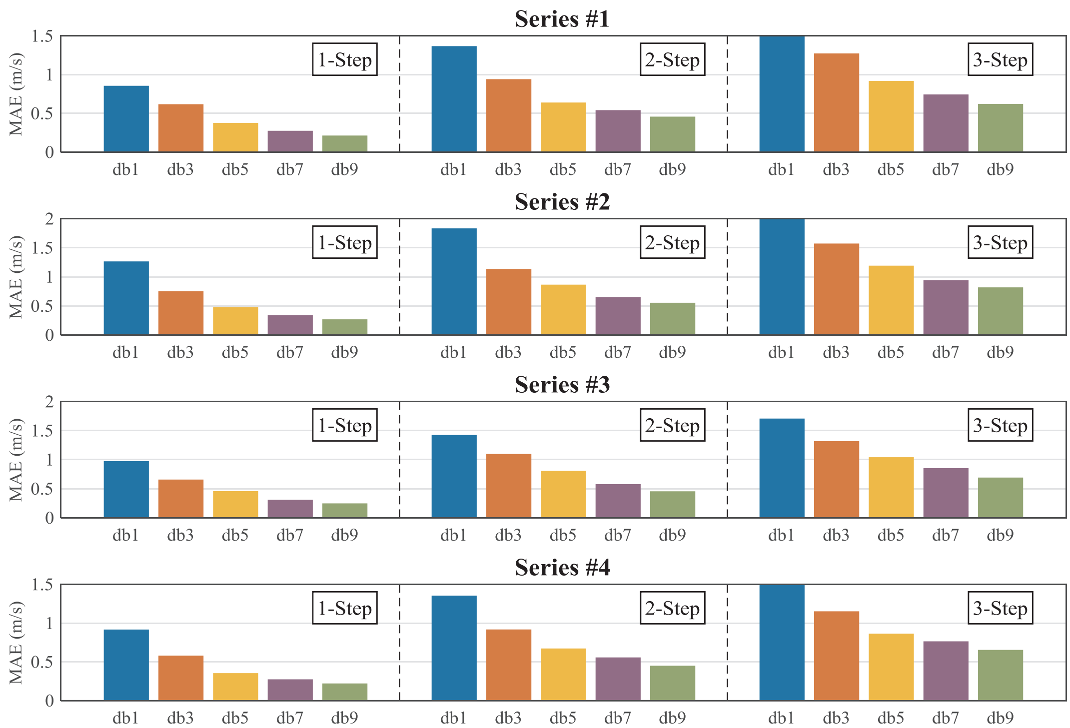

Fig. 10. The MAE of the WPD-Adaboost-ELM model with different mother wavelets.

Table 3
The decomposition errors in training part of the series #1 with different mother wavelets.

Wavelet	MAE	MAPE(%)	RMSE
db1	6.1355e − 15	4.4332e − 14	6.6596e − 15
db3	2.8273e − 11	2.1607e − 10	3.4747e − 11
db5	7.5916e − 12	5.8809e − 11	9.4905e − 12
db7	6.8112e − 12	5.2248e − 11	8.4715e − 12
db9	1.5938e − 10	1.2161e − 9	1.9861e − 10

- In step 1, the optimization training data is processed by WPD-AdaBoost.MRT -ORELM model with different mother wavelets to obtain several base optimized forecasting data.
- In step 2, the MOGWO is employed to find the optimal coefficients of all base optimization forecasting data. The optimization model reduces the bias and the variance between the combined individual optimization forecasting data and the optimization testing data.
- In step 3, the training data is processed in the same way as in step 1 to obtain the base forecasting data.
- In step 4, the optimal coefficients obtained in step 2 are utilized to combine all the base forecasting data to obtain the final forecasting data.

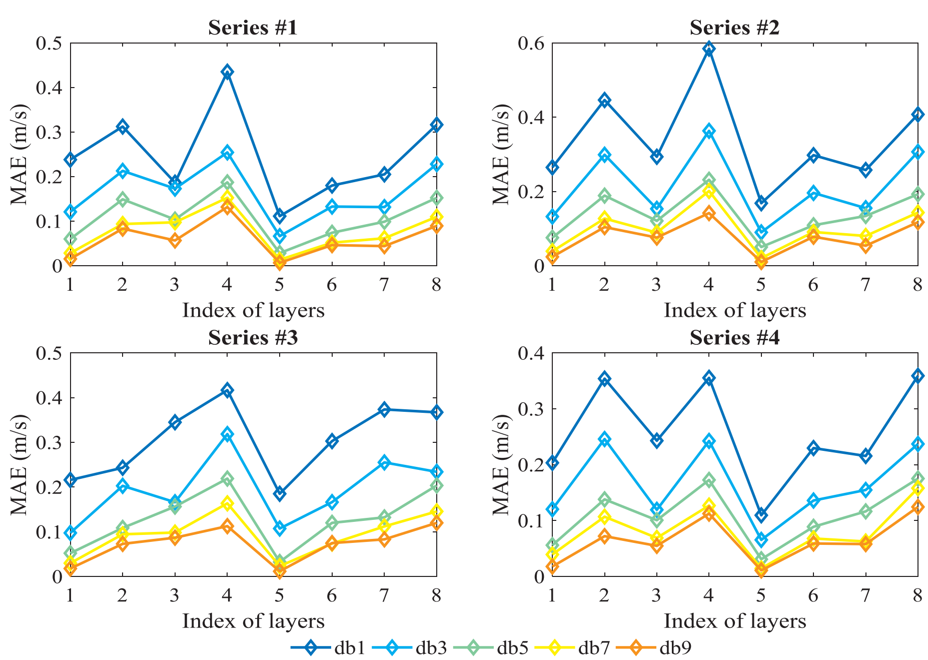

Fig. 11. The 1-step layer-wise MAE with wavelets of different vanishing moments.

