



Wind speed prediction by a swarm intelligence based deep learning model via signal decomposition and parameter optimization using improved chimp optimization algorithm

Leiming Suo^a, Tian Peng^{a,b,*}, Shihao Song^a, Chu Zhang^a, Yuhang Wang^a, Yongyan Fu^a, Muhammad Shahzad Nazir^a

^a Faculty of Automation, Huaiyin Institute of Technology, Huaiyin, 223003, China

^b Jiangsu Permanent Magnet Motor Engineering Research Center, Huaiyin Institute of Technology, Huaiyin, 223003, China

ARTICLE INFO

Handling Editor: Jesse L. The

Keywords:

Wind speed prediction
TVFEMD
Chimp optimization algorithm
BiGRU
Deep learning

ABSTRACT

Accurate prediction of wind speed plays a very important role in the stable operation of wind power plants. In this study, the goal is to establish a hybrid wind speed prediction model based on Time Varying Filtering based Empirical Mode Decomposition (TVFEMD), Fuzzy Entropy (FE), Partial Autocorrelation Function (PACF), improved Chimp Optimization Algorithm (IChOA) and Bi-directional Gated Recurrent Unit (BiGRU). Firstly, the original wind speed data was decomposed by TVFEMD to obtain modal components, and FE aggregation is used to decrease the computational complexity. Secondly, the components are processed by PACF to extract important input features. Thirdly, the BiGRU parameters are optimized using IChOA which is an improved version of ChOA. Finally, the optimized BiGRU is used to predict the decomposed components, and the predicted components are summed to obtain the final prediction result. In this experiment, the proposed model is used to predict the data of four months of a year from Station 46,060 of National Data Buoy Center, and the performance of eight benchmark models is analyzed. Experimental results show that TVFEMD and PACF can improve the prediction accuracy of the model. IChOA is feasible to optimize the parameters of BiGRU and can improve the prediction performance.

1. Introduction

As the economy and society continue to develop, the consumption of fossil energy and the pollution problems caused by it in recent years have grown more serious, and clean energy has gradually received attention from many nations [1]. As a sustainable clean energy, wind energy has developed rapidly. However, due to the intermittent, random and uncontrollable nature of wind energy itself, coupled with the influence of factors such as climate and terrain, how to connect large-scale wind power to the grid has become the key to wind power generation [2, 3]. The essence of wind power generation is to use wind energy to drive wind turbines, thereby converting wind energy into electricity [4]. If the accurate prediction of wind speed is achieved, the carrying capacity of the power grid can be adjusted, the loss of wind turbines and power grids can be reduced, and the Annual Energy Production (AEP) can also be improved [5]. Therefore, it is crucial to establish a stable and reliable

model that can predict wind speed for a long time to assist the smooth operation of wind power grid connection [6].

In order to increase the stability and accuracy of wind speed prediction models, scientists have proposed and improved various prediction models. Among them, prediction models can be divided into traditional physical prediction models, statistical prediction models, and emerging artificial intelligence models [7,8]. In the traditional forecasting model, the physical prediction model is based on various physical parameters of the prediction location (such as terrain, climate, season, etc.) to establish a wind speed prediction model. Numerical weather prediction (NWP) based on physical models, dating back to 1920, is represented by the High Resolution Integrated Predictive System (IFS) model from the European Centre for Medium-Term Weather Prediction (ECMWF) [9]. This type of model substitutes the collected meteorological parameters into the NWP, which needs to consider the geographical boundaries, climate, wind turbine operating conditions and other conditions, and the calculation volume is large and the process

* Corresponding author. Faculty of Automation, Huaiyin Institute of Technology, Huaiyin, 223003, China.

E-mail addresses: suoleiming@foxmail.com (L. Suo), hushydropt@126.com (T. Peng), ssh2413359338@163.com (S. Song), zhangchuhust@foxmail.com (C. Zhang), violin_5805@163.com (Y. Wang), 17696933553@163.com (Y. Fu), msn_bhutta88@yahoo.com (M.S. Nazir).

List of abbreviations	
AEP	Annual Energy Production
ANN	Artificial Neural Networks
AR	Autoregressive
ARMA	Autoregressive Moving Average
BP	Back Propagation
BiGRU	Bi-directional Gated Recurrent Unit
BiLSTM	Bi-directional Long Short-Term Memory
CF	capacity factor
ChOA	Chimp Optimization Algorithm
CNN	Convolutional Neural Network
R	Correlation coefficient
ECMWF	European Centre for Medium-Term Weather Prediction
ELM	Extreme Learning Machine
FE	Fuzzy Entropy
GRU	Gated Recurrent Unit
HT	Hilbert Transform
IChOA	Improved Chimp Optimization Algorithm
IFS	Integrated Predictive System
LSSVM	Least Squares Support Vector Machine
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOALO	Multi-Objective Ant Lion Algorithm
NDBC	National Data Buoy Center
NWP	Numerical Weather Prediction
PACF	Partial Autocorrelation Function
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
SVM	Support Vector Machine
TVFEMD	Time Varying Filtering based Empirical Mode Decomposition
TSDC	Time-Shifted Data Correction
VMD	Variational Mode Decomposition
WOA	Whale Optimization Algorithm
WPD	Wind power density
WS	Wind Speed

is complex, so it is difficult to achieve efficient and accurate wind speed prediction [10].

With the development of statistical prediction models, the limitations of physical prediction models are effectively solved. Statistical prediction models only use mathematical statistical methods to establish a specific mapping management of wind speed, without considering physical factors such as climate, seasons and topography, so they are called black box models or data-driven models [11]. Statistical prediction models include the persistence method, the Kalman filter method, and the Autoregressive Moving Average model (ARMA) et. The persistence method refers to the real value of the preceding moment as the forecast value of the next moment, which is very suitable for short-term prediction, but it is difficult to predict the wind speed in the medium and long term. Kalman filter method refers to the use of linear equations to build a wind speed prediction model, which is simple and feasible, but can only handle linear processes that conform to gaussian distribution, and cannot cope with complex and variable wind speed environments [12]. ARMA is a time series forecasting method that uses linear Autoregressive (AR) to process and predict future values of time series [13]. The model structure of this method is simple and the calculation efficiency is high, but it is difficult to cope with the prediction of gusts and sudden winds [14].

Compared with statistical models, machine learning models exhibit stronger nonlinear processing capabilities in the field of wind speed prediction [15,16]. Machine learning models such as Artificial Neural Networks (ANN), Back Propagation (BP) neural network, Support Vector Machine (SVM), Least Squares Support Vector Machine (LSSVM) and other machine learning models have achieved relatively good results in the field of prediction. For example, Wang et al. [17] used a combination of Extreme Learning Machine (ELM) and AdaBoost algorithms to measure from nearby locations and build a short-term wind speed prediction model. Experiments showed that the wind speed prediction model they built achieved better results than ARMA. Li and Jin [18] used Variational Mode Decomposition (VMD) to decompose the original wind speed data, then used LSSVM to build a wind speed prediction model, and used Multi-Objective Ant Lion Algorithm (MOALO) to optimize LSSVM, which also achieved good results. Filik et al. [19] take wind speed, temperature, and air pressure at station as multi-dimensional inputs to ANN, and a very sensitive ultra-short-term wind speed prediction model is established. Bashar et al. [20] combined feed-forward Back-propagation with ANN to train a large number of data points to form a wind speed prediction model with high accuracy.

Although machine learning models run fast and are easy to operate,

and can solve the problem of weak nonlinear processing capabilities of traditional models, but machine learning models also have trouble in dealing with high fluctuations in wind speed data is prone to overfitting. At the same time, with the continuous upgrading of computer computing power, scientists have turned to deep learning models with larger, more accurate and robust calculations.

In recent years, deep learning models have been continuously developed and have been widely used by experts and scholars in various fields [21]. The more popular deep learning models are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) neural network etc. Experts and scholars continue to develop new deep learning models, and develop models such as Bi-directional Gated Recurrent Unit (BiGRU) and Bi-directional Long Short-Term Memory (BiLSTM) on the basis of several models. Shen et al. [22] used a combination of LSTM and CNN to predict wind speed to ensure the safety of unmanned sailing ships. Lim et al. [23] used Time-Shifted Data Correction (TSDC) and LSTM to build hybrid models for short-term wind speed predictions during typhoons. Zhu et al. [24] used the VMD-BiGRU model to predict the rubber futures time, after research comparison, it was found that BiGRU, a bidirectional neural network, has significant advantages in fitting performance and predicting trends.

In the process of studying the predictive model, experts have found that the accuracy of the prediction model can be greatly improved by the decomposition-prediction-aggregation framework. Yan et al. [25] processed the original wind speed data by EEMD, and then predicted the decomposed data through the LSTM model, which greatly improved the model prediction accuracy. Yang et al. [26] used improved singular spectrum to decompose data, and experimental results show that this method can effectively improve the accuracy of artificial neural network models. Xiong et al. [27] proposed a TVFEMD-PSR-NNCT wind speed prediction model, and experimental results show that the stability of wind speed data can be improved by Time Varying Filtering based Empirical Mode Decomposition (TVFEMD) treatment. In addition, they used whale optimization algorithms to optimize weight coefficients and improve prediction accuracy.

However, in the face of the increasingly uncertain variables of machine learning models and deep learning models, such as coefficients, ranges and weights, intelligent optimization algorithms need to be introduced instead of traditional manual way to adjust parameters. Ding et al. [28] used WOA to improve the thresholds and weights of KELM to improve the accuracy of KELM. Hua et al. [29] used IASO to optimize ELM when predicting short-term wind speed, and compared it with

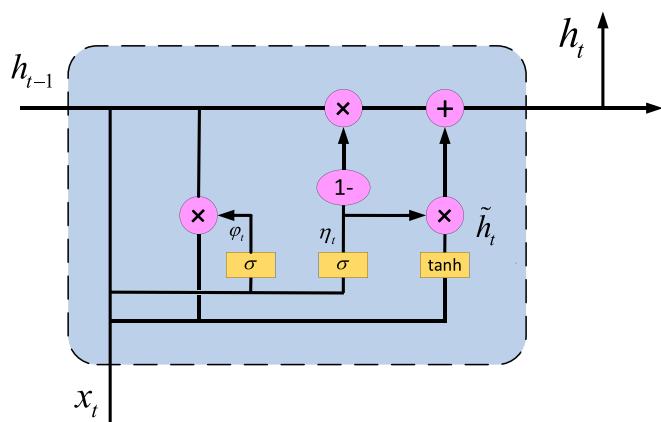


Fig. 1. The structure of GRU.

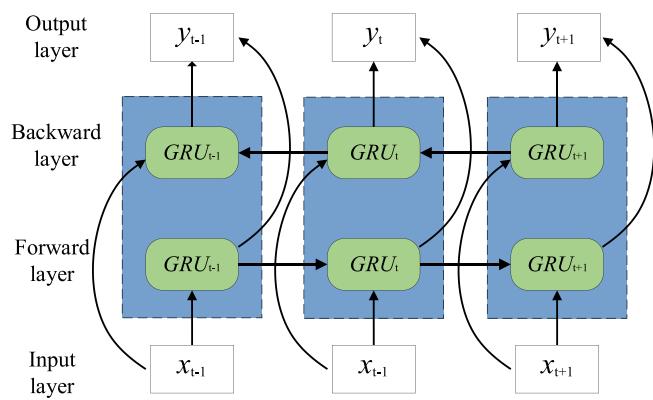


Fig. 2. The structure of BiGRU.

other models to reflect the performance of the optimization algorithm. Bo et al. [30] compared the optimization effect of Chimp optimization Algorithm (ChOA) and other intelligent optimization algorithms in optimizing model parameters, and proposed an efficient photovoltaic power generation prediction model.

In addition, the impact of wind energy potential, wind power density (WPD) and wind turbine capacity factor on wind power systems should also be taken into account. Ulatia et al. [31] used the air density in different seasons (mainly in the middle and high latitudes) will have a great impact on the stable operation of the wind power system. Carreno-Madinabeitia et al. [32] added temporal and spatial changes in wind density (WPD) and capacity factor (CF) in the wind power system, which will increase the revenue of offshore wind turbines by about 20% over 11 years. Therefore, in order to balance the many influencing factors of wind power system, this paper will choose the wind speed information of different seasons for wind speed prediction.

