Energy

A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short - term wind speed forecasting

--Manuscript Draft--

Manuscript Number:	EGY-D-22-00224R2
Article Type:	Full length article
Keywords:	Artificial Intelligence; Data preprocessing; Combined forecasting system; Multi-objective optimization
Corresponding Author:	Yining An CHINA
First Author:	Jianzhou Wang
Order of Authors:	Jianzhou Wang
	Yining An
	Zhiwu Li
	Haiyan Lu
Abstract:	Accurate wind speed prediction has become increasingly important in wind power generation. However, the lack of efficient data preprocessing techniques and integration strategies has been a big obstacle to the development of wind power forecasting system. Therefore, a novel and advanced combined forecasting system comprising a data preprocessing, an integration strategy and several single models is designed in this study. The proposed model not only eliminates the impact of noise, but also integrates several single-model forecasting results through a weight optimization operator. In addition, the uncertain prediction of wind speed is also discussed in detail. The results show that: (a) The MAPE values of the proposed model are 2.8645%, 2.1843% and 2.8727% respectively for the point prediction. (b) The FICP values of the proposed model are 85.1697, 89.5410 and 88.0111 respectively at the significant level $\alpha = 0.05$ for the uncertainty forecasting. The AWD values are 0.0559, 0.0400 and 0.0361 and the FINAW values are 0.0478, 0.0404 and 0.0390. It is reasonable to conclude that the proposed system can effectively boost the precision and stability of wind speed forecasting and provide a new approach for the exploitation of wind energy.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Declaration of interests

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

Highlights

- A combined forecasting system including a data preprocessing, a combined and uncertainty prediction module is designed.
- An advanced data preprocessing technique is intended to remove the noise.
- A combined prediction strategy including optimal sub-model selection and weight optimization operators is proposed.
- The Pareto optimality of the solutions is theoretically proven.
- Interval prediction increases the accuracy and certainty of prediction results.

Dear Editors:

We would like to submit the enclosed manuscript entitled "**A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting**" in **Energy**. No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

We deeply appreciate your consideration of our manuscript, and we look forward to receiving comments from the reviewers. If you have any queries, please don't hesitate to contact me at the address below.

Thank you and best regards.

Yours sincerely,

Yining An

Corresponding author:

Name: Yining An

E-mail: anyining202012@163.com

Credit Author Statement

Jianzhou Wang: Conceptualization, Software, Writing – original draft. **Yining An:** Methodology, Supervision, Writing – review & editing. **Zhiwu Li:** Visualization, Software, Validation. **Haiyan Lu:** Validation, Formal analysis.

Dear editors and reviewers:

Thank you very much for e-mailing us the comments raised by the respected reviewers. The manuscript **ID EGY-D-22-00224 "A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting"** has been revised taking into account all of the helpful comments and suggestions. The details of the comments raised, the answers and the actions taken are presented here. All the changes made in the new revised manuscript have been marked in yellow. We appreciate for respected editors/reviewers' warm work earnestly, and hope that the correction will meet with approval. We look forward to hearing from you.

Best regards,
Jianzhou Wang

Comment raised by respected Reviewer 1:

Comment 1: In the Introduction section, it starts with 1.2. The heading of 1.1 is missing.

Response: Thank you very much for your valuable advice and it is helpful to improve the quality of our paper. We have corrected the error in the heading. Please see the revised manuscript in section 1.1 in 3th page for more details.

****

Comment 2: In Section 2, the description of all algorithms should be given, rather than just that of denoising and optimization algorithm. This section should be re-written.

Response: Thank you for your suggestions. We have added the description of algorithms used in the study, with the additions shown below. Please see the revised manuscript in section 2.2 in 7th-9th page for more details. Thank you again for your valuable advice.

LSTM (long short-term memory) is a special type of RNNs. The key components of LSTM are memory cells and gates. The forget gate determines the number of the unit state $\boldsymbol{C}(\boldsymbol{t}-1)$ remaining at current moment $\boldsymbol{C}(\boldsymbol{t})$ and is calculated as $\overline{\overline{\boldsymbol{F}}}_{\boldsymbol{\sigma},\bar{\boldsymbol{b}},\bar{\boldsymbol{\omega}}}(\boldsymbol{t})=\boldsymbol{\sigma}(\bar{\boldsymbol{b}}_{\boldsymbol{f}}+\{\boldsymbol{x}_{\boldsymbol{t}},\boldsymbol{h}_{\boldsymbol{t}-1}\}*\bar{\boldsymbol{\omega}}_{\boldsymbol{f}})$. The number of inputs $\boldsymbol{x}(\boldsymbol{t})$ stored in the cell state $\boldsymbol{C}(\boldsymbol{t})$ is determined in the calculation process of the input gate $\overline{\overline{\boldsymbol{I}}}_{\boldsymbol{\sigma},\bar{\boldsymbol{b}},\bar{\boldsymbol{\omega}}}(\boldsymbol{t})=\boldsymbol{\sigma}(\bar{\boldsymbol{b}}_{\boldsymbol{i}}+\{\boldsymbol{x}_{\boldsymbol{t}},\boldsymbol{h}_{\boldsymbol{t}-1}\}*\bar{\boldsymbol{\omega}}_{\boldsymbol{i}})$. Finally, the output gate controls the number of unit states $\boldsymbol{C}(\boldsymbol{t})$ to be output to the present output value $\boldsymbol{h}(\boldsymbol{t})$ and is given by $\overline{\overline{\mathrm{O}}}_{\boldsymbol{\sigma},\bar{\boldsymbol{b}},\bar{\boldsymbol{\omega}}}(\boldsymbol{t})=\boldsymbol{\sigma}(\bar{\boldsymbol{\omega}}_{\boldsymbol{o}}*\{\boldsymbol{h}_{\boldsymbol{t}-1},\boldsymbol{x}_{\boldsymbol{t}}\}+\bar{\boldsymbol{b}}_{\boldsymbol{o}})$ where $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{\omega}}$ are bias and weight, respectively. The calculation principle of GRU (gated recurrent unit) is very similar to LSTM. The difference is that GRU combines the forgetting gate and input gate in LSTM algorithm into update gate, so GRU consists of two gates, the update gate and the reset gate.

The hidden layer of CNN (convolutional neural network) consists of convolution C_L, pooling P_L and full connection F_L. The convolution process is represented as $C_{x,y} = \sum_i^{p*q} \bar{\omega}_i v_i$, which is the sum of the product of the kernel weight and the brightness of the corresponding element in the input image. The main goal of pooling is to reduce the feature space of the maps. The fully connected layer is greatly important in combining the extracted features to obtain the output.

TCN (temporal convolutional networks) is suitable for sequence model construction under causal constraints, that is, the output $\{\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_t\}$ can only be predicted based on the past observation $\{x_0, x_1, \ldots, x_t\}$. Therefore, TCN can be designed as a nonlinear function with the mapping $f : X^T \to Y^T$. In addition, TCN also adds the dilated convolution and residual block to better extract historical information.

QRNN (quasi-recurrent neural networks) is the LSTM acceleration algorithm, including two components of convolution and pooling. In the convolution operation, the output of the input, forget and output gates can be expressed as $Z = \mathbf{tanh}(W_z * X)$, $F = \sigma(W_f * X)$ and $O = \sigma(W_O * X)$, where $\mathbf{X}$ is the input, W_z, W_f and W_O are the convolution filters. There are three ways the pooling process can reduce the number of features, named f-pooling, fo-pooling and ifo-pooling based on the number of gates used.

ANFIS (adaptive neuro-fuzzy inference system) is a neural network based on fuzzy reasoning. The first layer is a fuzzy layer. $\bar{\bar{O}}^1_{i=1,2} = \mu_{A_i}(x)$ and $\bar{\bar{O}}^1_{j=1,2} = \mu_{B_i}(y)$ are output functions where x and y are the inputs with the respective membership functions μ_{A_i} and μ_{B_i}. The second layer calculates the weights of each membership function according to the previous outputs. The third layer normalizes the weights as $\bar{\bar{O}}^3_i = \bar{\bar{\omega}}_i = \bar{\omega}_i / (\bar{\omega}_1 + \bar{\omega}_2)$. The fourth layer provides the output of the rule inference $\bar{\bar{O}}^4_i = \bar{\bar{\omega}}_i \bar{\bar{F}}_i$. The last layer generates the sum of each output $\bar{\bar{O}}^5_i = \sum \bar{\bar{\omega}}_i \bar{\bar{F}}_i = \left(\sum \bar{\bar{\omega}}_i \bar{\bar{F}}_i\right) / \sum \bar{w}_i$.

ELM (extreme learning machine) is a kind of feedforward neural network that includes input, hidden and output layers. The thresholds b and input weights ω are randomly generated. The input and output are $\vec{X} = (x_i^1, x_i^2, \ldots x_i^n) \in \mathbb{S}^n$ and $\vec{T} = (t_i^1, t_i^2, \ldots t_i^m) \in \mathbb{S}^m$. The hidden layer output is $\bar{\bar{H}}^{\bar{\omega},b}_{i=1,\ldots,n}(x) = \bar{\bar{F}}(x^T \bar{\omega}_i + b_i)$,where $\bar{\bar{F}}(\bullet)$ is the activation function. The final output is $\bar{\bar{O}}^{\omega,b,\theta}_n(x) = \sum_{i=1}^n \bar{\bar{H}}_i(x)\bar{\theta}(i)$, where the output weights are $\bar{\theta} = (\bar{\theta}_1, \cdots, \bar{\theta}_n)$.

BPNN (back propagation neural network) consists of input, hidden and output layers, and the calculation is: $y_i = f_I\left(\mu_j + \sum_{m=t-n}^{t-1} \mu_{jm} y_m\right)(0 \le \mu_j, \mu_{jm} \le 1)$, $y_t = f_0\left(\lambda_0 + \sum_{j=1}^{I} \lambda_{0j} y_j\right)(0 \le \lambda_0, \lambda_{0j} \le 1)$, where y_m and y_j are the inputs of the input and hidden layers, y_t is the predicted value at time **t**. N and I are the number of nodes in the input and the hidden layers. In the hidden and output layers, μ_j and λ_0 are the threshold values, μ_{jm} and λ_{0j} are weights, f_I and f_0 are activation functions.

On the basis of BPNN structure, ENN (elman neural network) adds a continuation layer to the hidden layer as a delay operator to achieve the purpose of memory. The

learning process is: $x(k) = f(w_1 x_c(k) + w_2(u(k-1)))$, $y(k) = g(w_3 x(k))$, where y is the m-dimensional output vector, x is the unit vector of n-dimensional middle layer, u is the r-dimensional input vector, $w_i (i = 1, 2, 3)$ is the connection weight of each layer, $g(\bullet)$ and $f(\bullet)$ are activation functions of output and middle layer neurons respectively.

GRNN (general regression neural network) is a kind of radial basis neural network, which is composed of input, pattern, summation and output layers. $p_i = \exp[-(X - X_i)^T (X - X_i) / 2\sigma^2]$ $i = 1, 2, \ldots, n$ is the neuron transfer function of the pattern layer, where X is the input variable and X_i is the learning sample of the i-th neuron. The output of neuron j can be calculated as $y_i = S_{Nj} / S_D$, where S_{Nj} and S_D are the arithmetic sum and weighted sum of neurons at the pattern layer.

****

Furthermore, there is another reviewer's comment in the attachment. We have also carefully revised according to these suggestions.

Comment raised by respected Reviewer 2:

Thank you again for your valuable suggestions and it is helpful to improve the quality of our paper.

****

Furthermore, there is another reviewer's comment in the attachment. We have also carefully revised according to these suggestions.

Comment raised by respected Reviewer 3:

Comment 1: Introduction: The reviewer suggested that the author add the following content: The reason why wind speed prediction technology can effectively improve the stability of wind power system.

Response: Thank you very much for your valuable advice and it is helpful to improve the quality of our paper. We have added the reason why wind speed prediction technology can effectively improve the stability of wind power system in the introduction section, and the corresponding modifications are listed as follows. Please see the revised manuscript in section 1 in 3th page for more details.

Owing to the rapid growth of the world economy, traditional energy sources such as natural gas and oil are being consumed in large quantities. Therefore, the development of sustainable green energy resources has received increasing attention [1]. Wind energy is a renewable energy resource with clean and effective characteristics that play an irreplaceable role in wind power generation. However, the instability and nonlinearity of wind speed limit the development of wind power and bring many obstacles to the wind power grid. Accurate wind speed prediction technology can

reduce the impact of wind speed characteristics, which not only helps power grid operators and decision makers to timely plan and dispatch the power system, but also reduces the failure risk of wind power system and improves power quality [2]. Hence, accurate wind speed prediction technology can effectively improve the stability of wind power generation system [3].

References

[1] Zhang Z, Qin H, Liu Y, Wang Y, Yao L, Li Q, et al. Long Short-Term Memory Network based on Neighborhood Gates for processing complex causality in wind speed prediction. Energy Conversion and Management. 2019; 192:37-51.
[2] Zhang Y, Pan G, Chen B, Han J, Zhao Y, Zhang C. Short-term wind speed prediction model based on GA-ANN improved by VMD. Renewable Energy. 2020; 156:1373-88.
[3] Wang J, Wang Y, Li Z, Li H, Yang H. A combined framework based on data preprocessing, neural networks and multi-tracker optimizer for wind speed prediction. Sustainable Energy Technologies and Assessments. 2020; 40:100757.

****

Comment 2: 1.2. Previous literature: First, I think the title number should be 1.1, not 1.2, and the same goes for 1.3. Then, it is suggested that the author divide the decomposition technique, prediction method and ensemble method into several paragraphs, so that scholars can read and understand them more clearly.

Response: Thank you for your suggestions. We have corrected the error in the heading, and we have also divided the content of section 1.1 into several paragraphs to make the reader read more clearly. Please see the revised manuscript in section 1.1 in 3th and 4th page for more details.

****

Comment 3: 2. Design of the combined forecasting model: I think the following sentence is not very good: "The data denoising algorithm and multi-objective grey wolf optimization algorithm are presented in this part." The reviewer suggested that the authors change this part into a brief introduction of their proposed model.

Response: Thank you very much for your valuable advice and it is helpful to improve the quality of our paper. We have revised the introduction at the beginning of section 2. The corresponding modifications are as follows. Please see the revised manuscript in section 2 in 7th page for more details.

The proposed prediction model in this study is mainly composed of data preprocessing technology, multiple single models, multi-objective weight optimization operator and uncertainty prediction, which is used to improve the prediction accuracy and quantify the uncertainty of prediction results. The corresponding algorithm and theoretical introduction are presented in this part.

****

Comment 4: 2.1 Data denoising strategy: Details of the authors' techniques and modeling to avoid data leaks can be added here.

Response: Thank you for your suggestions. We have added details of the techniques and modeling to avoid data leakage in section 2.1. The corresponding modifications are as follows. Please see the revised manuscript in section 2.1 in 7th page for more details.

Before data denoising, we process the original data to prevent data leakage. Specifically, data leakage can cause the model to look accurate, but when applied to real life, the model can become very inaccurate, which is mainly divided into feature leakage and training data leakage. For feature leakage, features generally have a strong correlation with target variables. In this study, prevention of training data leakage is of greater concern to us. Before dividing the datasets, preprocessing the whole data set will lead to data leakage and make the information of the test set appear in the training set. Therefore, after dividing the datasets, we use decomposition technology for training set and test set respectively to prevent data leakage. The corresponding data denoising strategy is presented as follows.

****

Comment 5: Some Figures, such as Fig. 4, Reviewers believe that the author could adjust it to be more beautiful or clearer, which would help readers further understand the meaning of them.

Response: Thank you very much for your valuable advice and it is helpful to improve the quality of our paper. We have adjusted the figures in the study to make them more clear and beautiful and easy for readers to understand. The modified figures are listed below.

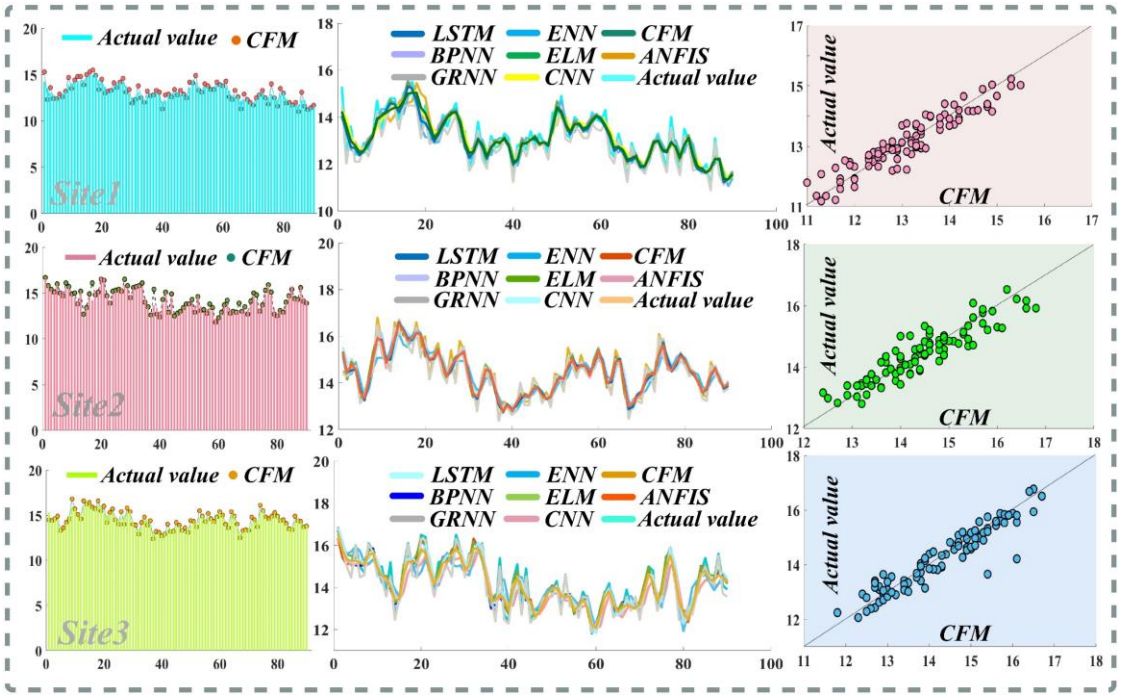

Fig.4 Results of CFM and denoising models

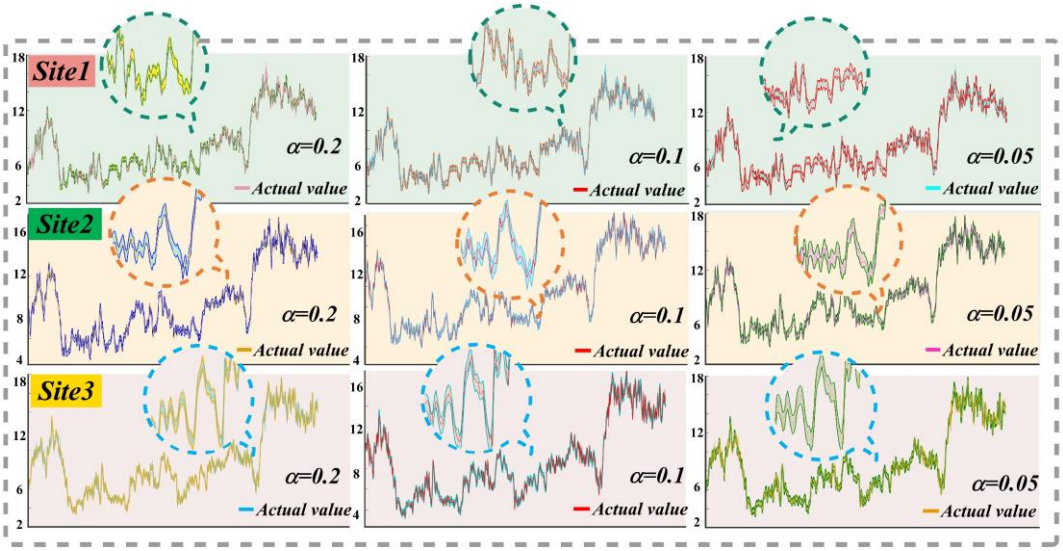

Fig.5 The interval forecasting results

****

Comment 6: As for the boldface and italic formats in the article, the reviewer suggests that the author carefully check and maintain consistency.

Response: Thank you for your suggestions. We have carefully checked the boldface and italic formats of this study to ensure the consistency of symbols. Thank you again for your valuable advice.

****

Revised Manuscript with No Changes Marked
Click here to view linked References

A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting

Jianzhou Wang[a], Yining An[b,*] , Zhiwu Li[a] , Haiyan Lu[c]

[a] *Macao Institute of Systems Engineering, Macau University of Science and Technology, Macao, 999078, China*

[b] *School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China*

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

** Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China*

Tel.: +86 17734579525

E-mail address: anyining202012@163.com

Abstract

Accurate wind speed prediction has become increasingly important in wind power generation. However, the lack of efficient data preprocessing techniques and integration strategies has been a big obstacle to the development of wind power forecasting system. Therefore, a novel and advanced combined forecasting system comprising a data preprocessing, an integration strategy and several single models is designed in this study. The proposed model not only eliminates the impact of noise, but also integrates several single-model forecasting results through a weight optimization operator. In addition, the uncertain prediction of wind speed is also discussed in detail. The results show that: (a) The MAPE values of the proposed model are 2.8645%, 2.1843% and 2.8727% respectively for the point prediction. (b) The FICP values of the proposed model are 85.1697, 89.5410 and 88.0111 respectively at the significant level α = 0.05 for the uncertainty forecasting. The AWD values are 0.0559, 0.0400 and 0.0361 and the FINAW values are 0.0478, 0.0404 and 0.0390. It is reasonable to conclude that the proposed system can effectively boost the precision and stability of wind speed forecasting and provide a new approach for the exploitation of wind energy.

Keywords: *Artificial intelligence; Data preprocessing; Combined forecasting model; Multi-objective optimization;*

1. Introduction

Owing to the rapid growth of the world economy, traditional energy sources such as natural gas and oil are being consumed in large quantities. Therefore, the development of sustainable green energy resources has received increasing attention [1]. Wind energy is a renewable energy resource with clean and effective characteristics that play an irreplaceable role in wind power generation. However, the instability and nonlinearity of wind speed limit the development of wind power and bring many obstacles to the wind power grid. Accurate wind speed prediction technology can reduce the impact of wind speed characteristics, which not only helps power grid operators and decision makers to timely plan and dispatch the power system, but also reduces the failure risk of wind power system and improves power quality [2]. Hence, accurate wind speed prediction technology can effectively improve the stability of wind power generation system [3].

1.1 Previous literature

At present, the wind speed forecasting approaches adopted in a large number of wind speed prediction studies include: (i) physical models, (ii) statistical models, (iii) Artificial intelligence models. Physical models are more suitable for long-term prediction, but have obvious defects for short-term prediction [4, 5]. Systematic errors can be easily generated in the predictions and the direction near the ground that result in differences in the predicted power generation when physical models are used [6, 7]. Dong et al. presented a hybrid model based on K-means cluster and general regression neural network (K-means-GRNN) of numerical weather forecasts; however, the practical application of the K-means-GRNN was difficult due to its high requirements on computation and information [8]. In contrast, statistical models have lower requirements for datasets [9]. Statistical models are usually linear models, such as the autoregressive moving average (ARMA) [10], the autoregressive integrated moving average model (ARIMA) [11, 12]. Movahed et al. predicted the development of cancer cells based on ARMA and Auto-Regressive (AR), and improved the forecasting accuracy [13]. However, for sequences with random and nonlinear characteristics, the statistical models were difficult to mine the information accurately and effectively [14].

Aiming at the deficiency of the above models, many researchers have performed in-depth studies on artificial intelligence (AI) prediction models [15, 16], which mainly include support vector machines (SVM) [17, 18], deep learning [19], and artificial neural networks (ANN) [20, 21]. AI prediction methods have strong generalization abilities, and fast calculation speeds, however, these single AI models may be affected by the initial parameters and are prone to fall into local optima during computations [22]. Considering the deficiency of the above approaches, combined models have been widely concerned, which often contain data preprocessing techniques and optimization algorithms [23].

On the one hand, because of the fluctuation and instability in wind speed, the use of data preprocessing techniques such as empirical mode decomposition (EMD) [24] and ensemble empirical mode decomposition (EEMD) [25] has been explored. In fact, disadvantages such as modes mix and boundary effect exist in EMD, although EEMD improves the shortcomings of EMD, it causes residual white noise [26]. Thus, an advanced strategy named improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) is proposed in this study, which can add special white noise to obtain accurate modal values and improve model accuracy.

On the other traditional single models may not be able to balance accuracy and stability in prediction [27]. Wang et al. designed a hybrid model of wavelet neural network optimized by genetic algorithm (GA-WNN) based on variational mode decomposition (VMD) for wind speed forecasting, however, it remains to be explored that both accuracy and stability can be satisfied [28]. Based on this, this paper adopts the multi-objective grey wolf optimizer (MOGWO), which can combine multiple single models and optimize the assigned weight through to obtain stable and accurate results. Moreover, many studies ignore the reliability and uncertainty of wind speed prediction, which will cause potential risks in practical applications [29]. Thus, the uncertainty of wind speed prediction based on interval prediction has been discussed in this study. **Table 1** summarizes the common model types in the literature in recent years.

1.2 Contribution

Through the above analysis, a novel combined predictive framework is built based on neural networks, deep learning approaches, and multi-objective optimization to improve the forecasting accuracy. The developed system consists of four modules: a data preprocessing module, a combination prediction module, an uncertainty prediction module and an assessment module. Modal decomposition is adopted in the data preprocessing module decomposition of the original sequence to eliminate noise and obtain a smooth reconstructed sequence. The combination prediction module is designed to combine several single prediction models and optimize the assigned weight to obtain stable and accurate results, and interval prediction is used for quantifying uncertainty and improving model reliability. Nine metrics are proposed to evaluate the prediction performance in the last module. The contributions and innovations are summarized below:

(1) **A novel and advanced wind speed combined forecasting model (CFM) that includes data preprocessing and combined prediction and assessment is developed in this work.** Considering the uncertainty and fluctuation of the initial sequence, the prediction system given can overcome these shortcomings and achieve accurate and reliable prediction performance.

(2) **An advanced data preprocessing technique based on the decomposition and ensemble theory is intended to remove the noise in the initial sequence.** An advanced data preprocessing technique is chosen to reduce fluctuations and uncertainty to obtain a smooth sequence and improve the forecasting accuracy.

(3) **A new combined prediction strategy that includes the selection of several sub-models to minimize the combination error and weight optimization operators is proposed.** Four sub-models are used to predict the original sequence, and their prediction results are integrated by the multi-objective grey wolf optimizer (MOGWO) to gain more precise and stable predictive effect.

(4) **The Pareto optimality of the solutions from the combined system is theoretically proven.** The proof ensures that the optimal weight vector is obtained in the combined system through the leader selection mechanism and superior to those generated by the individual models, thus improving the prediction performance of the proposed system.

(5) **An integrated and detailed assessment system was built to assess the point prediction (PP) and interval prediction (IP) results of CFM.** In the assessment module, nine metrics and several sites are used in the experiment with intervals of 10-min, 20-min and 30-min. In addition, the multi-step prediction and rolling input steps further are investigated to further ensure the accuracy and reliability of CFM.

The overall structure of this paper is as follows. The approaches used in this study

are introduced in Section 2, the flow of CFM described in detail in Section 3, the experimental processes presented in Section 4, and the discussion and conclusion are given in Sections 5 and 6, respectively.

Table 1 Various studies on wind speed.