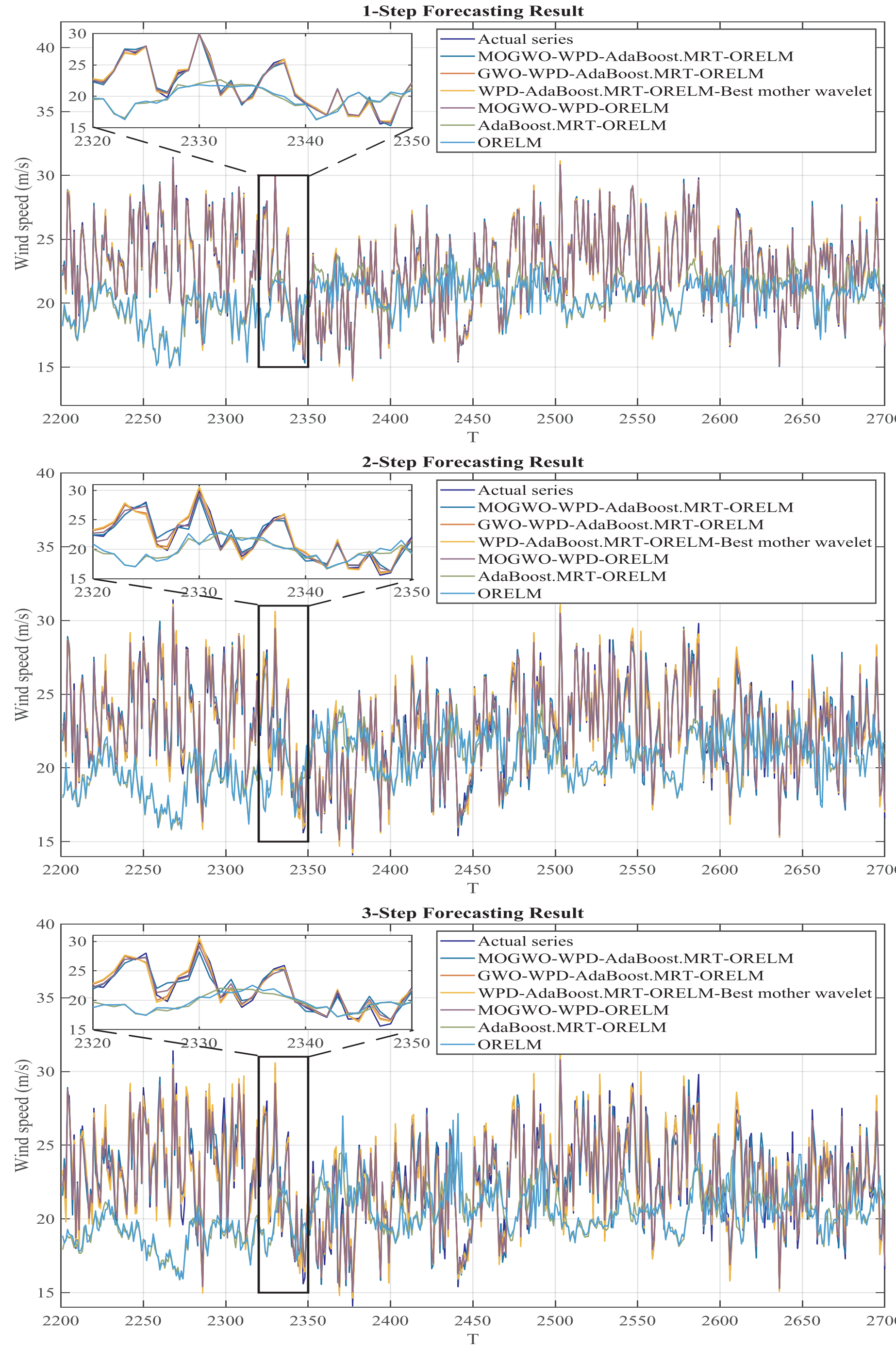

Fig. 12. The forecasting results of the series #1.

3. Methodology

3.1. Wavelet packet decomposition

The WPD is a signal decomposing method to process the wind speed data and extract the features of the original wind speed data. The wavelet transform is divided into two main types: the CWT (Continuous Wavelet Transform) and the DWT (Discrete Wavelet Transform).

The CWT is formulized as follow:

$$\mathrm{CWT}_X(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} X(t)\Psi\left(\frac{t-b}{a}\right)dt \tag{3.1}$$

where $X(t)$ is the original signal, $\Psi\left(\frac{t-b}{a}\right)$ is the mother wavelet. a and b determine the scale and translation of the wavelet function, respectively.

The DWT uses discrete a and b to dispose discrete signals. The DWT can be formulized as follows:

$$\mathrm{DWT}_X(m,n) = 2^{-\left(\frac{m}{2}\right)}\sum_{-\infty}^{\infty} X(t)\Psi\left(\frac{t-n*2^m}{2^m}\right) \tag{3.2}$$

The scale and translational variables are expressed as follows:

$$\begin{cases} a = 2^m \\ b = n2^m \end{cases} \tag{3.3}$$

where m and n are integers.

In the study, the Mallat algorithm [33] is adopted to decompose

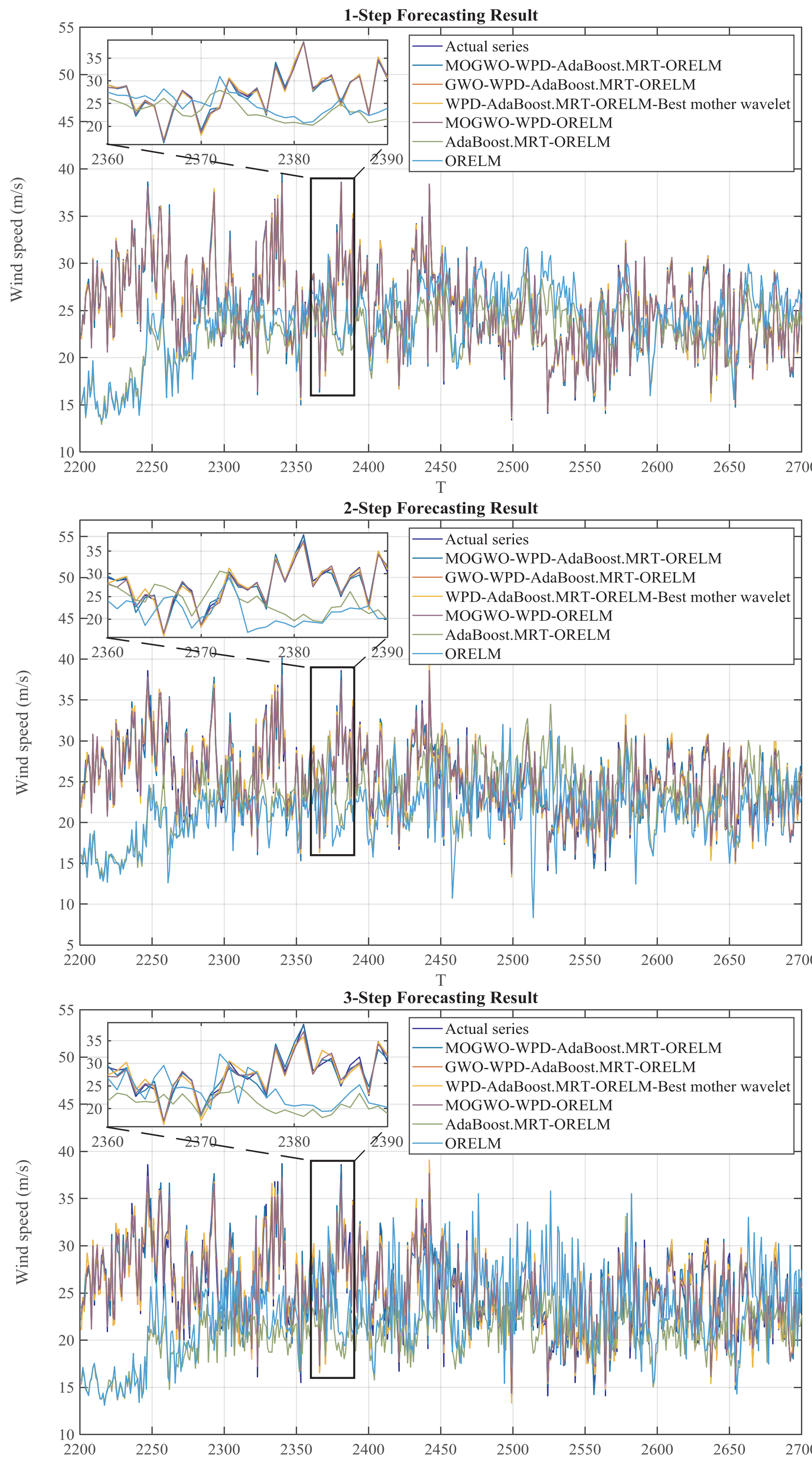

Fig. 13. The forecasting results of the series #2.