Inspired by the above research content, this study proposes a novel hybrid wind speed prediction model based on TVFEMD, Partial Auto-correlation Function (PACF), ChOA and BiGRU to address the shortcomings of traditional wind speed prediction. In order to improve the prediction efficiency of the overall model, this study uses the decomposition-prediction-aggregation structure to construct the wind speed prediction model. At the decomposition level, the original wind speed data is processed using TVFEMD. This method inherits the advantages of high frequency separation performance and the stability under low sampling rates, and this method has strong anti-noise interference and robustness to wind speed data processing. At the level of feature extraction, PACF is used to extract the important components in the signals to reduce the data dimension, and improve the operation efficiency of the overall model. To make full use of the temporal characteristics of the wind speed data, the BiGRU model is used to predict the wind speed data. This model takes sufficient of the past and future information to improve the wind speed prediction and performed well on time series prediction. To address the influence of subjective factors when selecting the model parameters, the ChOA algorithm is used to

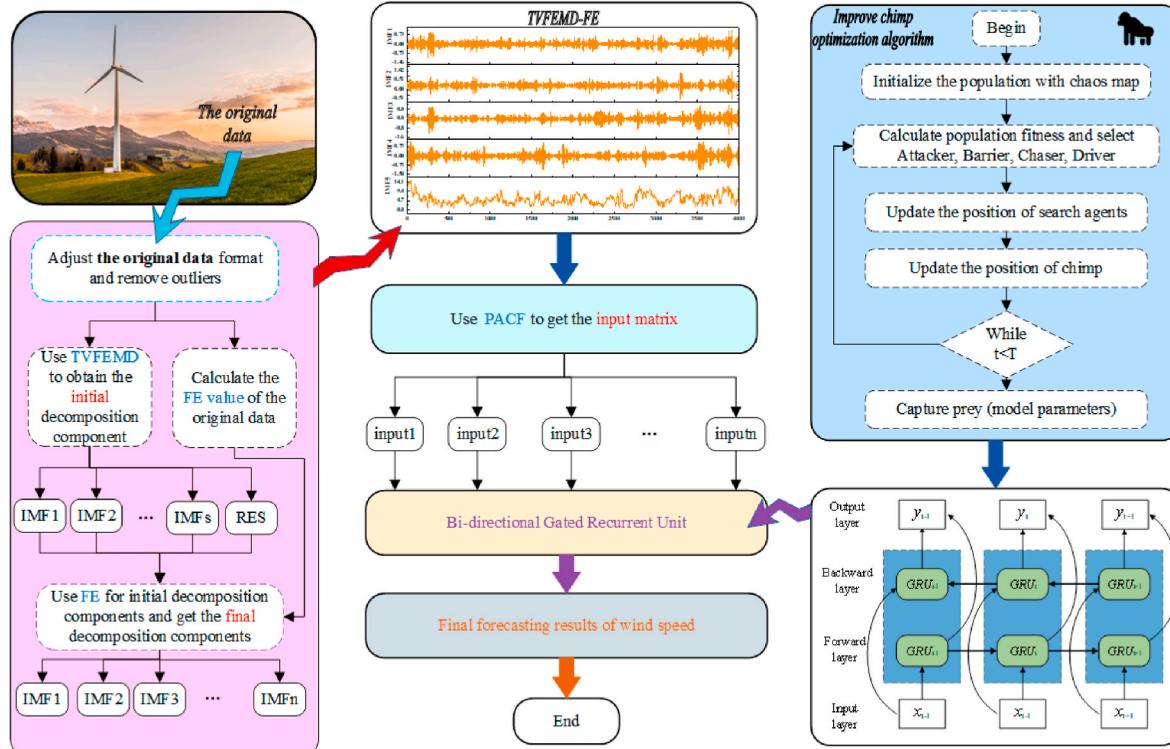


Fig. 3. The structure of TVFEMD-PACF-ICHOA-BiGRU model.

Table 1

Basic information of the four monthly wind speed data sets.

Months	Data Set	Data length	Max	Min	Mean	Kurtosis	Std-Dev	Skewness	I_T
Jan.	All(m/s)	4000	15	0.3	5.3456	3.6474	2.5142	0.7712	0.4702
	Training(m/s)	2800	15	0.3	4.9124	5.2049	2.3926	1.1481	0.4869
	Testing(m/s)	1200	12.9	0.7	6.3565	2.3007	2.5017	0.1098	0.3934
Apr.	All(m/s)	4000	15.5	0.1	3.8113	4.2296	2.9263	1.3008	0.7677
	Training(m/s)	2800	15.5	0.1	4.0555	3.4286	3.2606	1.1301	0.8038
	Testing(m/s)	1200	9.6	0.1	3.2417	3.3693	1.8105	0.8162	0.5583
Jul.	All(m/s)	4000	12.6	0.1	4.1095	3.4348	2.2269	0.8097	0.5418
	Training(m/s)	2800	10.8	0.1	4.0812	3.0865	1.9436	0.5855	0.4761
	Testing(m/s)	1200	12.6	0.1	4.1753	2.9385	2.7777	0.8873	0.6649
Oct.	All(m/s)	4000	19.8	0.1	5.8603	3.4491	3.6729	0.7735	0.6266
	Training(m/s)	2800	19.8	0.1	6.2004	3.0775	3.7924	0.6547	0.6115
	Testing(m/s)	1200	18.8	0.1	5.0669	4.9185	3.2434	1.0456	0.6398

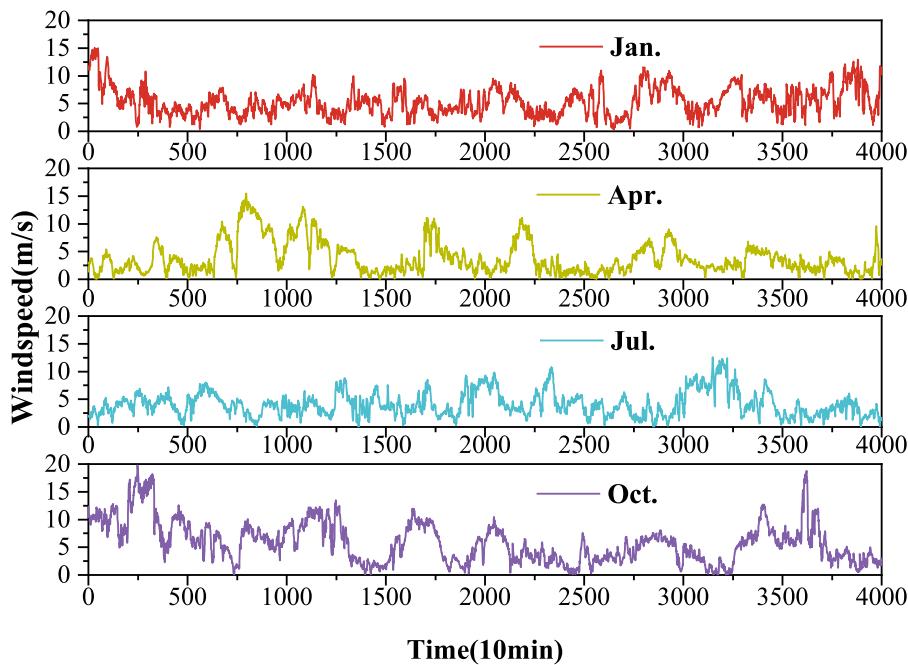


Fig. 4. Wind speed data of the four months.

optimize the BiGRU model parameters. Specific steps of the model are as follows: Firstly, the components are obtained by using TVFEMD to decompose original wind speed data, and then the components are reconstructed by FE (Fuzzy Entropy). Secondly, each decomposed component is transferred into a multidimensional input matrix, and then the important components of the multidimensional matrix are extracted by PACF. Thirdly, the improved Chimp optimization Algorithm (IChOA) is combined with BiGRU to get an IChOA-BiGRU model for time series forecasting. Finally, the predictions for each component are summed to give the ultimate wind speed value.

The proposed novel hybrid model takes full advantage of the advantages of each approach and overcomes their limitations. The innovations and contributions of the proposed model are as follows:

- (1) Based on the advantages of mixed models in prediction, this paper cleverly combines deep learning models and auxiliary methods, such as signal decomposition, feature selection and evolutionary algorithms. Through these methods, a new hybrid wind speed prediction model (namely, TVFEMD-PACF-IChOA-BiGRU) is constructed for wind speed prediction.
- (2) In order to reduce the volatility and complexity of the original wind speed signal, TVFEMD is used to decompose the original wind speed data. At the same time, in order to solve the situation

that the computation burden is too heavy due to excessive decomposition during the decomposition process, the FE value of the wind speed signal is calculated and used to aggregate the decomposed signals to improve the operation efficiency of the model.

- (3) Each one-dimensional IMF is turned into a useable multidimensional matrix for deep learning modelling. However, too high a dimension will reduce the efficiency of model operation and prediction accuracy, and it is necessary to filter out most of the invalid data in the matrix. The important components in multidimensional data are extracted through PACF to improve information utilization.
- (4) The core of the model proposed in this paper is a deep learning model optimized by evolutionary algorithms, namely IChOA-BiGRU. This paper cleverly exploits the advantages of deep learning models in prediction and adopts two improvement strategies to enhance ChOA to optimize the model parameters of BiGRU. In order to improve the optimization efficiency of ChOA, a new population iteration method based on ranking mechanism is proposed and integrated with ChOA. In addition, chaos initialization is adopted for population initiation of IChOA.
- (5) Real wind speed data from four representative months are used to compare the effectiveness of the proposed model. Four error

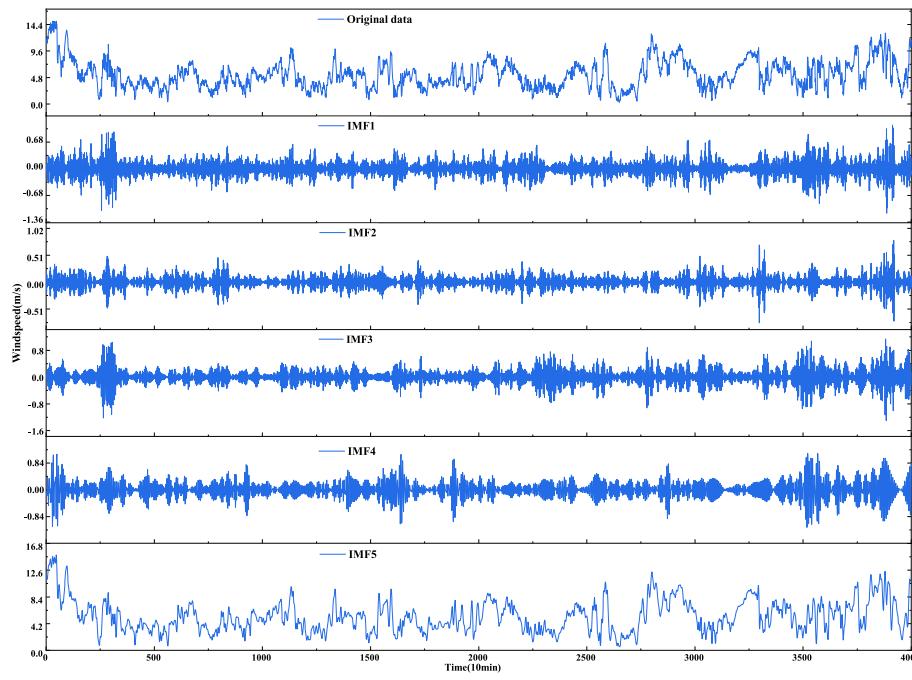


Fig. 5. Decomposition diagram of wind speed data in January.

Table 2
Comparison of wind speed prediction indexes in January.

Models	RMSE	MAE	MAPE	R
BP	0.86076	0.65834	0.13930	0.87933
GRU	0.87801	0.63972	0.14907	0.87444
BiGRU	0.80694	0.56030	0.11679	0.89395
ChOA-BiGRU	0.75725	0.53731	0.11558	0.90661
EMD-ChOA-BiGRU	0.464	0.34886	0.06820	0.96493
EMD-PACF-BiGRU	0.44222	0.33527	0.06910	0.96815
EMD-PACF-ChOA-BiGRU	0.42978	0.31589	0.06324	0.96992
TVFEMD-PACF-ChOA-BiGRU	0.25972	0.20619	0.04511	0.98901
TVFEMD-PACF-IChOA-BiGRU	0.2452	0.19095	0.04485	0.99021

indicators and percentage improvement indicators are provided for full verification.

In this paper, Chapter 2 introduces the research background of TVFEMD, ChOA, and BiGRU, and how to improve ChOA and combine it with the BiGRU model; Chapter 3 introduces and analyzes the processing of data; Chapter 4 analyzes the prediction results of the main models; Chapter 5 summarizes the results of the experiment.