Method	Published year	Contribution	Disadvantages
NWP-K means-GRNN [8]	2016	NWP and ANNs are combined and the parameters of clustering are discussed in detail.	high requirements on computation and information. Lack of comparison with other ANNs models.
ARMA-AR [13]	2021	The parameter values of AR and ARMA models are discussed.	Not suitable for non-linear data.
VMD-phase-space reconstruction (PSR) - BPNN, ELM, ENN-multi objective multi verse optimization (MOMVO) [15]	2021	A two-stage preprocessing way VMD-PSR is to analyze noise. MOMOVO is applied to combine BPNN, ELM and ENN.	The change of key parameter values needs to discuss.
ICEEMDAN- ARIMA, BPNN, ENN, GRNN, ELM-Modified MODA [23]	2020	The modes mix is solved in EMD and EEMD. Modified MODA enhances the optimization of weight coefficient.	The predictive power of selected single models is not fully demonstrated.
Flexible ensemble patch transformation (EPT)-CEEMDAN-CNN [26]	2022	EPT-CEEMDAN enhances the detection of local patterns embedded. Multiple deep learning models are compared.	Robustness and precision are not well balanced.
VMD-GA-WNN [28]	2017	VMD and GA are used to eliminate data noise and optimize WNN.	The optimization effect of single objective is limited.
CEEMDAN-ELM, RBF, GRNN, BPNN-MOGWO [29]	2021	MOGWO is adopted to integrate ELM, RBF, GRNN, BPNN.	Parameter setting of optimization algorithm is not explained.
ICEEMDAN-SVM- whale optimization algorithm (WOA) [30]	2017	WOA is used to optimize the parameters of SVM.	Fewer contrast models and optimization effect of single objective is limited.
VMD-BPNN, random vector functional link network (RVFL), ANFIS, GRNN-multi objective salp swarm algorithm (MSSA) and support vector regression (SVR) [31]	2021	Using the hybrid algorithm MSSA-SVR assigns coefficients to single models and the Pareto optimal solution of the optimization algorithm is analyzed.	Model parameter values are not displayed.
CEEMDAN-ELM, GRNN, ARIMA, BPNN, ENN- MOGOA[32]	2019	MOGOA can optimize individual model weights.	Lack of discussion on parameter setting and comparison of other deep learning models.

2. Design of the combined forecasting model

The proposed prediction model in this study is mainly composed of data preprocessing technology, multiple single models, multi-objective weight optimization operator and uncertainty prediction, which is used to improve the prediction accuracy and quantify the uncertainty of prediction results. The corresponding algorithm and theoretical introduction are presented in this part.

2.1 Data denoising strategy

Before data denoising, we process the original data to prevent data leakage. Specifically, data leakage can cause the model to look accurate, but when applied to real life, the model can become very inaccurate, which is mainly divided into feature leakage and training data leakage. For feature leakage, features generally have a strong correlation with target variables. In this study, prevention of training data leakage is of greater concern to us. Before dividing the datasets, preprocessing the whole data set will lead to data leakage and make the information of the test set appear in the training set. Therefore, after dividing the datasets, we use decomposition technology for training set and test set respectively to prevent data leakage. The corresponding data denoising strategy is presented as follows.

The ICEEMDAN (ICE) is adopted to decompose the initial sequence, which is characterized by instability and fluctuation [30, 33]. ICE, which is based on CEEMDAN, can further reduce the noise and aliasing in IMF, which has a good decomposition ability. Suppose $\tilde{\boldsymbol{\kappa}}$ is the original sequence, and the operators $\overline{\overline{\boldsymbol{D}}}_j(\cdot)$, $\overline{\overline{\overline{\boldsymbol{L}}}}(\cdot)$, and $\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}$ are introduced. $\overline{\overline{\boldsymbol{D}}}_j(\cdot)$ generates the j-th mode decomposed from the initial time series, $\overline{\overline{\overline{\boldsymbol{L}}}}(\cdot)$ produces the local average of the original sequence $\tilde{\boldsymbol{\kappa}}$, and $\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}$ denotes white gaussian noise with μ=0 and unit variance σ^2. The specific processes in ICE are as follows:

The local mean of the i-th realizations $\tilde{\boldsymbol{\kappa}}^i(\boldsymbol{t}) = \tilde{\boldsymbol{\kappa}} + \chi_0 \cdot \overline{\overline{\boldsymbol{D}}}_1(\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}(\boldsymbol{t}))$ is first calculated to obtain the first residue $\boldsymbol{R}_1(\boldsymbol{t})$: $\boldsymbol{R}_1(\boldsymbol{t}) = \left\{\overline{\overline{\overline{\boldsymbol{L}}}}(\tilde{\boldsymbol{\kappa}}^i(\boldsymbol{t}))\right\}$, where χ_i is an operation item to remove noise and $\{\bullet\}$ is the tool of the averaging process. The first mode can then be computed as $\boldsymbol{IM}_1 = \tilde{\boldsymbol{\kappa}}(\boldsymbol{t}) - \boldsymbol{R}_1(\boldsymbol{t})$.

The second residue $\boldsymbol{R}_2$ can be computed as $\boldsymbol{R}_1(\boldsymbol{t}) + \chi_1 \overline{\overline{\boldsymbol{D}}}_2(\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}(\boldsymbol{t}))$, and the second mode $\boldsymbol{IM}_2$ is obtained:

$$\boldsymbol{IM}_2 = \boldsymbol{R}_1(\boldsymbol{t}) - \boldsymbol{R}_2(\boldsymbol{t}) = \boldsymbol{R}_1(\boldsymbol{t}) - \left\{\overline{\overline{\overline{\boldsymbol{L}}}}(\chi_1 \overline{\overline{\boldsymbol{D}}}_2(\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}(\boldsymbol{t})) + \boldsymbol{R}_1(\boldsymbol{t}))\right\} \tag{1}$$

For $\boldsymbol{j} = 3, 4, \ldots, \boldsymbol{n}$, the j-th residue is calculated as:

$$\boldsymbol{R}_j(\boldsymbol{t}) = \left\{\overline{\overline{\overline{\boldsymbol{L}}}}(\chi_{j-1} \overline{\overline{\boldsymbol{D}}}_j(\overline{\overline{\overline{\boldsymbol{\varpi}}}}^{(i)}(\boldsymbol{t})) + \boldsymbol{R}_{j-1}(\boldsymbol{t}))\right\} \tag{2}$$

$\boldsymbol{IM}_j = \boldsymbol{R}_{j-1}(\boldsymbol{t}) - \boldsymbol{R}_j(\boldsymbol{t})$ can also be obtained. After obtaining all the modes, the original sequenc $\tilde{\boldsymbol{\kappa}}$ can be reconstructed as $\boldsymbol{\kappa}' = \sum \boldsymbol{R}_j(\boldsymbol{t}) + \overline{\tilde{\boldsymbol{R}}}(\boldsymbol{t})$, where $\overline{\tilde{\boldsymbol{R}}}$ is the residual sequence.

2.2 Neural network models

LSTM (long short-term memory) is a special type of RNNs. The key components of LSTM are memory cells and gates. The forget gate determines the number of the

unit state $C(t-1)$ remaining at current moment $C(t)$ and is calculated as $\overline{\overline{F}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{b}_f+\{x_t,h_{t-1}\}*\bar{\omega}_f)$. The number of inputs $x(t)$ stored in the cell state $C(t)$ is determined in the calculation process of the input gate $\overline{\overline{I}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{b}_i+\{x_t,h_{t-1}\}*\bar{\omega}_i)$. Finally, the output gate controls the number of unit states $C(t)$ to be output to the present output value $h(t)$ and is given by $\overline{\overline{O}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{\omega}_o*\{h_{t-1},x_t\}+\bar{b}_o)$ where $\bar{b}$ and $\bar{\omega}$ are bias and weight, respectively. The calculation principle of GRU (gated recurrent unit) is very similar to LSTM. The difference is that GRU combines the forgetting gate and input gate in LSTM algorithm into update gate, so GRU consists of two gates, the update gate and the reset gate.

The hidden layer of CNN (convolutional neural network) consists of convolution C_L, pooling P_L and full connection F_L. The convolution process is represented as $C_{x,y}=\sum_i^{p*q}\bar{\omega}_i v_i$, which is the sum of the product of the kernel weight and the brightness of the corresponding element in the input image. The main goal of pooling is to reduce the feature space of the maps. The fully connected layer is greatly important in combining the extracted features to obtain the output.

TCN (temporal convolutional networks) is suitable for sequence model construction under causal constraints, that is, the output $\{\hat{y}_0,\hat{y}_1,\ldots,\hat{y}_t\}$ can only be predicted based on the past observation $\{x_0,x_1,\ldots,x_t\}$. Therefore, TCN can be designed as a nonlinear function with the mapping $f: X^T\rightarrow Y^T$. In addition, TCN also adds the dilated convolution and residual block to better extract historical information.

QRNN (quasi-recurrent neural networks) is the LSTM acceleration algorithm, including two components of convolution and pooling. In the convolution operation, the output of the input, forget and output gates can be expressed as $Z=\tanh(W_z*X)$, $F=\sigma(W_f*X)$ and $O=\sigma(W_O*X)$, where $\mathbf{X}$ is the input, W_z, W_f and W_O are the convolution filters. There are three ways the pooling process can reduce the number of features, named f-pooling, fo-pooling and ifo-pooling based on the number of gates used.

ANFIS (adaptive neuro-fuzzy inference system) is a neural network based on fuzzy reasoning. The first layer is a fuzzy layer. $\overline{\overline{O}}^1_{i=1,2}=\mu_{A_i}(x)$ and $\overline{\overline{O}}^1_{j=1,2}=\mu_{B_i}(y)$ are output functions where x and y are the inputs with the respective membership functions μ_{A_i} and μ_{B_i}. The second layer calculates the weights of each membership function according to the previous outputs. The third layer normalizes the weights as $\overline{\overline{O}}^3_i=\overline{\bar{\omega}}_i=\bar{\omega}_i/(\bar{\omega}_1+\bar{\omega}_2)$. The fourth layer provides the output of the rule inference $\overline{\overline{O}}^4_i=\overline{\bar{\omega}}_i\overline{\overline{F}}_i$. The last layer generates the sum of each output $\overline{\overline{O}}^5_i=\sum\overline{\bar{\omega}}_i\overline{\overline{F}}_i=\left(\sum\overline{\bar{\omega}}_i\overline{\overline{F}}_i\right)/\sum\bar{w}_i$.

ELM (extreme learning machine) is a kind of feedforward neural network that includes input, hidden and output layers. The thresholds b and input weights ω are randomly generated. The input and output are $\overrightarrow{X}=(x_i^1,x_i^2,\ldots x_i^n)\in\mathbb{S}^n$ and $\overrightarrow{T}=(t_i^1,t_i^2,\ldots t_i^m)\in\mathbb{S}^m$. The hidden layer output is $\overline{\overline{H}}^{\bar{\omega},b}_{i=1,\ldots,n}(x)=\overline{\overline{F}}(x^T\bar{\omega}_i+b_i)$,where $\overline{\overline{F}}(\bullet)$ is the activation function. The final output is $\overline{\overline{O}}^{\omega,b,\theta}_n(x)=\sum_{i=1}^n\overline{\overline{H}}_i(x)\bar{\theta}(i)$, where the output weights are $\bar{\theta}=(\bar{\theta}_1,\cdots,\bar{\theta}_n)$.

BPNN (back propagation neural network) consists of input, hidden and output layers, and the calculation is: $y_i = f_I\left(\mu_j + \sum_{m=t-n}^{t-1} \mu_{jm} y_m\right)(0 \le \mu_j, \mu_{jm} \le 1)$, $y_t = f_0\left(\lambda_0 + \sum_{j=1}^{I} \lambda_{0j} y_j\right)(0 \le \lambda_0, \lambda_{0j} \le 1)$, where y_m and y_j are the inputs of the input and hidden layers, y_t is the predicted value at time **t**. N and I are the number of nodes in the input and the hidden layers. In the hidden and output layers, μ_j and λ_0 are the threshold values, μ_{jm} and λ_{0j} are weights, f_I and f_0 are activation functions.

On the basis of BPNN structure, ENN (elman neural network) adds a continuation layer to the hidden layer as a delay operator to achieve the purpose of memory. The learning process is: $x(k) = f(w_1 x_c(k) + w_2(u(k-1)))$, $y(k) = g(w_3 x(k))$, where y is the m-dimensional output vector, x is the unit vector of n-dimensional middle layer, u is the r-dimensional input vector, $w_i(i = 1,2,3)$ is the connection weight of each layer, $g(\bullet)$ and $f(\bullet)$ are activation functions of output and middle layer neurons respectively.

GRNN (general regression neural network) is a kind of radial basis neural network, which is composed of input, pattern, summation and output layers. $p_i = \exp[-(X - X_i)^T (X - X_i)/2\sigma^2]$ $i = 1,2,\ldots,n$ is the neuron transfer function of the pattern layer, where X is the input variable and X_i is the learning sample of the i-th neuron. The output of neuron j can be calculated as $y_i = S_{Nj} / S_D$, where S_{Nj} and S_D are the arithmetic sum and weighted sum of neurons at the pattern layer.

2.3 Multi objective grey wolf optimizer

Mirjalili et al. first proposed the MOGWO algorithm, who were inspired by social leadership and hunting technique of grey wolves [29, 34]. The alpha wolf α_w is defined as the fittest solution, and the beta wolf β_w and delta wolf δ_w, the second and third best solutions, respectively. The other candidate approaches are the omega wolves ω_w. α_w, β_w, and δ_w play a key role in capturing prey, and ω_w follow α_w, β_w, and δ_w to catch prey.

Definition 1 Encircling process. During the hunting of prey, the grey wolf location is updated as $\overline{\overline{L}}_{wolf}^{t'=\alpha,\beta,\delta}(t+1) = \overline{\overline{L}}_{prey}^{t}(t) - H \otimes \triangle \overline{\overline{D}}^{t}$, where $\overline{\overline{L}}_{prey}^{t}(t)$ and $\overline{\overline{L}}_{wolf}^{t'=\alpha,\beta,\delta}$ represent the location of prey and grey wolf, respectively, and $\triangle \overline{\overline{D}}^{t=\alpha,\beta,\delta} = \left| C \otimes \overline{\overline{L}}_{prey}^{t}(t) - \overline{\overline{L}}_{wolf}^{t}(t) \right|$ is the distance between the grey wolf and its prey. C and H are coefficient vectors that are calculated as $C = 2 \otimes \gamma_2$ and $H = 2\varsigma \otimes \gamma_1 - \varsigma$, respectively, where ς is the convergence factor, which decreases linearly from 2 to 0 with the number of iterations, and Υ is a random vector $\Upsilon = [\gamma_1, \gamma_2] \in [0,1]$.

Definition 2 Hunting behavior. Suppose that α_w, β_w and δ_w have a better command of the potential location of the prey and their distance from the prey is: $\triangle \overline{\overline{D}}^{t=\alpha,\beta,\delta} = \left| C^{t} \otimes \overline{\overline{L}}_{prey}^{t} - \overline{\overline{L}}_{wolf} \right|$. The current location of the wolves is then obtained as $\overline{\overline{L}}_{wolf}^{t=\alpha,\beta,\delta} = \overline{\overline{L}}_{prey}^{t} - H^{t} \otimes \triangle \overline{\overline{D}}^{t}$, and the location of ω_w can be calculated as $\overline{\overline{L}}_{wolf}^{\omega} = \sum_{t=\alpha,\beta,\delta} \overline{\overline{L}}_{wolf}^{t} / 3$.

Definition 3 Exploitation. When $H \in [-1,1]$ is random vector, the wolves can move at will between themselves and the prey. This narrows down the estimated location of the prey provided by α_w, β_w and δ_w.

Definition 4 Searching prey. When $|H| \succ 1$, the wolf is forced to separate from its prey while $|H| \prec 1$ helps to converge towards the prey. Moreover, the random vector $C \in [0,2]$ facilitates the discovery of new solutions. $C \succ 1$ indicates that the location of the wolf has a great influence on the prey and $C \prec 1$ indicates that the location of the wolf has little effect on the prey.

Definition 5 Archive. The archive is a store tool that saves or retrieves non-dominated Pareto optimal solutions.

Definition 6 Leader selection mechanism. This selection mechanism adopts the roulette wheel approach of $P_i = c / N_i$ to choose the non-dominated pareto optimal solution where c is a fixed value and $c \succ 1$. N denotes the sum of Pareto optimal solutions obtained.

2.4 Multi objective problems (Mop)

In multi-objective problems, it is critical to find vector solutions. The concept of Pareto dominance is designed to facilitate the choice of the vector solutions. The Pareto optimal solution is defined as follows:

(1) $\vec{\varpi}^1$ is considered to dominate $\vec{\varpi}^2$, i.e., $\vec{\varpi}^1 \succ \vec{\varpi}^2$, if and only if $\vec{\varpi}^1$ and $\vec{\varpi}^2$ meet: $[\forall p \in \{1,2,\ldots t\}, \overline{\overline{F}}_t(\vec{\varpi}^1) \le \overline{\overline{F}}_t(\vec{\varpi}^2)] \wedge [\exists q \in \{1,2,\ldots t\}, \overline{\overline{F}}_q(\vec{\varpi}^1) \prec \overline{\overline{F}}_q(\vec{\varpi}^2)]$, where $\overline{\overline{F}}$ is a function and t denotes a vector of numbers.

(2) The solution $\vec{\varpi}^1$ is called the Pareto optimal solution if the conditional equality $/\exists \vec{\varpi}^i \in \mathbb{S}, \vec{\varpi}^i \prec \vec{\varpi}^1$ is satisfied, where $\mathbb{S}$ denotes solution space.

(3) The Pareto optimal solution set $\overline{\overline{\mathrm{P}}}^s$ includes all the Pareto optimal solutions and can be expressed as $\overline{\overline{\mathrm{P}}}^s = \{\vec{\varpi} \mid /\exists \vec{\varpi}^i \in \mathbb{S}, \vec{\varpi} \succ \vec{\varpi}^i\}$.

To improve the precision and stability of the combined system, the MOGWO objective function is defined as:

$$\min \begin{cases} \overline{\overline{F}}^1 = (1/Q)\sum_i^Q \left| (P^i - A^i)/A^i \right| \times 100\% \\ \overline{\overline{F}}^2 = std(P^i - A^i)\ i = 1,\ldots,Q \end{cases} \tag{3}$$

where $\overline{\overline{F}}$ is the objective function, P and A represent the forecasted and true values.

2.5 The Pareto solution proof

The objectives in MOPs may often conflict with one another. The solutions in such problems can usually be expressed as the Pareto solution (λ^*). For MOPs, the Pareto solution (λ^*) that satisfies $/\exists \vartheta \in \Omega_{s.t.} FIT(\lambda^*) \prec FIT(\vartheta^*)$ is adopted rather than the accurate solution. *Arc* is a tool for storing the non-dominated solutions (λ'); however, the storage capacity of *Arc* is limited and has the upper bound of $\bar{\Upsilon}$, that is, $\Upsilon \prec \bar{\Upsilon}$. Hence, in the process of updating *Arc*, it is necessary to compare the new solution λ'

and the present non-dominated solutions $\{\hat{\lambda}_1^p, \hat{\lambda}_2^p, \ldots, \hat{\lambda}_\gamma^p\}$ in ***Arc*** . When $\lambda' \succ \hat{\lambda}_i^p$ or $\lambda' \sim \hat{\lambda}_i^p$, $i \in [1, \gamma]$, λ' is added into ***Arc*** and $\hat{\lambda}_i^p$ is removed or ranked behind λ' .Once the ***Arc*** capacity has reached the upper bound $\gamma = \bar{\gamma}$, the most crowded segments will be removed with the deletion probability $Pd_i = \gamma_i / c$ ($c \succ 1$) . Then, the roulette-wheel technique is employed to compare the present solution with λ' with the probability $Pd_i = \gamma_i / c$ ($c \succ 1$) to determine the Pareto optimal solution λ' . The pseudocode is listed as **Table 2**.

Table 2 Pseudo code.

Algorithm: ***MOGWO***

Input:

$m_{tr}^{(0)} = (m^{(0)}(1), m^{(0)}(2), \ldots, m^{(0)}(a))$ –training data

$(m^{(0)}(a+1), m^{(0)}(a+2), \ldots, m^{(0)}(a+i))$ –testing data

Output:

$(\hat{n}^{(0)}(a+1), \hat{n}^{(0)}(a+2), \ldots, \hat{n}^{(0)}(a+i))$ –forecasting data

Parameters:

F_i – the fitness function of *i*-th wolf

$Iter_{max}$ – maximum iteration times.

t – present iteration times.

d –dimensions numbers.

n –wolves numbers

$\vec{L}_i$ – the position of *i*-th wolf

/*Set up the parameters of ***MOGWO***.*/

/* Random initialization of the n wolves $\vec{L}_i$ population (*i*=1,2…n).*/

FOR EACH i=1:n DO

/*Calculate the F_i using the process of ranking.*/

END FOR

/*Decide the best search agent $\vec{L}_{wolf}^{\iota=\alpha,\beta,\delta}$.*/

WHILE (t< $Iter_{max}$) DO

  FOR EACH i=1:n DO

    /*Update ς , H , C and choose a wolf randomly from the archive.*/

    $C = 2 \otimes \gamma_2$, $H = 2\varsigma \otimes \gamma_1 - \varsigma$

    /* In the archive choose the elite from by roulette wheel */

    IF($|H| \prec 1$) ***THEN***

      /* Update the location from the search proxy at present .*/

      $\triangle \vec{D}^{\iota=\alpha,\beta,\delta} = | C^{\iota} \otimes \vec{L}_{prey}^{\iota} - \vec{L}_{wolf} |$

    ELSE ($|H| \geq 1$) ***THEN***

      /*Update the location from the search proxy at present.*/

      $\vec{L}_{wolf}^{\iota=\alpha,\beta,\delta} = \vec{L}_{prey}^{\iota} - H^{\iota} \otimes \triangle \vec{D}^{\iota}$

    END IF

  END FOR

  /* Compute all wolves' objective values.*/

  /* Search the non-dominated solutions.*/

  /* Renew the file depending on the non-dominated solutions.*/

  IF this file is up to its limit ***DO***

    /* To contain new solutions, remove some solutions from the file and use roulette wheel with $P_i = N_i / c$ (c>**1**).*/

25 *END IF*
26 *IF* the file capacity exceeds boundary *DO*
27 /*Renew the boundary to hold the new ones.*/
28 $\overline{L}_{wolf}^{=\omega} = \sum_{\iota=\alpha,\beta,\delta} \overline{L}_{wolf}^{=\iota} / 3$
29 *END IF*
30 *T=t+1*
31 **END WHILE**
32 **RETURN** *L**

2.6 Uncertainty prediction

Interval prediction is proposed to quantify the uncertainty. According to maximum likelihood estimation (MLE), Weibull, Gamma, Rayleigh and Lognormal distribution are adopted to fit the data. Interval prediction results of CFM is testified and compared with basic models, the upper and lower bound is expressed as: $Up^{i-th} = Fr^{i-th} + Dis_{1-\alpha/2} \bullet \sigma^{*}_{Dis} / \sqrt{n}$, $Lo^{i-th} = Fr^{i-th} - Dis_{\alpha/2} \bullet \sigma^{*}_{Dis} / \sqrt{n}$, where $Dis_{1-\alpha/2}$ and $Dis_{\alpha/2}$ are the critical values of optimal distribution and α is the significance level (α= 0.05, 0.1 and 0.2 in this study).

3. Flow of the combined forecasting system

In this study, a strategy of combining interval prediction with point prediction is adopted. A smooth sequence is first obtained using the ICE data preprocessing technique. Then, some single neural models are applied to forecast a stable sequence and obtain the forecasted values $\overrightarrow{F} = (\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4)$. In addition, the optimization operator is used to provide the most reasonable weights $\overrightarrow{W} = (\omega_1, \omega_2, \omega_3, \omega_4)$ to combine the forecasted values in $\overrightarrow{F}$. Finally, confidence intervals of 80%, 90%, and 95% are set for the experiment. The main framework is illustrated in **Fig.1**.

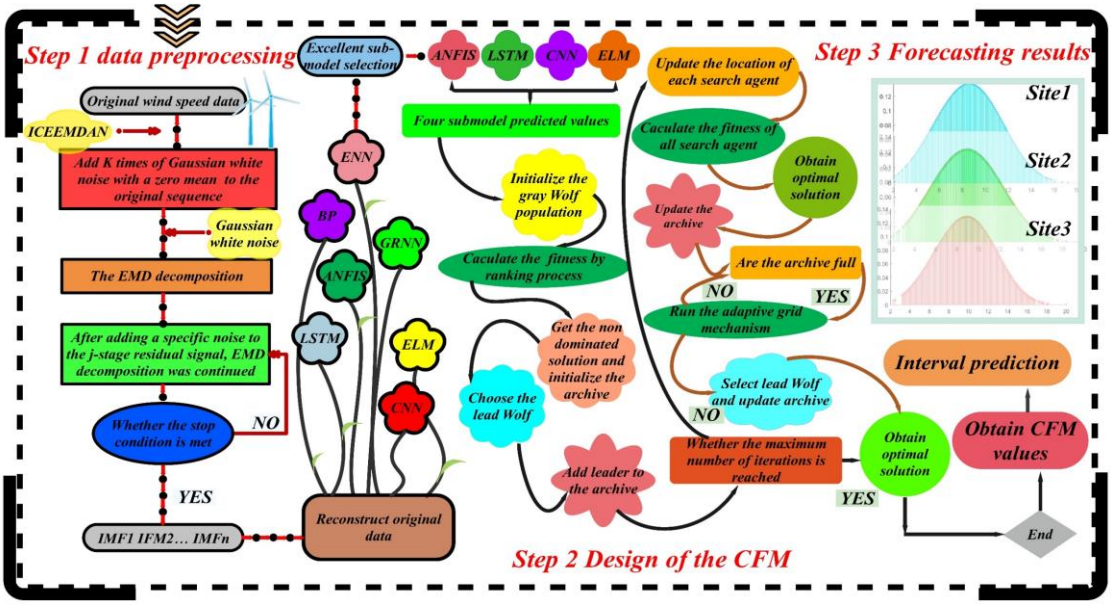

Fig.1 The main structure of this paper

3.1 Information of datasets

A wind power plant in Penglai (**37°48′N 120°45′E**), Shandong Province, China, is

used as the example. Seven sites are chosen to verify the performance of the proposed system, which include 10-min, 20-min and 30-min intervals. Besides, seven sites have the same sample size of 2220. The first 2000 samples are used for training, and the remaining 220 are used for testing. The details of the dataset are presented in **Table 3** and the features of the dataset are also visually depicted in **Fig. 2**.

Table 3 Datasets.