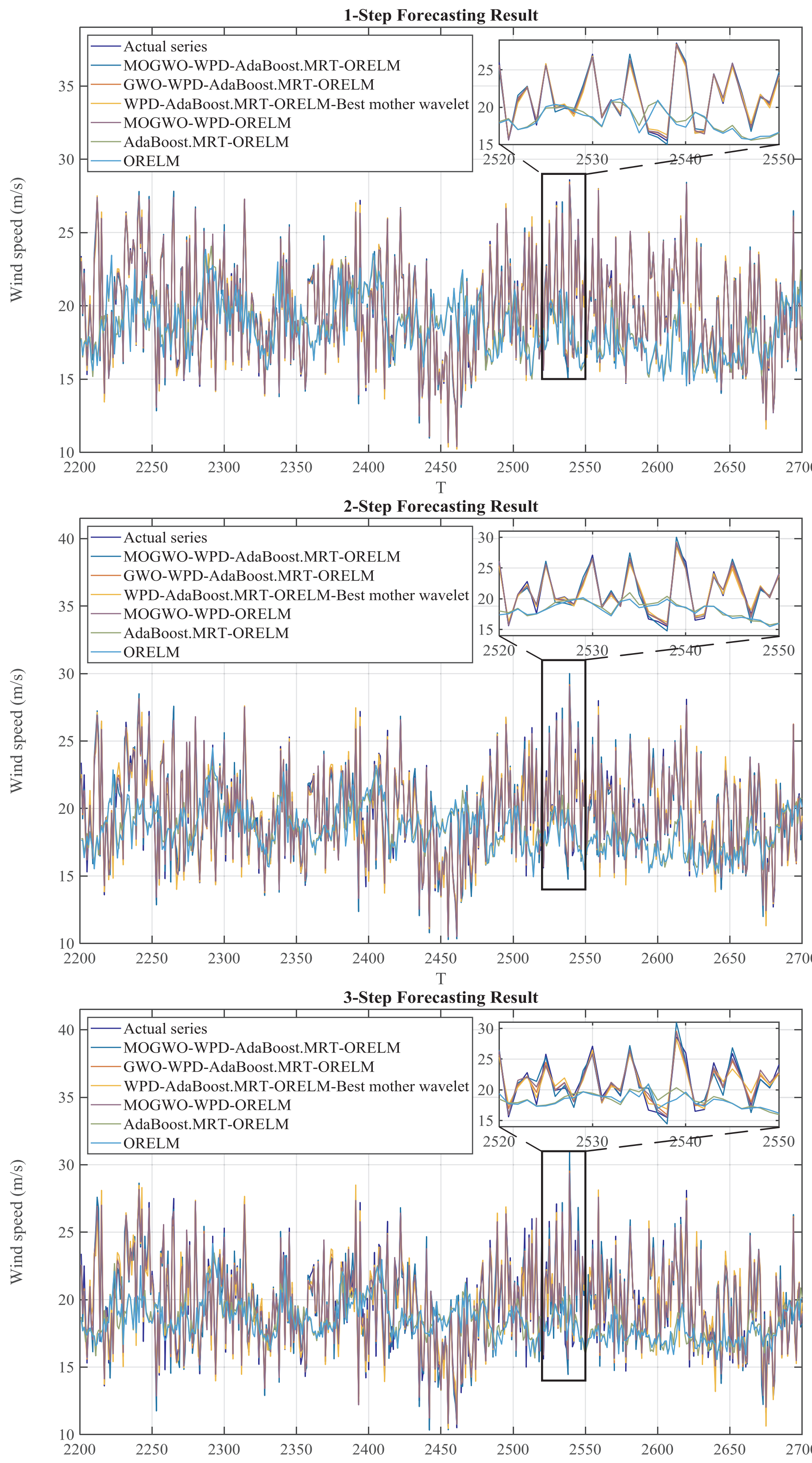

Fig. 14. The forecasting results of the series #3.

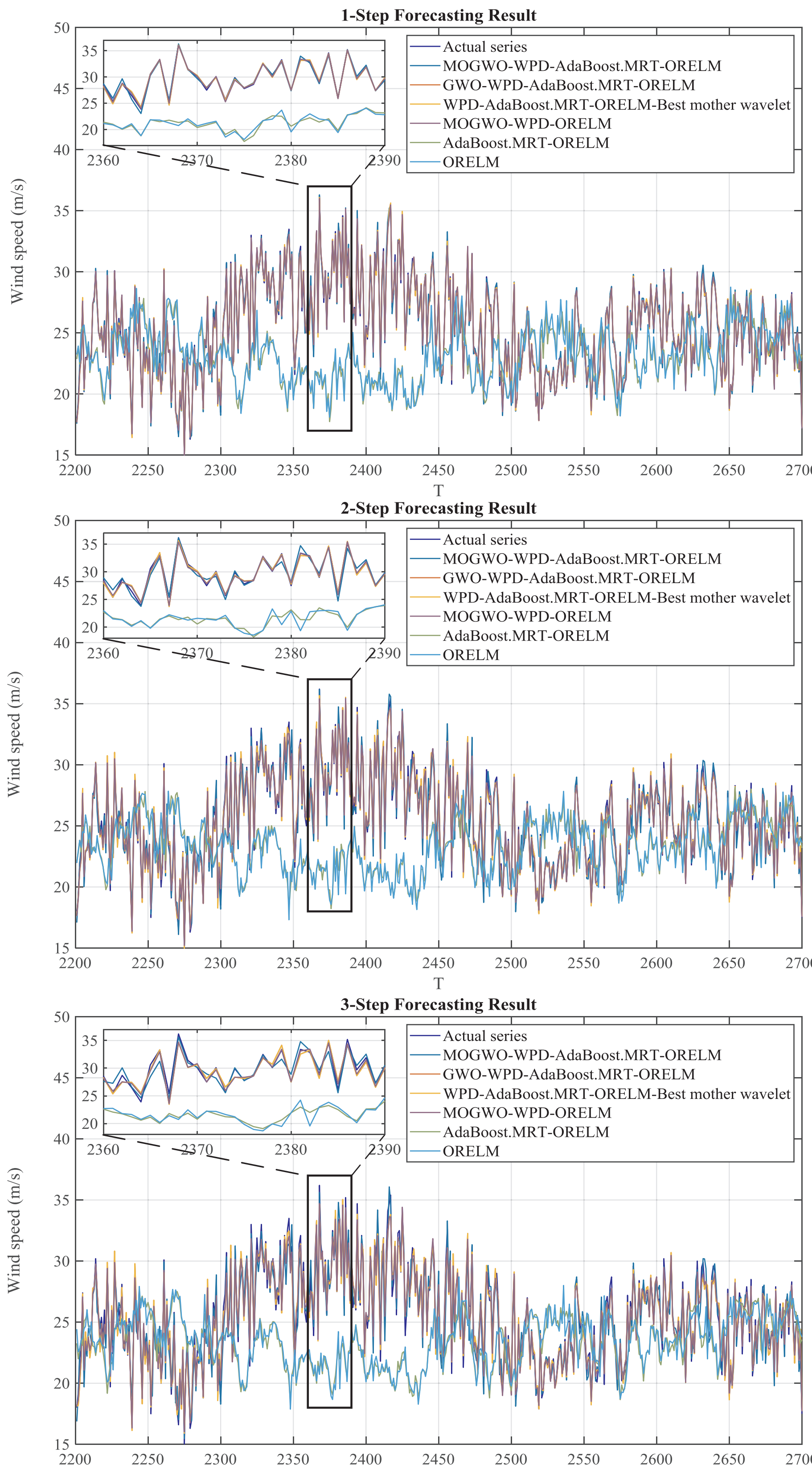

Fig. 15. The forecasting results of the series #4.