2. Methods

2.1. Time Varying Filtering based Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is a novel time-domain signal processing method based on the Hilbert transform, which has a good signal-to-noise ratio and time-frequency focus [33]. EMD is used for smoothing out timing signals and is essentially an adaptive data processing method that can theoretically be applied to the decomposition of any type of timing signal. When processing a timing signal, EMD first decomposes the original signal data into a series of IMFs and residuals, and then performs Hilbert transform (HT) on each IMF component.

Due to the adaptability of EMD, it is not necessary to choose the decomposition function by itself.

However, EMD has defects such as mode aliasing and endpoint effect when decomposing timing signals, TVFEMD proposed which improved

by EMD to eliminate this problem [34]. The TVFEMD screening process uses non-uniform B-spline approximations as time-varying filters. The local end-of-time frequency of TVFEMD is adaptively designed according to the homeopathic amplitude and frequency information, and the cut-off frequency rearrangement method is introduced to solve the intermittent problem. The TVFEMD method's computation procedure is as follows:

1. Perform HT on the original wind speed data $\sigma(k)$, and the result is $R(k)$.

$$AF(k) = \sqrt{\sigma(k)^2 + R(k)^2} \quad (1)$$

$$PF(k) = \arctan[\sigma(k) / R(k)] \quad (2)$$

where $AF(k)$ is the instantaneous amplitude, $PF(k)$ is the instantaneous phase.

2. $AF(k_{\max})$ and $AF(k_{\min})$ is the local upper limit and lower limit of the $AF(k)$. Then the upper limit point set and lower limit point set are interpolated to obtain curves $\tau_1(k)$ and $\tau_2(k)$. $\omega_1(k)$ and $\omega_2(k)$ are calculated by Eq. (3) and Eq. (4).

$$\omega_1(k) = \frac{\tau_1(k) + \tau_2(k)}{2} \quad (3)$$

$$\omega_2(k) = \frac{\tau_1(k) - \tau_2(k)}{2} \quad (4)$$

3. Interpolate $AF^2(k_{\max})\mu'(k_{\max})$ and $AF^2(k_{\min})\mu'(k_{\min})$, respectively to obtain $\gamma_1(k)$, $\gamma_2(k)$. The instantaneous frequency component $\mu'_1(t)$, $\mu'_2(t)$ are calculated by Eq. (5) and Eq. (6).

$$\mu'_1(k) = \frac{\gamma_1(k)}{2\omega_1^2(k) - 2\omega_1(k)\omega_2(k)} + \frac{\gamma_2(k)}{2\omega_1^2(k) + 2\omega_1(k)\omega_2(k)} \quad (5)$$

$$\mu'_2(k) = \frac{\gamma_1(k)}{2\omega_2^2(k) - 2\omega_1(k)\omega_2(k)} + \frac{\gamma_2(k)}{2\omega_2^2(k) + 2\omega_1(k)\omega_2(k)} \quad (6)$$

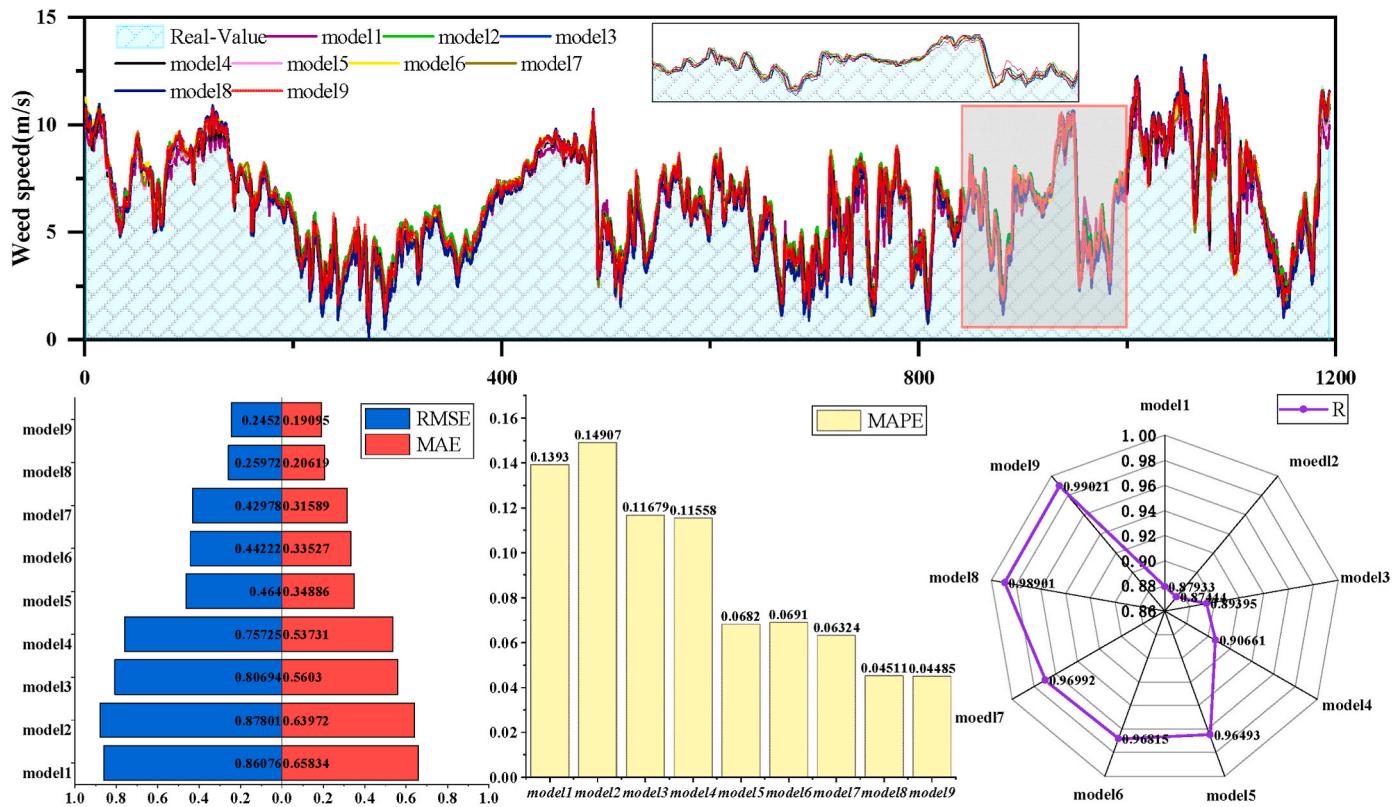


Fig. 6. Error analysis of January dataset.

4. The local cut-off frequency $\mu_{bis}'(k)$ is calculated according Eq. (7):

$$\mu_{bis}'(k) = \frac{\mu_1'(k) + \mu_2'(k)}{2} \quad (7)$$

5. Calculate $\varphi(k) = \cos[\int \mu_{bis}'(k)d(k)]$, the extreme point of $\varphi(k)$ is used as the node, and B-spline interpolation is used to further filter $\tau(k)$, the approximation result is $\alpha(k)$.

6. When $\rho(k) \leq r$, $\tau(k)$ is IMF. Otherwise, $R(k) = R(k) - \alpha(k)$, and repeat Steps 1-6. $\rho(k)$ is calculated according Eq. (8):

$$\rho(k) = \frac{B_L(k)}{\mu_{\alpha(k)}} \quad (8)$$

where $B_L(t)$ represent the instantaneous bandwidth, $\mu_{\alpha(k)}$ represent the weighted average about the instantaneous frequency, $r = 0.1$.

2.2. Subseries aggregation based on Fuzzy Entropy for TVFEMD

The concept of FE is based on sample entropy, and the fuzzy index characterizes the similarity of the two signals and has a strong ability to resist noise interference. Entropy is a measure of the probability of generating a new pattern when the complexity and dimensions of a time series change. FE uses the membership function and mean algorithm as the measurement criterion, takes the fault feature vector as the sample information, and effectively classifies the fuzzy similarity measure. This way can solve the problem that the fault characteristics of other types of entropy division are too single on the time series scale, and are relatively stable. FE can effectively solve the problem of aliasing in the signal decomposition process, and at the same time, the stability of the frequency component and the magnitude of the mutual dependence in the

time series can be controlled by controlling the size of the FE value. The FE formula is shown Eq. (9).

$$FE(m, s, L) = \ln \psi^m(s) - \ln \psi^{m+1}(s) \quad (9)$$

where m is the reconstruction dimension, s is the similarity tolerance, L is the data length of wind speed.

2.3. Improved chimp optimization algorithm

2.3.1. Chimp optimization algorithm

The ChOA is a new inspired optimization algorithm proposed by Khishe and Mosavi [35] in 2020. This algorithm is similar to traditional animal algorithms such as PSO and Whale Optimization Algorithm (WOA). It simulates the advantages of chimp's high efficiency and speed in the hunting process to find the optimal solution. At the same time, the ChOA adopts a unique population classification mode, divides the population into 4 types, and uses different strategies for the optimization process, which greatly increases the precision and accuracy of the algorithm.

According to the division of labor of chimp in the group, the population is divided into 4 categories: ordinary members responsible for driving (Driver) and blocking (Barrier) prey; young adult chimp responsible for the main process of chasing (Chaser); dominant (Attacker) and preferential enjoyment leader of the prey. Chimps have the ability to think independently in the group, and under certain circumstances appear chaotic hunting behavior. The specific process of the algorithm is as follows :

- (1) Division of labor. According to different division of labor in hunting, chimps are divided into 4 types according to their functions. In the algorithm, it is necessary to spatially divide the chimp hunters into four types, and the leader of them will lead the chimp group to hunt.

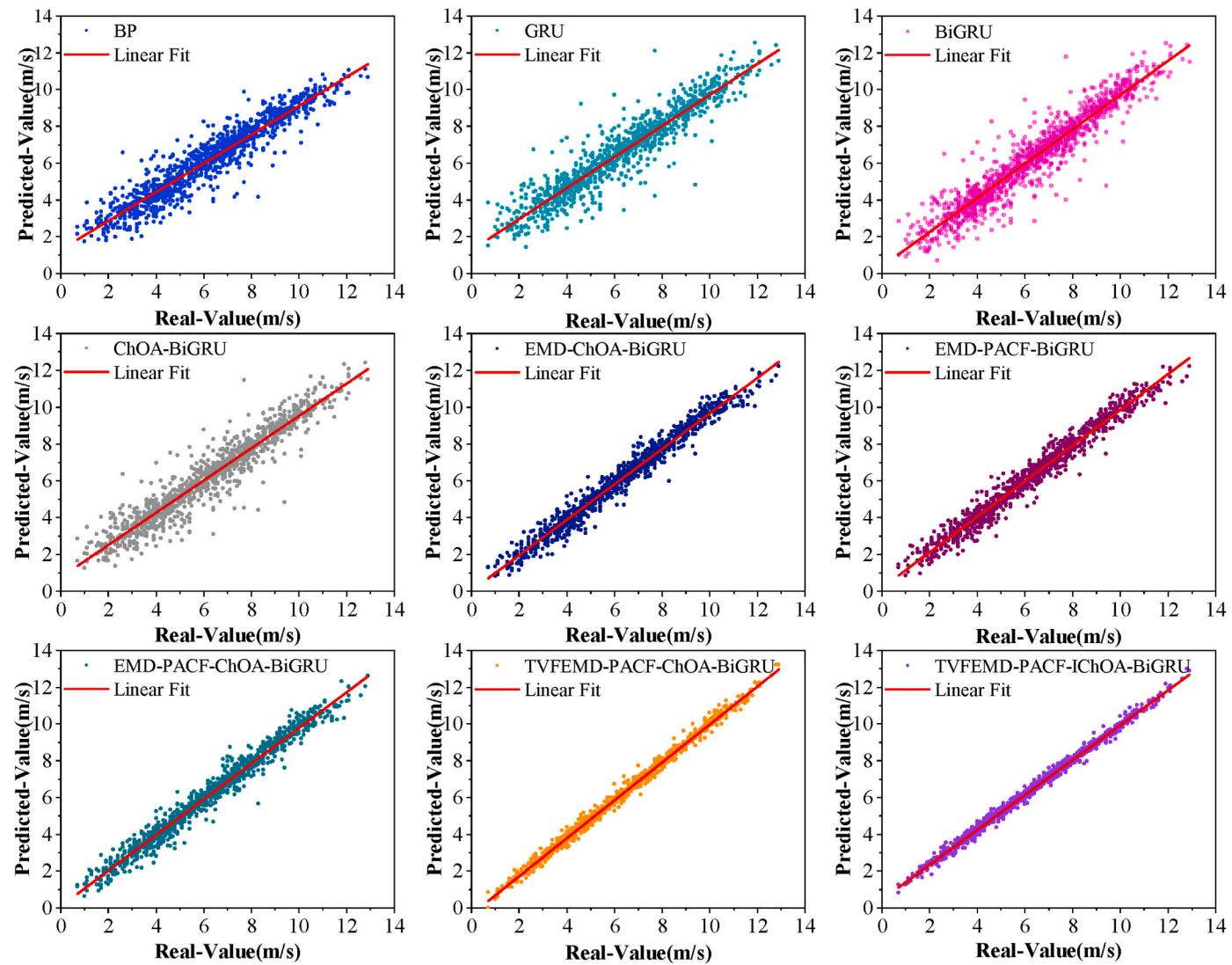


Fig. 7. Scatter plots of January dataset.