SET	Sample	Size	Indicators(m/s)			
			Maximun	Minimun	Mean	Std.
	Total samples	2220	17.2000	2.0000	8.5566	2.8012
Site1: 10-min	**Train-set**	2000	16.1000	2.0000	8.1810	2.5671
	Test-set	220	17.2000	5.1000	11.9707	2.5207
	Total samples	2220	18.1000	2.3000	9.6626	2.9661
Site2: 10-min	**Train-set**	2000	18.1000	2.3000	9.2664	2.7311
	Test-set	220	17.5000	6.6000	13.2641	2.5688
	Total samples	2220	18.4000	2.0000	9.0219	3.1068
Site3: 10-min	**Train-set**	2000	18.4000	2.0000	8.6368	2.8735
	Test-set	220	17.6000	5.4000	12.5236	2.9601
	Total samples	2220	16.7000	0.9000	6.7181	2.7697
Site4: 20-min	**Train-set**	2000	16.7000	0.9000	6.6773	2.7774
	Test-set	220	12.9000	2.6000	7.0886	2.6763
	Total samples	2220	18.2000	0.8000	7.8146	3.0395
Site5: 20-min	**Train-set**	2000	18.2000	0.8000	7.7815	3.0340
	Test-set	220	15.6000	2.1000	8.1159	3.0796
	Total samples	2220	18.2000	1.3000	8.4859	3.2925
Site6: 30-min	**Train-set**	2000	17.9000	1.3000	8.4406	3.2510
	Test-set	220	18.2000	2.8000	8.8977	3.6307
	Total samples	2220	17.5000	1.0000	7.5488	3.1403
Site7: 30-min	**Train-set**	2000	17.5000	1.0000	7.5082	3.0817
	Test-set	220	17.2000	2.7000	7.9182	3.6172

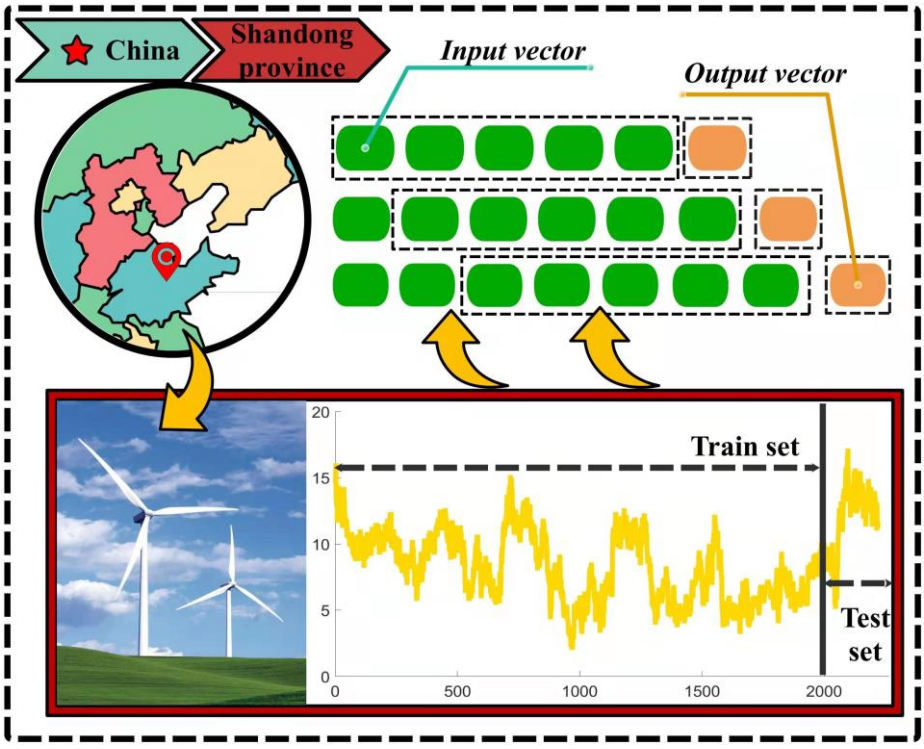

Fig.2 Wind speed data

3.2 Metrics

In the point prediction, six commonly used indicators are applied to evaluate model performance, namely **RMSE, MAE, MAPE, R^2, IA** and **PE**, which can better reflect the prediction accuracy. The smaller the index values of **RMSE, MAE** and

MAPE are, the higher the prediction accuracy of the models are, while $\mathbf{R^2}$ and **IA** are opposite. Three indicators named **FINAW, AWD,** and **FICP** are designed for interval forecasting. **FINAW** calculates the width of the interval. **AWD** is used to estimate the degree of deviation of the interval. **FICP** measures the extent to which ranges cover true values. The detailed information of metrics is presented in **Table 4.**

Table 4 Metrics.

Metric	Expression	Definition
RMSE	$\mathbf{RMSE} = \sqrt{(1/Q)\sum_{i=1}^{Q}(F_i - A_i)^2}$	Root mean square error
MAE	$\mathbf{MAE} = (1/Q)\sum_{i=1}^{Q}\lvert F_i - A_i\rvert$	Mean absolute error
MAPE	$\mathbf{MAPE} = (1/Q)\sum_{i=1}^{Q}\lvert(F_i - A_i)/A_i\rvert \times 100\%$	Absolute percentage error mean
$\mathbf{R^2}$	$\mathbf{R}^2 = 1 - \frac{\sum_{i=1}^{Q}(F_i - A_i)^2}{\sum_{i=1}^{Q}(F_i - \bar{A})^2}$	Coefficient of determination
IA	$\mathbf{IA} = 1 - \frac{\sum_{i=1}^{Q}(A_i - F_i)^2}{\sum_{i=1}^{Q}(\lvert A_i - \bar{A}\rvert + \lvert \bar{A} - F_i\rvert)^2}$	Consistency of the predicted results
PE	$\mathbf{PE}(i) = (F_i - A_i)/A_i \times 100\%$	The percentage of the predicted value in a specified error range
FICP	$\mathbf{FICP} = \left(\sum_{i=1}^{Q} C_i / Q\right) \times 100\%$	Coverage probability of the forecast interval
FINAW	$\mathbf{FINAW} = \sum_{i=1}^{Q}(U_i - L_i)/FR$	Normalized averaged width of the forecast interval
AWD	$\mathbf{AWD}_i = \begin{cases} (A_i - U_i)/(U_i - L_i) & A_i \succ U_i \\ 0 & A_i \in [L_i, U_i] \\ (L_i - A_i)/(U_i - L_i) & A_i \prec L_i \end{cases}$ $\mathbf{AWD} = \left(\sum_{i=1}^{Q} \mathbf{AWD}_i\right)/FR$	**AWD*i*** denotes the cumulative deviation of the forecast interval and **AWD** is the mean of cumulative deviation

Note: F_i, A_i represents the forecasting and actual value of *i*-th. In the point prediction evaluation index, ±5%, ±10% and ±15% error range are chosen to calculate the index PE; In the interval prediction evaluation index, if A_i belongs $[U_i, L_i]$, then c_i=1, otherwise c_i=0; FR means the range of predicted values.

3.3 Model parameter setting

The parameters for the neural network approaches are presented in **Table 5** and those for MOGWO and ICE in **Table 6**.

Table 5 Parameters of single models.

Model	Symbol	Meaning	Value	Setting reasons
BPNN	Tg	Training goal	0.00004	Preset
	Tf_{hid}	Activation function of hidden	tansig	Preset
ENN	E_{max}	The maximum epochs	1000	Trial- error method
	Tg	Training goal	0.00004	Preset
ELM	Tf	Activation function of hidden	sig	Preset
ANFIS	C_{ANf}	The number of the cluster	10	Preset
	E_{max}	The maximum epochs	1500	Trial- error method
GRNN	S_{GRnn}	Spread	1	Preset
LSTM	Lr	Training learning rate	0.005	Preset
	E_{max}	The maximum epochs	2000	Trial- error method
CNN	Lr	Training learning rate	0.005	Preset
	B_{min}	MiniBatchsize	16	Trial- error method
GRU	H_L	Numbers of hidden layers	40	Trial- error method

	E_{max}	The maximum epochs	2000	Preset
TCN	H_L	Numbers of hidden layers	50	Trial- error method
	E_{max}	The maximum epochs	500	Preset

Table 6 Parameters of ICE and MOGWO.

Model	Symbol	Meaning	Value	Setting reasons
ICE	$Iter_{max}$	The maximum number of iterations	1000	Trial- error method
	NR	Realization Number	100	Trial- error method
MOGWO	$Iter_{max}$	The maximum number of iterations	100	Trial- error method
	As	Archive size	100	Trial- error method
	P_s	Population size	200	Trial- error method
	α	Grid inflation	0.1	Preset
	g	Sum of grids per dimension	10	Preset
	β	Leader selection pressure	3	Preset

4. Experiment results

In this section, the results of the four comparative experiments and two validation tests to evaluate CFM are shown.

4.1 Selection of excellent sub models

Seven single predictive models named ANFIS [35, 36], LSTM [37, 38], CNN [39, 40], ELM [41-43], BPNN [44], GRNN [45, 46], and ENN [47] were chosen to forecast the original sequence in this study. To choose the most suitable models for improving the accuracy of CFM, the SEM strategy was developed [48]. The strategy is introduced below:

(1) The MAPE, MAE, and RMSE of every sub-model are calculated for one hundred iterations.

(2) Each index value is normalized as, for example,

$$\overline{MAE}_i = ((MAE_i - \min_{1\le i\le Q}(MAE)) / (\max_{1\le i\le Q}(MAE) - \min_{1\le i\le Q}(MAE)) \quad (4)$$

(3) Weights are assigned to each index value and the SEM value is calculated as $SEM_i = (1/3) * \overline{MAE}_i + (1/3) * \overline{MAPE}_i + (1/3) * \overline{RMSE}_i$.

The SEM values are listed in **Table 7**. A smaller SEM value indicates a higher model precision. Based on the results, ANFIS, LSTM, CNN, and ELM were chosen as the sub-models.

Table 7 Results of SEM.

SEM	Model						
	GRNN	**LSTM**	**ELM**	**BPNN**	**ANFIS**	**CNN**	**ENN**
Site1_SEM	0.3300	0.3021	0.3797	0.2458	0.2871	0.2590	0.2589
Site2_SEM	0.2814	0.1587	0.2240	0.3657	0.3125	0.2200	0.4129
Site3_SEM	0.3251	0.4011	0.2134	0.2948	0.2250	0.3100	0.3201
SEM*	0.3122	**0.2873**	**0.2724**	0.3021	**0.2749**	**0.2630**	0.3306

Note: * mean the average of SEM values of all sites.

4.2 Experiment Ⅰ: Comparison between CFM and basic models

To demonstrate the performance of CFM, seven individual models were used for comparison in Experiment Ⅰ. The results are as shown in **Table 8** and **Fig. 3**.

In **Table 8**, CFM exhibited the smallest error with MAPE, MAE and RMSE values of 2.8645%, 0.3217 and 0.4114 at site1. At site2, CFM exhibited the best performance with the MAPE value of 2.1843% compared to the other models, which had an average MAPE value of 4.8747%. The superior precision of CFM was thus verified. The IA and

R^2 can better reflect the agreement between the actual and the forecast data. At site1, CFM showed a better performance with the values of $R^2$and IA exceeding 97%, which was a 10% improvement in IA and R^2 over the best model, GRNN.

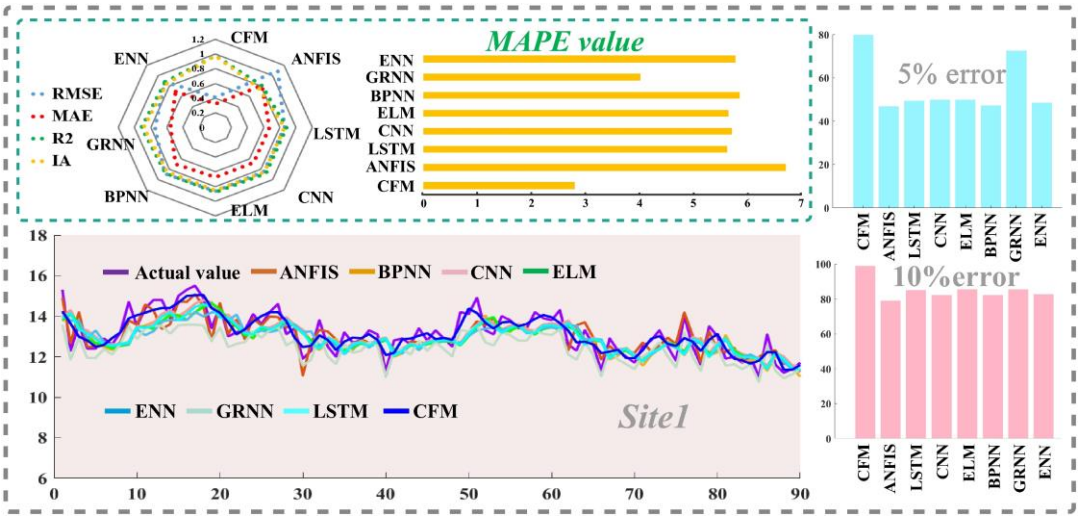

Fig.3 Results of CFM and basic models

4.3 Experiment Ⅱ: Comparison between CFM and single model based on ICE

Experiment Ⅱ was designed to contrast the accuracy of CFM and other models based on ICE. The results are presented in **Table 9** and **Fig.4**. The following conclusions can be drawn from the results:

It is obvious that CFM achieved the best prediction results at site1, with MAPE, MAE and RMSE values of 2.8645%, 0.3217, and 0.4114. For site2, the worst-performing model was ICE-ENN with MAPE exceeding 3.1044%. By contrast, the corresponding index value for CFM was 1% lower than that of ICE-ENN. For site3, CFM still exhibited the excellent performance with R^2 exceeding 97%. The prediction error PE of CFM in the range of 5%, 10%, and 15% was greater than 80%, 90% and 95%, respectively. Obviously, CFM had more excellent prediction accuracy.

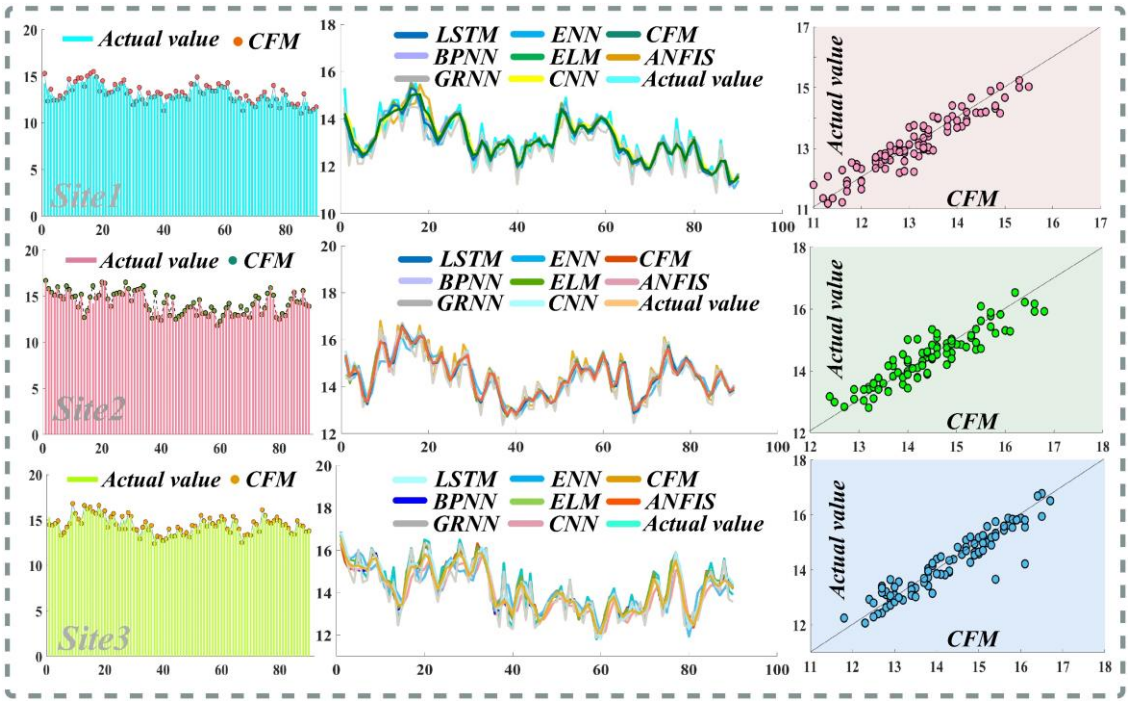

Fig.4 Results of CFM and denoising models

4.4 Experiment Ⅲ: Comparison between CFM and models with other denoising strategy

In this section, three denoising strategies, namely, CEE, EMD, and EEMD, were chosen to verify the predictive abilities of CFM, and the detailed results are presented in **Table 10**. The following conclusions can be drawn:

For site1, for the worst-performing EEMD-based model, the values of MAPE, MAE and RMSE were 4.6335%, 0.6592 and 0.6368, which far exceeded the corresponding values for CFM. For site2, $\boldsymbol{MAPE_{EMD}}$=3.3387% was obtained for the best classical model, which was based on EMD. In comparison, the smaller value of $\boldsymbol{MAPE_{CFM}}$=2.1843% was obtained for CFM. Finally, for site3, the prediction errors of the comparison models in the range of 5% were almost less than 80%. It is not difficult to find that CFM performed better with $\boldsymbol{PE}_{\boldsymbol{CFM}}^{5\%}$ exceeding 80%.

4.5 Experiment Ⅳ: Comparison between CFM and other optimization strategies

In this part, three optimization strategies called MOGOA, MOALO and MODA were designed to evaluate the predictive precision of CFM. The calculation results are listed in **Table 11**. It can be easily seen that the CFM had excellent performance. For example, for site1, the $\boldsymbol{MAPE_{CFM}}$=2.8645% of CFM was 0.2%, 0.02% and 0.05% lower than those of MOGOA, MOALO, and MODA, respectively. In addition, only the R^2 of CFM exceeded 95%, while its IA exceeded 97%. It can be found from the above evaluation index values that CFM has better performance in forecasting.

Table 8 The performances of CFM and basic models.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	ANFIS	6.8914	1.0924	0.8126	0.8113	0.7627	46.82	79.09	89.55
	LSTM	5.7710	0.8551	0.6705	0.8844	0.8553	49.55	85.00	95.00
	CNN	5.8641	0.8492	0.6731	0.8860	0.8576	50.00	82.27	95.45
	ELM	5.7888	0.8667	0.6737	0.8812	0.8512	50.00	85.45	96.36
	BPNN	6.0018	0.8832	0.6972	0.8767	0.8455	47.27	82.27	95.00
	GRNN	4.1138	0.7486	0.5544	0.9114	0.8879	72.73	85.45	99.09
	ENN	5.9280	0.8843	0.6943	0.8764	0.8449	48.64	82.73	96.36
	TCN	5.5340	0.8412	0.6571	0.8920	0.8588	50.50	86.67	96.69
	GRU	5.9284	0.8800	0.6877	0.8776	0.8867	49.55	80.91	95.45
	QRNN	5.4369	0.8391	0.6489	0.9012	0.8955	52.33	86.00	96.89
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	ANFIS	4.9404	0.8776	0.6535	0.8827	0.8550	62.73	88.64	96.82
	LSTM	4.9317	0.8474	0.6516	0.8907	0.8647	59.55	88.18	97.27
	CNN	5.0023	0.8375	0.6506	0.8932	0.8682	58.64	88.18	96.36
	ELM	4.8449	0.8369	0.6365	0.8934	0.8681	61.82	87.73	97.27
	BPNN	4.8209	0.8350	0.6341	0.8939	0.8690	61.36	88.18	95.91
	GRNN	4.1655	0.6751	0.5865	0.9306	0.9134	74.55	83.24	94.32
	ENN	5.4175	0.9370	0.7183	0.8663	0.8345	57.27	85.91	96.36
	TCN	4.7569	0.8250	0.6477	0.8985	0.8799	60.60	89.71	97.92
	GRU	4.9899	0.8448	0.6504	0.8914	0.8990	58.64	88.18	96.82
	QRNN	4.9210	0.8409	0.6450	0.8920	0.8644	59.82	87.65	96.34
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	ANFIS	7.0114	1.0389	0.8457	0.8762	0.7934	38.64	75.00	93.18
	LSTM	6.6676	0.9739	0.8076	0.8913	0.8181	43.64	78.64	96.36
	CNN	6.6651	0.9511	0.7841	0.8963	0.8275	42.73	80.91	93.64
	ELM	6.5754	0.9697	0.7944	0.8922	0.8198	41.82	78.64	95.91
	BPNN	6.4957	0.9626	0.7789	0.8938	0.8227	43.18	80.00	94.55
	GRNN	3.2736	0.5712	0.4483	0.9626	0.9370	65.00	74.25	93.24
	ENN	6.4089	0.9378	0.7616	0.8992	0.8319	45.45	81.82	95.91
	TCN	6.7258	0.9841	0.8130	0.8842	0.8032	42.30	76.82	95.20
	GRU	6.5127	0.9622	0.7770	0.8939	0.8778	44.55	82.27	94.09
	QRNN	6.5433	0.9678	0.7820	0.8912	0.8461	44.32	81.00	94.15

Note: The table lists the predicted results of CFM and basic models. The blacked part represents the evaluation index value of CFM. The equation of metrics are defined as: $\mathbf{PE}(i) = (F_i - A_i)/A_i \times 100\%$, $\mathbf{MAE} = (1/Q)\sum_{i=1}^{Q}|A_i - F_i|$, $\mathbf{MAPE} = (1/Q)\sum_{i=1}^{Q}|(A_i - F_i)/A_i| \times 100\%$, $\mathbf{R^2} = -\sum_{i=1}^{Q}(A_i - F_i)^2 / \sum_{i=1}^{Q}(F_i - \overline{A})^2 + 1$, $\mathbf{RMSE} = ((1/Q)\sum_{i=1}^{Q}(A_i - F_i)^2)^{\wedge(1/2)}$, $\mathbf{IA} = -\sum_{i=1}^{Q}(A_i - F_i)^2 / \sum_{i=1}^{Q}(|A_i + \overline{A}| + |F_i - \overline{A}|)^2 + 1$. In addition, GRU, TCN and QRNN are introduced in experiments Ⅰ and Ⅱ to make a more complete comparison.

Table 9 The performances of CFM and the denoising models.

SET	Model	MAPE	RMSE	MAE	R^2	IA	PE ±5%	PE ±10%	PE ±15%
	CFM	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	ICE-ANFIS	3.3003	0.5527	0.4017	0.9517	0.9400	77.73	96.82	99.55
	ICE- LSTM	2.8753	0.4367	0.3409	0.9698	0.9625	83.18	95.09	99.00
	ICE-CNN	3.6981	0.5433	0.4330	0.9533	0.9420	73.18	96.82	98.32
	ICE-ELM	2.8631	0.4293	0.3375	0.9709	0.9638	85.45	95.02	97.20
Site1: 10-min	**ICE-BPNN**	2.9085	0.4329	0.3423	0.9704	0.9632	85.00	96.09	99.00
	ICE-GRNN	2.9884	0.4068	0.3790	0.9738	0.9672	95.91	99.09	99.55
	ICE-ENN	3.5534	0.5371	0.4172	0.9544	0.9432	73.18	97.73	98.50
	ICE-TCN	2.8910	0.4340	0.3545	0.9621	0.9654	84.20	97.23	98.80
	ICE-GRU	3.6746	0.5441	0.4301	0.9532	0.9571	70.91	96.36	98.20
	ICE-QRNN	2.8741	0.4220	0.3398	0.9717	0.9638	84.89	97.89	99.00
	CFM	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	ICE-ANFIS	2.2827	0.3748	0.2978	0.9786	0.9737	90.00	93.40	98.30
	ICE- LSTM	2.3995	0.3989	0.3153	0.9758	0.9702	88.64	90.15	92.31
	ICE-CNN	3.0983	0.5160	0.4076	0.9595	0.9501	78.18	97.09	99.00
	ICE-ELM	2.2209	0.3599	0.2901	0.9303	0.9757	92.73	95.60	97.43
Site2: 10-min	**ICE-BPNN**	2.2658	0.3714	0.2954	0.9590	0.9742	79.09	98.64	99.55
	ICE-GRNN	1.0858	0.2071	0.1529	0.9435	0.9919	89.55	94.55	96.60
	ICE-ENN	3.1044	0.5411	0.4099	0.9554	0.9451	79.09	96.64	98.55
	ICE-TCN	2.3876	0.3847	0.3119	0.9722	0.9690	89.74	95.83	98.00
	ICE-GRU	3.1606	0.5159	0.4108	0.9595	0.9633	80.45	97.27	98.50
	ICE-QRNN	2.2531	0.3687	0.2937	0.9723	0.9714	91.52	97.03	98.11
	CFM	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	ICE-ANFIS	3.1424	0.4798	0.3782	0.9736	0.9561	81.36	97.73	98.50
	ICE- LSTM	3.3935	0.5259	0.4132	0.9683	0.9472	79.09	97.27	99.00
	ICE-CNN	4.6616	0.7052	0.5686	0.9430	0.9046	59.09	93.64	98.64
	ICE-ELM	3.0488	0.4652	0.3629	0.9752	0.9588	82.73	97.73	98.60
Site3: 10-min	**ICE-BPNN**	3.0550	0.4682	0.3647	0.9749	0.9582	85.45	96.09	97.25
	ICE-GRNN	2.9030	0.2869	0.2178	0.9709	0.9542	83.45	97.09	98.44
	ICE-ENN	4.0672	0.6429	0.4986	0.9526	0.9210	67.73	94.55	98.09
	ICE-TCN	3.4184	0.5315	0.4274	0.9630	0.9426	80.05	96.30	98.17
	ICE-GRU	4.1967	0.6376	0.5032	0.9534	0.9469	63.64	91.82	99.09
	ICE-QRNN	3.0251	0.4591	0.3563	0.9680	0.9536	82.29	97.81	99.00

Table 10 The forecasting results of CFM and other denoising strategy models.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	CEE	4.6156	0.5576	0.6120	0.7671	0.8889	70.00	92.22	94.30
	EEMD	4.6335	0.6368	0.6562	0.7853	0.8294	76.67	85.56	92.28
	EMD	3.7898	0.5845	0.4958	0.7826	0.9341	78.89	87.45	91.32
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	CEE	4.4341	0.7932	0.6434	0.9440	0.8842	72.22	94.44	99.09
	EEMD	4.5079	0.8711	0.6603	0.8088	0.8597	83.33	96.67	97.78
	EMD	3.3387	0.5709	0.4832	0.8602	0.9401	78.89	94.33	96.22
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	CEE	5.6386	0.9667	0.8056	0.9309	0.9276	52.22	94.44	97.42
	EEMD	4.3859	0.5918	0.4114	0.8848	0.9100	86.01	87.78	96.67
	EMD	2.8755	0.5203	0.4102	0.8148	0.9500	84.44	98.89	99.01

Table 11 The forecasting results of CFM and combined models using other algorithms.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	MOALO	3.0294	0.4986	0.3570	0.9214	0.9429	82.78	97.44	98.12
	MODA	2.8897	0.4387	0.3570	0.9414	0.9629	83.78	94.55	98.82
	MOGOA	2.9192	0.4544	0.3781	0.9083	0.9603	81.56	93.22	97.56
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	MOALO	2.1900	0.4074	0.3187	0.9435	0.9725	92.22	93.50	98.12
	MODA	2.1948	0.4083	0.3102	0.9428	0.9725	84.32	94.31	97.66
	MOGOA	2.2048	0.3999	0.3285	0.9407	0.9717	85.56	92.31	96.45
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	MOALO	2.9797	0.5791	0.3655	0.9430	0.9577	87.78	97.89	99.00
	MODA	2.9075	0.4800	0.3700	0.9424	0.9575	82.43	94.57	98.75
	MOGOA	2.9708	0.4894	0.3804	0.9362	0.9558	84.44	92.34	98.88

4.6 Uncertainty forecasting

An interval estimate is based on a point estimate and can increase the prediction reliability and certainty [31]. In this subsection, four distributions were chosen to fit the sequence based on maximum likelihood estimation (MLE). They are Weibull, Gamma, Rayleigh and Lognormal distribution, which are shown in **Table 12**. After analysis, the Weibull distribution was adopted for the three sites, and the detailed parameter settings are presented in **Table 13**. Then, three indicators named FINAW, AWD, and FICP were designed for interval forecasting. The detailed results are depicted in **Fig.5** and **Table 14**.