Table 4
The forecasting results of the comparison analysis.

Series	Step	MAE (m/s)	MAPE (%)	RMSE (m/s)	MAE (m/s)	MAPE (%)	RMSE (m/s)
#1		MOGWO-WPD-AdaBoost.MRT-ORELM			GWO-WPD-AdaBoost.MRT-ORELM		
	1-step	0.1579	0.7033	0.2001	0.1831	0.8159	0.232
	2-setp	0.3391	1.5162	0.429	0.3924	1.7534	0.4951
	3-step	0.4771	2.1311	0.5998	0.5403	2.4089	0.6762
		WPD-AdaBoost.MRT-ORELM-Best mother wavelet			MOGWO-WPD-ORELM		
	1-step	0.205	0.9137	0.2577	0.1769	0.7909	0.221
	2-setp	0.4488	2.0015	0.5612	0.3633	1.627	0.4577
	3-step	0.6127	2.7236	0.7592	0.511	2.2857	0.6437
		AdaBoost.MRT-ORELM			ORELM		
	1-step	3.4901	17.6571	4.4197	3.5139	17.8428	4.4188
	2-setp	3.5992	18.2022	4.4952	3.557	17.8812	4.4514
	3-step	3.7189	18.8803	4.5948	3.7815	19.1362	4.6395
#2		MOGWO-WPD-AdaBoost.MRT-ORELM			GWO-WPD-AdaBoost.MRT-ORELM		
	1-step	0.1871	0.7558	0.24	0.2572	1.042	0.3272
	2-setp	0.4062	1.6429	0.5158	0.5422	2.1931	0.684
	3-step	0.668	2.6755	0.8441	0.8068	3.2231	1.0249
		WPD-AdaBoost.MRT-ORELM-Best mother wavelet			MOGWO-WPD-ORELM		
	1-step	0.2572	1.042	0.3272	0.2151	0.8818	0.2832
	2-setp	0.5422	2.1931	0.684	0.4389	1.7799	0.5493
	3-step	0.8068	3.2231	1.0249	0.713	2.8817	0.8828
		AdaBoost.MRT-ORELM			ORELM		
	1-step	4.9367	23.862	6.319	5.0908	23.671	6.4573
	2-setp	5.2497	25.0178	6.6846	5.6446	29.5062	7.1501
	3-step	5.8068	30.4782	7.2722	5.5793	27.1052	7.0977
#3		MOGWO-WPD-AdaBoost.MRT-ORELM			GWO-WPD-AdaBoost.MRT-ORELM		
	1-step	0.1691	0.8812	0.2133	0.1704	0.8883	0.2153
	2-setp	0.3261	1.6876	0.4142	0.3304	1.7096	0.4202
	3-step	0.5437	2.8173	0.6779	0.5495	2.8476	0.6853
		WPD-AdaBoost.MRT-ORELM-Best mother wavelet			MOGWO-WPD-ORELM		
	1-step	0.2365	1.2437	0.3064	0.1924	1.0034	0.2437
	2-setp	0.4442	2.3224	0.5668	0.3556	1.8407	0.4593
	3-step	0.6804	3.5284	0.8493	0.5792	2.9965	0.7398
		AdaBoost.MRT-ORELM			ORELM		
	1-step	3.164	17.311	3.9659	3.2079	17.7222	4.0385
	2-setp	3.089	16.8481	3.8693	3.1248	17.2079	3.9394
	3-step	3.0368	16.5689	3.7888	3.0537	16.7773	3.8378
#4		MOGWO-WPD-AdaBoost.MRT-ORELM			GWO-WPD-AdaBoost.MRT-ORELM		
	1-step	0.164	0.6532	0.213	0.2128	0.8467	0.2741
	2-setp	0.3492	1.3862	0.4426	0.4415	1.7576	0.5619
	3-step	0.5411	2.1529	0.6745	0.646	2.581	0.807
		WPD-AdaBoost.MRT-ORELM-Best mother wavelet			MOGWO-WPD-ORELM		
	1-step	0.2128	0.8467	0.2741	0.1793	0.7128	0.2355
	2-setp	0.4415	1.7576	0.5619	0.3629	1.4359	0.4663
	3-step	0.646	2.581	0.807	0.5437	2.1633	0.6859
		AdaBoost.MRT-ORELM			ORELM		
	1-step	4.2849	19.2699	5.3731	4.2778	19.1644	5.3664
	2-setp	4.1493	18.4801	5.2814	4.1927	18.7187	5.3337
	3-step	4.165	18.4883	5.2622	4.1236	18.2394	5.2342

digital series into several subseries. The three-layer binary tree of the WPD is shown in Fig. 2.

The mother wavelet function is the key parameter of the WPD. The naming of wavelet functions uses an alphanumeric scheme based on the function family and the vanishing moments. The vanishing moment of a wavelet $\psi(t)$ is N_V, when below expression is satisfied for $\forall\ p \in [1, 2, ..., N_V]$:

$$\int t^p \psi(t) dt = 0 \tag{3.4}$$

In the previous studies, the Daubechies function family was a commonly used function family. Mi et al. [34], Tascikaraoglu et al. [35] and Azimi et al. [36] both used the db4 function. Yu et al. [37] used the db6 function.

The magnitude-frequency characteristics of the wavelet low-pass filters with the different vanishing moments and the ideal filter are illustrated in Fig. 3. From Fig. 3, it can be seen that the low-pass filter gradually approximates to the ideal filter with the increase of vanishing moments. This phenomenon indicates that the mother wavelet with larger vanishing moments could prevent the linkage of higher frequency components and accurately decompose the frequency band of the subseries.

3.2. AdaBoost.MRT

In the multi-step wind speed forecasting, the outputs of the predictors contain multiple variables under MIMO (Multi-Input and Multi-Output) forecasting strategy [38]. The AdaBoost.MRT [29] is upgraded from classical AdaBoost algorithm, which could handle multiple output variables. The pseudo code of the AdaBoost.MRT can be expressed as Algorithm 1.

Algorithm 1 (*AdaBoost.MRT*).

Input:
- Training set $\mathbf{S} = \{(\mathbf{x}_n, \mathbf{y}_n)|n = 1, 2, ..., N\}$, where N is number of instances.
- Output vector $\mathbf{y}_n = \{y_n^r | r = 1, 2, ..., R\}$, where R is number of variables.
- Weak learner algorithm WL_t.
- The number of example M.
- Sampling weight distribution $\mathbf{D_t}(i) = 1/M$ for all examples.
- Output error distribution $\mathbf{D}_\mathbf{t}^{(r)}(i) = 1/M$ for all output variables of each example.
- Maximum number of iterations T.
- Threshold vector $\mathbf{\Phi} = \{\phi^r | r = 1...R\}$ for classifying forecasting example as correct or incorrect.

Output
- Output forecasting function $F_A(x)$.