Table 3
Comparison of wind speed prediction indexes in April.

Models	RMSE	MAE	MAPE	R
BP	0.48545	0.36415	0.24914	0.92800
GRU	0.49922	0.36751	0.24054	0.92386
BiGRU	0.46187	0.36414	0.26964	0.93482
ChOA-BiGRU	0.40913	0.30342	0.20877	0.94886
EMD-ChOA-BiGRU	0.26617	0.20594	0.09737	0.97835
EMD-PACF-BiGRU	0.23663	0.17980	0.10089	0.98289
EMD-PACF-ChOA-BiGRU	0.22177	0.16398	0.09180	0.98497
TVFEMD-PACF-ChOA-BiGRU	0.14277	0.11657	0.07966	0.99377
TVFEMD-PACF-IChOA-BiGRU	0.10312	0.07842	0.04307	0.99675

(2) Drive and chase phases. In the process of chimp hunting, the chimps need to judge the direction and distance of the next step according to the distance between themselves and their prey.

$$d = |cx_{\text{prey}}(t) - mx_{\text{chimp}}(t)| \quad (10)$$

$$x_{\text{chimp}}(t+1) = x_{\text{prey}}(t) - ad \quad (11)$$

where Eqs. (10) and (11) are the position analysis formulas of each chimp. Among them, d is the distance between the prey and the chimp; t is the current number of iterations; $x_{\text{prey}}(t)$ is the position of the prey

Table 4
Comparison of wind speed prediction indexes in July.

Models	RMSE	MAE	MAPE	R
BP	0.62376	0.46144	0.24717	0.94966
GRU	0.65168	0.49452	0.25072	0.94505
BiGRU	0.58459	0.41568	0.18919	0.95578
ChOA-BiGRU	0.554628	0.38931	0.15434	0.96020
EMD-ChOA-BiGRU	0.44293	0.34363	0.13389	0.97461
EMD-PACF-BiGRU	0.38712	0.28661	0.11116	0.98061
EMD-PACF-ChOA-BiGRU	0.36001	0.27796	0.10958	0.98323
TVFEMD-PACF-ChOA-BiGRU	0.19451	0.16037	0.06429	0.99510
TVFEMD-PACF-IChOA-BiGRU	0.1817	0.14446	0.08163	0.99572

currently; $x_{\text{chimp}}(t)$ is the current position of the chimp; a , m and c are the coefficient vectors. Among them, $a = 2fr_1 - f$, $c = 2r_2$, $m = \text{Chaotic_value}$ (chaotic vector based on chaotic map); f is a nonlinear reduction process from 2 to 0, which is affected by the number of iterations; r_1, r_2 are random vectors in the range of [0,1]. At the same time, since each chimp rounds up differently, its position update requires iterations using different formulas.

(3) Encirclement and capture process. Each chimp determines the process of hunting independently according to its division of labor,

Table 5
Comparison of wind speed prediction indexes in October.

Models	RMSE	MAE	MAPE	R
BP	0.69788	0.52058	0.34240	0.95387
GRU	0.73741	0.50124	0.15339	0.94849
BiGRU	0.66567	0.48904	0.14835	0.95803
ChOA-BiGRU	0.60813	0.44754	0.26297	0.96497
EMD-ChOA-BiGRU	0.50872	0.36334	0.13824	0.97548
EMD-PACF-BiGRU	0.45669	0.34312	0.11250	0.98024
EMD-PACF-ChOA-BiGRU	0.40391	0.30097	0.11005	0.98454
TVFEMD-PACF-ChOA-BiGRU	0.31852	0.26389	0.18578	0.99039
TVFEMD-PACF-IChOA-BiGRU	0.27542	0.23067	0.13361	0.99281

that is, the position vector between each chimp and its prey. After the four kinds of chimps determine their position vector with the prey, each chimp updates its position according to the best chimp position, and the position of the prey is estimated by the best chimp individual position, which is shown Eq. (12), Eq. (13) and Eq. (14).

$$\left\{ \begin{array}{l} d_{Attac\ ker} = |c_1 x_{Attac\ ker} - m_1 x| \\ d_{Barrier} = |c_2 x_{Barrier} - m_2 x| \\ d_{Chaser} = |c_3 x_{Chaser} - m_3 x| \\ d_{Driver} = |c_4 x_{Driver} - m_4 x| \end{array} \right. \quad (12)$$

$$\left\{ \begin{array}{l} x_1 = x_{Attac\ ker} - a_1 d_{Attac\ ker} \\ x_2 = x_{Barrier} - a_2 d_{Barrier} \\ x_3 = x_{Chaser} - a_3 d_{Chaser} \\ x_4 = x_{Driver} - a_4 d_{Driver} \end{array} \right. \quad (13)$$

$$x(t+1) = \frac{x_1 + x_2 + x_3 + x_4}{4} \quad (14)$$

where $d_{Attac\ ker}$, $d_{Barrier}$, d_{Chaser} and d_{Driver} re the distances from the prey at the current stage of attacking chimp, blocking chimp, chasing chimp, and driving chimp, respectively; $x_{Attac\ ker}$, $x_{Barrier}$, x_{Chaser} and x_{Driver} are the position vectors of attacking chimp, blocking chimp, chasing chimp,

and driving chimp relative to their prey. $a_1 \sim a_4$, $c_1 \sim c_4$, $m_1 \sim m_4$ are the vector coefficients of the four chimps, respectively. Four chimps work together to decide the next best location $x(t+1)$.

(4) Chaotic hunting behavior. Driven by instinct and other factors, chimps may also break out of the current territory to hunt, which avoid the ChOA fall into local optimum and slow convergence when solving high-dimensional problems. During the optimization process, the chimp normal position update or the position update through the chaotic model is selected with a probability of 50%, which is shown Eq. (15).

$$x_{chimp}(t+1) = \begin{cases} x_{prey}(t) - ad & (\mu \leq 0.5) \\ Chaotic_value & (\mu > 0.5) \end{cases} \quad (15)$$

where μ is a random number in the range of $[0, 1]$.

2.3.2. Improvements of ChOA

Heuristic algorithms such as PSO and WOA generally have the problems of easily falling into local optimum and low convergence speed, and the ChOA is also unavoidable. For this reason, a chaotic initialization step is introduced in the initialization process of ChOA to improve the initial optimization efficiency of the algorithm; and the iterative update method of the chimp algorithm is optimized to improve the convergence accuracy and speed of the algorithm.

2.3.2.1. Chaos initialization. The introduction of chaotic initialization in the process of chimp population initialization can make the four chimp populations conduct a more extensive preliminary search and improve the search efficiency of the algorithm. Using the chaotic initialization strategy at the beginning of each iteration also helps the algorithm to escape the dilemma of falling into the local optimum. In the original chimp algorithm, initialization is randomly generated based on the dimensions and number of input parameters. The formula is:

$$Position = rand * (ub - lb) + lb \quad (16)$$

where, the *Position* is generated by the original ChOA, *rand* is a random

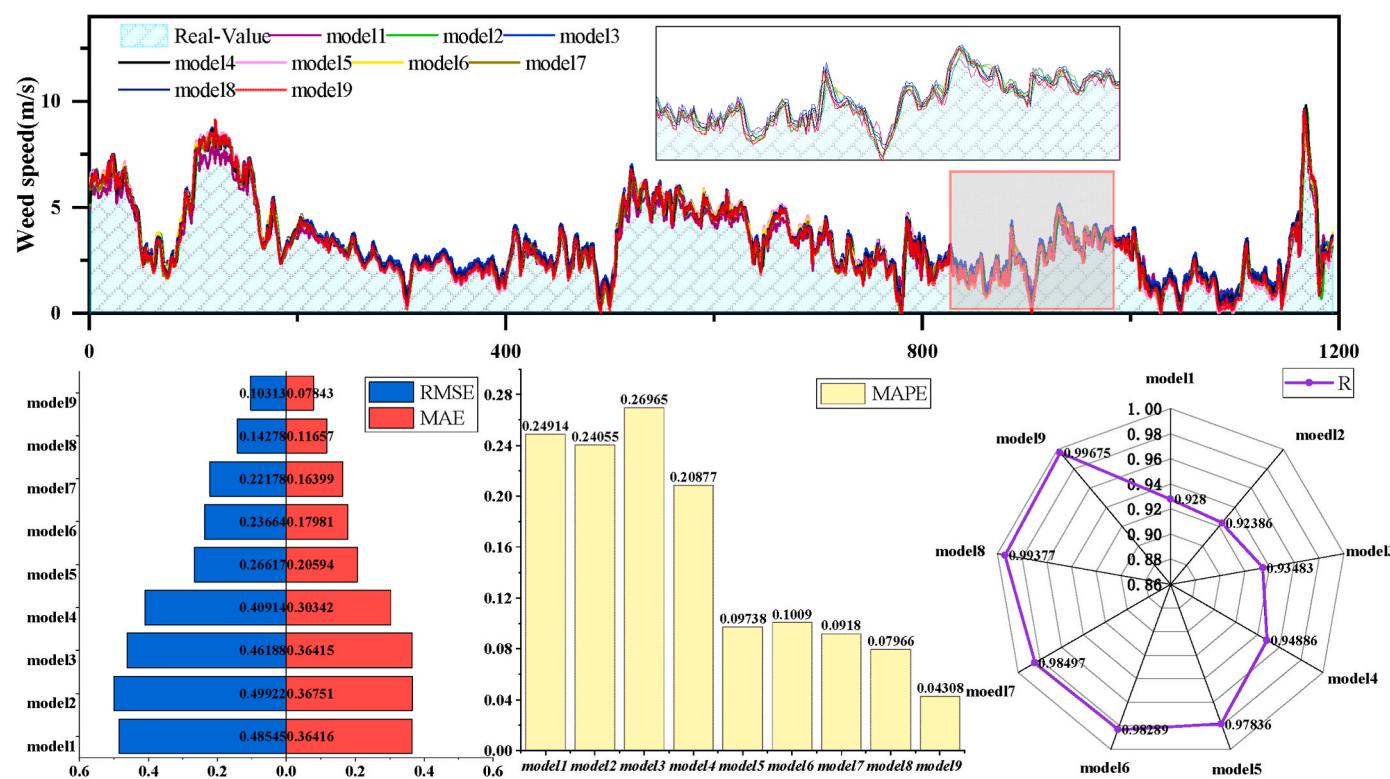


Fig. 8. Error analysis of April dataset.