The indicators were evaluated for the probabilities P_1=95%, P_2=90%, and P_3=80%. These probabilities can be expressed as $(1-\alpha)\times 100\%$. It is clear that the prediction accuracy at the significance level of α=0.05 was superior to the accuracies at α=0.1 and α=0.2. At the significance level of α=0.05, the FICP for CFM at every site exceeded 85%. In addition, at α=0.1, the FICP for CFM was between 80% to 85%. The similar values of AWD and FINAW at different significance levels implies that most of the actual values fell within the predicted ranges and that CFM achieved excellent performance.

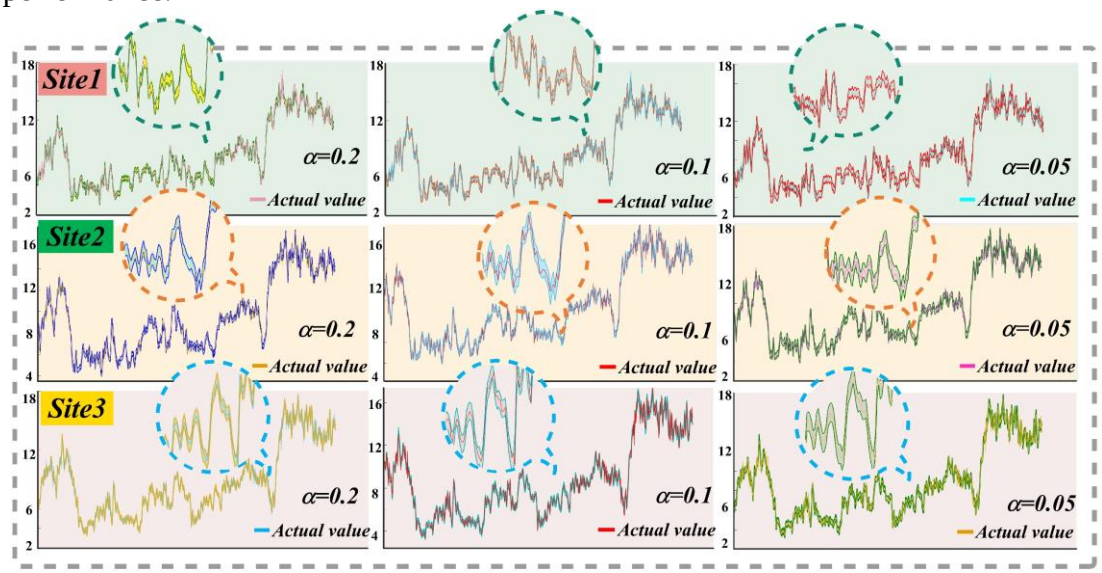

Fig.5 The interval forecasting results

5. Discussion

In this section, the results of improvement ratio, hyperparametric analysis, and the exploration of input and multi-step forecasting, operation time, comparative analysis and practical applicability are discussed to further analyze the above experimental results.

5.1 Improvement ratio

The aim of the improvement indicator is to quantify the improvement in the predictive precision of CFM. $\mathbf{IR}_{\mathbf{MAPE}}$, $\mathbf{IR}_{\mathbf{MAE}}$ and $\mathbf{IR}_{\mathbf{RMSE}}$ are designed to represent the improvement ratio. The details are shown in **Table 15**. For example, the models based on ICE for site1 had the values of $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ANFIS}$ =9.2670%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{LSTM}$ =4.1438%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{CNN}$ =19.0262%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ELM}$ =4.5895%, which shows that the advantages of each model were

combined in CFM. For the different denoising methods, it is clear from the average value of 34% for $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ave}$ in these models that ICE played a key role in improving the forecasting accuracy. The various optimization strategies achieved an average value of 7.3309% for $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ave}$. This implies that MOGWO could improve forecasting ability. There is no doubt that CFM was superior to the single models and the combined strategy achieved excellent prediction results.

5.2 Hyperparametric analysis

To investigate the influence of key parameter changes on the proposed model, only one key parameter was changed each time while the other parameters remained the same in this section. We changed the values of five parameters in the ICE and MOGWO algorithms. For ICE, the maximum iteration parameters values were [500, 3000] and the number realizations were [50, 100, 300]. For MOGWO, the archive size parameters values were [200,300,500], the maximum iteration parameter values were [50,150, 200], and the population sizes were [50,70,120]. The calculation results are shown in **Table 16**. With the change of parameter values, the precision of the model also changed. For example, at site1, the number realizations were set to 50,100 and 300. When 100 was selected, the proposed model achieved good MAPE values with 2.8645%, 2.1843% and 2.8727% in three sets. In general, it is not difficult to find that the parameter values selected in this paper perform better.

5.3 Influence of input step and multi-step prediction

This section adds sites 4-7 to better discuss the performance of the proposed model. Specifically, sites 4-5 use wind speed data at 20-minute intervals and sites 6-7 apply wind speed data at 30-minute intervals. The sample size and experimental mechanism are consistent with the above. Considering the neatness and symmetry of the table content, in sections 5.3.1 and 5.3.2, site1 (10-min), site2 (10-min), site4 (20-min), site5 (20-min), site6 (30-min) and site7 (30-min) are used to test the performance of the proposed model.

5.3.1 Influence of input step

To analyze the influence of different input steps on prediction more accurately, the input steps $\boldsymbol{\lambda}$ were adjusted to 3, 4 and 6, and a comparative test was performed to further explore the performance of the proposed forecasting system. From **Table 17**, it is easy to see that for the 10-min sites, when the input step $\boldsymbol{\lambda}$ was 3, the prediction errors of $\mathbf{MAPE}_{\boldsymbol{\lambda}=3}^{site1}$ = 3.0833% and $\mathbf{MAPE}_{\boldsymbol{\lambda}=3}^{site2}$ = 2.4134% were larger at those of other steps. For the 20-min sites, $\boldsymbol{PE}_{\lambda=3}^{5\%}$ and $\boldsymbol{PE}_{\lambda=4}^{5\%}$ are less than 70%, which implies that better forecasting results were obtained by CFM compared to $\boldsymbol{PE}_{\lambda=5}^{5\%}$=69.33% at site4 and $\boldsymbol{PE}_{\lambda=5}^{5\%}$=72.22% at site5. The prediction accuracy for the 30-min sites was obviously inferior to those of the previously mentioned sites; however, the input step 5 still performed well. From the above analysis, it can be concluded that as the time interval increased, the prediction accuracy decreased continuously, but the best performance was still obtained when the input step was 5.

5.3.2 Influence of multi-step prediction

The multistep forecasting ability of the proposed system is explored in the subsection. The details of the two-and three-step predictions are presented in **Table 18**.

For site1, which had 10-min intervals, the smallest errors of $\mathbf{MAPE}_{CFM}^{step\,2}$ =4.4556%, $\mathbf{MAE}_{CFM}^{step\,2}$ =0.5848 were achieved by CFM compared to the other models, which had the average errors with MAPE and MAE values of 5.0829% and 0.7960 for two-step predictions. In addition, the accuracy of the three-step prediction was obviously lower than that of the two-step prediction, but the CFM still achieved the good result with $\boldsymbol{R}_{CFM}^{2}$ greater than 90%.

The accuracy decreased with longer interval times. For example, for the 20-min predictions by CFM, the MAPE values in two-step was 6.5225% at site4, which exceeded the values for the 10-min sites. For the 30-min predictions, the prediction error $\mathbf{MAPE}_{CFM}^{step\,2}$ increased to 9% at site6. These results indicate that the prediction system may be more suitable for short-term wind speed prediction.

5.4 Operation time

Table 19 shows the mean operation time of all models used in this paper. It is clear that the proposed system had the longest calculation time, at 264.8361s. For basic models, the running time of BPNN was 0.4111, which was the least time in all models. The computation time of models based on ICE denoising algorithm was also short, which was 57.9170s, 19.2999s, 65.4855s, 36.4029s, 20.4532s, 18.9124s and 22.8657s. In addition, when applying other noise reduction algorithms or optimization algorithms, the time spent on combined models increased significantly, which was in the 100s to 200s range. Although the time consumed of the proposed model was longer than other models, it had superior predictive power. Its time was within the acceptable range and did not affect its practical application greatly.

Table 19 Computing time.

Model	Computation time(s)	Model	Computation time(s)
CFM	264.8361	**MOALO**	197.8581
ICE-ANFIS	57.9170	**MODA**	207.2554
ICE-ELM	19.2999	**MOGOA**	198.3225
ICE-LSTM	65.4855	**BPNN**	0.4111
ICE-CNN	36.4029	**ENN**	1.4821
ICE-GRNN	20.4532	**ELM**	3.2169
ICE-BPNN	18.9124	**LSTM**	38.3067
ICE-ENN	22.8657	**CNN**	18.7405
EMD	137.0275	**ANFIS**	40.2649
EEMD	110.2881	**GRNN**	10.1546
CEE	203.0657		

5.5 Comparative analysis

Through the experiments above, the validity of the proposed model is proved. However, the results based only on the experimental mechanism in this paper may not be convincing. Therefore, in order to be fair, this section compares the proposed model with other similar studies. Specifically, Niu's model [32], Shao's model [29]and Liu's model [23]are used for comparative study. Niu et al. designed the combined model based on data preprocessing and optimization algorithm for wind speed forecasting, which obtained high accuracy with the smallest MAPE value of 2.89%. However, Niu et al. ignored the importance of uncertain predictions and the influence of parameters on model performance was not considered. Shao et al. used the same MOGWO algorithm and proposed a combined model based on decomposition-integration, but also lacked the exploration of uncertainty prediction and the influence of model

parameters. Liu et al. not only applied the improved optimization algorithm to point prediction and interval prediction of wind speed, but also considered the influence of different seasons on wind speed prediction. However, Liu's model lacked comparison with other literatures, and the results of parameter adjustment were not fully demonstrated. **Table 20** presents the comparison results of this study with other studies.

There is no unified comparison standard for model accuracy among different studies [49]. Therefore, to verify the accuracy of the proposed model, the evaluation index and datasets in this paper are applied in the models of Niu, Shao and Liu. Meanwhile, parameter settings are basically consistent with those in the original literature. Although the error between these models was not large, it was enough to show that the proposed model is superior to the models from comparative studies. For example, the MAPE value of the proposed model was 2.8645%, while Niu, Shao and Liu had values of 3.4181%, 4.0200% and 3.1835% respectively. The proposed model also had outstanding advantages in other indicators. Therefore, it can be reasonably concluded that the proposed model has excellent predictive ability.

5.6 Practical applicability

Efficient and timely wind speed prediction plays an important role in wind power generation system, which can not only meet people's demand for electricity, but also maximize economic benefits. Meanwhile, the proposed model in this paper can also be applied to other fields:

(a) Air quality forecasting. Deep learning technology has been widely used in air monitoring, this proposed prediction system based on deep learning methods can be applied to the monitoring of air quality, for example, to timely predict air pollutant concentrations and to provide reasonable suggestions for travelling and improving the environment.

(b) Traffic forecasting. In recent years, the prediction of traffic flow has become the focus of attention. This proposed system can sensitively capture the change of traffic flow, so as to reduce traffic congestion, relieve people's travel pressure and provide support for the development of intelligent transportation.

(c) Financial market forecasting. This system can also be used in the prediction of future trends in stocks, funds, etc, timely providing technical support for discovery of potential financial risks.

6. Conclusion and future work

Wind energy has received increasing attention as a renewable energy source to address the shortages in the energy market. Nevertheless, the fluctuation and instability of wind speed still present difficulties for wind speed forecasting. An advanced wind speed prediction system was developed in this study. The ICE denoising strategy was first used to eliminate noise to obtain stationary sequences based on the de-composition and ensemble theory. Several single models were then adopted to predict the processed data. The optimal weights and final forecasting results were obtained using the MOGWO operator. The Pareto optimality of the MOGWO solutions was theoretically proven to ensure that the optimal weight vector can be obtained in the combined system. The point prediction results show that the MAPE values of the proposed model were 2.8645%, 2.1843% and 2.8727% respectively. Besides, for uncertainty forecasting, the FICP values of the proposed model were 85.1697, 89.5410 and 88.0111 respectively at the significant level α = 0.05. The AWD values were 0.0559, 0.0400 and 0.0361 and the FINAW values were 0.0478, 0.0404 and 0.0390. It is clear that the proposed system

exhibits remarkable accuracy and stability performance. This system can hence provide accurate and real-time wind power information and contribute to wind power generation.

There are still some areas that need to be improved in future studies, for example: (1) In order to further improve the prediction accuracy, the influence of temperature, wind direction and other variables on wind speed prediction needs to be further discussed. (2) More advanced deep learning algorithms can be applied to future research to improve the accuracy of developed system. (3) Using more efficient preprocessing methods deals with unstable, random wind speed data. In addition, the effects of different preprocessing methods need to be further explored.

Table 12 Data distribution fitting.

Distribution	Site1		Site2		Site3	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Weibull	**0.9892**	**0.0299**	**0.9932**	**0.0239**	**0.9971**	**0.0156**
Gamma	0.9866	0.0334	0.9914	0.0267	0.9932	0.0238
Rayleigh	0.8903	0.0955	0.8579	0.1087	0.9066	0.0881
Lognormal	0.9827	0.0380	0.9872	0.0326	0.9857	0.0345

Table 13 Distribution parameters for three sites.

SET	Distribution	Parameter	
		λ	k
Site1: 10-min	**Weibull**	9.5454	3.3830
Site2: 10-min	**Weibull**	10.7378	3.6000
Site3: 10-min	**Weibull**	10.0906	3.1837

Note: λ and k are scale and shape parameter.

Table 14 Interval forecasting results.

Model	Expectation probability	Site1: 10-min			Site2: 10-min			Site3: 10-min		
		FICP	AWD	FINAW	FICP	AWD	FINAW	FICP	AWD	FINAW
CFM	**95%**	85.1697	0.0559	0.0478	89.5410	0.0400	0.0404	88.0111	0.0361	0.0390
	90%	83.8456	0.0646	0.0364	84.6134	0.0519	0.0339	81.1961	0.0641	0.0327
	80%	80.3296	0.0796	0.0305	81.0431	0.0659	0.0264	78.2684	0.0965	0.0255

Note: This table lists the interval forecasting results of CFM, which can further show the accuracy of CFM. The formula of **FICP, FINAW** and **AWD** are: $\mathbf{FICP} = \left(\sum_{i=1}^{Q} C_i / M\right) \times 100\%$, $\mathbf{FINAW} = \sum_{i=1}^{Q} (U_i - L_i) / FR$, if $A_i \succ U_i$, then $\mathbf{AWD}_i = (A_i - U_i) / (U_i - L_i)$, if $A_i \in [L_i, U_i]$, then $\mathbf{AWD}_i = 0$, if $A_i \prec L_i$, then $\mathbf{AWD}_i = (L_i - A_i) / (U_i - L_i)$. Thus, $\mathbf{AWD} = \left(\sum_{i=1}^{Q} \mathbf{AWD}_i\right) / FR$, FR represents the range of predicted values.

Table 15 The improvement percentages of CFM and all other models.

Model	Site1: 10-min			Site2:10-min			Site3: 10-min		
	IR_{MAPE}	IR_{MAE}	IR_{RMSE}	IR_{MAPE}	IR_{MAE}	IR_{RMSE}	IR_{MAPE}	IR_{MAE}	IR_{RMSE}
ICE-ANFIS	9.2670	2.4780	12.9021	4.3124	6.1470	6.4870	8.5813	8.3642	8.3569
ICE-LSTM	4.1438	14.8928	10.2339	8.9694	0.2655	0.0573	15.3468	0.8169	1.1439
ICE-CNN	19.0262	9.5438	11.4006	29.4999	22.4435	22.6521	38.3755	27.9343	26.2731
ICE-ELM	4.5895	16.0755	12.1302	1.6459	8.9708	10.8772	5.7750	12.9224	11.7576
ICE-BPNN	2.9585	14.4446	11.1923	3.5967	6.9986	7.4574	5.9677	12.3657	11.0349
ICE-GRNN	43.3893	40.3992	18.3251	11.1729	16.7673	32.7546	39.3307	28.1278	11.1886
ICE-ENN	15.7277	6.1089	10.3648	29.6390	22.8787	26.2376	29.3698	17.8166	19.1285
ANFIS	56.5473	51.7972	55.9334	55.7865	51.6294	54.5255	59.0283	51.5459	49.9590
LSTM	48.1109	41.5767	43.7026	55.7087	51.4914	52.9036	56.9152	49.2542	46.6145
CNN	48.9351	41.8068	43.3142	56.3344	51.4124	52.3438	56.8993	47.7365	45.3357
ELM	48.2710	41.8598	44.4566	54.9156	50.3353	52.3116	56.3115	48.4144	46.3876
BPNN	50.1068	43.8167	45.4931	54.6910	50.1487	52.2037	55.7754	47.3885	45.9918
GRNN	27.2087	29.3473	35.6913	47.5616	46.1051	40.8872	12.2467	8.5824	8.9760
ENN	49.4851	43.5831	45.5588	59.681	55.9942	57.4058	55.1765	46.1910	44.5643
CEE	35.1219	35.9979	36.4612	50.7386	50.8682	49.6846	49.0531	49.1288	46.2175
EEMD	46.8444	48.2038	48.6125	51.5454	52.1242	54.1845	55.0851	55.0361	52.3802
EMD	20.9860	20.9976	17.6387	34.5769	34.5880	30.0906	12.0987	15.1083	10.0810
MOALO	9.7114	9.7189	9.7564	2.0716	2.4013	3.0105	11.5294	12.1131	8.5268
MODA	9.7014	9.7088	9.7415	1.8440	1.8924	2.7919	10.1714	10.7422	8.3018
MOGOA	2.5798	3.5893	5.9401	2.7982	2.4714	2.0948	7.5586	7.7358	6.2351

Note: This indicator aims to verify the improvement of predictive precision of the CFM, the calculation formulas are: $IR_{MAPE} = (MAPE_{other} - MAPE_{CFM}) / MAPE_{other} * 100\%$, $IR_{MAE} = (MAE_{other} - MAE_{CFM}) / MAE_{other} * 100\%$ and $IR_{RMSE} = (RMSE_{other} - RMSE_{CFM}) / RMSE_{other} * 100\%$.

Table 16 Prediction results of different parameter values.

ICEMMDAN		MOGWO			Site1: 10-min			Site2: 10-min			Site3: 10-min		
Realization Number	**Iteration Number**	**Archive Size**	**Iteration Number**	**Population Number**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**
50	1000	100	100	200	2.9678	0.4785	0.3871	2.3445	0.4406	0.3419	3.4458	0.5905	0.4871
100*	**1000**	**100**	**100**	**200**	**2.8645**	**0.4114**	**0.3217**	**2.1843**	**0.3991**	**0.3111**	**2.8727**	**0.4299**	**0.3598**
300	1000	100	100	200	3.1287	0.4938	0.4068	2.2759	0.4194	0.3270	2.8058	0.5124	0.4011
100	500	100	100	200	3.0425	0.4854	0.3982	2.2270	0.4249	0.3254	2.8128	0.5050	0.3969
100	1000	100	100	200	3.1475	0.5275	0.4170	2.3009	0.4153	0.3307	2.7843	0.4883	0.3927
100	3000	100	100	200	3.2805	0.5445	0.4366	2.2667	0.4157	0.3287	2.9734	0.5320	0.4249
100	1000	200	100	200	2.9896	0.4805	0.3911	2.1855	0.3993	0.3163	2.8776	0.5202	0.4105
100	1000	300	100	200	2.9920	0.4808	0.3914	2.1862	0.3995	0.3164	2.8780	0.5206	0.4106
100	1000	500	100	200	2.9887	0.4804	0.3910	2.1851	0.3993	0.3162	2.8778	0.5205	0.4105
100	1000	100	50	200	2.9947	0.4812	0.3917	2.1859	0.3994	0.3163	2.8807	0.5207	0.4109
100	1000	100	150	200	2.9973	0.4818	0.3921	2.1847	0.3992	0.3162	2.8680	0.5192	0.4091
100	1000	100	200	200	2.9934	0.4812	0.3916	2.1862	0.3995	0.3164	2.8886	0.5217	0.4121
100	1000	100	100	50	3.0007	0.4823	0.3925	2.1847	0.3992	0.3162	2.8854	0.5215	0.4116
100	1000	100	100	70	2.9911	0.4807	0.3913	2.1844	0.3991	0.3161	2.8808	0.5211	0.4110
100	1000	100	100	120	2.9945	0.4813	0.3917	2.1862	0.3995	0.3164	2.8677	0.5192	0.4091

Note: * represents the parameter settings selected in this study.

Table 17 The forecasting results of CFM with different rolling input steps.

Rolling-input	Site1: 10-min				Site2: 10-min			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	3.0833	2.8955	**2.8645**	3.0694	2.4134	2.2406	**2.1843**	2.3213
RMSE	0.4967	0.4699	**0.4114**	0.4998	0.4405	0.4163	**0.3991**	0.4232
MAE	0.4025	0.3800	**0.3217**	0.4062	0.3494	0.3268	**0.3111**	0.3320
R^2	0.7710	0.7950	**0.9749**	0.7681	0.7976	0.8193	**0.9639**	0.8133
IA	0.9524	0.9574	**0.9754**	0.9518	0.9643	0.9681	**0.9942**	0.9672
PE (±5%)	78.89	81.11	**80.00**	80.00	88.00	88.89	**92.22**	87.78
PE (±10%)	98.89	97.21	**98.89**	97.89	95.64	94.45	**94.23**	96.25
PE (±15%)	99.01	98.46	**99.00**	99.01	98.72	99.12	**98.20**	98.85
Rolling-input	**Site4: 20-min**				**Site5: 20-min**			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	5.0079	5.2326	**4.8669**	5.1209	5.2472	5.1036	**4.1947**	4.9909
RMSE	0.3490	0.3561	**0.3328**	0.3477	0.3703	0.3751	**0.2999**	0.3647
MAE	0.2476	0.2569	**0.2428**	0.2528	0.2789	0.2748	**0.2204**	0.2692
R^2	0.9155	0.9120	**0.9231**	0.9161	0.9278	0.9259	**0.9526**	0.9299
IA	0.9635	0.9620	**0.9667**	0.9638	0.9621	0.9610	**0.9750**	0.9630
PE (±5%)	68.89	0.6222	**69.33**	65.56	58.89	61.11	**72.22**	62.22
PE (±10%)	82.22	0.8556	**82.22**	84.44	91.11	87.88	**92.22**	91.11
PE (±15%)	93.33	0.9333	**97.78**	95.56	97.78	96.67	**96.67**	96.67
Rolling-input	**Site6: 30-min**				**Site7: 30-min**			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	5.5761	5.6650	**5.8810**	5.7209	5.8572	6.2507	**5.0589**	5.8738
RMSE	0.4880	0.4965	**0.4962**	0.4944	0.4518	0.4617	**0.3946**	0.4516
MAE	0.3982	0.4033	**0.4121**	0.4070	0.3445	0.3620	**0.3120**	0.3499
R2	0.9739	0.9730	**0.9730**	0.9732	0.9755	0.9744	**0.9813**	0.9755
IA	0.9440	0.9419	**0.9421**	0.9425	0.9483	0.9459	**0.9606**	0.9423
PE (±5%)	52.22	50.00	**54.44**	54.44	51.11	44.44	**57.78**	55.56
PE (±10%)	86.67	86.67	**84.44**	84.44	82.22	82.22	**92.22**	78.89
PE (±15%)	96.67	97.78	**96.67**	96.67	95.56	93.33	**96.67**	92.22

Table 18 The multi-step forecasting results of CFM and denoising models.

SET	Model	STEP2								STEP3							
							PE								PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%	MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	ICE-ANFIS	5.5741	0.9808	0.6722	0.8469	0.8108	58.18	85.00	95.91	7.0662	1.0365	0.8328	0.8284	0.7869	41.82	77.27	95.45
	ICE-LSTM	4.8109	0.7103	0.5638	0.9197	0.9005	59.09	88.64	98.64	6.6049	0.9362	0.7582	0.8600	0.8265	42.73	77.73	93.18
	ICE-CNN	5.4072	0.8220	0.6498	0.8925	0.8656	52.27	88.18	98.18	6.9907	0.9489	0.7874	0.8561	0.8240	41.36	74.55	90.91
	ICE-ELM	4.5395	0.6710	0.5277	0.9284	0.9112	63.64	89.55	99.09	6.7105	0.9762	0.7816	0.8477	0.8106	42.73	79.09	95.00
	CFM	**4.4556**	**0.7306**	**0.5848**	**0.8821**	**0.8967**	**56.67**	**92.22**	**98.89**	**5.3861**	**0.8419**	**0.6998**	**0.9198**	**0.8629**	**47.78**	**86.67**	**97.78**
Site2: 10-min	ICE-ANFIS	4.0231	0.6975	0.5311	0.9256	0.9086	68.64	93.18	99.09	5.9385	1.0245	0.7839	0.8382	0.8019	50.45	82.27	93.18
	ICE-LSTM	4.5031	0.7722	0.5976	0.9088	0.8880	59.55	91.82	98.18	5.9602	1.0106	0.7814	0.8425	0.8075	50.45	81.82	94.55
	ICE-CNN	5.2576	0.9032	0.7025	0.8752	0.8458	55.00	86.82	98.64	7.2749	1.2357	0.9857	0.8646	0.7084	49.55	72.73	90.91
	ICE-ELM	3.7898	0.6692	0.5035	0.9315	0.9159	69.09	95.00	99.55	5.7117	0.9487	0.7467	0.8612	0.8307	51.82	85.91	96.82
	CFM	**3.6019**	**0.4861**	**0.3826**	**0.9336**	**0.9564**	**88.89**	**95.45**	**99.11**	**5.1879**	**0.9662**	**0.7645**	**0.8446**	**0.8269**	**56.67**	**87.78**	**97.78**
Site4: 20-min	ICE-ANFIS	7.1044	0.5657	0.4441	0.9550	0.9205	46.36	74.55	88.18	10.1619	0.8281	0.6233	0.9032	0.8294	37.73	60.45	79.09
	ICE-LSTM	7.3513	0.5861	0.4581	0.9516	0.9145	46.36	76.36	89.55	10.2983	0.8190	0.6215	0.9053	0.8328	40.91	62.73	77.72
	ICE-CNN	8.3737	0.6659	0.5158	0.9376	0.8896	42.27	68.64	84.55	11.7013	0.9045	0.6820	0.8845	0.7974	37.27	60.91	72.73
	ICE-ELM	6.8610	0.5554	0.4302	0.9566	0.9233	48.64	77.73	88.64	10.0260	0.7985	0.6108	0.9100	0.8413	36.36	61.82	80.91
	CFM	**6.5225**	**0.5131**	**0.4617**	**0.9531**	**0.8942**	**54.44**	**73.33**	**89.22**	**9.3751**	**0.7594**	**0.6072**	**0.9062**	**0.8591**	**40.00**	**67.78**	**83.33**
Site5: 20-min	ICE-ANFIS	6.6738	0.7163	0.5199	0.9455	0.8806	49.09	78.18	92.27	9.7967	0.9733	0.7354	0.8992	0.7794	35.00	61.36	78.18
	ICE-LSTM	6.7348	0.7048	0.5177	0.9473	0.8843	51.82	76.36	92.37	9.7545	0.9507	0.7142	0.9038	0.7891	35.45	63.64	79.55
	ICE-CNN	7.6882	0.8352	0.5981	0.9260	0.8365	47.27	72.27	85.91	10.6698	1.0898	0.8183	0.8737	0.7205	31.36	75.45	58.64
	ICE-ELM	6.6218	0.6915	0.5120	0.9493	0.8887	48.18	80.00	91.09	9.6863	0.9514	0.7133	0.9037	0.7888	38.18	62.27	80.45
	CFM	**6.5292**	**0.5714**	**0.4248**	**0.9265**	**0.9095**	**50.00**	**77.78**	**92.78**	**9.0242**	**0.8712**	**0.6935**	**0.8937**	**0.7897**	**36.44**	**73.55**	**81.89**
Site6: 30-min	ICE-ANFIS	7.9222	0.8152	0.6341	0.9495	0.8530	38.64	71.82	89.09	11.0788	1.1669	0.8808	0.8967	0.6981	32.73	56.82	74.09
	ICE-LSTM	7.5799	0.7742	0.6062	0.9545	0.8674	42.27	73.18	90.45	11.1169	1.1475	0.8822	0.9001	0.7077	30.45	56.82	74.09
	ICE-CNN	8.8575	1.1007	0.7957	0.9079	0.7283	37.73	66.82	83.64	11.8460	1.2170	0.9456	0.8876	0.6740	28.64	54.55	68.64
	ICE-ELM	7.6516	0.7838	0.6102	0.9533	0.8641	39.09	72.27	89.09	10.8049	1.1544	0.8622	0.8989	0.7040	35.45	57.27	74.55
	CFM	**8.1519**	**0.6920**	**0.5413**	**0.9463**	**0.8865**	**43.89**	**72.22**	**86.67**	**10.5601**	**1.0711**	**0.8671**	**0.8670**	**0.7281**	**36.89**	**58.00**	**74.44**
Site7: 30-min	ICE-ANFIS	12.3948	1.1695	0.8588	0.8957	0.6809	26.82	54.55	71.82	12.3948	1.1695	0.8588	0.8957	0.6809	26.82	54.55	71.82
	ICE-LSTM	12.1247	1.0916	0.8220	0.9091	0.7221	28.64	55.00	71.36	12.1247	1.0916	0.8220	0.9091	0.7221	28.64	55.00	71.36
	ICE-CNN	12.6388	1.1586	0.8763	0.8976	0.6894	27.27	49.55	68.64	12.6388	1.1586	0.8763	0.8976	0.6894	27.27	49.55	68.64
	ICE-ELM	11.6304	1.0483	0.7875	0.9162	0.7438	31.82	56.36	72.27	11.6304	1.0484	0.7875	0.9162	0.7438	31.82	56..36	72.27
	CFM	**9.5321**	**1.0092**	**0.7905**	**0.8692**	**0.7401**	**45.56**	**51.11**	**77.78**	**10.5321**	**1.0092**	**0.7965**	**0.8692**	**0.7401**	**38.56**	**55.11**	**72.78**

Table 20 Comparison results.