Algorithm

```
for t = 1: T do
    Sample M examples from S using weight D_t with replacement, where M < N.
    Train WL_t with M examples, and build regression model h_t(x) → y.
    Compute the errors of all M examples for each output variable:
        Er_t^(r)(i) = |h_t^r(x_i) − y_i^r| / σ_t^r, where σ_t^r is the standard deviation of (h_t^r(x_i) − y_i^r).
    Calculate the misclassification error rate of all M examples for each output variable:
        ε_t^(r) = Σ_{i:Er_t^r > φ^r} D_t^(r)(i).
    Calculate the weight updating parameter: β_t^r = (ε_t^(r))^n, where n is the power coefficient.
        It is suggested that n = 2 or n = 3 [39].
    Update the output error distribution D_t^(r) as:
        D_{t+1}^(r)(i) = D_t^(r)(i)/Z_t × { β_t^r   Er_t^(r)(i) < φ^r ; 1   otherwise },
        where Z_t is the normalization factor.
    Update the output error distribution D_t as:
        D_{t+1}(i) = (1/R) Σ_{k=1}^r D_t^(r=k)(i).
end for
Output forecasting function F_A(x) = Σ_{t=1}^T h_t(x)/T.
return F_A(x)
```

In this study, T is set to 25 and M is set to $0.9N$ after tuning, respectively.

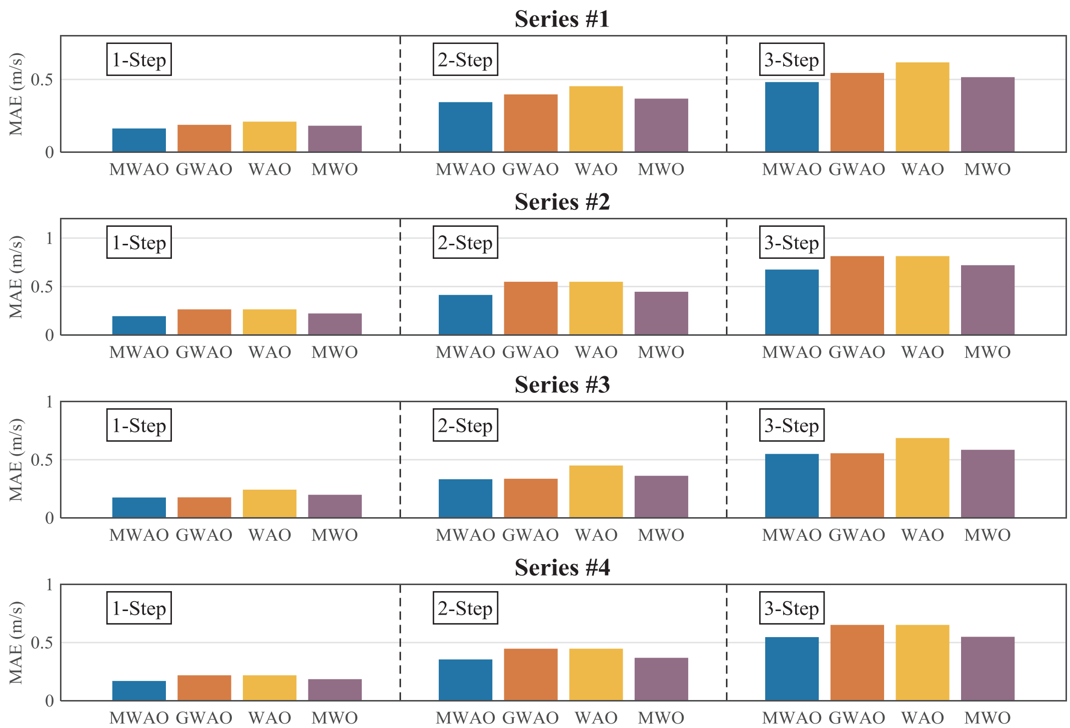

Fig. 16. The MAE of the former four models in the comparison analysis.

3.3. Forecasting algorithm

The ELM [40] algorithm has good generalization performance and it needs less computing time than the traditional algorithms. The ELM is widely used in various fields, such as the wind speed forecasting [28], the stock market forecasting [41], the turbojet engine modeling [42], etc. The accuracy of the ELM can be affected by the outliers. To cope with the overfitting of outliers, the ORELM is proposed to avoid overfitting. The objective function and the constraint of the ORELM with l_2-norm is formulized as follows:

$$\min_{\beta} \|\mathbf{e}\|_1 + \frac{1}{C}\|\beta\|_2 \quad s.\ t.\quad \mathbf{y}-\mathbf{H}\beta = \mathbf{e} \tag{3.5}$$

where C is the regularization parameter, $\mathbf{e}$ is the error series of the ORELM.

3.4. Multi-objective optimization algorithm

Multi-objective optimization is a way to find the optimal solution [43]. This kind of optimization can also be combined with some popular signal decomposition [44]. Multi-objective optimization is widely used in different fields, such as the mechanical engineering [45], the civil engineering [46], the wind speed forecasting [47], etc. In the study, the adopted multi-objective optimization can be explained as follows:

$$\begin{aligned} \min:\quad & \mathrm{F}(\mathbf{x}) = \{f_q(\mathbf{x})|q = 1, 2, ..., Q\} \\ s.\ t.\quad & \mathrm{g}_m(\mathbf{x}) \geqslant 0,\ m = 1, 2, ..., M \\ & h_p(\mathbf{x}) \geqslant 0,\ p = 1, 2, ..., P \\ & L_n \leqslant x_n \leqslant U_n,\ n = 1, 2, ..., N \end{aligned} \tag{3.6}$$

where $F(\mathbf{x})$ is the multi-objective function, Q is the number of the objective functions, M is the quantity of the inequality constraints, P is the quantity of the equality constraints, g_m is the inequality constraint, h_i is the equality constraint and $[\mathrm{L}_i, \mathrm{U}_i]$ is the boundary of the variable quantity.

The GWO [48] is built based on the social leadership and the hunting strategy of the grey wolves. In the study, the following equations are adopted to describe the surrounding behaviors of the grey wolves when they hunt the target.

$$D_\alpha = |C_1 \cdot X_\alpha - X| \tag{3.7}$$

$$D_\beta = |C_1 \cdot X_\beta - X| \tag{3.8}$$

$$D_\delta = |C_1 \cdot X_\delta - X| \tag{3.9}$$

$$X_1 = X_\alpha - A_1 \cdot (\mathrm{D}_\alpha) \tag{3.10}$$

$$X_2 = X_\beta - A_2 \cdot (\mathrm{D}_\beta) \tag{3.11}$$

$$X_3 = X_\delta - A_3 \cdot (\mathrm{D}_\delta) \tag{3.12}$$

$$X(\mathrm{t}+1) = \frac{X_1 + X_2 + X_3}{3} \tag{3.13}$$

where X_α, X_β, X_δ are the positions of the alpha, beta and delta wolfs, X is position of a certain grey wolf, A is a random vector varying from -1 to 1, and C is a random vector varying from 0 to 2.

To realize the MOGWO computation in the study, two additional mechanisms in the GWO (i.e., archive mechanism and leader selection mechanism), are used to assist the MOGWO to find the global optimal solution.