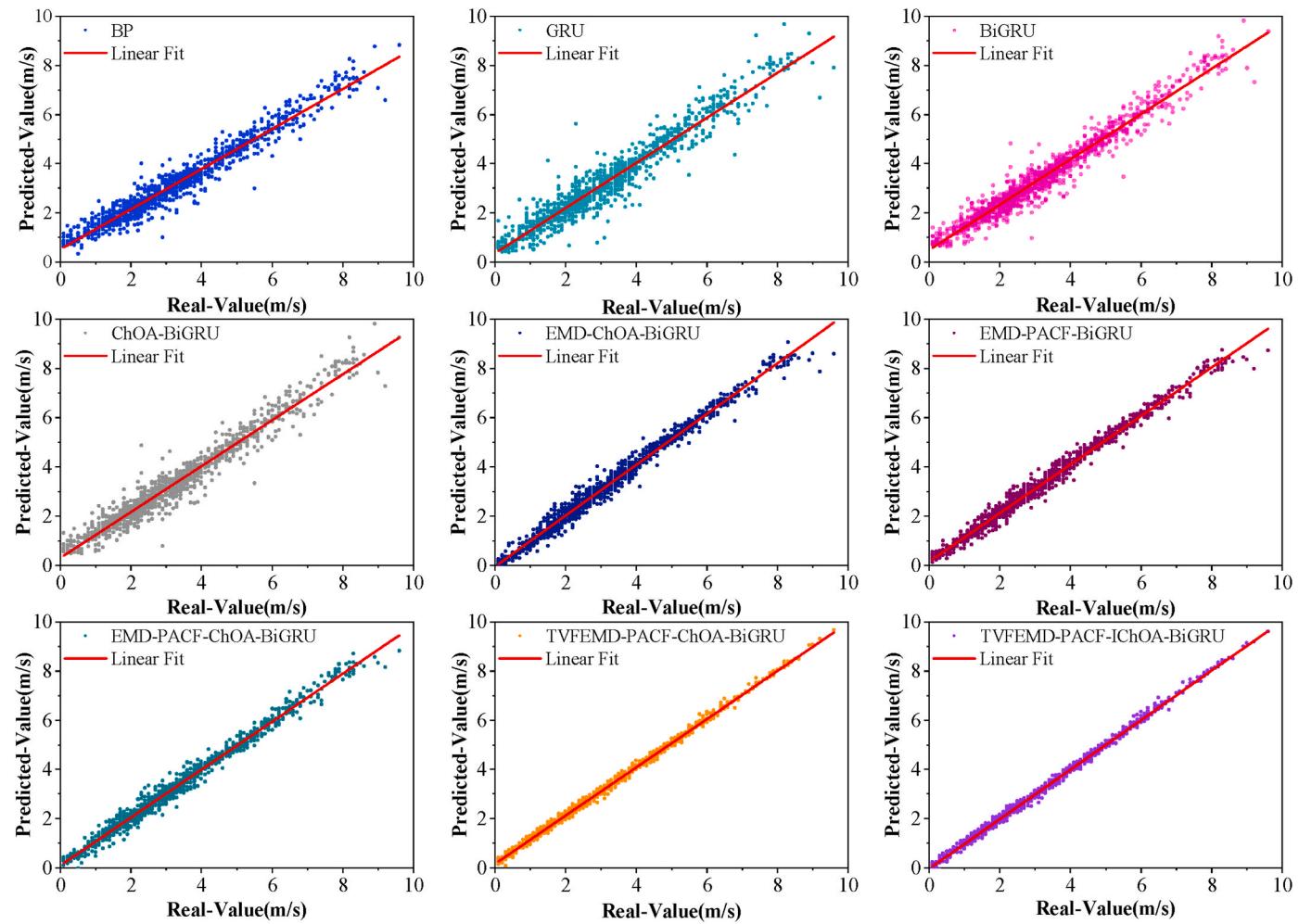


Fig. 9. Scatter plots of April dataset.

between 0 and 1, ub is the upper limit of the chimp population position, and lb is the lower limit of the chimp population position.

Chaos initialization includes Chebyshev, Gauss, Logistic, Circle, etc. When optimizing problems of different dimensions, different types of chaotic maps can be used to achieve the best optimization process. Therefore, using the chaotic mapping principle to initialize the population can make the initial population have the characteristics of diversity, and establish a foundation for further effective global search. This initialization process adopts a typical chaotic mapping system, namely, Logistic mapping, and its expression is:

$$\begin{cases} k(n+1) = \mu k(n)(1 - k(n)), 0 < k(n) < 1 \\ 0 < \mu \leq 4, i = 0, 1, 2, \dots \end{cases} \quad (17)$$

where μ is the control variable.

In the Logistic mapping, the initial individual is completely in a chaotic state, and then all individuals of the chimp are generated by a step-by-step mapping method. The specific steps are as follows:

- (1) Generate a completely chaotic D-dimensional vector according to the data dimension D, and use it as the first individual of a chimp population. The 4 populations are independent of each other, which are recorded as $K_1^{(1)} = (k_1, k_2, \dots, k_d), K_1^{(2)} = (k_1, k_2, \dots, k_d), K_1^{(3)} = (k_1, k_2, \dots, k_d), K_1^{(4)} = (k_1, k_2, \dots, k_d), k_i \in (0, 1), 1 < i < d$.
- (2) According to the number of individuals N of the chimp population, iterate separately in 4 populations, and each population iterates $N - 1$ times. That is $K_i^{(1)}, K_i^{(2)}, K_i^{(3)}, K_i^{(4)}$.

- (3) After completing the initialization iteration process of all chimp populations, map to the solution space according to the following formula:

$$P_{id} = ld + (1 + k_{id})(ud - ld) / 2 \quad (18)$$

where P_{id} is an individual of a chimp population, and at this step, the entire process of chimp population initialization is completed.

2.3.2.2. Ranking mechanism optimization. The chimp optimization algorithm adopts a new method of position update, that is, only the corresponding optimal position is updated during the hunting process of the chimp group. Although the hunting method simplifies the optimization process and improves the efficiency, it will cause the algorithm to converge too slowly, and it is easier to fall into the local optimum. It iterates as follows:

$$\begin{cases} Attac\ ker = x(t+1), fit > fit_{Attac\ ker} \\ Barrier = x(t+1), fit_{Attac\ ker} > fit > fit_{Barrier} \\ Chaser = x(t+1), fit_{Barrier} > fit > fit_{Chaser} \\ Driver = x(t+1), fit_{Chaser} > fit > fit_{Driver} \end{cases} \quad (19)$$

where $Attac\ ker$, $Barrier$, $Chaser$ and $Driver$ represent the best positions corresponding to the 4 populations in the previous iteration, respectively, $fit_{Attac\ ker}$, $fit_{Barrier}$, fit_{Chaser} and fit_{Driver} represent the fitness corresponding to the 4 population individuals calculated in the previous iteration; $x(t+1)$ is the position of an individual in the current iteration, fit is the fitness of $x(t+1)$.

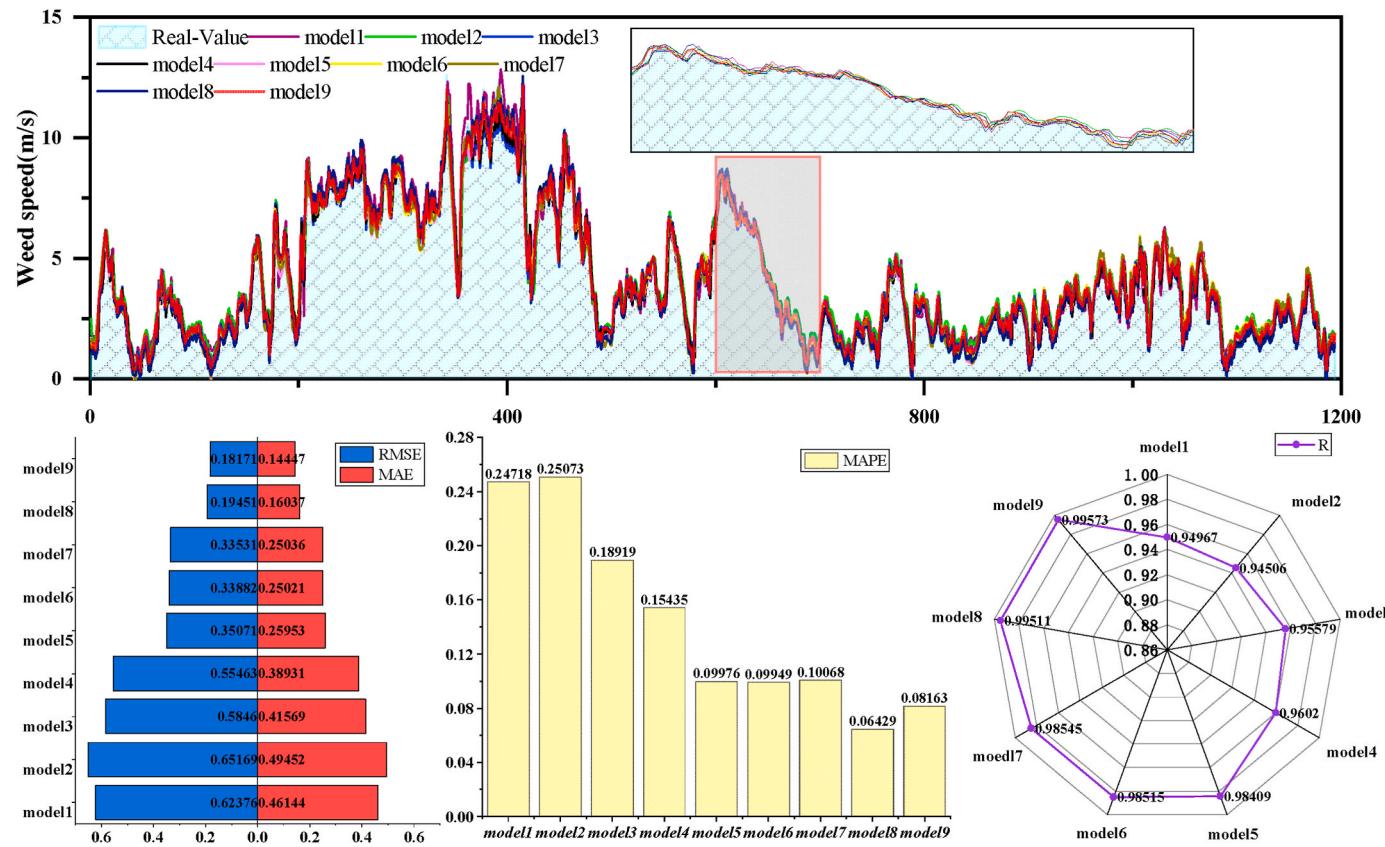


Fig. 10. Error analysis of July dataset.

After adjustment, the iteratively optimized ChOA process is as follow Eq. (20):

(d) When the new chimp hunting ability is between Chase and Driver ($fit_{Chaser} > fit > fit_{Driver}$), the new chimp will become the new

$$\left\{ \begin{array}{l} Attac\ ker = x(t+1), Barrier = Attac\ ker, Chaser = Barrier, Driver = Chaser, fit > fit_{Attac\ ker} \\ Barrier = x(t+1), Chaser = Barrier, Driver = Chaser, fit_{Attac\ ker} > fit > fit_{Barrier} \\ Chaser = x(t+1), Driver = Chaser, fit_{Barrier} > fit > fit_{Chaser} \\ Driver = x(t+1), fit_{Chaser} > fit > fit_{Driver} \end{array} \right. \quad (20)$$

where is the ranking mechanism optimization, there will be a total of 5 situations:

- (a) When a new chimp $x(t+1)$ appears with a stronger hunting ability than the Attacker ($fit > fit_{Attac\ ker}$), the new chimp will become a new Attacker, while the original Attacker will become a new Barrier, the original Barrier will become a new Chaser, the original Chase will become a new driver, and the original Driver will be eliminated.
- (b) When the new chimp hunting ability is between Attack and Barrier ($fit_{Attac\ ker} > fit > fit_{Barrier}$), the new chimp will become the new Barrier, the original Barrier and Chase will be downgraded to Chase and Driver, and the original Driver will be eliminated.
- (c) When the new chimp hunting ability is between Barrier and Chaser ($fit_{Barrier} > fit > fit_{Chaser}$), the new chimp will become the new Chaser, the original Chaser downgraded to Driver, and the original Driver will be eliminated.

Driver, and the original Driver will be eliminated.

(e) When the new chimp hunting ability was weaker than the Driver ($fit < fit_{Driver}$), the Attacker, Barrier, Chaser, and Driver remained unchanged.

2.4. Bi-directional Gated Recurrent Unit

2.4.1. Gated Recurrent Unit

GRU is a recurrent neural network proposed by Cho et al. [31], which is a variant of the LSTM [32] network. There are a total of three gates in an LSTM: a forget gate, an input gate, and an output gate. In order to improve the efficiency of model and simplify the network structure, GRU combines the input gate and the forget gate to form a new update gate η_t . The structure of GRU is shown in Fig. 1. The update gate η_t controls the amount of data that the front door memory information can continue to retain up to the current moment. Meanwhile, another gate of GRU is called a reset gate φ_t , which controls the amount of data to forget about how much past information to forget. GRU can be expressed as Eq. (21).