Authors	Published year	Algorithm	Results	Similarity	Difference
Niu et al. [32]	2019	1. CEEMDAN 2. BPNN+GRNN+ENN+ELM +ARIMA 3. MOGOA	The combined model obtained high accuracy, the MAPE values of five data sets are 3.14%, 2.89%, 3.43%, 4.06% and 3.62% respectively.	Preprocessing methods, several single models and optimization algorithms are used. Different denoising strategies and optimization methods are compared.	Uncertain predictions and the influence of parameters on model performance are not discussed. The operation time of models is ignored.
Shao et al. [29]	2021	1. CEEMDAN 2. BPNN+GRNN+RBF+ ELM 3. MOGWO	Several experiments verify the effectiveness of the proposed model with the smallest MAPE value of 2.03%.	The same optimization algorithm is used. Different denoising strategies and optimization methods are compared.	Uncertain predictions and the influence of model parameters are not discussed. The operation time of models is ignored. Lack comparison with other literatures.
Liu et al. [23]	2019	1. ICEEMDAN 2. BPNN+GRNN+ENN+ELM +ARIMA 3. MMODA	The proposed model can improve the prediction accuracy. MAPE values of multi-step prediction are 3.15%, 4.41% and 5.02% respectively.	The same denoising strategy is applied. Multiple single models are applied. Different denoising strategies and optimization methods are compared.	The optimization algorithm is improved. Wind speed forecasts for different seasons are discussed. The result of parameter adjustment is not discussed in detail.

SET	Model	MAPE	RMSE	MAE	R^2	IA	PE ±5%	PE ±10%	PE ±15%
Site1:10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	Niu's model	3.4181	0.4898	0.4322	0.7773	0.9541	32.44	58.42	75.28
	Shao's model	4.0200	0.6628	0.5380	0.7222	0.9149	33.58	61.06	78.36
	Liu's model	3.1835	0.4345	0.4064	0.8247	0.9639	30.97	56.37	73.01
Site2:10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	Niu's model	2.3214	0.4077	0.3372	0.8267	0.9696	32.29	59.41	77.06
	Shao's model	3.0082	0.5493	0.4388	0.7853	0.9445	37.63	66.10	82.22
	Liu's model	1.8987	0.2917	0.2703	0.9113	0.9845	90.93	93.11	98.94
Site3:10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	Niu's model	3.1730	0.5150	0.4480	0.8186	0.9515	38.32	51.23	79.02
	Shao's model	3.8116	0.6726	0.5498	0.7906	0.9163	39.57	54.24	72.78
	Liu's model	3.1508	0.4774	0.4442	0.8441	0.9584	40.91	60.58	88.60

Data availability

Because of the sensitive issues involved in the survey, the data would not be shared.

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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 71671029).

Appendix

Table A List of terminologies

NWP	numerical weather prediction	**GPR**	gaussian process regression
ARIMA	autoregressive integrated moving average	**ARMA**	autoregressive moving average
AI	artificial intelligence	**PSO**	particle swarm optimization
ANFIS	adaptive neuro-fuzzy inference system	**EEMD**	ensemble empirical mode decomposition
SSA	singular spectrum analysis	**WA**	wavelet transform
GM	gray prediction model	**CFM**	combined forecasting model
CEE	complete ensemble empirical mode decomposition with adaptive noise	**ICE**	improved complete ensemble empirical mode decomposition with adaptive noise
MOGWO	multi objective grey wolf optimizer	**PP**	point prediction
IP	interval prediction	**AR**	autoregressive model
ANN	artificial neural networks	**SVM**	support vector machines
LSTM	long short-term memory	**CNN**	convolutional neural network
BPNN	back propagation neural network	**ELM**	extreme learning machine
GRNN	general regression neural network	**ENN**	elman neural network
SEM	selection of excellent sub models	**AWD**	cumulative breadth error
RMSE	root mean square error	**MAE**	mean absolute error
MAPE	mean absolute percentage error	**R^2**	goodness of fit
IA	index of agreement of predictive results	**RE**	relative error
FICP	forecasting interval coverage probability	**FINAW**	prediction interval standardized mean breadth
MOGOA	multi-objective grasshopper optimization algorithm	**MOALO**	multi-objective antlion algorithm
MODA	multi-objective dragonfly algorithm	**EMD**	empirical mode decomposition
GRU	gated recurrent unit	**TCN**	temporal convolutional networks
QRNN	quasi-recurrent neural networks	**RNN**	recurrent neural network
WNN	wavelet neural network	**VMD**	variational mode decomposition
RBF	radial basis function	**PSR**	phase-space reconstruction
MOMVO	multi objective multi verse optimization	**RVFL**	random vector functional link network
WOA	whale optimization algorithm	**EPT**	ensemble patch transformation
SVR	support vector regression	**MSSA**	multi objective salp swarm algorithm

References

[1] Zhang Z, Qin H, Liu Y, Wang Y, Yao L, Li Q, et al. Long Short-Term Memory Network based on Neighborhood Gates for processing complex causality in wind speed prediction. Energy Conversion and Management. 2019;192:37-51. https://doi.org/10.1016/j.enconman.2019.04.006

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

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

[4] Allen DJ, Tomlin AS, Bale CSE, Skea A, Vosper S, Gallani ML. A boundary layer scaling technique for estimating near-surface wind energy using numerical weather prediction and wind map data. Applied Energy. 2017;208:1246-57. https://doi.org/10.1016/j.apenergy.2017.09.029

[5] Chu J, Yuan L, Pan L, Liu Q, Yan J, Liu Y. NWP Combination Correction Model Based on Variable-weight Stacking Algorithm. Energy Procedia. 2019;158:6309-14. https://doi.org/10.1016/j.egypro.2019.01.408

[6] Li L-L, Chang Y-B, Tseng M-L, Liu J-Q, Lim MK. Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. Journal of Cleaner Production. 2020;270:121817. https://doi.org/10.1016/j.jclepro.2020.121817

[7] Hoolohan V, Tomlin AS, Cockerill T. Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. Renewable Energy. 2018;126:1043-54. https://doi.org/10.1016/j.renene.2018.04.019

[8] Dong L, Wang L, Khahro SF, Gao S, Liao X. Wind power day-ahead prediction with cluster analysis of NWP. Renewable and Sustainable Energy Reviews. 2016;60:1206-12. https://doi.org/10.1016/j.rser.2016.01.106

[9] Liu Z, Jiang P, Wang J, Zhang L. Ensemble forecasting system for short-term wind speed forecasting based on optimal sub-model selection and multi-objective version of mayfly optimization algorithm. Expert Systems with Applications. 2021;177:114974. https://doi.org/10.1016/j.eswa.2021.114974

[10] Grivel E, Diversi R, Merchan F. Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. Digital Signal Processing. 2021;116:103089. https://doi.org/10.1016/j.dsp.2021.103089

[11] Wang L, Tao R, Hu H, Zeng Y-R. Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. Renewable Energy. 2021;164:642-55. https://doi.org/10.1016/j.renene.2020.09.108

[12] Liu H, Tian H-q, Li Y-f. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Applied Energy. 2012;98:415-24. https://doi.org/10.1016/j.apenergy.2012.04.001

[13] Modaresi Movahed T, Jalaly Bidgoly H, Khoshgoftar Manesh MH, Mirzaei HR. Predicting cancer cells progression via entropy generation based on AR and ARMA models. International Communications in Heat and Mass Transfer. 2021;127:105565. https://doi.org/10.1016/j.icheatmasstransfer.2021.105565

[14] Fu W, Fang P, Wang K, Li Z, Xiong D, Zhang K. Multi-step ahead short-term wind speed

forecasting approach coupling variational mode decomposition, improved beetle antennae search algorithm-based synchronous optimization and Volterra series model. Renewable Energy. 2021;179:1122-39. https://doi.org/10.1016/j.renene.2021.07.119
[15] Wang Y, Wang J, Li Z, Yang H, Li H. Design of a combined system based on two-stage data preprocessing and multi-objective optimization for wind speed prediction. Energy. 2021;231:121125. https://doi.org/10.1016/j.energy.2021.121125
[16] Pousinho HMI, Mendes VMF, Catalão JPS. A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal. Energy Conversion and Management. 2011;52:397-402. https://doi.org/10.1016/j.enconman.2010.07.015
[17] Jiang P, Liu Z, Wang J, Zhang L. Decomposition-selection-ensemble forecasting system for energy futures price forecasting based on multi-objective version of chaos game optimization algorithm. Resources Policy. 2021;73:102234. https://doi.org/10.1016/j.resourpol.2021.102234
[18] Huang W, Liu H, Zhang Y, Mi R, Tong C, Xiao W, et al. Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. Applied Soft Computing. 2021;109:107541. https://doi.org/10.1016/j.asoc.2021.107541
[19] An F-P, Ma X-m, Bai L. Image fusion algorithm based on unsupervised deep learning-optimized sparse representation. Biomedical Signal Processing and Control. 2022;71:103140. https://doi.org/10.1016/j.bspc.2021.103140
[20] Gnatowski M, Buchaniec S, Brus G. The prediction of the polarization curves of a solid oxide fuel cell anode with an artificial neural network supported numerical simulation. International Journal of Hydrogen Energy. 2021. https://doi.org/10.1016/j.ijhydene.2021.09.100
[21] Du P, Wang J, Yang W, Niu T. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renewable Energy. 2018;122:533-50. https://doi.org/10.1016/j.renene.2018.01.113
[22] Wang J, Zhang L, Wang C, Liu Z. A regional pretraining-classification-selection forecasting system for wind power point forecasting and interval forecasting. Applied Soft Computing. 2021:107941. https://doi.org/10.1016/j.asoc.2021.107941
[23] Liu Z, Jiang P, Zhang L, Niu X. A combined forecasting model for time series: Application to short-term wind speed forecasting. Applied Energy. 2020;259:114137. https://doi.org/10.1016/j.apenergy.2019.114137
[24] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. Renewable Energy. 2021;165:573-94. https://doi.org/10.1016/j.renene.2020.11.002
[25] Li H, Liu T, Wu X, Li S. Research on test bench bearing fault diagnosis of improved EEMD based on improved adaptive resonance technology. Measurement. 2021;185:109986. https://doi.org/10.1016/j.measurement.2021.109986
[26] Li D, Jiang F, Chen M, Qian T. Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks. Energy. 2022;238:121981. https://doi.org/10.1016/j.energy.2021.121981
[27] Zhang X, Wang J, Gao Y. A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. Energy Economics. 2019;81:899-913. https://doi.org/10.1016/j.eneco.2019.05.026
[28] Wang D, Luo H, Grunder O, Lin Y. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renewable Energy. 2017;113:1345-58. https://doi.org/10.1016/j.renene.2017.06.095

[29] Shao Y, Wang J, Zhang H, Zhao W. An advanced weighted system based on swarm intelligence optimization for wind speed prediction. Applied Mathematical Modelling. 2021;100:780-804. https://doi.org/10.1016/j.apm.2021.07.024
[30] Xu Y, Yang W, Wang J. Air quality early-warning system for cities in China. Atmospheric Environment. 2017;148:239-57. https://doi.org/10.1016/j.atmosenv.2016.10.046
[31] Wang J, Cheng Z. Wind speed interval prediction model based on variational mode decomposition and multi-objective optimization. Applied Soft Computing. 2021;113:107848. https://doi.org/10.1016/j.asoc.2021.107848
[32] Niu X, Wang J. A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. Applied Energy. 2019;241:519-39. https://doi.org/10.1016/j.apenergy.2019.03.097
[33] He Y, Tsang KF. Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. Energy Reports. 2021;7:6473-88. https://doi.org/10.1016/j.egyr.2021.09.115
[34] Mirjalili S, Saremi S, Mirjalili SM, Coelho LdS. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. Expert Systems with Applications. 2016;47:106-19. https://doi.org/10.1016/j.eswa.2015.10.039
[35] Onyelowe KC, Shakeri J, Amini-Khoshalann H, Salahudeen AB, Arinze EE, Ugwu HU. Application of ANFIS hybrids to predict coefficients of curvature and uniformity of treated unsaturated lateritic soil for sustainable earthworks. Cleaner Materials. 2021;1:100005. https://doi.org/10.1016/j.clema.2021.100005
[36] Mahdevari S, Khodabakhshi MB. A hybrid PSO-ANFIS model for predicting unstable zones in underground roadways. Tunnelling and Underground Space Technology. 2021;117:104167. https://doi.org/10.1016/j.tust.2021.104167
[37] Ozkok FO, Celik M. A hybrid CNN-LSTM model for high resolution melting curve classification. Biomedical Signal Processing and Control. 2022;71:103168. https://doi.org/10.1016/j.bspc.2021.103168
[38] Shi X, Li Y, Yang Y, Sun B, Qi F. Multi-models and dual-sampling periods quality prediction with time-dimensional K-means and state transition-LSTM network. Information Sciences. 2021;580:917-33. https://doi.org/10.1016/j.ins.2021.09.056
[39] Elmaz F, Eyckerman R, Casteels W, Latré S, Hellinckx P. CNN-LSTM architecture for predictive indoor temperature modeling. Building and Environment. 2021;206:108327. https://doi.org/10.1016/j.buildenv.2021.108327
[40] İnik Ö, Altıok M, Ülker E, Koçer B. MODE-CNN: A fast converging multi-objective optimization algorithm for CNN-based models. Applied Soft Computing. 2021;109:107582. https://doi.org/10.1016/j.asoc.2021.107582
[41] Fetimi A, Dâas A, Benguerba Y, Merouani S, Hamachi M, Kebiche-Senhadji O, et al. Optimization and prediction of safranin-O cationic dye removal from aqueous solution by emulsion liquid membrane (ELM) using artificial neural network-particle swarm optimization (ANN-PSO) hybrid model and response surface methodology (RSM). Journal of Environmental Chemical Engineering. 2021;9:105837. https://doi.org/10.1016/j.jece.2021.105837
[42] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM)

forecasts using a GPR (Gaussian Process Regression) model. Energy. 2015;93:41-56. https://doi.org/10.1016/j.energy.2015.08.045

[43] Bardhan A, Samui P, Ghosh K, Gandomi AH, Bhattacharyya S. ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. Applied Soft Computing. 2021;110:107595. https://doi.org/10.1016/j.asoc.2021.107595

[44] Zhang Y, Aslani F, Lehane B. Compressive strength of rubberized concrete: Regression and GA-BPNN approaches using ultrasonic pulse velocity. Construction and Building Materials. 2021;307:124951. https://doi.org/10.1016/j.conbuildmat.2021.124951

[45] Izonin I, Tkachenko R, Gregus ml M, Zub K, Tkachenko P. A GRNN-based Approach towards Prediction from Small Datasets in Medical Application. Procedia Computer Science. 2021;184:242-9. https://doi.org/10.1016/j.procs.2021.03.033

[46] Wang J, Zhou Y, Li Z. Hour-ahead photovoltaic generation forecasting method based on machine learning and multi objective optimization algorithm. Applied Energy. 2022;312:118725. https://doi.org/10.1016/j.apenergy.2022.118725

[47] Liu J, Meng X, Ma Y, Liu X. Introduce canopy temperature to evaluate actual evapotranspiration of green peppers using optimized ENN models. Journal of Hydrology. 2020;590:125437. https://doi.org/10.1016/j.jhydrol.2020.125437

[48] Jiang P, Liu Z, Niu X, Zhang L. A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. Energy. 2021;217:119361. https://doi.org/10.1016/j.energy.2020.119361

[49] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Transactions on Power Systems. 1989;4:1484-91. https://doi.org/10.1109/59.41700

Revised Manuscript with changes Marked

A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting

Jianzhou Wang[a], Yining An[b,*] , Zhiwu Li[a] , Haiyan Lu[c]

[a] *Macao Institute of Systems Engineering, Macau University of Science and Technology, Macao, 999078, China*

[b] *School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China*

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

** Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China*

Tel.: +86 17734579525

E-mail address: anyining202012@163.com

Abstract

Accurate wind speed prediction has become increasingly important in wind power generation. However, the lack of efficient data preprocessing techniques and integration strategies has been a big obstacle to the development of wind power forecasting system. Therefore, a novel and advanced combined forecasting system comprising a data preprocessing, an integration strategy and several single models is designed in this study. The proposed model not only eliminates the impact of noise, but also integrates several single-model forecasting results through a weight optimization operator. In addition, the uncertain prediction of wind speed is also discussed in detail. The results show that: (a) The MAPE values of the proposed model are 2.8645%, 2.1843% and 2.8727% respectively for the point prediction. (b) The FICP values of the proposed model are 85.1697, 89.5410 and 88.0111 respectively at the significant level α = 0.05 for the uncertainty forecasting. The AWD values are 0.0559, 0.0400 and 0.0361 and the FINAW values are 0.0478, 0.0404 and 0.0390. It is reasonable to conclude that the proposed system can effectively boost the precision and stability of wind speed forecasting and provide a new approach for the exploitation of wind energy.

Keywords: *Artificial intelligence; Data preprocessing; Combined forecasting model; Multi-objective optimization;*

1. Introduction

Owing to the rapid growth of the world economy, traditional energy sources such as natural gas and oil are being consumed in large quantities. Therefore, the development of sustainable green energy resources has received increasing attention [1]. Wind energy is a renewable energy resource with clean and effective characteristics that play an irreplaceable role in wind power generation. However, the instability and nonlinearity of wind speed limit the development of wind power and bring many obstacles to the wind power grid. Accurate wind speed prediction technology can reduce the impact of wind speed characteristics, which not only helps power grid operators and decision makers to timely plan and dispatch the power system, but also reduces the failure risk of wind power system and improves power quality [2]. Hence, accurate wind speed prediction technology can effectively improve the stability of wind power generation system [3].

1.1 Previous literature

At present, the wind speed forecasting approaches adopted in a large number of wind speed prediction studies include: (i) physical models, (ii) statistical models, (iii) Artificial intelligence models. Physical models are more suitable for long-term prediction, but have obvious defects for short-term prediction [4, 5]. Systematic errors can be easily generated in the predictions and the direction near the ground that result in differences in the predicted power generation when physical models are used [6, 7]. Dong et al. presented a hybrid model based on K-means cluster and general regression neural network (K-means-GRNN) of numerical weather forecasts; however, the practical application of the K-means-GRNN was difficult due to its high requirements on computation and information [8]. In contrast, statistical models have lower requirements for datasets [9]. Statistical models are usually linear models, such as the autoregressive moving average (ARMA) [10], the autoregressive integrated moving average model (ARIMA) [11, 12]. Movahed et al. predicted the development of cancer cells based on ARMA and Auto-Regressive (AR), and improved the forecasting accuracy [13]. However, for sequences with random and nonlinear characteristics, the statistical models were difficult to mine the information accurately and effectively [14].

Aiming at the deficiency of the above models, many researchers have performed in-depth studies on artificial intelligence (AI) prediction models [15, 16], which mainly include support vector machines (SVM) [17, 18], deep learning [19], and artificial neural networks (ANN) [20, 21]. AI prediction methods have strong generalization abilities, and fast calculation speeds, however, these single AI models may be affected by the initial parameters and are prone to fall into local optima during computations [22]. Considering the deficiency of the above approaches, combined models have been widely concerned, which often contain data preprocessing techniques and optimization algorithms [23].

On the one hand, because of the fluctuation and instability in wind speed, the use of data preprocessing techniques such as empirical mode decomposition (EMD) [24] and ensemble empirical mode decomposition (EEMD) [25] has been explored. In fact, disadvantages such as modes mix and boundary effect exist in EMD, although EEMD improves the shortcomings of EMD, it causes residual white noise [26]. Thus, an advanced strategy named improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) is proposed in this study, which can add special white noise to obtain accurate modal values and improve model accuracy.

On the other traditional single models may not be able to balance accuracy and stability in prediction [27]. Wang et al. designed a hybrid model of wavelet neural network optimized by genetic algorithm (GA-WNN) based on variational mode decomposition (VMD) for wind speed forecasting, however, it remains to be explored that both accuracy and stability can be satisfied [28]. Based on this, this paper adopts the multi-objective grey wolf optimizer (MOGWO), which can combine multiple single models and optimize the assigned weight through to obtain stable and accurate results. Moreover, many studies ignore the reliability and uncertainty of wind speed prediction, which will cause potential risks in practical applications [29]. Thus, the uncertainty of wind speed prediction based on interval prediction has been discussed in this study. **Table 1** summarizes the common model types in the literature in recent years.

1.2 Contribution

Through the above analysis, a novel combined predictive framework is built based on neural networks, deep learning approaches, and multi-objective optimization to improve the forecasting accuracy. The developed system consists of four modules: a data preprocessing module, a combination prediction module, an uncertainty prediction module and an assessment module. Modal decomposition is adopted in the data preprocessing module decomposition of the original sequence to eliminate noise and obtain a smooth reconstructed sequence. The combination prediction module is designed to combine several single prediction models and optimize the assigned weight to obtain stable and accurate results, and interval prediction is used for quantifying uncertainty and improving model reliability. Nine metrics are proposed to evaluate the prediction performance in the last module. The contributions and innovations are summarized below:

(1) **A novel and advanced wind speed combined forecasting model (CFM) that includes data preprocessing and combined prediction and assessment is developed in this work.** Considering the uncertainty and fluctuation of the initial sequence, the prediction system given can overcome these shortcomings and achieve accurate and reliable prediction performance.

(2) **An advanced data preprocessing technique based on the decomposition and ensemble theory is intended to remove the noise in the initial sequence.** An advanced data preprocessing technique is chosen to reduce fluctuations and uncertainty to obtain a smooth sequence and improve the forecasting accuracy.

(3) **A new combined prediction strategy that includes the selection of several sub-models to minimize the combination error and weight optimization operators is proposed.** Four sub-models are used to predict the original sequence, and their prediction results are integrated by the multi-objective grey wolf optimizer (MOGWO) to gain more precise and stable predictive effect.

(4) **The Pareto optimality of the solutions from the combined system is theoretically proven.** The proof ensures that the optimal weight vector is obtained in the combined system through the leader selection mechanism and superior to those generated by the individual models, thus improving the prediction performance of the proposed system.

(5) **An integrated and detailed assessment system was built to assess the point prediction (PP) and interval prediction (IP) results of CFM.** In the assessment module, nine metrics and several sites are used in the experiment with intervals of 10-min, 20-min and 30-min. In addition, the multi-step prediction and rolling input steps further are investigated to further ensure the accuracy and reliability of CFM.

The overall structure of this paper is as follows. The approaches used in this study

are introduced in Section 2, the flow of CFM described in detail in Section 3, the experimental processes presented in Section 4, and the discussion and conclusion are given in Sections 5 and 6, respectively.

Table 1 Various studies on wind speed.

Method	Published year	Contribution	Disadvantages
NWP-K means-GRNN [8]	2016	NWP and ANNs are combined and the parameters of clustering are discussed in detail.	high requirements on computation and information. Lack of comparison with other ANNs models.
ARMA-AR [13]	2021	The parameter values of AR and ARMA models are discussed.	Not suitable for non-linear data.
VMD-phase-space reconstruction (PSR) - BPNN, ELM, ENN-multi objective multi verse optimization (MOMVO) [15]	2021	A two-stage preprocessing way VMD-PSR is to analyze noise. MOMOVO is applied to combine BPNN, ELM and ENN.	The change of key parameter values needs to discuss.
ICEEMDAN- ARIMA, BPNN, ENN, GRNN, ELM-Modified MODA [23]	2020	The modes mix is solved in EMD and EEMD. Modified MODA enhances the optimization of weight coefficient.	The predictive power of selected single models is not fully demonstrated.
Flexible ensemble patch transformation (EPT)-CEEMDAN-CNN [26]	2022	EPT-CEEMDAN enhances the detection of local patterns embedded. Multiple deep learning models are compared.	Robustness and precision are not well balanced.
VMD-GA-WNN [28]	2017	VMD and GA are used to eliminate data noise and optimize WNN.	The optimization effect of single objective is limited.
CEEMDAN-ELM, RBF, GRNN, BPNN-MOGWO [29]	2021	MOGWO is adopted to integrate ELM, RBF, GRNN, BPNN.	Parameter setting of optimization algorithm is not explained.
ICEEMDAN-SVM- whale optimization algorithm (WOA) [30]	2017	WOA is used to optimize the parameters of SVM.	Fewer contrast models and optimization effect of single objective is limited.
VMD-BPNN, random vector functional link network (RVFL), ANFIS, GRNN-multi objective salp swarm algorithm (MSSA) and support vector regression (SVR) [31]	2021	Using the hybrid algorithm MSSA-SVR assigns coefficients to single models and the Pareto optimal solution of the optimization algorithm is analyzed.	Model parameter values are not displayed.
CEEMDAN-ELM, GRNN, ARIMA, BPNN, ENN- MOGOA[32]	2019	MOGOA can optimize individual model weights.	Lack of discussion on parameter setting and comparison of other deep learning models.

2. Design of the combined forecasting model

The proposed prediction model in this study is mainly composed of data preprocessing technology, multiple single models, multi-objective weight optimization operator and uncertainty prediction, which is used to improve the prediction accuracy and quantify the uncertainty of prediction results. The corresponding algorithm and theoretical introduction are presented in this part.

2.1 Data denoising strategy

Before data denoising, we process the original data to prevent data leakage. Specifically, data leakage can cause the model to look accurate, but when applied to real life, the model can become very inaccurate, which is mainly divided into feature leakage and training data leakage. For feature leakage, features generally have a strong correlation with target variables. In this study, prevention of training data leakage is of greater concern to us. Before dividing the datasets, preprocessing the whole data set will lead to data leakage and make the information of the test set appear in the training set. Therefore, after dividing the datasets, we use decomposition technology for training set and test set respectively to prevent data leakage. The corresponding data denoising strategy is presented as follows.