4. Comparison analysis

4.1. Wind speed data sets

To verify the effectiveness of the proposed model, four wind speed time series, collected in Xinjiang Province, China, are provided in this study. The series #1, #2 and series #3, #4 were collected at different sites. The series #1, #3 were collected in 2017, and the series #2, #4 were collected in 2015. Four wind speed time series are shown in Fig. 4. Each wind speed time series contains 2700 data. The 1st ~ 800th data are selected as the optimization training data, the 801th ~ 1200th data are chosen as the optimization testing data, the 1201th ~ 2200th data are selected as the training data and the 2201th ~ 2700th data are defined as the testing data. The statistical parameters of the series are

Table 5
The improvement percentage of the comparison models by the proposed model.

Series	Step	P_{MAE} (%)	P_{MAPE} (%)	P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)	P_{RMSE} (%)
#1		GWO-WPD-AdaBoost.MRT-ORELM			WPD-AdaBoost.MRT-ORELM-Best mother wavelet		
	1-step	13.7856	13.8046	13.7391	22.9742	23.031	22.3576
	2-setp	13.5901	13.5323	13.3504	24.4501	24.249	23.5638
	3-step	11.6879	11.5329	11.2935	22.1211	21.7558	20.9968
		MOGWO-WPD-ORELM			AdaBoost.MRT-ORELM		
	1-step	10.7406	11.0865	9.4403	95.4767	96.0171	95.4728
	2-setp	6.6709	6.8132	6.2848	90.5798	91.6704	90.4571
	3-step	6.6281	6.7666	6.82	87.1696	88.7127	86.9457
		ORELM					
	1-step	95.5074	96.0586	95.4718			
	2-setp	90.4678	91.5209	90.3633			
	3-step	87.3822	88.8636	87.0713			
#2		GWO-WPD-AdaBoost.MRT-ORELM			WPD-AdaBoost.MRT-ORELM-Best mother wavelet		
	1-step	27.2469	27.4681	26.6431	27.2469	27.4681	26.6431
	2-setp	25.0968	25.0867	24.5921	25.0968	25.0867	24.5921
	3-step	17.2027	16.9907	17.6452	17.2027	16.9907	17.6452
		MOGWO-WPD-ORELM			AdaBoost.MRT-ORELM		
	1-step	12.9824	14.2931	15.2311	96.2093	96.8328	96.2015
	2-setp	7.4703	7.6977	6.1041	92.2633	93.4331	92.2836
	3-step	6.3158	7.1563	4.3864	88.4962	91.2216	88.3929
		ORELM					
	1-step	96.3241	96.8072	96.2829			
	2-setp	92.8046	94.432	92.7859			
	3-step	88.0271	90.1292	88.1076			
#3		GWO-WPD-AdaBoost.MRT-ORELM			WPD-AdaBoost.MRT-ORELM-Best mother wavelet		
	1-step	0.7673	0.7949	0.9215	28.4711	29.1427	30.3871
	2-setp	1.2997	1.2896	1.434	26.5966	27.3356	26.9343
	3-step	1.053	1.0632	1.0785	20.0849	20.1532	20.1839
		MOGWO-WPD-ORELM			AdaBoost.MRT-ORELM		
	1-step	12.0978	12.1749	12.453	94.6544	94.9094	94.621
	2-setp	8.3138	8.3201	9.8191	89.4443	89.9836	89.2961
	3-step	6.1326	5.9811	8.3696	82.0958	82.9964	82.1075
		ORELM					
	1-step	94.7277	95.0275	94.7176			
	2-setp	89.5653	90.193	89.4868			
	3-step	82.1945	83.2076	82.3362			
#4		GWO-WPD-AdaBoost.MRT-ORELM			WPD-AdaBoost.MRT-ORELM-Best mother wavelet		
	1-step	22.9228	22.859	22.2934	22.9228	22.859	22.2934
	2-setp	20.8984	21.1334	21.234	20.8984	21.1334	21.234
	3-step	16.2458	16.5865	16.4179	16.2458	16.5865	16.4179
		MOGWO-WPD-ORELM			AdaBoost.MRT-ORELM		
	1-step	8.5291	8.3657	9.5536	96.1717	96.6105	96.0362
	2-setp	3.7803	3.4583	5.0719	91.5834	92.499	91.6191
	3-step	0.4806	0.4811	1.6632	87.0093	88.3552	87.1821
		ORELM					
	1-step	96.1653	96.5918	96.0313			
	2-setp	91.6705	92.5946	91.7014			
	3-step	86.8789	88.1963	87.1134			

shown in Table 1.

4.2. Performance evaluation

In the study, the MAE, MAPE and RMSE are employed to evaluate the forecasting accuracy of every involved model. The three error evaluation indexes are formulized as follows:

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

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n} (Y_t - \widetilde{Y}_t)^2} \tag{4.2}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{Y_t - \widetilde{Y}_t}{Y_t}\right| \tag{4.3}$$

where Y_t is the measured data, $\widetilde{Y}_t$ is the forecasted data, and n is the number of forecasted data.

The optimization objective functions are bias and variance in the study. Given the squared loss as $L(Y_t, \widetilde{Y}_t) = (Y_t - \widetilde{Y}_t)^2$, the loss values can be decomposed into three parts: bias, variance and noise [49]. The details can be given as follows:

$$\begin{aligned} E(L(\widetilde{Y}_t - Y_t)) &= E((\widetilde{Y}_t - Y_t)^2) \\ &= \mathrm{E}((\mathrm{E}(Y_t) - \widetilde{Y}_t)^2) + \mathrm{E}((Y_t - \mathrm{E}(Y_t))^2) \\ &= \mathrm{E}((\widetilde{Y}_t - \mathrm{E}(\widetilde{Y}_t) + \mathrm{E}(\widetilde{Y}_t) - E(Y_t))^2) + Noise \\ &= \mathrm{E}((\mathrm{E}(\widetilde{Y}_t) - E(Y_t))^2) + \mathrm{E}((\widetilde{Y}_t - \mathrm{E}(\widetilde{Y}_t))^2) + Noise \\ &= Bias + Var + Noise \end{aligned} \tag{4.4}$$

where $E(Y_t)$ is the expectation of the true value Y_t and $E(\widetilde{Y}_t)$ is the expectation of the true value $\widetilde{Y}_t$ through several experiments.

4.3. Analysis of the proposed model

4.3.1. Analysis of convergence

The proposed ensemble model consists of the MOGWO, WPD and AdaBoost.MRT-ORELM. The section analyzes the convergence of the MOGWO and AdaBoost.MRT-ORELM, because the WPD does not have the iterative process. To demonstrate the convergence of the MOGWO, the average objective functions of all solutions in the Pareto optimal set during the optimization process are shown in Fig. 5. The final Pareto fonts are shown in Fig. 6. Besides, the training process of the AdaBoost.MRT-ORELM is shown in Fig. 7.

According to Figs. 5–7, it can be summarized that:

(a) The MOGWO and AdaBoost.MRT-ORELM have absolute convergence. All training curves drop dramatically at the beginning and fluctuate slightly at the end. The final Pareto fonts of MOGWO are hyperbolic.
(b) The algorithm could converge at different sites and different years. The proposed model has good convergence.
(c) The guarantee of convergence under non-linear extreme variations is achieved at three levels: (1) The WPD decomposes non-linear extreme variations series into more stable subseries; (2) The ORELM, improved by the AdaBoost.MRT, is robust to outliers and has first-rate nonlinear fitting capacity of extreme variation data; (3) if the WPD-AdaBoost.MRT-ORELM model doesn't converge, the MOGWO can adjust the coefficients adaptively to guarantee the high-accuracy.