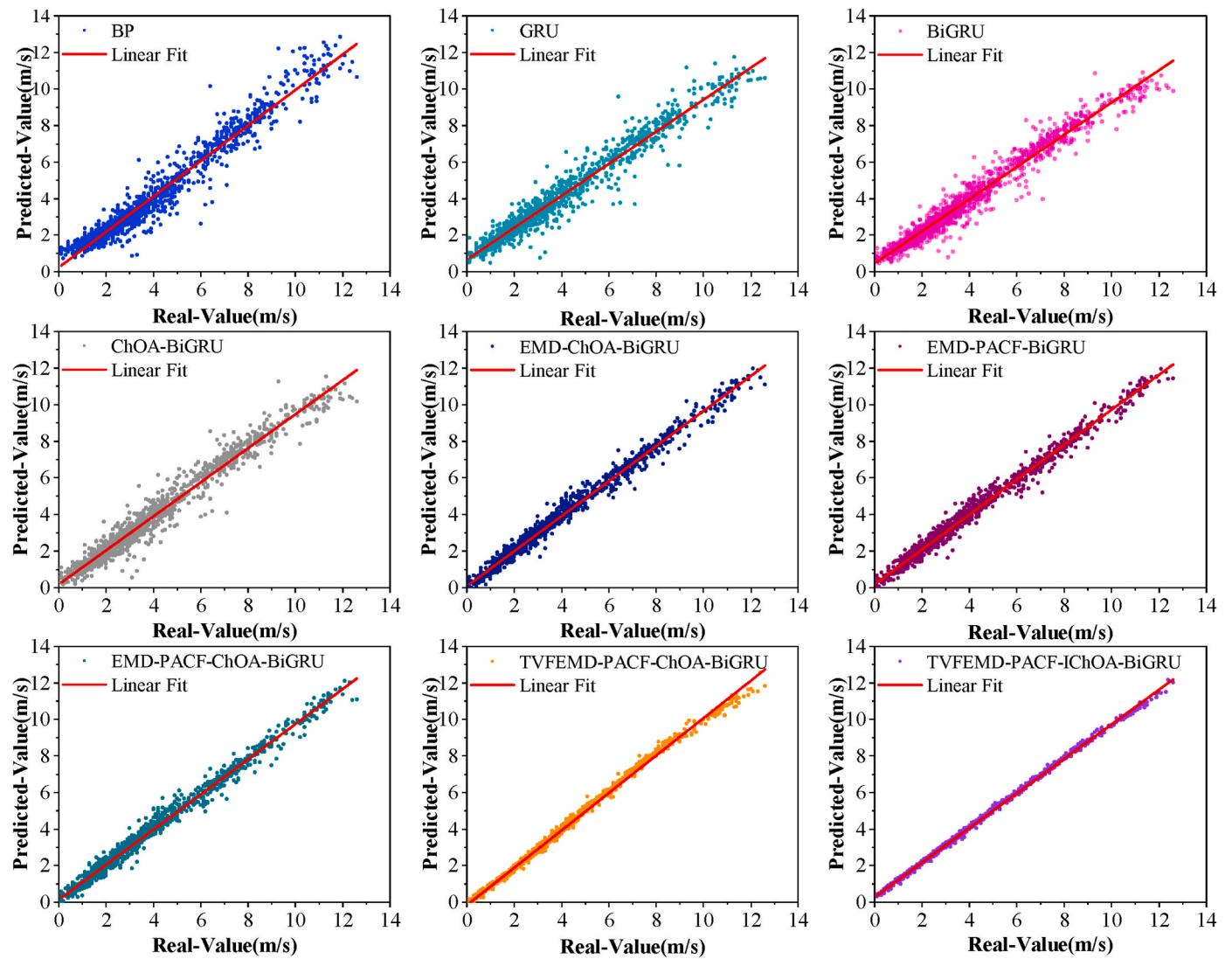


Fig. 11. Scatter plots of July dataset.

$$\begin{cases} \varphi_t = \sigma(\text{Weight}_\varphi \cdot [h_{t-1}, x_t]) \\ \eta_t = \sigma(\text{Weight}_\eta \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t = \tanh(\text{Weight}_h \cdot [r_t \times h_{t-1}, \tilde{h}_t]) \\ h_t = (E - \eta_t) \times h_{t-1} + \eta_t \times \tilde{h}_t \end{cases} \quad (21)$$

2.4.2. Bi-directional Gated Recurrent Unit

In traditional neural networks, the state is always output from front to back, that is, the use of information is always one-way. In order to overcome the limitations of the traditional GRU, BiGRU is used to process the obtained data information from both positive and negative directions [36]. In the forward layer of BiGRU neural network, the forward layer node of GRU read the vectors of input in forward order (from x_{t-1} to x_{t+1}) and calculate the forward hidden layer state corresponding to each vector. Similarly, in the reverse layer of BiGRU neural network, the each reverse layer node of GRU read the input vectors in reverse order (from x_{t+1} to x_{t-1}) and compute the reverse hidden layer state corresponding to each vector. Fig. 2 is the structure of BiGRU.

Fig. 2 shows the structural principles of BiGRU, GRU_{t-1} , GRU_t and GRU_{t+1} represents a single GRU node, x_{t-1} , x_t , x_{t+1} represents input, and y_{t-1} , y_t , y_{t+1} represents output.

3. Flowchart of the TVFEMD-PACF-IChOA-BiGRU model

The steps of the proposed TVFEMD-PACF-IChOA-BiGRU model are as follows:

Step 1. Wind speed data is collected at 10 min intervals and the outliers of the dataset are removed.

Step 2. Preprocess the data set obtained in **Step 1**.

Step 2.1. Apply TVFEMD on the original wind speed data to get the initial decomposition components, and calculate the FE values. Aggregate initial decomposition components according to FE values to obtain final decomposition components.

Step 2.2. PACF is used to process the final decomposed components, sort the importance, and select the data with high importance to reduce information redundancy. Then the data set is divided into training set and test set. After extracting the data features to obtain a multidimensional matrix, input the matrix into the IChOA-BiGRU model.

Step 3. Train the IChOA-BiGRU models

Step 3.1. To improve ChOA, Logistic mapping initialization is used instead of the original random initialization, and the ranking mechanism optimization is used to change the update method of population position. The IChOA is applied to optimize the learning rate and the number

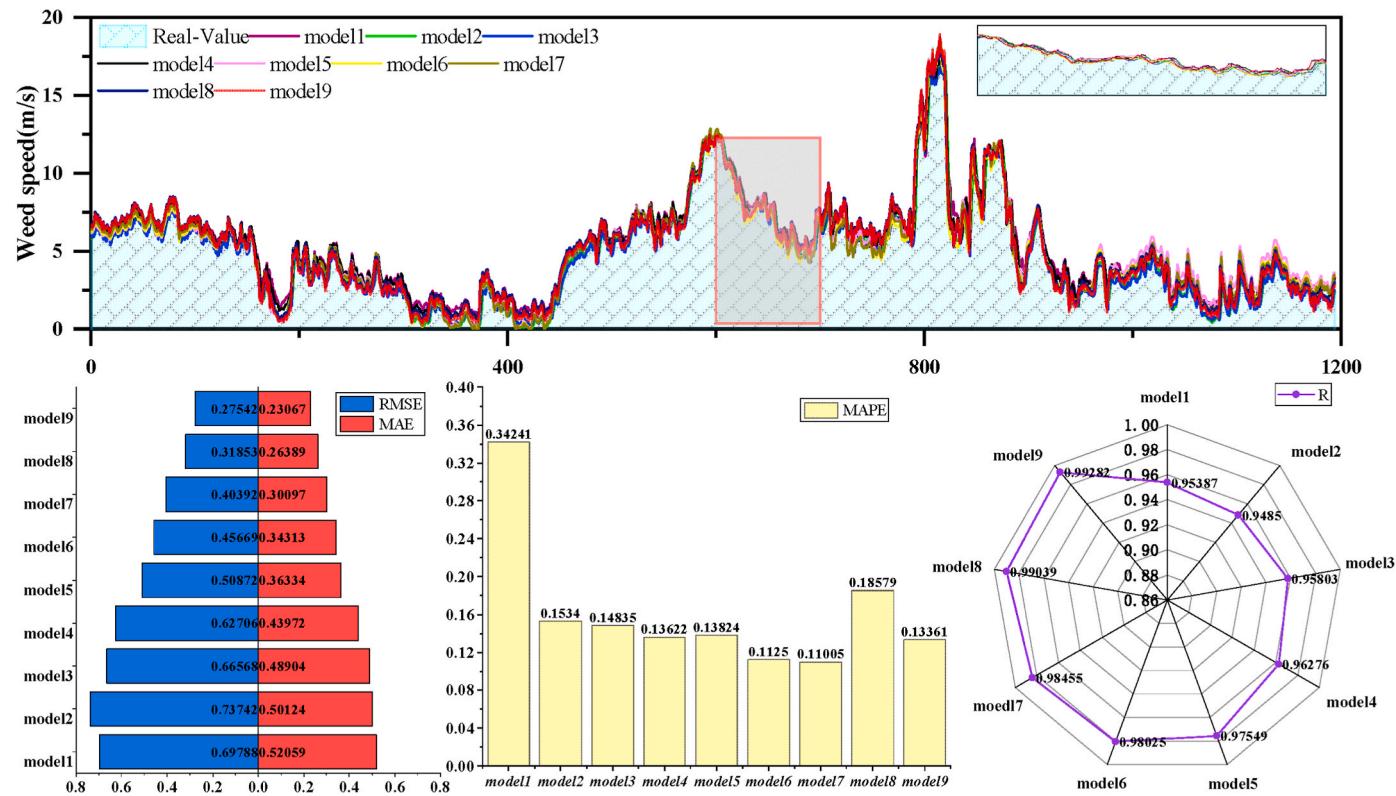


Fig. 12. Error analysis of October dataset.

of hidden layers of BiGRU.

Step 3.2. Use the multidimensional matrix obtained from [Step 2](#) to train the IChOA-BiGRU model and obtain the best model parameters.

Step 4. Input the test set data into the trained IChOA-BiGRU model, and the predicted values of each component are superimposed to obtain the final wind speed predicted value.

The flowchart of the TVFEMD-PACF-IChOA-BiGRU model is shown in [Fig. 3](#).

4. Case study

4.1. Data source and processing

This wind speed data of this experiment are obtained from National Data Buoy Center (NDBC), station 46,060. The latitude and longitude coordinates of this station are 60.586 N 146.787 W (60°35'10" N 146°47'12" W). The site is located near Prince William Sound and belongs to the Kenai Peninsula in southern Alaska. The area has a temperate marine climate, which is affected by the marine monsoon all year round and is rich in wind resources, which is very suitable for wind power generation and wind speed observation. In addition, Aizpurua-Etxezarreta et al. [37] found in the study that the freezing degree in this region has decreased significantly for a long time, and the Annual Energy Production (AEP) loss in this region has decreased year by year, which is conducive to the stable operation of wind power generation. The station collects wind speed data at 10min intervals, and the anemometer height of this station are 4.9 m above site elevation. In order to demonstrate the accuracy and universality of model prediction, this study selects four months in Total: January, April, July, and October in 2020. Among them, there were 4446 sampling points in January 4320 sampling points in April 4044 sampling points in July, and 4422 sampling points in October, for a total of 17,232 sampling points. 70% of the data for each month is taken as training set and 30% as test set. [Table 1](#) is

the basic information of the four monthly wind speed data sets, and [Fig. 4](#) is the wind speed data of the four months.

The turbulence intensity formula in [Table 1](#) is calculated as follows

$$I_T = \frac{\sigma}{V}$$

Where I_T represents the turbulence intensity, σ represents the standard deviation of the wind speed, and V represents the average wind speed of 10min of the wind speed.

In this experiment, the obtained wind speed data is decomposed by using the TVFEMD method, and the wind speed data with large fluctuations is decomposed into multiple components with gradually decreasing fluctuations (IMF1, IMF2, ... IMFn). The bandwidth threshold set 0.1 to complete the energy-decreasing decomposition of the wind speed data. Meanwhile, in order to improve the efficiency of data processing, the subcomponents are aggregated by FE. The decomposed subcomponents of the January data are shown in [Fig. 5](#).