The ICEEMDAN (ICE) is adopted to decompose the initial sequence, which is characterized by instability and fluctuation [30, 33]. ICE, which is based on CEEMDAN, can further reduce the noise and aliasing in IMF, which has a good decomposition ability. Suppose $\tilde{\boldsymbol{\kappa}}$ is the original sequence, and the operators $\overline{\overline{\boldsymbol{D}}}_j(\cdot)$, $\overline{\overline{\boldsymbol{L}}}(\cdot)$, and $\overline{\overline{\boldsymbol{\varpi}}}^{(i)}$ are introduced. $\overline{\overline{\boldsymbol{D}}}_j(\cdot)$ generates the j-th mode decomposed from the initial time series, $\overline{\overline{\boldsymbol{L}}}(\cdot)$ produces the local average of the original sequence $\tilde{\boldsymbol{\kappa}}$, and $\overline{\overline{\boldsymbol{\varpi}}}^{(i)}$ denotes white gaussian noise with μ=0 and unit variance σ^2. The specific processes in ICE are as follows:

The local mean of the i-th realizations $\tilde{\boldsymbol{\kappa}}^i(t) = \tilde{\boldsymbol{\kappa}} + \chi_0 \cdot \overline{\overline{\boldsymbol{D}}}_1(\overline{\overline{\boldsymbol{\varpi}}}^{(i)}(t))$ is first calculated to obtain the first residue $\boldsymbol{R}_1(t)$: $\boldsymbol{R}_1(t) = \left\{\overline{\overline{\boldsymbol{L}}}(\tilde{\boldsymbol{\kappa}}^i(t))\right\}$, where χ_i is an operation item to remove noise and $\{\bullet\}$ is the tool of the averaging process. The first mode can then be computed as $\boldsymbol{IM}_1 = \tilde{\boldsymbol{\kappa}}(t) - \boldsymbol{R}_1(t)$.

The second residue $\boldsymbol{R}_2$ can be computed as $\boldsymbol{R}_1(t) + \chi_1 \overline{\overline{\boldsymbol{D}}}_2(\overline{\overline{\boldsymbol{\varpi}}}^{(i)}(t))$, and the second mode $\boldsymbol{IM}_2$ is obtained:

$$\boldsymbol{IM}_2 = \boldsymbol{R}_1(t) - \boldsymbol{R}_2(t) = \boldsymbol{R}_1(t) - \left\{\overline{\overline{\boldsymbol{L}}}(\chi_1 \overline{\overline{\boldsymbol{D}}}_2(\overline{\overline{\boldsymbol{\varpi}}}^{(i)}(t)) + \boldsymbol{R}_1(t))\right\} \tag{1}$$

For $\boldsymbol{j} = 3, 4, \ldots, \boldsymbol{n}$, the j-th residue is calculated as:

$$\boldsymbol{R}_j(t) = \left\{\overline{\overline{\boldsymbol{L}}}(\chi_{j-1} \overline{\overline{\boldsymbol{D}}}_j(\overline{\overline{\boldsymbol{\varpi}}}^{(i)}(t)) + \boldsymbol{R}_{j-1}(t))\right\} \tag{2}$$

$\boldsymbol{IM}_j = \boldsymbol{R}_{j-1}(t) - \boldsymbol{R}_j(t)$ can also be obtained. After obtaining all the modes, the original sequenc $\tilde{\boldsymbol{\kappa}}$ can be reconstructed as $\boldsymbol{\kappa}' = \sum \boldsymbol{R}_j(t) + \overline{\tilde{\boldsymbol{R}}}(t)$, where $\overline{\tilde{\boldsymbol{R}}}$ is the residual sequence.

2.2 Neural network models

LSTM (long short-term memory) is a special type of RNNs. The key components of LSTM are memory cells and gates. The forget gate determines the number of the

unit state $C(t-1)$ remaining at current moment $C(t)$ and is calculated as $\overline{\overline{F}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{b}_f+\{x_t,h_{t-1}\}*\bar{\omega}_f)$. The number of inputs $x(t)$ stored in the cell state $C(t)$ is determined in the calculation process of the input gate $\overline{\overline{I}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{b}_i+\{x_t,h_{t-1}\}*\bar{\omega}_i)$. Finally, the output gate controls the number of unit states $C(t)$ to be output to the present output value $h(t)$ and is given by $\overline{\overline{O}}_{\sigma,\bar{b},\bar{\omega}}(t)=\sigma(\bar{\omega}_o*\{h_{t-1},x_t\}+\bar{b}_o)$ where $\bar{b}$ and $\bar{\omega}$ are bias and weight, respectively. The calculation principle of GRU (gated recurrent unit) is very similar to LSTM. The difference is that GRU combines the forgetting gate and input gate in LSTM algorithm into update gate, so GRU consists of two gates, the update gate and the reset gate.

The hidden layer of CNN (convolutional neural network) consists of convolution C_L, pooling P_L and full connection F_L. The convolution process is represented as $C_{x,y}=\sum_i^{p*q}\bar{\omega}_i v_i$, which is the sum of the product of the kernel weight and the brightness of the corresponding element in the input image. The main goal of pooling is to reduce the feature space of the maps. The fully connected layer is greatly important in combining the extracted features to obtain the output.

TCN (temporal convolutional networks) is suitable for sequence model construction under causal constraints, that is, the output $\{\hat{y}_0,\hat{y}_1,\ldots,\hat{y}_t\}$ can only be predicted based on the past observation $\{x_0,x_1,\ldots,x_t\}$.Therefore, TCN can be designed as a nonlinear function with the mapping $f:X^T\rightarrow Y^T$. In addition, TCN also adds the dilated convolution and residual block to better extract historical information.

QRNN (quasi-recurrent neural networks) is the LSTM acceleration algorithm, including two components of convolution and pooling. In the convolution operation, the output of the input, forget and output gates can be expressed as $Z=\tanh(W_z*X)$, $F=\sigma(W_f*X)$ and $O=\sigma(W_O*X)$, where $\mathbf{X}$ is the input, W_z, W_f and W_O are the convolution filters. There are three ways the pooling process can reduce the number of features, named f-pooling, fo-pooling and ifo-pooling based on the number of gates used.

ANFIS (adaptive neuro-fuzzy inference system) is a neural network based on fuzzy reasoning. The first layer is a fuzzy layer. $\overline{\overline{O}}^1_{i=1,2}=\mu_{A_i}(x)$ and $\overline{\overline{O}}^1_{j=1,2}=\mu_{B_i}(y)$ are output functions where x and y are the inputs with the respective membership functions μ_{A_i} and μ_{B_i}. The second layer calculates the weights of each membership function according to the previous outputs. The third layer normalizes the weights as $\overline{\overline{O}}^3_i=\overline{\overline{\omega}}_i=\bar{\omega}_i/(\bar{\omega}_1+\bar{\omega}_2)$. The fourth layer provides the output of the rule inference $\overline{\overline{O}}^4_i=\overline{\overline{\omega}}_i\overline{\overline{F}}_i$. The last layer generates the sum of each output $\overline{\overline{O}}^5_i=\sum\overline{\overline{\omega}}_i\overline{\overline{F}}_i=\left(\sum\overline{\overline{\omega}}_i\overline{\overline{F}}_i\right)/\sum\bar{w}_i$.

ELM (extreme learning machine) is a kind of feedforward neural network that includes input, hidden and output layers. The thresholds b and input weights ω are randomly generated. The input and output are $\overline{X}=(x_i^1,x_i^2,\ldots x_i^n)\in\mathbb{S}^n$ and $\vec{T}=(t_i^1,t_i^2,\ldots t_i^m)\in\mathbb{S}^m$. The hidden layer output is $\overline{\overline{H}}^{\bar{\omega},b}_{i=1,\ldots,n}(x)=\overline{\overline{F}}(x^T\bar{\omega}_i+b_i)$,where $\overline{\overline{F}}(\bullet)$ is the activation function. The final output is $\overline{\overline{O}}^{\omega,b,\theta}_n(x)=\sum_{i=1}^n\overline{\overline{H}}_i(x)\bar{\theta}(i)$, where the output weights are $\bar{\theta}=(\bar{\theta}_1,\cdots,\bar{\theta}_n)$.

BPNN (back propagation neural network) consists of input, hidden and output layers, and the calculation is: $y_i = f_I\left(\mu_j + \sum_{m=t-n}^{t-1} \mu_{jm} y_m\right)(0 \le \mu_j, \mu_{jm} \le 1)$, $y_t = f_0\left(\lambda_0 + \sum_{j=1}^{I} \lambda_{0j} y_j\right)(0 \le \lambda_0, \lambda_{0j} \le 1)$, where y_m and y_j are the inputs of the input and hidden layers, y_t is the predicted value at time **t**. N and I are the number of nodes in the input and the hidden layers. In the hidden and output layers, μ_j and λ_0 are the threshold values, μ_{jm} and λ_{0j} are weights, f_I and f_0 are activation functions.

On the basis of BPNN structure, ENN (elman neural network) adds a continuation layer to the hidden layer as a delay operator to achieve the purpose of memory. The learning process is: $x(k) = f(w_1 x_c(k) + w_2(u(k-1)))$, $y(k) = g(w_3 x(k))$, where y is the m-dimensional output vector, x is the unit vector of n-dimensional middle layer, u is the r-dimensional input vector, $w_i(i = 1,2,3)$ is the connection weight of each layer, $g(\bullet)$ and $f(\bullet)$ are activation functions of output and middle layer neurons respectively.

GRNN (general regression neural network) is a kind of radial basis neural network, which is composed of input, pattern, summation and output layers. $p_i = \exp[-(X - X_i)^T (X - X_i)/2\sigma^2]$ $i = 1,2,\ldots,n$ is the neuron transfer function of the pattern layer, where X is the input variable and X_i is the learning sample of the i-th neuron. The output of neuron j can be calculated as $y_i = S_{Nj} / S_D$, where S_{Nj} and S_D are the arithmetic sum and weighted sum of neurons at the pattern layer.

2.3 Multi objective grey wolf optimizer

Mirjalili et al. first proposed the MOGWO algorithm, who were inspired by social leadership and hunting technique of grey wolves [29, 34]. The alpha wolf α_w is defined as the fittest solution, and the beta wolf β_w and delta wolf δ_w, the second and third best solutions, respectively. The other candidate approaches are the omega wolves ω_w. α_w, β_w, and δ_w play a key role in capturing prey, and ω_w follow α_w, β_w, and δ_w to catch prey.

Definition 1 Encircling process. During the hunting of prey, the grey wolf location is updated as $\vec{L}_{wolf}^{t'=\alpha,\beta,\delta}(t+1) = \vec{L}_{prey}^{t}(t) - H \otimes \Delta\vec{D}^{t}$, where $\vec{L}_{prey}^{t}(t)$ and $\vec{L}_{wolf}^{t'=\alpha,\beta,\delta}$ represent the location of prey and grey wolf, respectively, and $\Delta\vec{D}^{t=\alpha,\beta,\delta} = \left|C \otimes \vec{L}_{prey}^{t}(t) - \vec{L}_{wolf}^{t}(t)\right|$ is the distance between the grey wolf and its prey. C and H are coefficient vectors that are calculated as $C = 2 \otimes \gamma_2$ and $H = 2\varsigma \otimes \gamma_1 - \varsigma$, respectively, where ς is the convergence factor, which decreases linearly from 2 to 0 with the number of iterations, and Υ is a random vector $\Upsilon = [\gamma_1, \gamma_2] \in [0,1]$.

Definition 2 Hunting behavior. Suppose that α_w, β_w and δ_w have a better command of the potential location of the prey and their distance from the prey is: $\Delta\vec{D}^{t=\alpha,\beta,\delta} = \left|C^{t} \otimes \vec{L}_{prey}^{t} - \vec{L}_{wolf}^{t}\right|$. The current location of the wolves is then obtained as $\vec{L}_{wolf}^{t=\alpha,\beta,\delta} = \vec{L}_{prey}^{t} - H^{t} \otimes \Delta\vec{D}^{t}$, and the location of ω_w can be calculated as $\vec{L}_{wolf}^{\omega} = \sum_{t=\alpha,\beta,\delta} \vec{L}_{wolf}^{t} / 3$.

Definition 3 Exploitation. When $H \in [-1,1]$ is random vector, the wolves can move at will between themselves and the prey. This narrows down the estimated location of the prey provided by α_w, β_w and δ_w.

Definition 4 Searching prey. When $|H| \succ 1$, the wolf is forced to separate from its prey while $|H| \prec 1$ helps to converge towards the prey. Moreover, the random vector $C \in [0,2]$ facilitates the discovery of new solutions. $C \succ 1$ indicates that the location of the wolf has a great influence on the prey and $C \prec 1$ indicates that the location of the wolf has little effect on the prey.

Definition 5 Archive. The archive is a store tool that saves or retrieves non-dominated Pareto optimal solutions.

Definition 6 Leader selection mechanism. This selection mechanism adopts the roulette wheel approach of $P_i = c / N_i$ to choose the non-dominated pareto optimal solution where c is a fixed value and $c \succ 1$. N denotes the sum of Pareto optimal solutions obtained.

2.4 Multi objective problems (Mop)

In multi-objective problems, it is critical to find vector solutions. The concept of Pareto dominance is designed to facilitate the choice of the vector solutions. The Pareto optimal solution is defined as follows:

(1) $\vec{\varpi}^1$ is considered to dominate $\vec{\varpi}^2$, i.e., $\vec{\varpi}^1 \succ \vec{\varpi}^2$, if and only if $\vec{\varpi}^1$ and $\vec{\varpi}^2$ meet: $[\forall p \in \{1,2,\ldots t\}, \overline{\overline{F}}_t(\vec{\varpi}^1) \le \overline{\overline{F}}_t(\vec{\varpi}^2)] \wedge [\exists q \in \{1,2,\ldots t\}, \overline{\overline{F}}_q(\vec{\varpi}^1) \prec \overline{\overline{F}}_q(\vec{\varpi}^2)]$, where $\overline{\overline{F}}$ is a function and t denotes a vector of numbers.

(2) The solution $\vec{\varpi}^1$ is called the Pareto optimal solution if the conditional equality $/\exists \vec{\varpi}^i \in \mathbb{S}, \vec{\varpi}^i \prec \vec{\varpi}^1$ is satisfied, where $\mathbb{S}$ denotes solution space.

(3) The Pareto optimal solution set $\overline{\overline{\mathrm{P}}}^s$ includes all the Pareto optimal solutions and can be expressed as $\overline{\overline{\mathrm{P}}}^s = \{\vec{\varpi} \mid /\exists \vec{\varpi}^i \in \mathbb{S}, \vec{\varpi} \succ \vec{\varpi}^i\}$.

To improve the precision and stability of the combined system, the MOGWO objective function is defined as:

$$\min \begin{cases} \overline{\overline{F}}^1 = (1/Q)\sum_i^Q \left| (P^i - A^i)/A^i \right| \times 100\% \\ \overline{\overline{F}}^2 = std(P^i - A^i)\ i = 1,\ldots,Q \end{cases} \tag{3}$$

where $\overline{\overline{F}}$ is the objective function, P and A represent the forecasted and true values.

2.5 The Pareto solution proof

The objectives in MOPs may often conflict with one another. The solutions in such problems can usually be expressed as the Pareto solution (λ^*). For MOPs, the Pareto solution (λ^*) that satisfies $/\exists \vartheta \in \Omega_{s.t.} FIT(\lambda^*) \prec FIT(\vartheta^*)$ is adopted rather than the accurate solution. *Arc* is a tool for storing the non-dominated solutions (λ'); however, the storage capacity of *Arc* is limited and has the upper bound of $\bar{\gamma}$, that is, $\gamma \prec \bar{\gamma}$. Hence, in the process of updating *Arc*, it is necessary to compare the new solution λ'

and the present non-dominated solutions $\{\hat{\lambda}_1^p, \hat{\lambda}_2^p, \ldots, \hat{\lambda}_\gamma^p\}$ in ***Arc*** . When $\lambda' \succ \hat{\lambda}_i^p$ or $\lambda' \sim \hat{\lambda}_i^p$, $i \in [1, \gamma]$, λ' is added into ***Arc*** and $\hat{\lambda}_i^p$ is removed or ranked behind λ' .Once the ***Arc*** capacity has reached the upper bound $\gamma = \bar{\gamma}$, the most crowded segments will be removed with the deletion probability $Pd_i = \gamma_i / c$ ($c \succ 1$) . Then, the roulette-wheel technique is employed to compare the present solution with λ' with the probability $Pd_i = \gamma_i / c$ ($c \succ 1$) to determine the Pareto optimal solution λ' . The pseudocode is listed as **Table 2**.

Table 2 Pseudo code.

Algorithm: ***MOGWO***

Input:

$m_{tr}^{(0)} = (m^{(0)}(1), m^{(0)}(2), \ldots, m^{(0)}(a))$ –training data

$(m^{(0)}(a+1), m^{(0)}(a+2), \ldots, m^{(0)}(a+i))$ –testing data

Output:

$(\hat{n}^{(0)}(a+1), \hat{n}^{(0)}(a+2), \ldots, \hat{n}^{(0)}(a+i))$ –forecasting data

Parameters:

F_i – the fitness function of *i*-th wolf

$Iter_{max}$ – maximum iteration times.

t – present iteration times.

d –dimensions numbers.

n –wolves numbers

$\vec{L}_i$ – the position of *i*-th wolf

/*Set up the parameters of ***MOGWO***.*/

/* Random initialization of the n wolves $\vec{L}_i$ population (*i*=1,2…n).*/

FOR EACH i=1:n DO

/*Calculate the F_i using the process of ranking.*/

END FOR

/*Decide the best search agent $\vec{L}_{wolf}^{\,t=\alpha,\beta,\delta}$.*/

WHILE (t< $Iter_{max}$**) DO**

  FOR EACH i=1:n DO

    /*Update ς , H , C and choose a wolf randomly from the archive.*/

    $C = 2 \otimes \gamma_2$, $H = 2\varsigma \otimes \gamma_1 - \varsigma$

    /* In the archive choose the elite from by roulette wheel */

    IF$(|H| \prec 1)$ ***THEN***

      /* Update the location from the search proxy at present .*/

      $\vec{\Delta D}^{\,t=\alpha,\beta,\delta} = |C^t \otimes \vec{L}_{prey}^{\,t} - \vec{L}_{wolf}|$

    ELSE $(|H| \geq 1)$ ***THEN***

      /*Update the location from the search proxy at present.*/

      $\vec{L}_{wolf}^{\,t=\alpha,\beta,\delta} = \vec{L}_{prey}^{\,t} - H^t \otimes \vec{\Delta D}^{\,t}$

    END IF

  END FOR

  /* Compute all wolves' objective values.*/

  /* Search the non-dominated solutions.*/

  /* Renew the file depending on the non-dominated solutions.*/

  IF this file is up to its limit ***DO***

    /* To contain new solutions, remove some solutions from the file and use

roulette wheel with $P_i = N_i/c$ (c>1).*/

25 *END IF*

26 *IF* the file capacity exceeds boundary *DO*

/*Renew the boundary to hold the new ones.*/

28 $\vec{L}_{wolf}^{=\omega} = \sum_{\iota=\alpha,\beta,\delta} \vec{L}_{wolf}^{=\iota} / 3$

29 *END IF*

30 *T=t+1*

31 **END WHILE**

32 **RETURN** *L**

2.6 Uncertainty prediction

Interval prediction is proposed to quantify the uncertainty. According to maximum likelihood estimation (MLE), Weibull, Gamma, Rayleigh and Lognormal distribution are adopted to fit the data. Interval prediction results of CFM is testified and compared with basic models, the upper and lower bound is expressed as: $Up^{i-th} = Fr^{i-th} + Dis_{1-\alpha/2} \bullet \sigma^*_{Dis} / \sqrt{n}$, $Lo^{i-th} = Fr^{i-th} - Dis_{\alpha/2} \bullet \sigma^*_{Dis} / \sqrt{n}$, where $Dis_{1-\alpha/2}$ and $Dis_{\alpha/2}$ are the critical values of optimal distribution and α is the significance level (α= 0.05, 0.1 and 0.2 in this study).

3. Flow of the combined forecasting system

In this study, a strategy of combining interval prediction with point prediction is adopted. A smooth sequence is first obtained using the ICE data preprocessing technique. Then, some single neural models are applied to forecast a stable sequence and obtain the forecasted values $\vec{F} = (\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4)$. In addition, the optimization operator is used to provide the most reasonable weights $\vec{W} = (\omega_1, \omega_2, \omega_3, \omega_4)$ to combine the forecasted values in $\vec{F}$. Finally, confidence intervals of 80%, 90%, and 95% are set for the experiment. The main framework is illustrated in **Fig.1**.

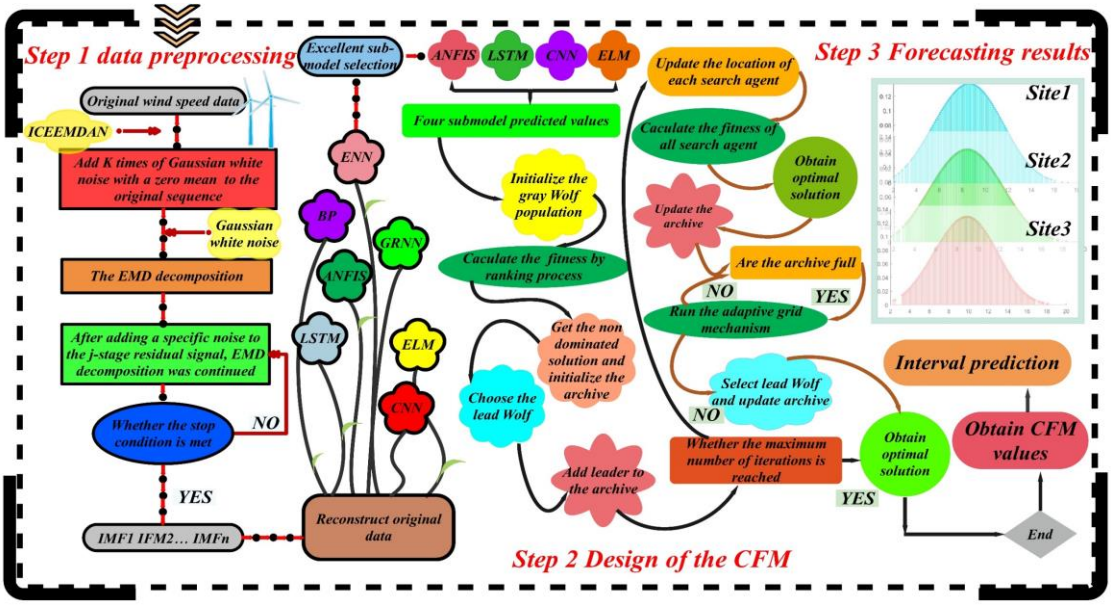

Fig.1 The main structure of this paper

3.1 Information of datasets

A wind power plant in Penglai (**37°48′N 120°45′E**), Shandong Province, China, is used as the example. Seven sites are chosen to verify the performance of the proposed system, which include 10-min, 20-min and 30-min intervals. Besides, seven sites have the same sample size of 2220. The first 2000 samples are used for training, and the remaining 220 are used for testing. The details of the dataset are presented in **Table 3** and the features of the dataset are also visually depicted in **Fig. 2**.

Table 3 Datasets.

SET	Sample	Size	Indicators(m/s)			
			Maximun	Minimun	Mean	Std.
Site1: 10-min	Total samples	2220	17.2000	2.0000	8.5566	2.8012
	Train-set	2000	16.1000	2.0000	8.1810	2.5671
	Test-set	220	17.2000	5.1000	11.9707	2.5207
Site2: 10-min	Total samples	2220	18.1000	2.3000	9.6626	2.9661
	Train-set	2000	18.1000	2.3000	9.2664	2.7311
	Test-set	220	17.5000	6.6000	13.2641	2.5688
Site3: 10-min	Total samples	2220	18.4000	2.0000	9.0219	3.1068
	Train-set	2000	18.4000	2.0000	8.6368	2.8735
	Test-set	220	17.6000	5.4000	12.5236	2.9601
Site4: 20-min	Total samples	2220	16.7000	0.9000	6.7181	2.7697
	Train-set	2000	16.7000	0.9000	6.6773	2.7774
	Test-set	220	12.9000	2.6000	7.0886	2.6763
Site5: 20-min	Total samples	2220	18.2000	0.8000	7.8146	3.0395
	Train-set	2000	18.2000	0.8000	7.7815	3.0340
	Test-set	220	15.6000	2.1000	8.1159	3.0796
Site6: 30-min	Total samples	2220	18.2000	1.3000	8.4859	3.2925
	Train-set	2000	17.9000	1.3000	8.4406	3.2510
	Test-set	220	18.2000	2.8000	8.8977	3.6307
Site7: 30-min	Total samples	2220	17.5000	1.0000	7.5488	3.1403
	Train-set	2000	17.5000	1.0000	7.5082	3.0817
	Test-set	220	17.2000	2.7000	7.9182	3.6172

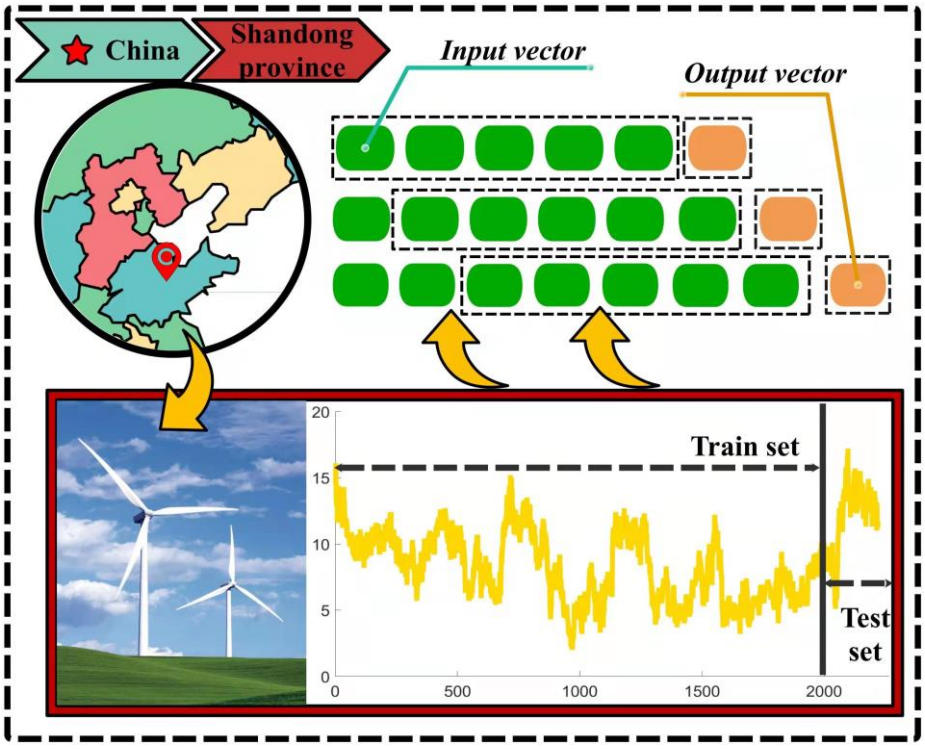

Fig.2 Wind speed data

3.2 Metrics

In the point prediction, six commonly used indicators are applied to evaluate

model performance, namely **RMSE, MAE, MAPE, R^2**, **IA** and **PE**, which can better reflect the prediction accuracy. The smaller the index values of **RMSE, MAE** and **MAPE** are, the higher the prediction accuracy of the models are, while R^2 and **IA** are opposite. Three indicators named **FINAW, AWD,** and **FICP** are designed for interval forecasting. **FINAW** calculates the width of the interval. **AWD** is used to estimate the degree of deviation of the interval. **FICP** measures the extent to which ranges cover true values. The detailed information of metrics is presented in **Table 4.**

Table 4 Metrics.