4.3.2. Analysis of series length

An important hypothesis of the machine learning theory is that all training data and testing data obey the same distribution. The effecting performance of training data quantity on MAE is shown in Fig. 8, where the testing data quantity is 500. The effecting performance of testing data quantity on MAE is shown in Fig. 9, where the training data quantity is 1000. To guarantee the fairness, the optimized coefficients used in training data quantity test are set to be fixed. Both the optimized coefficients and trained AdaBoost.MRT-ORELM models are fixed in the testing data set.

According to Figs. 8 and 9, it can be summarized that:

(a) The training data quantity does not have significant influences on the model performance. According to the classical machine learning theory, larger training data means clearer description of the data distribution. The model performance increases with the quantity of training data. However, the trend in Fig. 8 is contrary to the theory, which indicates that the time-variation reduces the benefits brought by large quantities of training data. To achieve good model performance, the quantity of training data should be properly increased in consideration of the time-variation and time-consumption.
(b) The impacts of testing data quantity on the model performance are not significant as well. Although the time-variation increases as the test data quantity increases, the performance fluctuates slightly. This indicates that the proposed model is robust and the model is effective in a long time range.
(c) In the previous studies, Xiao et al. [46] used 1500 data for training and 500 data for testing. Qureshi et al. used 1000 data for training and validation and 500 data for testing [19]. Song used 2160 data for training and 432 data for testing [50]. Khosravi et al. used 840 data for training and 361 data for testing [51]. Based on previous studies, the quantities of training and testing series in this study are selected to 1000 and 500, respectively. The quantities of training and testing data do not have significant influences on the model performance.

4.3.3. The comparison of different mother wavelets

The forecasting performance of the base predictors with different mother wavelets is shown in Table 2. Fig. 10 presents the MAE index with different vanishing moments.

According to Table 2 and Figs. 3, 10, it can be summarized that:

(a) The forecasting performance increases with the vanishing moments of mother wavelets. The mother wavelet with the best forecasting performance is db9, while the one with the worst forecasting performance is db1.
(b) According to the band partition and forecasting performance of different mother wavelets in Figs. 3 and 10. From Figs. 3 and 10, it can be concluded that the forecasting performance decreases with the increase of the vanishing moments of mother wavelets.

Besides the forecasting errors, the decomposition errors in the proposed hybrid forecasting framework are also important. The decomposition errors are shown in Table 3. From Table 3, it can be seen that the decomposition errors vary greatly with the mother wavelet.

The 1-step MAE error of each layer is shown in Fig. 11. The forecasting errors of each layer decrease with the increase of the vanishing moment. The phenomenon indicates the forecasting error of each layer is the main component of the final forecasting errors. The forecasting error of each layer is the main component of the final forecasting errors.

4.4. Comparison analysis

In the section, five models are provided as the comparison models, which are the GWO-WPD-AdaBoost.MRT-ORELM model, the WPD-AdaBoost.MRT-ORELM -Best mother wavelet, the MOGWO-WPD-ORELM model, the AdaBoost.MRT -ORELM model and the ORELM model. The forecasting wind speed series of all models are shown in Figs. 12–15. The forecasting performance of the six models is given in Table 4. Fig. 16 illustrates the MAE of the former four models, where abscissa labels are abbreviations of models. The MWAO represents the MOGWO-WPD-AdaBoost.MRT-ORELM model, the GWAO represents the GWO-WPD-AdaBoost.MRT-ORELM, the WAO represents the WPD-AdaBoost.MRT -ORELM-Best mother wavelet and MWO represents the MOGWO-WPD -ORELM. The percentage of improvement is shown in Table 5.

According to Tables 4, 5 and Figs. 12–16, it can be summarized that:

(a) The proposed model has outstanding performance in series #1 ∼ 4, which indicates the proposed model can be adapted in different sites and different years to achieve high-accuracy prediction. Taking 1-step forecasting as example, the MAE, MAPE and RMSE errors of the proposed model in series #1 are 0.1579 m/s, 0.7033% and 0.2001 m/s, respectively. The MAE, MAPE and RMSE errors of the proposed model in series #2 are 0.1871 m/s, 0.7558% and 0.24 m/s, respectively. The MAE, MAPE and RMSE errors of the proposed model in series #3 are 0.1691 m/s, 0.8812% and 0.2133 m/s, respectively. The MAE, MAPE and RMSE errors of the proposed model in series #4 are 0.164 m/s, 0.6532% and 0.213 m/s, respectively.
(b) The proposed model significantly outperforms the AdaBoost.MRT-ORELM and ORELM. It indicates that the proposed ensemble model could enhance performance of each single model. Taking 1-step forecasting of series #1 as example, the MAE, MAPE and RMSE errors of the proposed model are 0.1579 m/s, 0.7033% and 0.2001 m/s, respectively. The MAE, MAPE and RMSE errors of the AdaBoost.MRT-ORELM are 3.4901 m/s, 17.6571% and 4.4197 m/s, respectively. The MAE, MAPE and RMSE errors of the ORELM are 3.5139 m/s, 17.8428% and 4.4188 m/s, respectively. Compared with the proposed model, the improvement percentages of the MAE, MAPE and RMSE errors for the AdaBoost.MRT-ORELM model are 95.4767%, 90.5798% and 87.1696%, respectively. Compared with the proposed model, the improvement percentages of the MAE, MAPE and RMSE errors for the ORELM model are 95.5074%, 90.4678% and 87.3822%, respectively.
(c) The proposed model outperforms the MOGWO-WPD-ORELM significantly, which indicates that the AdaBoost.MRT can enhance the performance of the proposed model. Taking 1-step forecasting of series #2 as example, the MAE, MAPE and RMSE errors of the proposed model are 0.1871 m/s, 0.7558% and 0.24 m/s, respectively. The MAE, MAPE and RMSE errors of the MOGWO-WPD-ORELM model are 0.2572 m/s, 1.042% and 0.3272 m/s,

respectively. Compared with the proposed model, the improvement percentages of the MAE, MAPE and RMSE errors for the MOGWO-WPD-ORELM model are 12.9824%, 14.2931% and 15.2311%, respectively.

(d) The proposed model outperforms the GWO-WPD-AdaBoost.MRT-ORELM and WPD-AdaBoost.MRT-ORELM-Best mother wavelet significantly, which indicates that the multi-objective optimizer can enhance the performance. Taking 1-step forecasting of series #1 as example, the MAE, MAPE and RMSE errors of the proposed model are 0.1691 m/s, 0.8812% and 0.2133 m/s, respectively. The MAE, MAPE and RMSE errors of the GWO-WPD-AdaBoost.MRT-ORELM model are 0.1704 m/s, 0.8883% and 0.2153 m/s, respectively. The MAE, MAPE and RMSE errors of the WPD-AdaBoost.MRT-ORELM-Best mother wavelet model are 0.2365 m/s, 1.2437% and 0.3064 m/s, respectively. Compared with the proposed model, the improvement percentages of the MAE, MAPE and RMSE errors for the GWO-WPD-AdaBoost.MRT-ORELM model are 0.7673%, 0.7949% and 0.9215%, respectively. Compared with the proposed model, the improvement percentages of the MAE, MAPE and RMSE errors of the WPD-AdaBoost.MRT -ORELM-Best mother wavelet model are 28.4711%, 29.1427% and 30.3871%, respectively.