4.2. Performance metrics

In this experiment, four different error indicators are used to calculate the performance of the model described above, the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Correlation coefficient (R), and the Mean Absolute Percentage Error (MAPE) [38].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{pi} - Y_{Ai})^2} \quad (22)$$

$$R = \frac{\sum_{i=1}^n (Y_{pi} - \bar{Y}_{pi})(Y_{Ai} - \bar{Y}_{Ai})}{\sqrt{\sum_{i=1}^n (Y_{pi} - \bar{Y}_{pi})^2} \sqrt{\sum_{i=1}^n (Y_{Ai} - \bar{Y}_{Ai})^2}} \quad (23)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{pi} - Y_{Ai}| \quad (24)$$

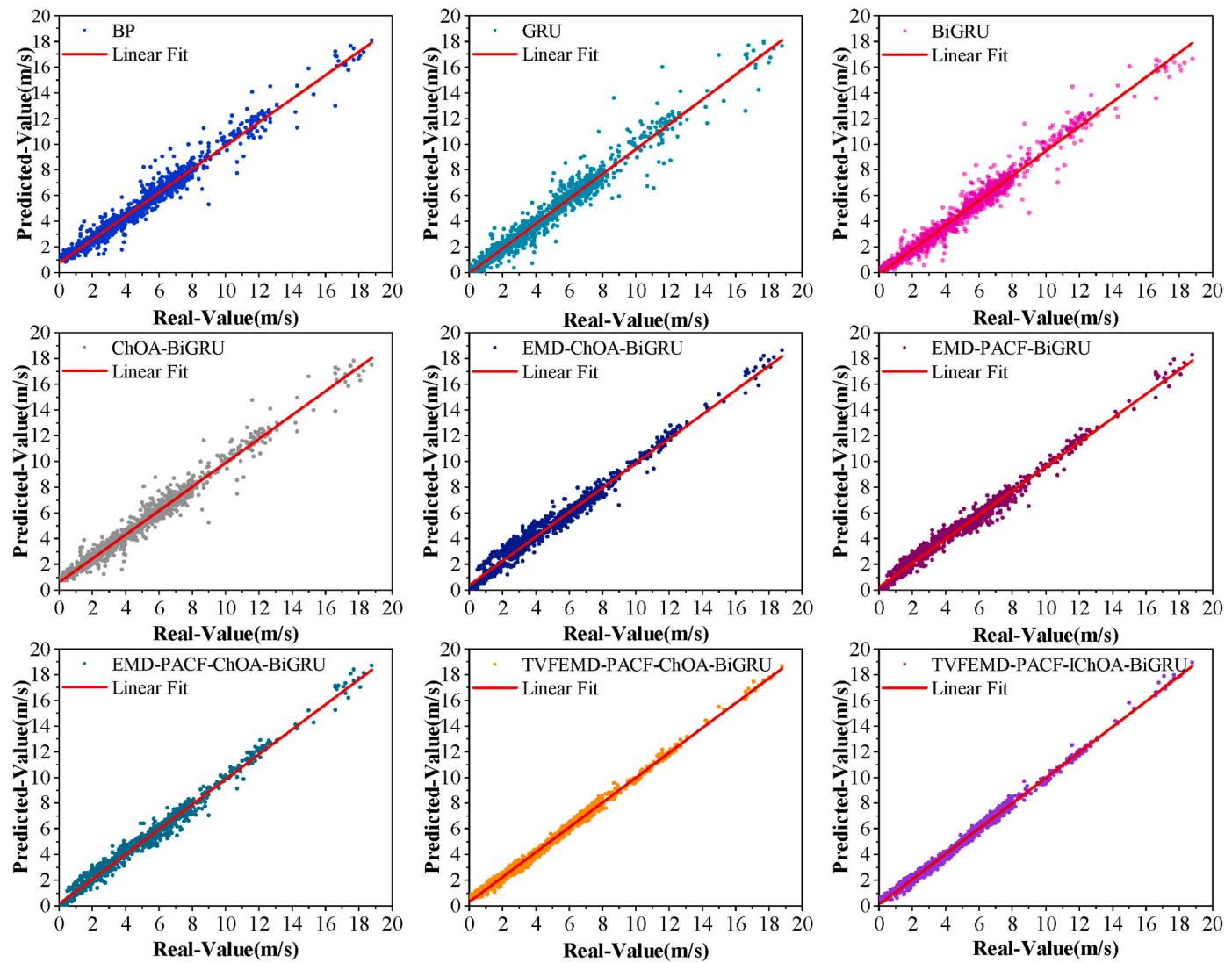


Fig. 13. Scatter plots of October dataset.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_{p_i} - Y_{A_i}}{Y_{A_i}} \right| \cdot \frac{100}{n} \quad (25)$$

where Y_{A_i} represent the actual value of the i th training sample, Y_{p_i} represent the predicted value of the i th sample, \bar{Y}_{A_i} and \bar{Y}_{p_i} represent the average of predicted and actual values, and n represent the length of samples.

5. Comparisons results

During the experiment, the collected data for 4 months were tested with 12 models. Through the comparison of single models and hybrid models, the effectiveness of the proposed hybrid prediction model is verified. First, the BP, GRU and BiGRU single models are control groups. Then, combined models with feature selection and evolution algorithms, the effect of individual optimization is displayed. Finally, signal decomposition is combined with other improvements to further illustrate the advantages of the hybrid model. These models include: BP, GRU, BiGRU, ChOA-BiGRU, IChOA-BiGRU, PACF-ChOA-BiGRU, EMD-PACF-ChOA-BiGRU, TVFEMD-PACF-ChOA-BiGRU, TVFEMD-PACF-IChOA-BiGRU.

The programming language used in this experiment is Matlab. The model parameters are set as follows: the training epochs of BP is 100, the

goal is 0.005, the learning rate is 0.02; the training epochs of GRU and BiGRU is 50; the iterations of ChOA are 20, the population of ChOA is 20; the training epochs range of improved BiGRU is [1,150], the learning rate range is [1e-4, 5e-1] [39].

5.1. Comparison of the prediction results in January

To evaluate the performance of the predictive model, the test results of the models are analyzed with four commonly used metrics, namely, RMSE, MAE, MAPE and R, while RMSE, MAE, and MAPE follow the smaller the better, and R should be as close as 1 as possible.

Models 1 to 9 represent BP, GRU, BiGRU, ChOA-BiGRU, IChOA-BiGRU, PACF-ChOA-BiGRU, EMD-PACF-ChOA-BiGRU, TVFEMD-PACF-ChOA-BiGRU, TVFEMD-PACF-IChOA-BiGRU, respectively. Table 2 is the comparison of wind speed prediction indexes in January. Fig. 6 and Fig. 7 are the error analysis and scatter plots of January, respectively.

By comparing the model predictors of Table 3, Figs. 6 and 7, the following conclusions can be drawn:

- (a) The TVFEMD-PACF-IChOA-BiGRU model proposed in this study has the most accurate prediction results compared with other control models. The RMSE of this model is 0.2452 m/s, MAE is 0.19095 m/s, and MAPE is 0.04485, respectively. The error value is the smallest compared with other models. The R of this model is

- 0.99021, which demonstrated that the proposed model has the highest degree of fit compared with other models, so the accuracy of the proposed model is the best.
- (b) By comparing the three basic models of BP, GRU and BiGRU, the BiGRU model can obtain the best effect. Compared with BP and GRU, BiGRU reduced RMSE by 6.25% and 8.09%, respectively, indicating the performance effect of BiGRU in the base models.
 - (c) In the experiment, the RMSE of BiGRU, ChOA-BiGRU, EMD-ChOA-BiGRU, EMD-PACF-ChOA-BiGRU was 0.80694 m/s, 0.75724 m/s, 0.464 m/s, and 0.42978 m/s, respectively. Comparing the RMSE of ChOA-BiGRU and BiGRU, EMD-PACF-ChOA-BiGRU and EMD-PACF-BiGRU, the improvement of ChOA to the model was 6.16% and 2.81%, respectively. The comparison of these two groups verifies that ChOA can optimize model parameters, improve model operation efficiency and calculation accuracy.
 - (d) The RMSE of the TVFEMD-PACF-IChOA-BiGRU and TVFEMD-PACF-ChOA-BiGRU models proposed in this study are 0.2452 m/s and 0.25972 m/s. The experimental results show the effect of improvement about ChOA on model parameter optimization.
 - (e) Through the comparison of ChOA-BiGRU and EMD-ChOA-BiGRU, the optimization effect of ChOA on the model under different improvements is verified.
 - (f) Compared with EMD-PACF-ChOA-BiGRU, TVFEMD-PACF-ChOA-BiGRU has improved the RMSE, MAE, MAPE and R by 39.57%, 34.73%, 28.66%, and 1.93%, respectively. This paper verifies that the data decomposition has improved the effect of the model significantly, and further verifies the effectiveness of TVFEMD on the model.

5.2. Comparison results of April, July and October

To further verify the stability and accuracy of the model, experiments were conducted on the site for another three months: April, July, October. January, April, July and October represent the spring, summer, autumn and winter of the site. The evaluation indicators are still RMSE, MAE, MAP, R to ensure the uniformity of the controls. Table 3 ~ 5 are the experimental results of April (Table 3), July (Table 4), and October (Table 5), and the error graphs and scatter plots of April (Figs. 8 and 9), July (Figs. 10 and 11), and October (Figs. 12 and 13) are given in Figs. 8–13, respectively.

The following conclusions can be drawn from Figs. 8 ~ 13 and Tables 3 ~ 5: (a) The TVFEMD-PACF-IChOA-BiGRU model proposed in this study is superior to other control models in four error indicators compared with other control models. When using RMSE as a reference index: (a) TVFEMD-PACF-IChOA-BiGRU can achieve 60%–80% optimization effect when compared with BP, GRU, BiGRU. (b) By observing the scatter plot and histogram, it can be known that in the control group, the GRU has the worst degree of dispersion and the largest error value, and the model proposed by this experiment has the smallest degree of dispersion and the smallest error value; GRU had the worst fit, and the model proposed in this experiment had the best fit. (c) Comparing the forecast curves, it can be concluded that the underlying model without optimization is difficult to adapt to large fluctuations and cannot keep up with peaks and valleys in time; After training, the prediction trend of the model proposed in this experiment is basically consistent with the real value. In summary, the experimental results show that the model proposed in this experiment performs best.

5.3. Comparison of the results in model1~model9

To observe the difference from the model1~model9, this study was analyzed the improvements percentage of RMSE, MAE, MAPE, and R [40]. Taking the models proposed in this paper as a reference, the other 8 models are evaluated separately to further reflect the excellent effects of the proposed models. The formula for percentage improvement is as

follows:

$$P_1 = \frac{RMSE_x - RMSE_y}{RMSE_x} * 100\% \quad (26)$$

$$P_2 = \frac{MAE_x - MAE_y}{MAE_x} * 100\% \quad (27)$$

$$P_3 = \frac{MAPE_x - MAPE_y}{MAPE_x} * 100\% \quad (28)$$

$$P_4 = \frac{R_y - R_x}{R_y} * 100\% \quad (29)$$

where $P_1 \sim P_4$ represent the percentage indicators of RMSE, MAE, MAPE, and R.

Tables 6–9 represent the percentage indicators for January (Table 6), April (Table 7), July (Table 8), and October (Table 9) for four months, respectively.

It can be inferred from the thorough review of Tables 5 ~ 8 that: a) Compared with BiGRU, RMSE, MAE, MAPE, and R in January are decreased by 69.61%, 65.91%, 61.59% and 10.76% on average, respectively (see Table 6). At the same time, RMSE, MAE, MAPE, and R are all decreased significantly in April, July and October, indicating that hybrid improvement of the single model can improve the accuracy of

Table 6
Improvement percentages of January.

Models	January			
	P ₁ (%)	P ₂ (%)	P ₃ (%)	P ₄ (%)
BP	71.51	70.99	67.8	12.6
GRU	72.07	70.15	69.91	13.23
BiGRU	69.61	65.91	61.59	10.76
ChOA-BiGRU	67.61	64.46	61.19	9.22
EMD-ChOA-BiGRU	47.15	45.26	34.23	2.61
EMD-PACF-BiGRU	44.55	43.04	35.09	2.27
EMD-PACF-ChOA-BiGRU	42.94	39.55	29.07	2.09
TVFEMD-PACF-ChOA-BiGRU	5.59	7.38	0.57	0.12

Table 7
Improvement percentages of April.

Models	April			
	P ₁ (%)	P ₂ (%)	P ₃ (%)	P ₄ (%)
BP	78.75	78.46	82.7	7.4
GRU	79.34	78.66	82.09	7.88
BiGRU	77.67	78.46	84.02	6.62
ChOA-BiGRU	74.79	74.15	79.36	5.04
EMD-ChOA-BiGRU	61.25	61.91	55.76	1.88
EMD-PACF-BiGRU	56.42	56.38	57.3	1.4
EMD-PACF-ChOA-BiGRU	53.5	52.17	53.07	1.19
TVFEMD-PACF-ChOA-BiGRU	27.77	32.72	45.92	0.29

Table 8
Improvement percentages of July.