Metric	Expression	Definition
RMSE	$\mathbf{RMSE} = \sqrt{(1/Q)\sum_{i=1}^{Q}(F_i - A_i)^2}$	Root mean square error
MAE	$\mathbf{MAE} = (1/Q)\sum_{i=1}^{Q}\lvert F_i - A_i\rvert$	Mean absolute error
MAPE	$\mathbf{MAPE} = (1/Q)\sum_{i=1}^{Q}\lvert(F_i - A_i)/A_i\rvert \times 100\%$	Absolute percentage error mean
$\mathbf{R^2}$	$\mathbf{R}^2 = 1 - \frac{\sum_{i=1}^{Q}(F_i - A_i)^2}{\sum_{i=1}^{Q}(F_i - \bar{A})^2}$	Coefficient of determination
IA	$\mathbf{IA} = 1 - \frac{\sum_{i=1}^{Q}(A_i - F_i)^2}{\sum_{i=1}^{Q}(\lvert A_i - \bar{A}\rvert + \lvert \bar{A} - F_i\rvert)^2}$	Consistency of the predicted results
PE	$\mathbf{PE}(i) = (F_i - A_i)/A_i \times 100\%$	The percentage of the predicted value in a specified error range
FICP	$\mathbf{FICP} = \left(\sum_{i=1}^{Q} C_i / Q\right) \times 100\%$	Coverage probability of the forecast interval
FINAW	$\mathbf{FINAW} = \sum_{i=1}^{Q}(U_i - L_i)/FR$	Normalized averaged width of the forecast interval
AWD	$\mathbf{AWD}_i = \begin{cases} (A_i - U_i)/(U_i - L_i) & A_i \succ U_i \\ 0 & A_i \in [L_i, U_i] \\ (L_i - A_i)/(U_i - L_i) & A_i \prec L_i \end{cases}$ $\mathbf{AWD} = \left(\sum_{i=1}^{Q}\mathbf{AWD}_i\right)/FR$	$\mathbf{AWD}_i$ denotes the cumulative deviation of the forecast interval and **AWD** is the mean of cumulative deviation

Note: F_i, A_i represents the forecasting and actual value of i-th. In the point prediction evaluation index, ±5%, ±10% and ±15% error range are chosen to calculate the index PE; In the interval prediction evaluation index, if A_i belongs $[U_i, L_i]$, then c_i=1, otherwise c_i=0; FR means the range of predicted values.

3.3 Model parameter setting

The parameters for the neural network approaches are presented in **Table 5** and those for MOGWO and ICE in **Table 6**.

Table 5 Parameters of single models.

Model	Symbol	Meaning	Value	Setting reasons
BPNN	Tg	Training goal	0.00004	Preset
	Tf_{hid}	Activation function of hidden	tansig	Preset
ENN	E_{max}	The maximum epochs	1000	Trial- error method
	Tg	Training goal	0.00004	Preset
ELM	Tf	Activation function of hidden	sig	Preset
ANFIS	C_{ANf}	The number of the cluster	10	Preset
	E_{max}	The maximum epochs	1500	Trial- error method
GRNN	S_{GRnn}	Spread	1	Preset
LSTM	Lr	Training learning rate	0.005	Preset
	E_{max}	The maximum epochs	2000	Trial- error method

CNN	Lr	Training learning rate	0.005	Preset
	B_{min}	MiniBatchsize	16	Trial- error method
GRU	H_L	Numbers of hidden layers	40	Trial- error method
	E_{max}	The maximum epochs	2000	Preset
TCN	H_L	Numbers of hidden layers	50	Trial- error method
	E_{max}	The maximum epochs	500	Preset

Table 6 Parameters of ICE and MOGWO.

Model	Symbol	Meaning	Value	Setting reasons
ICE	$Iter_{max}$	The maximum number of iterations	1000	Trial- error method
	NR	Realization Number	100	Trial- error method
MOGWO	$Iter_{max}$	The maximum number of iterations	100	Trial- error method
	As	Archive size	100	Trial- error method
	P_s	Population size	200	Trial- error method
	α	Grid inflation	0.1	Preset
	g	Sum of grids per dimension	10	Preset
	β	Leader selection pressure	3	Preset

4. Experiment results

In this section, the results of the four comparative experiments and two validation tests to evaluate CFM are shown.

4.1 Selection of excellent sub models

Seven single predictive models named ANFIS [35, 36], LSTM [37, 38], CNN [39, 40], ELM [41-43], BPNN [44], GRNN [45, 46], and ENN [47] were chosen to forecast the original sequence in this study. To choose the most suitable models for improving the accuracy of CFM, the SEM strategy was developed [48]. The strategy is introduced below:

(1) The MAPE, MAE, and RMSE of every sub-model are calculated for one hundred iterations.

(2) Each index value is normalized as, for example,

$$\overline{MAE}_i = ((MAE_i - \min_{1\le i\le Q}(MAE)) / (\max_{1\le i\le Q}(MAE) - \min_{1\le i\le Q}(MAE)) \quad (4)$$

(3) Weights are assigned to each index value and the SEM value is calculated as $SEM_i = (1/3) * \overline{MAE}_i + (1/3) * \overline{MAPE}_i + (1/3) * \overline{RMSE}_i$.

The SEM values are listed in **Table 7**. A smaller SEM value indicates a higher model precision. Based on the results, ANFIS, LSTM, CNN, and ELM were chosen as the sub-models.

Table 7 Results of SEM.

SEM	Model						
	GRNN	**LSTM**	**ELM**	**BPNN**	**ANFIS**	**CNN**	**ENN**
Site1_SEM	0.3300	0.3021	0.3797	0.2458	0.2871	0.2590	0.2589
Site2_SEM	0.2814	0.1587	0.2240	0.3657	0.3125	0.2200	0.4129
Site3_SEM	0.3251	0.4011	0.2134	0.2948	0.2250	0.3100	0.3201
SEM*	0.3122	**0.2873**	**0.2724**	0.3021	**0.2749**	**0.2630**	0.3306

Note: * mean the average of SEM values of all sites.

4.2 Experiment Ⅰ: Comparison between CFM and basic models

To demonstrate the performance of CFM, seven individual models were used for comparison in Experiment Ⅰ. The results are as shown in **Table 8** and **Fig. 3**.

In **Table 8**, CFM exhibited the smallest error with MAPE, MAE and RMSE values of 2.8645%, 0.3217 and 0.4114 at site1. At site2, CFM exhibited the best performance

with the MAPE value of 2.1843% compared to the other models, which had an average MAPE value of 4.8747%. The superior precision of CFM was thus verified. The IA and R^2 can better reflect the agreement between the actual and the forecast data. At site1, CFM showed a better performance with the values of $R^2$and IA exceeding 97%, which was a 10% improvement in IA and R^2 over the best model, GRNN.

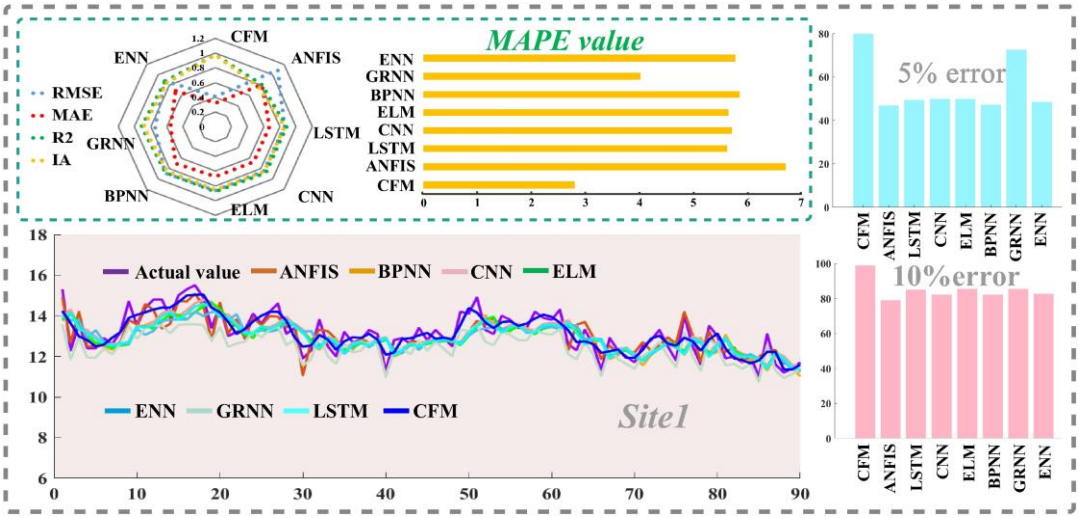

Fig.3 Results of CFM and basic models

4.3 Experiment Ⅱ: Comparison between CFM and single model based on ICE

Experiment Ⅱ was designed to contrast the accuracy of CFM and other models based on ICE. The results are presented in **Table 9** and **Fig.4**. The following conclusions can be drawn from the results:

It is obvious that CFM achieved the best prediction results at site1, with MAPE, MAE and RMSE values of 2.8645%, 0.3217, and 0.4114. For site2, the worst-performing model was ICE-ENN with MAPE exceeding 3.1044%. By contrast, the corresponding index value for CFM was 1% lower than that of ICE-ENN. For site3, CFM still exhibited the excellent performance with R^2 exceeding 97%. The prediction error PE of CFM in the range of 5%, 10%, and 15% was greater than 80%, 90% and 95%, respectively. Obviously, CFM had more excellent prediction accuracy.

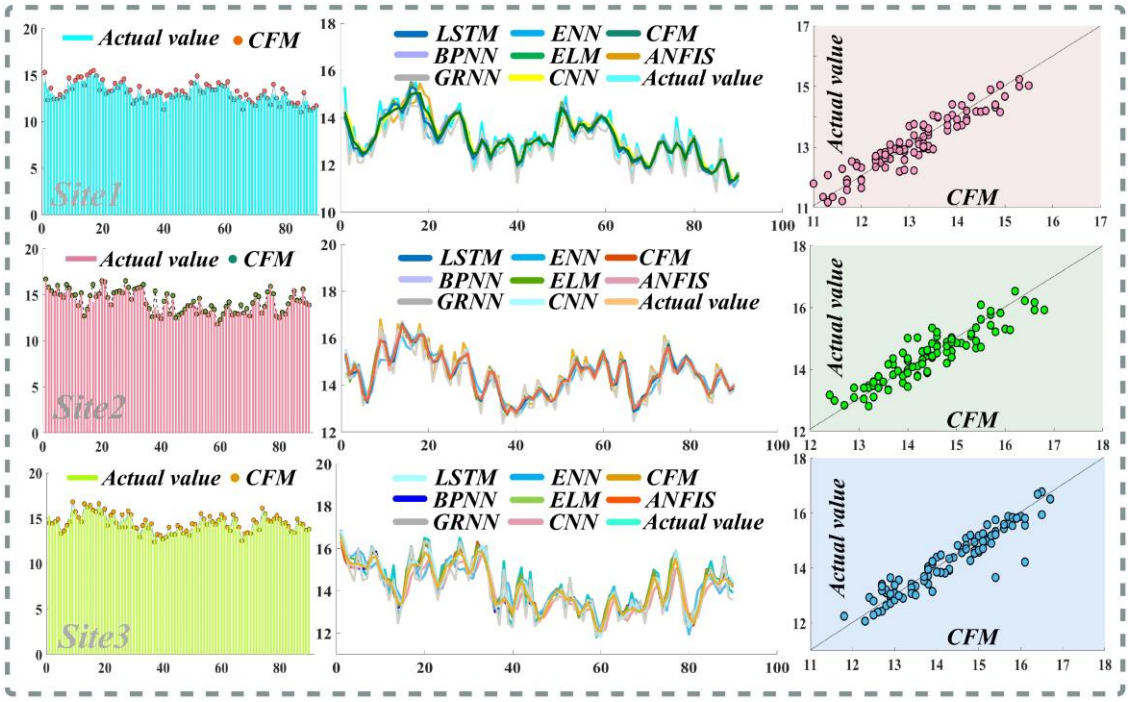

Fig.4 Results of CFM and denoising models

4.4 Experiment Ⅲ: Comparison between CFM and models with other denoising strategy

In this section, three denoising strategies, namely, CEE, EMD, and EEMD, were chosen to verify the predictive abilities of CFM, and the detailed results are presented in **Table 10**. The following conclusions can be drawn:

For site1, for the worst-performing EEMD-based model, the values of MAPE, MAE and RMSE were 4.6335%, 0.6592 and 0.6368, which far exceeded the corresponding values for CFM. For site2, $\boldsymbol{MAPE_{EMD}}$=3.3387% was obtained for the best classical model, which was based on EMD. In comparison, the smaller value of $\boldsymbol{MAPE_{CFM}}$=2.1843% was obtained for CFM. Finally, for site3, the prediction errors of the comparison models in the range of 5% were almost less than 80%. It is not difficult to find that CFM performed better with $\boldsymbol{PE}_{\boldsymbol{CFM}}^{5\%}$ exceeding 80%.

4.5 Experiment Ⅳ: Comparison between CFM and other optimization strategies

In this part, three optimization strategies called MOGOA, MOALO and MODA were designed to evaluate the predictive precision of CFM. The calculation results are listed in **Table 11**. It can be easily seen that the CFM had excellent performance. For example, for site1, the $\boldsymbol{MAPE_{CFM}}$=2.8645% of CFM was 0.2%, 0.02% and 0.05% lower than those of MOGOA, MOALO, and MODA, respectively. In addition, only the R^2 of CFM exceeded 95%, while its IA exceeded 97%. It can be found from the above evaluation index values that CFM has better performance in forecasting.

Table 8 The performances of CFM and basic models.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	ANFIS	6.8914	1.0924	0.8126	0.8113	0.7627	46.82	79.09	89.55
	LSTM	5.7710	0.8551	0.6705	0.8844	0.8553	49.55	85.00	95.00
	CNN	5.8641	0.8492	0.6731	0.8860	0.8576	50.00	82.27	95.45
	ELM	5.7888	0.8667	0.6737	0.8812	0.8512	50.00	85.45	96.36
	BPNN	6.0018	0.8832	0.6972	0.8767	0.8455	47.27	82.27	95.00
	GRNN	4.1138	0.7486	0.5544	0.9114	0.8879	72.73	85.45	99.09
	ENN	5.9280	0.8843	0.6943	0.8764	0.8449	48.64	82.73	96.36
	TCN	5.5340	0.8412	0.6571	0.8920	0.8588	50.50	86.67	96.69
	GRU	5.9284	0.8800	0.6877	0.8776	0.8867	49.55	80.91	95.45
	QRNN	5.4369	0.8391	0.6489	0.9012	0.8955	52.33	86.00	96.89
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	ANFIS	4.9404	0.8776	0.6535	0.8827	0.8550	62.73	88.64	96.82
	LSTM	4.9317	0.8474	0.6516	0.8907	0.8647	59.55	88.18	97.27
	CNN	5.0023	0.8375	0.6506	0.8932	0.8682	58.64	88.18	96.36
	ELM	4.8449	0.8369	0.6365	0.8934	0.8681	61.82	87.73	97.27
	BPNN	4.8209	0.8350	0.6341	0.8939	0.8690	61.36	88.18	95.91
	GRNN	4.1655	0.6751	0.5865	0.9306	0.9134	74.55	83.24	94.32
	ENN	5.4175	0.9370	0.7183	0.8663	0.8345	57.27	85.91	96.36
	TCN	4.7569	0.8250	0.6477	0.8985	0.8799	60.60	89.71	97.92
	GRU	4.9899	0.8448	0.6504	0.8914	0.8990	58.64	88.18	96.82
	QRNN	4.9210	0.8409	0.6450	0.8920	0.8644	59.82	87.65	96.34
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	ANFIS	7.0114	1.0389	0.8457	0.8762	0.7934	38.64	75.00	93.18
	LSTM	6.6676	0.9739	0.8076	0.8913	0.8181	43.64	78.64	96.36
	CNN	6.6651	0.9511	0.7841	0.8963	0.8275	42.73	80.91	93.64
	ELM	6.5754	0.9697	0.7944	0.8922	0.8198	41.82	78.64	95.91
	BPNN	6.4957	0.9626	0.7789	0.8938	0.8227	43.18	80.00	94.55
	GRNN	3.2736	0.5712	0.4483	0.9626	0.9370	65.00	74.25	93.24
	ENN	6.4089	0.9378	0.7616	0.8992	0.8319	45.45	81.82	95.91
	TCN	6.7258	0.9841	0.8130	0.8842	0.8032	42.30	76.82	95.20
	GRU	6.5127	0.9622	0.7770	0.8939	0.8778	44.55	82.27	94.09
	QRNN	6.5433	0.9678	0.7820	0.8912	0.8461	44.32	81.00	94.15

Note: The table lists the predicted results of CFM and basic models. The blacked part represents the evaluation index value of CFM. The equation of metrics are defined as: $\mathbf{PE}(i) = (F_i - A_i)/A_i \times 100\%$, $\mathbf{MAE} = (1/Q)\sum_{i=1}^{Q}|A_i - F_i|$, $\mathbf{MAPE} = (1/Q)\sum_{i=1}^{Q}|(A_i - F_i)/A_i| \times 100\%$, $\mathbf{R}^2 = -\sum_{i=1}^{Q}(A_i - F_i)^2 / \sum_{i=1}^{Q}(F_i - \overline{A})^2 + 1$, $\mathbf{RMSE} = ((1/Q)\sum_{i=1}^{Q}(A_i - F_i)^2)^{\wedge(1/2)}$, $\mathbf{IA} = -\sum_{i=1}^{Q}(A_i - F_i)^2 / \sum_{i=1}^{Q}(|A_i + \overline{A}| + |F_i - \overline{A}|)^2 + 1$. In addition, GRU, TCN and QRNN are introduced in experiments Ⅰ and Ⅱ to make a more complete comparison.

Table 9 The performances of CFM and the denoising models.

SET	Model	MAPE	RMSE	MAE	R^2	IA	PE ±5%	PE ±10%	PE ±15%
	CFM	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	ICE-ANFIS	3.3003	0.5527	0.4017	0.9517	0.9400	77.73	96.82	99.55
	ICE- LSTM	2.8753	0.4367	0.3409	0.9698	0.9625	83.18	95.09	99.00
	ICE-CNN	3.6981	0.5433	0.4330	0.9533	0.9420	73.18	96.82	98.32
	ICE-ELM	2.8631	0.4293	0.3375	0.9709	0.9638	85.45	95.02	97.20
Site1: 10-min	**ICE-BPNN**	2.9085	0.4329	0.3423	0.9704	0.9632	85.00	96.09	99.00
	ICE-GRNN	2.9884	0.4068	0.3790	0.9738	0.9672	95.91	99.09	99.55
	ICE-ENN	3.5534	0.5371	0.4172	0.9544	0.9432	73.18	97.73	98.50
	ICE-TCN	2.8910	0.4340	0.3545	0.9621	0.9654	84.20	97.23	98.80
	ICE-GRU	3.6746	0.5441	0.4301	0.9532	0.9571	70.91	96.36	98.20
	ICE-QRNN	2.8741	0.4220	0.3398	0.9717	0.9638	84.89	97.89	99.00
	CFM	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	ICE-ANFIS	2.2827	0.3748	0.2978	0.9786	0.9737	90.00	93.40	98.30
	ICE- LSTM	2.3995	0.3989	0.3153	0.9758	0.9702	88.64	90.15	92.31
	ICE-CNN	3.0983	0.5160	0.4076	0.9595	0.9501	78.18	97.09	99.00
	ICE-ELM	2.2209	0.3599	0.2901	0.9303	0.9757	92.73	95.60	97.43
Site2: 10-min	**ICE-BPNN**	2.2658	0.3714	0.2954	0.9590	0.9742	79.09	98.64	99.55
	ICE-GRNN	1.0858	0.2071	0.1529	0.9435	0.9919	89.55	94.55	96.60
	ICE-ENN	3.1044	0.5411	0.4099	0.9554	0.9451	79.09	96.64	98.55
	ICE-TCN	2.3876	0.3847	0.3119	0.9722	0.9690	89.74	95.83	98.00
	ICE-GRU	3.1606	0.5159	0.4108	0.9595	0.9633	80.45	97.27	98.50
	ICE-QRNN	2.2531	0.3687	0.2937	0.9723	0.9714	91.52	97.03	98.11
	CFM	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	ICE-ANFIS	3.1424	0.4798	0.3782	0.9736	0.9561	81.36	97.73	98.50
	ICE- LSTM	3.3935	0.5259	0.4132	0.9683	0.9472	79.09	97.27	99.00
	ICE-CNN	4.6616	0.7052	0.5686	0.9430	0.9046	59.09	93.64	98.64
	ICE-ELM	3.0488	0.4652	0.3629	0.9752	0.9588	82.73	97.73	98.60
Site3: 10-min	**ICE-BPNN**	3.0550	0.4682	0.3647	0.9749	0.9582	85.45	96.09	97.25
	ICE-GRNN	2.9030	0.2869	0.2178	0.9709	0.9542	83.45	97.09	98.44
	ICE-ENN	4.0672	0.6429	0.4986	0.9526	0.9210	67.73	94.55	98.09
	ICE-TCN	3.4184	0.5315	0.4274	0.9630	0.9426	80.05	96.30	98.17
	ICE-GRU	4.1967	0.6376	0.5032	0.9534	0.9469	63.64	91.82	99.09
	ICE-QRNN	3.0251	0.4591	0.3563	0.9680	0.9536	82.29	97.81	99.00

Table 10 The forecasting results of CFM and other denoising strategy models.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	CEE	4.6156	0.5576	0.6120	0.7671	0.8889	70.00	92.22	94.30
	EEMD	4.6335	0.6368	0.6562	0.7853	0.8294	76.67	85.56	92.28
	EMD	3.7898	0.5845	0.4958	0.7826	0.9341	78.89	87.45	91.32
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	CEE	4.4341	0.7932	0.6434	0.9440	0.8842	72.22	94.44	99.09
	EEMD	4.5079	0.8711	0.6603	0.8088	0.8597	83.33	96.67	97.78
	EMD	3.3387	0.5709	0.4832	0.8602	0.9401	78.89	94.33	96.22
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	CEE	5.6386	0.9667	0.8056	0.9309	0.9276	52.22	94.44	97.42
	EEMD	4.3859	0.5918	0.4114	0.8848	0.9100	86.01	87.78	96.67
	EMD	2.8755	0.5203	0.4102	0.8148	0.9500	84.44	98.89	99.01

Table 11 The forecasting results of CFM and combined models using other algorithms.

SET	Model						PE		
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	MOALO	3.0294	0.4986	0.3570	0.9214	0.9429	82.78	97.44	98.12
	MODA	2.8897	0.4387	0.3570	0.9414	0.9629	83.78	94.55	98.82
	MOGOA	2.9192	0.4544	0.3781	0.9083	0.9603	81.56	93.22	97.56
Site2: 10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	MOALO	2.1900	0.4074	0.3187	0.9435	0.9725	92.22	93.50	98.12
	MODA	2.1948	0.4083	0.3102	0.9428	0.9725	84.32	94.31	97.66
	MOGOA	2.2048	0.3999	0.3285	0.9407	0.9717	85.56	92.31	96.45
Site3: 10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	MOALO	2.9797	0.5791	0.3655	0.9430	0.9577	87.78	97.89	99.00
	MODA	2.9075	0.4800	0.3700	0.9424	0.9575	82.43	94.57	98.75
	MOGOA	2.9708	0.4894	0.3804	0.9362	0.9558	84.44	92.34	98.88

4.6 Uncertainty forecasting

An interval estimate is based on a point estimate and can increase the prediction reliability and certainty [31]. In this subsection, four distributions were chosen to fit the sequence based on maximum likelihood estimation (MLE). They are Weibull, Gamma, Rayleigh and Lognormal distribution, which are shown in **Table 12**. After analysis, the Weibull distribution was adopted for the three sites, and the detailed parameter settings are presented in **Table 13**. Then, three indicators named FINAW, AWD, and FICP were designed for interval forecasting. The detailed results are depicted in **Fig.5** and **Table 14**.

The indicators were evaluated for the probabilities P_1=95%, P_2=90%, and P_3=80%. These probabilities can be expressed as $(1-\alpha)\times 100\%$. It is clear that the prediction accuracy at the significance level of α=0.05 was superior to the accuracies at α=0.1 and α=0.2. At the significance level of α=0.05, the FICP for CFM at every site exceeded 85%. In addition, at α=0.1, the FICP for CFM was between 80% to 85%. The similar values of AWD and FINAW at different significance levels implies that most of the actual values fell within the predicted ranges and that CFM achieved excellent performance.

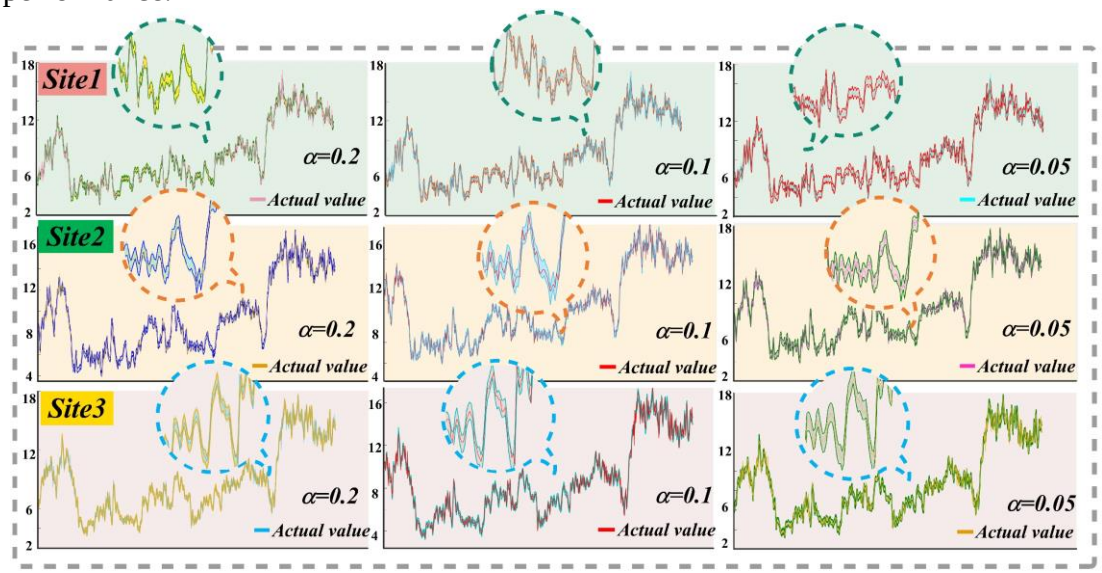

Fig.5 The interval forecasting results

5. Discussion

In this section, the results of improvement ratio, hyperparametric analysis, and the exploration of input and multi-step forecasting, operation time, comparative analysis and practical applicability are discussed to further analyze the above experimental results.