5. Energy applications

The utilization of wind energy has many potential applications, such as control of wind turbine system, integration of wind power, etc. Effective control of wind turbine system can avoid over-load and increase conversion efficiency. The major issue of control of wind turbine system is the time-delay of wind speed forecasting. Accurate forecasting of the wind speed can solve the time-delay issue to achieve dynamic control of wind turbine system [52]. Besides, the integration of wind power into power grid is indispensable during the utilization of wind energy. The high-performance wind speed forecasting can enhance the power grid stability and power quality with less reserve capacity. Due to the intermittency and volatility of the wind speed, the wind speed forecasting is difficult to be achieved. In this study, a novel ensemble model has been proposed to achieve accurate wind speed forecasting.

The proposed model was validated by four real wind speed series from different sites and years. The simulation results prove the good performance of the proposed model. Besides, the convergence of the proposed model is also considered. The proposed model can be expanded to forecast the wind speed with non-linear extreme variations.

6. Conclusion

In the study, a novel ensemble model, named as the MOGWO -WPD-AdaBoost.MRT -ORELM, is proposed to accurately forecast the wind speed. The MOGWO is used to assemble several WPD-AdaBoost.MRT-ORELM with different mother wavelets. Compared with the single ANNs of classical ensemble models, the base predictors in the proposed model are hybrid with different hyper-parameters, which enable the proposed ensemble model to achieve outstanding performance. The results show that: (a) the proposed ensemble model has good convergence; (b) the forecasting accuracy of the WPD-AdaBoost.MRT-ORELM increases with the increasing of the vanishing moment; (c) the proposed ensemble model has good multi-step wind speed forecasting results; and (d) the proposed ensemble model outperforms other benchmark models significantly including the base predictor with the best vanishing moment.

Acknowledgements

This study is fully supported by the National Natural Science Foundation of China (Grant No. 51308553), the Changsha Science & Technology Project (Grant No. KQ1707017), 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) and the engineering college 'double first-rate' supporting project of the Central South University. This study is partly supported by the National Natural Science Foundation of China (Grant No. U1534210).

References

[1] Liu H, Mi X, Li Y. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manage 2018;156:498–514.
[2] Soman SS, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. North American power symposium (NAPS), 2010. IEEE; 2010. p. 1–8.
[3] 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.
[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] El-Fouly TH, El-Saadany EF, Salama MM. One day ahead prediction of wind speed and direction. IEEE Trans Energy Convers 2008;23:191–201.
[6] El-Fouly T, El-Saadany E, Salama M. One day ahead prediction of wind speed using annual trends. 2006 IEEE power engineering society general meeting. IEEE; 2006. p. 7.
[7] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[8] Masseran N. Modeling the fluctuations of wind speed data by considering their mean and volatility effects. Renew Sustain Energy Rev 2016;54:777–84.
[9] 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.
[10] Zuluaga CD, Alvarez MA, Giraldo E. Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison. Appl Energy 2015;156:321–30.
[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] Agrawal A, Sandhu KS. Most influential parametrical and data needs for realistic wind speed prediction. Renew Energy 2016;94:452–65.
[13] Agrawal A, Sandhu K. Comparative study of stochastic wind speed prediction models. 2014 IEEE 6th India international conference on power electronics (IICPE). IEEE; 2014. p. 1–6.
[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] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87:2313–20.
[16] 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.
[17] Liu H, Duan Z, Han F-Z, Li Y-F. Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. Energy Convers Manage 2018;156:525–41.
[18] Xiao L, Dong Y, Dong Y. An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. Energy Convers Manage 2018;160:273–88.
[19] 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.
[20] 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.
[21] Qu Z, Zhang K, Mao W, Wang J, Liu C, Zhang W. Research and application of ensemble forecasting based on a novel multi-objective optimization algorithm for wind-speed forecasting. Energy Convers Manage 2017;154:440–54.
[22] Liu H, Mi X, Li Y. An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm. Renew Energy 2018;123:694–705.
[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] Lei C, Ran L. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. 2008 DRPT 2008 third international conference on electric utility deregulation and restructuring and power technologies. IEEE; 2008. p. 2525–9.
[25] Li C, Xiao Z, Xia X, Zou W, Zhang C. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. Appl Energy 2018;215:131–44.
[26] Peng T, Zhou J, Zhang C, Zheng Y. Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. Energy Convers Manage 2017;153:589–602.
[27] 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.

[28] Yu C, Li Y, Zhang M. Comparative study on three new hybrid models using Elman neural network and empirical mode decomposition based technologies improved by singular spectrum analysis for hour-ahead wind speed forecasting. Energy Convers Manage 2017;147:75–85.
[29] Kummer N, Najjaran H. Adaboost.MRT: Boosting regression for multivariate estimation. Artificial Intelligence Res 2014;3.
[30] Zhang K, Luo M. Outlier-robust extreme learning machine for regression problems. Neurocomputing 2015;151:1519–27.
[31] Calderbank A, Daubechies I, Sweldens W, Yeo B-L. Wavelet transforms that map integers to integers. Appl Comput Harmonic Anal 1998;5:332–69.
[32] Mirjalili S, Saremi S, Mirjalili SM, Coelho LDS. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. Expert Syst Appl 2016;47:106–19.
[33] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intelligence 1989;11:674–93.
[34] Mi X-W, Liu H, Li Y-F. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. Energy Convers Manage 2017;151:709–22.
[35] Tascikaraoglu A, Sanandaji BM, Poolla K, Varaiya P. Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform. Appl Energy 2016;165:735–47.
[36] Azimi R, Ghofrani M, Ghayekhloo M. A hybrid wind power forecasting model based on data mining and wavelets analysis. Energy Convers Manage 2016;127:208–25.
[37] Yu C, Li Y, Zhang M. An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on Elman neural network. Energy Convers Manage 2017;148:895–904.
[38] 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.
[39] Shrestha DL, Solomatine DP. Experiments with AdaBoost. RT, an improved boosting scheme for regression. Neural Comput 2006;18:1678–710.
[40] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manage 2018;159:54–64.
[41] Zhao J, Wang Z, Park DS. Online sequential extreme learning machine with forgetting mechanism. Neurocomputing 2012;87:79–89.
[42] Zhao Y-P, Hu Q-K, Xu J-G, Li B, Huang G, Pan Y-T. A robust extreme learning machine for modeling a small-scale turbojet engine. Appl Energy 2018;218:22–35.
[43] Liu H, Wu H, Li Y. Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. Energy Convers Manage 2018;161:266–83.
[44] 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.
[45] Kipouros T, Jaeggi DM, Dawes WN, Parks GT, Savill AM, Clarkson PJ. Biobjective design optimization for axial compressors using tabu search. AIAA J 2008;46:701–11.
[46] Xiao L, Qian F, Shao W. Multi-step wind speed forecasting based on a hybrid forecasting architecture and an improved bat algorithm. Energy Convers Manage 2017;143:410–30.
[47] Du P, Wang J, Guo Z, Yang W. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers Manage 2017;150:90–107.
[48] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Software 2014;69:46–61.
[49] Neville J, Jensen D. Bias/variance analysis for relational domains. International conference on inductive logic programming. Springer; 2007. p. 27–8.
[50] Song J, Wang J, Lu H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. Appl Energy 2018;215:643–58.
[51] Khosravi A, Machado L, Nunes R. Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. Appl Energy 2018;224:550–66.
[52] Riahy G, Abedi M. Short term wind speed forecasting for wind turbine applications using linear prediction method. Renew Energy 2008;33:35–41.