Models	July			
	P ₁ (%)	P ₂ (%)	P ₃ (%)	P ₄ (%)
BP	70.86	68.69	66.97	4.85
GRU	72.11	70.78	67.44	5.36
BiGRU	68.91	65.24	56.85	4.17
ChOA-BiGRU	67.23	62.89	47.11	3.69
EMD-ChOA-BiGRU	58.97	57.95	39.03	2.16
EMD-PACF-BiGRU	53.06	49.59	26.56	1.54
EMD-PACF-ChOA-BiGRU	49.52	48.02	25.5	1.27
TVFEMD-PACF-ChOA-BiGRU	6.58	9.91	-26.97	0.06

Table 9
Improvement percentages of October.

Models	October			
	P ₁ (%)	P ₂ (%)	P ₃ (%)	P ₄ (%)
BP	60.53	55.68	60.97	60.53
GRU	62.65	53.97	12.89	62.65
BiGRU	58.62	52.83	9.93	58.62
ChOA-BiGRU	54.71	48.45	49.19	54.71
EMD-ChOA-BiGRU	45.85	36.51	3.34	45.85
EMD-PACF-BiGRU	39.69	32.77	-18.76	39.69
EMD-PACF-ChOA-BiGRU	31.81	23.35	-21.4	31.81
TVFEMD-PACF-ChOA-BiGRU	13.53	12.58	28.08	13.53

prediction model; b) Comparing EMD-PACF-ChOA-BiGRU with TVFEMD-PACF-ChOA-BiGRU, TVFEMD-PACF-ChOA-BiGRU is significantly better than EMD-PACF-ChOA-BiGRU and is closer to the proposed model, indicating that TVFEMD is effective than EMD; c) The results show that the suggested model can process data information more effectively and predict wind speed data more accurately.

6. Conclusions

The instability of wind speed has a great impact on the integration of wind turbines and wind power into the grid. Therefore, accurate prediction of wind speed plays a very important role in the stable operation of wind power generation equipment. In this paper, a novel hybrid wind speed prediction model based on TVFEMD, PACF, IChOA and BiGRU, was established. Based on four months of data, 9 groups of experiments were conducted for each month and the following conclusions were reached:

- (1) The model proposed in this experiment has the smallest value and the smallest error in RMSE, MAE and MAPE. The R value is the largest and the fit is the highest. Therefore, the model has the best performance with more accurate wind speed prediction results compared with other benchmark models.
- (2) BiGRU is more adaptable to the nonlinear characteristics of wind speed than other base models. By PACF, the key factors in the wind speed data can be extracted, and thus make the data easier to be processed by the model.
- (3) By analyzing the wind speed prediction performance of the models, it can be concluded that the evolutionary algorithm IChOA can improve the parameters of the BiGRU, thereby improving the wind speed prediction effect of the proposed model.
- (4) Data decomposition can greatly improve the prediction effect of the model. After processing by EMD and TVFEMD, the wind speed prediction effect of the model has been greatly improved, and the effect of TVFEMD was more significant than that of EMD.

7. Further research

In this paper, a hybrid wind speed prediction model based on ChOA optimized deep learning model is established, which realizes the single-step prediction of wind speed and provides some reference for wind power generation. However, in future studies, there are still further improvements, including the following aspects:

- (1) The prediction model is only based on the simple wind speed data, without considering the influence of wind direction, air density, wind power density, turbulence intensity and many other factors on the wind turbine. In the actual operation, the wind power generation needs to be adjusted according to many parameters to stabilize the normal operation of the wind motor, and the impact of different heights on the wind power generation should also be considered.

- (2) When establishing the wind speed prediction model, the single-step wind speed prediction is only based on a single station. If conditions permit, multiple sites will be processed together to improve the prediction accuracy and wind power prediction efficiency. At the same time, in future research, multi-step or online wind speed prediction models can be built to meet the actual wind power generation demand.
- (3) In this paper, the wind speed prediction model only considers from the level of wind speed data, but fails to consider other factors such as climate and terrain. In future studies, physical models will be built to be combined/hybridized with the proposed model to improve the wind speed prediction accuracy.

Credit author statement

Leiming Suo: Conceptualization, Methodology, Software, Writing - review & editing. **Tian Peng:** Software, Visualization, Writing - Original draft. **Shihao Song:** Supervision, Writing - review & editing. **Chu Zhang:** Supervision, Software. **Yuhan Wang:** Supervision, Software. **Yongyan Fu:** Supervision, Writing - review & editing. **Muhammad Shahzad Nazir:** Visualization, Writing - review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the Natural Science Foundation of Jiangsu Province (No. BK20191052), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB470012), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB480007), the Construction system of Jiangsu Province of China (No. 2019ZD061), the Double-innovation Doctor Program of Jiangsu province (No. JSSCBS20201033 and No. JSSCBS20201037). Special thanks are given to the Huai'an Innovative Talents Project.

References

- [1] Liang T, Chai C, Sun H, Tan J. Wind speed prediction based on multi-variable Capsnet-BiLSTM-MOHHO for WPCCC. Energy 2022;250:123761.
- [2] Tian Z, Wang J. Variable frequency wind speed trend prediction system based on combined neural network and improved multi-objective optimization algorithm. Energy 2022;254:124249.
- [3] Zhang C, Ma H, Hua L, Sun W, Nazir MS, Peng T. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. Energy 2022;254:124250.
- [4] Huang Y, Zhang B, Pang H, Wang B, Lee KY, Xie J, et al. Spatio-temporal wind speed prediction based on Clayton Copula function with deep learning fusion. Renew Energy 2022;192:526–36.
- [5] Gonzalez-Arceo A, Zirion-Martinez de Musitu M, Ulatzua A, del Rio M, Garcia O. Calibration of reanalysis data against wind measurements for energy production estimation of building integrated savonius-type wind turbine. Appl Sci 2017;10.
- [6] Niu D, Sun L, Yu M, Wang K. Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model. Energy 2022;254:124384.
- [7] Li J, Song Z, Wang X, Wang Y, Jia Y. A novel offshore wind farm typhoon wind speed prediction model based on PSO-Bi-LSTM improved by VMD. Energy 2022; 251:123848.
- [8] Ren Y, Wen Y, Liu F, Zhang Y. A short-term wind speed prediction method based on interval type 2 fuzzy model considering the selection of important input variables. J Wind Eng Ind Aerod 2022;225:104990.
- [9] Yang D, Wang W, Hong T. A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting. Sol Energy 2022;232:263–74.

- [10] Hur S-h. Short-term wind speed prediction using Extended Kalman filter and machine learning. *Energy Rep* 2021;7:1046–54.
- [11] Ma H, Zhang C, Peng T, Nazir MS, Li Y. An integrated framework of gated recurrent unit based on improved sine cosine algorithm for photovoltaic power forecasting. *Energy* 2022;256:124650.
- [12] Yu C, Li Y, Chen Q, Lai X, Zhao L. Matrix-based wavelet transformation embedded in recurrent neural networks for wind speed prediction. *Appl Energy* 2022;324:119692.
- [13] Gaalman G, Disney SM, Wang X. When bullwhip increases in the lead time: an eigenvalue analysis of ARMA demand. *Int J Prod Econ* 2022;108623.
- [14] Rodríguez F, Alonso-Pérez S, Sánchez-Guardamino I, Galarza A. Ensemble forecaster based on the combination of time-frequency analysis and machine learning strategies for very short-term wind speed prediction. *Elec Power Syst Res* 2023;214:108863.
- [15] Duan J, Chang M, Chen X, Wang W, Zuo H, Bai Y, et al. A combined short-term wind speed forecasting model based on CNN-RNN and linear regression optimization considering error. *Renewable Energy*; 2022.
- [16] Del Ser J, Casillas-Perez D, Cornejo-Bueno L, Prieto-Godino L, Sanz-Justo J, Casanova-Mateo C, et al. Randomization-based machine learning in renewable energy prediction problems: critical literature review, new results and perspectives. *Appl Soft Comput* 2022;118:108526.
- [17] Wang L, Guo Y, Fan M, Li X. Wind speed prediction using measurements from neighboring locations and combining the extreme learning machine and the AdaBoost algorithm. *Energy Rep* 2022;8:1508–18.
- [18] Li R, Jin Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. *Appl Energy* 2018;228:2207–20.
- [19] Filik ÜB, Filik T. Wind speed prediction using artificial neural networks based on multiple local measurements in eskisehir. *Energy Proc* 2017;107:264–9.
- [20] Shboul B, Al-Arfi I, Michailos S, Ingham D, Ma L, Hughes KJ, et al. A new ANN model for hourly solar radiation and wind speed prediction: a case study over the north & south of the Arabian Peninsula. *Sustain Energy Technol Assessments* 2021;46:101248.
- [21] Han Y, Mi L, Shen L, Cai CS, Liu Y, Li K, et al. A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. *Appl Energy* 2022;312:118777.
- [22] Shen Z, Fan X, Zhang L, Yu H. Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network. *Ocean Eng* 2022;254:111352.
- [23] Lim J-Y, Kim S, Kim H-K, Kim Y-K. Long short-term memory (LSTM)-based wind speed prediction during a typhoon for bridge traffic control. *J Wind Eng Ind Aerod* 2022;220:104788.
- [24] Zhu Q, Zhang F, Liu S, Wu Y, Wang L. A hybrid VMD-BiGRU model for rubber futures time series forecasting. *Appl Soft Comput* 2019;84:105739.
- [25] Yan Y, Wang X, Ren F, Shao Z, Tian C. Wind speed prediction using a hybrid model of EEMD and LSTM considering seasonal features. *Energy Rep* 2022;8:8965–80.
- [26] Yang Q, Deng C, Chang X. Ultra-short-term/short-term wind speed prediction based on improved singular spectrum analysis. *Renew Energy* 2022;184:36–44.
- [27] Xiong D, Fu W, Wang K, Fang P, Chen T, Zou F. A blended approach incorporating TVFEMD, PSR, NNCT-based multi-model fusion and hierarchy-based merged optimization algorithm for multi-step wind speed prediction. *Energy Convers Manag* 2021;230:113680.
- [28] Ding Y, Chen Z, Zhang H, Wang X, Guo Y. A short-term wind power prediction model based on CEEMD and WOA-KELM. *Renew Energy* 2022;189:188–98.
- [29] Hua L, Zhang C, Peng T, Ji C, Shahzad Nazir M. Integrated framework of extreme learning machine (ELM) based on improved atom search optimization for short-term wind speed prediction. *Energy Convers Manag* 2022;252:115102.
- [30] Bo Q, Cheng W, Khishe M, Mohammadi M, Mohammed AH. Solar photovoltaic model parameter identification using robust niching chimp optimization. *Sol Energy* 2022;239:179–97.
- [31] Ulazia A, Sáenz J, Ibarra-Berastegui G, González-Rojo SJ, Carreno-Madinabeitia S. Global estimations of wind energy potential considering seasonal air density changes. *Energy* 2019;187:115938.
- [32] Carreno-Madinabeitia S, Ibarra-Berastegui G, Sáenz J, Ulazia A. Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900–2010). *Energy* 2021;226:120364.
- [33] Yang G, Yuan E, Wu W. Predicting the long-term CO₂ concentration in classrooms based on the BO-EMD-LSTM model. *Build Environ* 2022;224:109568.
- [34] Li Y, Peng T, Hua L, Ji C, Ma H, Nazir MS, et al. Research and application of an evolutionary deep learning model based on improved grey wolf optimization algorithm and DBN-ELM for AQI prediction. *Sustain Cities Soc* 2022;87:104209.
- [35] Khishe M, Mosavi MR. Chimp optimization algorithm. *Expert Syst Appl* 2020;149:113338.
- [36] Zhang C, Peng T, Nazir MS. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. *Elec Power Syst Res* 2022;213:108796.
- [37] Aizpurua-Etxezarreta M, Carreno-Madinabeitia S, Ulazia A, Sáenz J, Saenz-Aguirre A. Long-term freezing temperatures frequency change effect on wind energy gain (eurasia and north America, 1950–2019). *Sustainability* 2022;14:5630.
- [38] Tao Z, Zhang C, Xiong J, Hu H, Ji J, Peng T, et al. Evolutionary gate recurrent unit coupling convolutional neural network and improved manta ray foraging optimization algorithm for performance degradation prediction of PEMFC. *Appl Energy* 2023;336:120821.
- [39] Zhang C, Peng T, Nazir MS. A novel hybrid approach based on variational heteroscedastic Gaussian process regression for multi-step ahead wind speed forecasting. *Int J. Elect. Power Energy Syst* 2022;136:107717.
- [40] Xiong J, Peng T, Tao Z, Zhang C, Song S, Nazir MS. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction. *Energy* 2023;266:126419.