5.1 Improvement ratio

The aim of the improvement indicator is to quantify the improvement in the predictive precision of CFM. $\mathbf{IR}_{\mathbf{MAPE}}$, $\mathbf{IR}_{\mathbf{MAE}}$ and $\mathbf{IR}_{\mathbf{RMSE}}$ are designed to represent the improvement ratio. The details are shown in **Table 15**. For example, the models based on ICE for site1 had the values of $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ANFIS}$ =9.2670%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{LSTM}$ =4.1438%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{CNN}$ =19.0262%, $\overline{\overline{\mathbf{IR}}}_{MAPE}^{ELM}$ =4.5895%, which shows that the advantages of each model were

combined in CFM. For the different denoising methods, it is clear from the average value of 34% for $\overline{\overline{\mathbf{IR}}}^{ave}_{MAPE}$ in these models that ICE played a key role in improving the forecasting accuracy. The various optimization strategies achieved an average value of 7.3309% for $\overline{\overline{\mathbf{IR}}}^{ave}_{MAPE}$. This implies that MOGWO could improve forecasting ability. There is no doubt that CFM was superior to the single models and the combined strategy achieved excellent prediction results.

5.2 Hyperparametric analysis

To investigate the influence of key parameter changes on the proposed model, only one key parameter was changed each time while the other parameters remained the same in this section. We changed the values of five parameters in the ICE and MOGWO algorithms. For ICE, the maximum iteration parameters values were [500, 3000] and the number realizations were [50, 100, 300]. For MOGWO, the archive size parameters values were [200,300,500], the maximum iteration parameter values were [50,150, 200], and the population sizes were [50,70,120]. The calculation results are shown in **Table 16**. With the change of parameter values, the precision of the model also changed. For example, at site1, the number realizations were set to 50,100 and 300. When 100 was selected, the proposed model achieved good MAPE values with 2.8645%, 2.1843% and 2.8727% in three sets. In general, it is not difficult to find that the parameter values selected in this paper perform better.

5.3 Influence of input step and multi-step prediction

This section adds sites 4-7 to better discuss the performance of the proposed model. Specifically, sites 4-5 use wind speed data at 20-minute intervals and sites 6-7 apply wind speed data at 30-minute intervals. The sample size and experimental mechanism are consistent with the above. Considering the neatness and symmetry of the table content, in sections 5.3.1 and 5.3.2, site1 (10-min), site2 (10-min), site4 (20-min), site5 (20-min), site6 (30-min) and site7 (30-min) are used to test the performance of the proposed model.

5.3.1 Influence of input step

To analyze the influence of different input steps on prediction more accurately, the input steps $\boldsymbol{\lambda}$ were adjusted to 3, 4 and 6, and a comparative test was performed to further explore the performance of the proposed forecasting system. From **Table 17**, it is easy to see that for the 10-min sites, when the input step $\boldsymbol{\lambda}$ was 3, the prediction errors of $\mathbf{MAPE}^{site1}_{\boldsymbol{\lambda}=3}$ = 3.0833% and $\mathbf{MAPE}^{site2}_{\boldsymbol{\lambda}=3}$ = 2.4134% were larger at those of other steps. For the 20-min sites, $\boldsymbol{PE}^{5\%}_{\lambda=3}$ and $\boldsymbol{PE}^{5\%}_{\lambda=4}$ are less than 70%, which implies that better forecasting results were obtained by CFM compared to $\boldsymbol{PE}^{5\%}_{\lambda=5}$ =69.33% at site4 and $\boldsymbol{PE}^{5\%}_{\lambda=5}$ =72.22% at site5. The prediction accuracy for the 30-min sites was obviously inferior to those of the previously mentioned sites; however, the input step 5 still performed well. From the above analysis, it can be concluded that as the time interval increased, the prediction accuracy decreased continuously, but the best performance was still obtained when the input step was 5.

5.3.2 Influence of multi-step prediction

The multistep forecasting ability of the proposed system is explored in the subsection. The details of the two-and three-step predictions are presented in **Table 18**.

For site1, which had 10-min intervals, the smallest errors of $\mathbf{MAPE}_{CFM}^{step\,2}$ =4.4556%, $\mathbf{MAE}_{CFM}^{step\,2}$ =0.5848 were achieved by CFM compared to the other models, which had the average errors with MAPE and MAE values of 5.0829% and 0.7960 for two-step predictions. In addition, the accuracy of the three-step prediction was obviously lower than that of the two-step prediction, but the CFM still achieved the good result with $\boldsymbol{R}_{CFM}^{2}$ greater than 90%.

The accuracy decreased with longer interval times. For example, for the 20-min predictions by CFM, the MAPE values in two-step was 6.5225% at site4, which exceeded the values for the 10-min sites. For the 30-min predictions, the prediction error $\mathbf{MAPE}_{CFM}^{step\,2}$ increased to 9% at site6. These results indicate that the prediction system may be more suitable for short-term wind speed prediction.

5.4 Operation time

Table 19 shows the mean operation time of all models used in this paper. It is clear that the proposed system had the longest calculation time, at 264.8361s. For basic models, the running time of BPNN was 0.4111, which was the least time in all models. The computation time of models based on ICE denoising algorithm was also short, which was 57.9170s, 19.2999s, 65.4855s, 36.4029s, 20.4532s, 18.9124s and 22.8657s. In addition, when applying other noise reduction algorithms or optimization algorithms, the time spent on combined models increased significantly, which was in the 100s to 200s range. Although the time consumed of the proposed model was longer than other models, it had superior predictive power. Its time was within the acceptable range and did not affect its practical application greatly.

Table 19 Computing time.

Model	Computation time(s)	Model	Computation time(s)
CFM	264.8361	**MOALO**	197.8581
ICE-ANFIS	57.9170	**MODA**	207.2554
ICE-ELM	19.2999	**MOGOA**	198.3225
ICE-LSTM	65.4855	**BPNN**	0.4111
ICE-CNN	36.4029	**ENN**	1.4821
ICE-GRNN	20.4532	**ELM**	3.2169
ICE-BPNN	18.9124	**LSTM**	38.3067
ICE-ENN	22.8657	**CNN**	18.7405
EMD	137.0275	**ANFIS**	40.2649
EEMD	110.2881	**GRNN**	10.1546
CEE	203.0657		

5.5 Comparative analysis

Through the experiments above, the validity of the proposed model is proved. However, the results based only on the experimental mechanism in this paper may not be convincing. Therefore, in order to be fair, this section compares the proposed model with other similar studies. Specifically, Niu's model [32], Shao's model [29]and Liu's model [23]are used for comparative study. Niu et al. designed the combined model based on data preprocessing and optimization algorithm for wind speed forecasting, which obtained high accuracy with the smallest MAPE value of 2.89%. However, Niu et al. ignored the importance of uncertain predictions and the influence of parameters on model performance was not considered. Shao et al. used the same MOGWO algorithm and proposed a combined model based on decomposition-integration, but also lacked the exploration of uncertainty prediction and the influence of model

parameters. Liu et al. not only applied the improved optimization algorithm to point prediction and interval prediction of wind speed, but also considered the influence of different seasons on wind speed prediction. However, Liu's model lacked comparison with other literatures, and the results of parameter adjustment were not fully demonstrated. **Table 20** presents the comparison results of this study with other studies.

There is no unified comparison standard for model accuracy among different studies [49]. Therefore, to verify the accuracy of the proposed model, the evaluation index and datasets in this paper are applied in the models of Niu, Shao and Liu. Meanwhile, parameter settings are basically consistent with those in the original literature. Although the error between these models was not large, it was enough to show that the proposed model is superior to the models from comparative studies. For example, the MAPE value of the proposed model was 2.8645%, while Niu, Shao and Liu had values of 3.4181%, 4.0200% and 3.1835% respectively. The proposed model also had outstanding advantages in other indicators. Therefore, it can be reasonably concluded that the proposed model has excellent predictive ability.

5.6 Practical applicability

Efficient and timely wind speed prediction plays an important role in wind power generation system, which can not only meet people's demand for electricity, but also maximize economic benefits. Meanwhile, the proposed model in this paper can also be applied to other fields:

(a) Air quality forecasting. Deep learning technology has been widely used in air monitoring, this proposed prediction system based on deep learning methods can be applied to the monitoring of air quality, for example, to timely predict air pollutant concentrations and to provide reasonable suggestions for travelling and improving the environment.

(b) Traffic forecasting. In recent years, the prediction of traffic flow has become the focus of attention. This proposed system can sensitively capture the change of traffic flow, so as to reduce traffic congestion, relieve people's travel pressure and provide support for the development of intelligent transportation.

(c) Financial market forecasting. This system can also be used in the prediction of future trends in stocks, funds, etc, timely providing technical support for discovery of potential financial risks.

6. Conclusion and future work

Wind energy has received increasing attention as a renewable energy source to address the shortages in the energy market. Nevertheless, the fluctuation and instability of wind speed still present difficulties for wind speed forecasting. An advanced wind speed prediction system was developed in this study. The ICE denoising strategy was first used to eliminate noise to obtain stationary sequences based on the de-composition and ensemble theory. Several single models were then adopted to predict the processed data. The optimal weights and final forecasting results were obtained using the MOGWO operator. The Pareto optimality of the MOGWO solutions was theoretically proven to ensure that the optimal weight vector can be obtained in the combined system. The point prediction results show that the MAPE values of the proposed model were 2.8645%, 2.1843% and 2.8727% respectively. Besides, for uncertainty forecasting, the FICP values of the proposed model were 85.1697, 89.5410 and 88.0111 respectively at the significant level α = 0.05. The AWD values were 0.0559, 0.0400 and 0.0361 and the FINAW values were 0.0478, 0.0404 and 0.0390. It is clear that the proposed system

exhibits remarkable accuracy and stability performance. This system can hence provide accurate and real-time wind power information and contribute to wind power generation.

There are still some areas that need to be improved in future studies, for example: (1) In order to further improve the prediction accuracy, the influence of temperature, wind direction and other variables on wind speed prediction needs to be further discussed. (2) More advanced deep learning algorithms can be applied to future research to improve the accuracy of developed system. (3) Using more efficient preprocessing methods deals with unstable, random wind speed data. In addition, the effects of different preprocessing methods need to be further explored.

Table 12 Data distribution fitting.

Distribution	Site1		Site2		Site3	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Weibull	**0.9892**	**0.0299**	**0.9932**	**0.0239**	**0.9971**	**0.0156**
Gamma	0.9866	0.0334	0.9914	0.0267	0.9932	0.0238
Rayleigh	0.8903	0.0955	0.8579	0.1087	0.9066	0.0881
Lognormal	0.9827	0.0380	0.9872	0.0326	0.9857	0.0345

Table 13 Distribution parameters for three sites.

SET	Distribution	Parameter	
		λ	k
Site1: 10-min	**Weibull**	9.5454	3.3830
Site2: 10-min	**Weibull**	10.7378	3.6000
Site3: 10-min	**Weibull**	10.0906	3.1837

Note: λ and k are scale and shape parameter.

Table 14 Interval forecasting results.

Model	Expectation probability	Site1: 10-min			Site2: 10-min			Site3: 10-min		
		FICP	AWD	FINAW	FICP	AWD	FINAW	FICP	AWD	FINAW
CFM	**95%**	85.1697	0.0559	0.0478	89.5410	0.0400	0.0404	88.0111	0.0361	0.0390
	90%	83.8456	0.0646	0.0364	84.6134	0.0519	0.0339	81.1961	0.0641	0.0327
	80%	80.3296	0.0796	0.0305	81.0431	0.0659	0.0264	78.2684	0.0965	0.0255

Note: This table lists the interval forecasting results of CFM, which can further show the accuracy of CFM. The formula of **FICP, FINAW** and **AWD** are: $\mathbf{FICP} = \left(\sum_{i=1}^{Q} C_i / M\right) \times 100\%$, $\mathbf{FINAW} = \sum_{i=1}^{Q} (U_i - L_i) / FR$, if $A_i \succ U_i$, then $\mathbf{AWD}_i = (A_i - U_i) / (U_i - L_i)$, if $A_i \in [L_i, U_i]$, then $\mathbf{AWD}_i = 0$, if $A_i \prec L_i$, then $\mathbf{AWD}_i = (L_i - A_i) / (U_i - L_i)$. Thus, $\mathbf{AWD} = \left(\sum_{i=1}^{Q} \mathbf{AWD}_i\right) / FR$, FR represents the range of predicted values.

Table 15 The improvement percentages of CFM and all other models.

Model	Site1: 10-min			Site2:10-min			Site3: 10-min		
	IR_{MAPE}	IR_{MAE}	IR_{RMSE}	IR_{MAPE}	IR_{MAE}	IR_{RMSE}	IR_{MAPE}	IR_{MAE}	IR_{RMSE}
ICE-ANFIS	9.2670	2.4780	12.9021	4.3124	6.1470	6.4870	8.5813	8.3642	8.3569
ICE-LSTM	4.1438	14.8928	10.2339	8.9694	0.2655	0.0573	15.3468	0.8169	1.1439
ICE-CNN	19.0262	9.5438	11.4006	29.4999	22.4435	22.6521	38.3755	27.9343	26.2731
ICE-ELM	4.5895	16.0755	12.1302	1.6459	8.9708	10.8772	5.7750	12.9224	11.7576
ICE-BPNN	2.9585	14.4446	11.1923	3.5967	6.9986	7.4574	5.9677	12.3657	11.0349
ICE-GRNN	43.3893	40.3992	18.3251	11.1729	16.7673	32.7546	39.3307	28.1278	11.1886
ICE-ENN	15.7277	6.1089	10.3648	29.6390	22.8787	26.2376	29.3698	17.8166	19.1285
ANFIS	56.5473	51.7972	55.9334	55.7865	51.6294	54.5255	59.0283	51.5459	49.9590
LSTM	48.1109	41.5767	43.7026	55.7087	51.4914	52.9036	56.9152	49.2542	46.6145
CNN	48.9351	41.8068	43.3142	56.3344	51.4124	52.3438	56.8993	47.7365	45.3357
ELM	48.2710	41.8598	44.4566	54.9156	50.3353	52.3116	56.3115	48.4144	46.3876
BPNN	50.1068	43.8167	45.4931	54.6910	50.1487	52.2037	55.7754	47.3885	45.9918
GRNN	27.2087	29.3473	35.6913	47.5616	46.1051	40.8872	12.2467	8.5824	8.9760
ENN	49.4851	43.5831	45.5588	59.681	55.9942	57.4058	55.1765	46.1910	44.5643
CEE	35.1219	35.9979	36.4612	50.7386	50.8682	49.6846	49.0531	49.1288	46.2175
EEMD	46.8444	48.2038	48.6125	51.5454	52.1242	54.1845	55.0851	55.0361	52.3802
EMD	20.9860	20.9976	17.6387	34.5769	34.5880	30.0906	12.0987	15.1083	10.0810
MOALO	9.7114	9.7189	9.7564	2.0716	2.4013	3.0105	11.5294	12.1131	8.5268
MODA	9.7014	9.7088	9.7415	1.8440	1.8924	2.7919	10.1714	10.7422	8.3018
MOGOA	2.5798	3.5893	5.9401	2.7982	2.4714	2.0948	7.5586	7.7358	6.2351

Note: This indicator aims to verify the improvement of predictive precision of the CFM, the calculation formulas are: $IR_{MAPE} = (MAPE_{other} - MAPE_{CFM}) / MAPE_{other} * 100\%$, $IR_{MAE} = (MAE_{other} - MAE_{CFM}) / MAE_{other} * 100\%$ and $IR_{RMSE} = (RMSE_{other} - RMSE_{CFM}) / RMSE_{other} * 100\%$.

Table 16 Prediction results of different parameter values.

ICEMMDAN		MOGWO			Site1: 10-min			Site2: 10-min			Site3: 10-min		
Realization Number	**Iteration Number**	**Archive Size**	**Iteration Number**	**Population Number**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**
50	1000	100	100	200	2.9678	0.4785	0.3871	2.3445	0.4406	0.3419	3.4458	0.5905	0.4871
100*	**1000**	**100**	**100**	**200**	**2.8645**	**0.4114**	**0.3217**	**2.1843**	**0.3991**	**0.3111**	**2.8727**	**0.4299**	**0.3598**
300	1000	100	100	200	3.1287	0.4938	0.4068	2.2759	0.4194	0.3270	2.8058	0.5124	0.4011
100	500	100	100	200	3.0425	0.4854	0.3982	2.2270	0.4249	0.3254	2.8128	0.5050	0.3969
100	1000	100	100	200	3.1475	0.5275	0.4170	2.3009	0.4153	0.3307	2.7843	0.4883	0.3927
100	3000	100	100	200	3.2805	0.5445	0.4366	2.2667	0.4157	0.3287	2.9734	0.5320	0.4249
100	1000	200	100	200	2.9896	0.4805	0.3911	2.1855	0.3993	0.3163	2.8776	0.5202	0.4105
100	1000	300	100	200	2.9920	0.4808	0.3914	2.1862	0.3995	0.3164	2.8780	0.5206	0.4106
100	1000	500	100	200	2.9887	0.4804	0.3910	2.1851	0.3993	0.3162	2.8778	0.5205	0.4105
100	1000	100	50	200	2.9947	0.4812	0.3917	2.1859	0.3994	0.3163	2.8807	0.5207	0.4109
100	1000	100	150	200	2.9973	0.4818	0.3921	2.1847	0.3992	0.3162	2.8680	0.5192	0.4091
100	1000	100	200	200	2.9934	0.4812	0.3916	2.1862	0.3995	0.3164	2.8886	0.5217	0.4121
100	1000	100	100	50	3.0007	0.4823	0.3925	2.1847	0.3992	0.3162	2.8854	0.5215	0.4116
100	1000	100	100	70	2.9911	0.4807	0.3913	2.1844	0.3991	0.3161	2.8808	0.5211	0.4110
100	1000	100	100	120	2.9945	0.4813	0.3917	2.1862	0.3995	0.3164	2.8677	0.5192	0.4091

Note: * represents the parameter settings selected in this study.

Table 17 The forecasting results of CFM with different rolling input steps.

Rolling-input	Site1: 10-min				Site2: 10-min			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	3.0833	2.8955	**2.8645**	3.0694	2.4134	2.2406	**2.1843**	2.3213
RMSE	0.4967	0.4699	**0.4114**	0.4998	0.4405	0.4163	**0.3991**	0.4232
MAE	0.4025	0.3800	**0.3217**	0.4062	0.3494	0.3268	**0.3111**	0.3320
R^2	0.7710	0.7950	**0.9749**	0.7681	0.7976	0.8193	**0.9639**	0.8133
IA	0.9524	0.9574	**0.9754**	0.9518	0.9643	0.9681	**0.9942**	0.9672
PE (±5%)	78.89	81.11	**80.00**	80.00	88.00	88.89	**92.22**	87.78
PE (±10%)	98.89	97.21	**98.89**	97.89	95.64	94.45	**94.23**	96.25
PE (±15%)	99.01	98.46	**99.00**	99.01	98.72	99.12	**98.20**	98.85
Rolling-input	**Site4: 20-min**				**Site5: 20-min**			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	5.0079	5.2326	**4.8669**	5.1209	5.2472	5.1036	**4.1947**	4.9909
RMSE	0.3490	0.3561	**0.3328**	0.3477	0.3703	0.3751	**0.2999**	0.3647
MAE	0.2476	0.2569	**0.2428**	0.2528	0.2789	0.2748	**0.2204**	0.2692
R^2	0.9155	0.9120	**0.9231**	0.9161	0.9278	0.9259	**0.9526**	0.9299
IA	0.9635	0.9620	**0.9667**	0.9638	0.9621	0.9610	**0.9750**	0.9630
PE (±5%)	68.89	0.6222	**69.33**	65.56	58.89	61.11	**72.22**	62.22
PE (±10%)	82.22	0.8556	**82.22**	84.44	91.11	87.88	**92.22**	91.11
PE (±15%)	93.33	0.9333	**97.78**	95.56	97.78	96.67	**96.67**	96.67
Rolling-input	**Site6: 30-min**				**Site7: 30-min**			
	3	**4**	**5**	**6**	**3**	**4**	**5**	**6**
MAPE	5.5761	5.6650	**5.8810**	5.7209	5.8572	6.2507	**5.0589**	5.8738
RMSE	0.4880	0.4965	**0.4962**	0.4944	0.4518	0.4617	**0.3946**	0.4516
MAE	0.3982	0.4033	**0.4121**	0.4070	0.3445	0.3620	**0.3120**	0.3499
R2	0.9739	0.9730	**0.9730**	0.9732	0.9755	0.9744	**0.9813**	0.9755
IA	0.9440	0.9419	**0.9421**	0.9425	0.9483	0.9459	**0.9606**	0.9423
PE (±5%)	52.22	50.00	**54.44**	54.44	51.11	44.44	**57.78**	55.56
PE (±10%)	86.67	86.67	**84.44**	84.44	82.22	82.22	**92.22**	78.89
PE (±15%)	96.67	97.78	**96.67**	96.67	95.56	93.33	**96.67**	92.22

Table 18 The multi-step forecasting results of CFM and denoising models.

SET	Model	STEP2								STEP3							
		PE								PE							
		MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%	MAPE	RMSE	MAE	R^2	IA	±5%	±10%	±15%
Site1: 10-min	ICE-ANFIS	5.5741	0.9808	0.6722	0.8469	0.8108	58.18	85.00	95.91	7.0662	1.0365	0.8328	0.8284	0.7869	41.82	77.27	95.45
	ICE-LSTM	4.8109	0.7103	0.5638	0.9197	0.9005	59.09	88.64	98.64	6.6049	0.9362	0.7582	0.8600	0.8265	42.73	77.73	93.18
	ICE-CNN	5.4072	0.8220	0.6498	0.8925	0.8656	52.27	88.18	98.18	6.9907	0.9489	0.7874	0.8561	0.8240	41.36	74.55	90.91
	ICE-ELM	4.5395	0.6710	0.5277	0.9284	0.9112	63.64	89.55	99.09	6.7105	0.9762	0.7816	0.8477	0.8106	42.73	79.09	95.00
	CFM	**4.4556**	**0.7306**	**0.5848**	**0.8821**	**0.8967**	**56.67**	**92.22**	**98.89**	**5.3861**	**0.8419**	**0.6998**	**0.9198**	**0.8629**	**47.78**	**86.67**	**97.78**
Site2: 10-min	ICE-ANFIS	4.0231	0.6975	0.5311	0.9256	0.9086	68.64	93.18	99.09	5.9385	1.0245	0.7839	0.8382	0.8019	50.45	82.27	93.18
	ICE-LSTM	4.5031	0.7722	0.5976	0.9088	0.8880	59.55	91.82	98.18	5.9602	1.0106	0.7814	0.8425	0.8075	50.45	81.82	94.55
	ICE-CNN	5.2576	0.9032	0.7025	0.8752	0.8458	55.00	86.82	98.64	7.2749	1.2357	0.9857	0.8646	0.7084	49.55	72.73	90.91
	ICE-ELM	3.7898	0.6692	0.5035	0.9315	0.9159	69.09	95.00	99.55	5.7117	0.9487	0.7467	0.8612	0.8307	51.82	85.91	96.82
	CFM	**3.6019**	**0.4861**	**0.3826**	**0.9336**	**0.9564**	**88.89**	**95.45**	**99.11**	**5.1879**	**0.9662**	**0.7645**	**0.8446**	**0.8269**	**56.67**	**87.78**	**97.78**
Site4: 20-min	ICE-ANFIS	7.1044	0.5657	0.4441	0.9550	0.9205	46.36	74.55	88.18	10.1619	0.8281	0.6233	0.9032	0.8294	37.73	60.45	79.09
	ICE-LSTM	7.3513	0.5861	0.4581	0.9516	0.9145	46.36	76.36	89.55	10.2983	0.8190	0.6215	0.9053	0.8328	40.91	62.73	77.72
	ICE-CNN	8.3737	0.6659	0.5158	0.9376	0.8896	42.27	68.64	84.55	11.7013	0.9045	0.6820	0.8845	0.7974	37.27	60.91	72.73
	ICE-ELM	6.8610	0.5554	0.4302	0.9566	0.9233	48.64	77.73	88.64	10.0260	0.7985	0.6108	0.9100	0.8413	36.36	61.82	80.91
	CFM	**6.5225**	**0.5131**	**0.4617**	**0.9531**	**0.8942**	**54.44**	**73.33**	**89.22**	**9.3751**	**0.7594**	**0.6072**	**0.9062**	**0.8591**	**40.00**	**67.78**	**83.33**
Site5: 20-min	ICE-ANFIS	6.6738	0.7163	0.5199	0.9455	0.8806	49.09	78.18	92.27	9.7967	0.9733	0.7354	0.8992	0.7794	35.00	61.36	78.18
	ICE-LSTM	6.7348	0.7048	0.5177	0.9473	0.8843	51.82	76.36	92.37	9.7545	0.9507	0.7142	0.9038	0.7891	35.45	63.64	79.55
	ICE-CNN	7.6882	0.8352	0.5981	0.9260	0.8365	47.27	72.27	85.91	10.6698	1.0898	0.8183	0.8737	0.7205	31.36	75.45	58.64
	ICE-ELM	6.6218	0.6915	0.5120	0.9493	0.8887	48.18	80.00	91.09	9.6863	0.9514	0.7133	0.9037	0.7888	38.18	62.27	80.45
	CFM	**6.5292**	**0.5714**	**0.4248**	**0.9265**	**0.9095**	**50.00**	**77.78**	**92.78**	**9.0242**	**0.8712**	**0.6935**	**0.8937**	**0.7897**	**36.44**	**73.55**	**81.89**
Site6: 30-min	ICE-ANFIS	7.9222	0.8152	0.6341	0.9495	0.8530	38.64	71.82	89.09	11.0788	1.1669	0.8808	0.8967	0.6981	32.73	56.82	74.09
	ICE-LSTM	7.5799	0.7742	0.6062	0.9545	0.8674	42.27	73.18	90.45	11.1169	1.1475	0.8822	0.9001	0.7077	30.45	56.82	74.09
	ICE-CNN	8.8575	1.1007	0.7957	0.9079	0.7283	37.73	66.82	83.64	11.8460	1.2170	0.9456	0.8876	0.6740	28.64	54.55	68.64
	ICE-ELM	7.6516	0.7838	0.6102	0.9533	0.8641	39.09	72.27	89.09	10.8049	1.1544	0.8622	0.8989	0.7040	35.45	57.27	74.55
	CFM	**8.1519**	**0.6920**	**0.5413**	**0.9463**	**0.8865**	**43.89**	**72.22**	**86.67**	**10.5601**	**1.0711**	**0.8671**	**0.8670**	**0.7281**	**36.89**	**58.00**	**74.44**
Site7: 30-min	ICE-ANFIS	12.3948	1.1695	0.8588	0.8957	0.6809	26.82	54.55	71.82	12.3948	1.1695	0.8588	0.8957	0.6809	26.82	54.55	71.82
	ICE-LSTM	12.1247	1.0916	0.8220	0.9091	0.7221	28.64	55.00	71.36	12.1247	1.0916	0.8220	0.9091	0.7221	28.64	55.00	71.36
	ICE-CNN	12.6388	1.1586	0.8763	0.8976	0.6894	27.27	49.55	68.64	12.6388	1.1586	0.8763	0.8976	0.6894	27.27	49.55	68.64
	ICE-ELM	11.6304	1.0483	0.7875	0.9162	0.7438	31.82	56.36	72.27	11.6304	1.0484	0.7875	0.9162	0.7438	31.82	56..36	72.27
	CFM	**9.5321**	**1.0092**	**0.7905**	**0.8692**	**0.7401**	**45.56**	**51.11**	**77.78**	**10.5321**	**1.0092**	**0.7965**	**0.8692**	**0.7401**	**38.56**	**55.11**	**72.78**

Table 20 Comparison results.

Authors	Published year	Algorithm	Results	Similarity	Difference
Niu et al. [32]	2019	1. CEEMDAN 2. BPNN+GRNN+ENN+ELM +ARIMA 3. MOGOA	The combined model obtained high accuracy, the MAPE values of five data sets are 3.14%, 2.89%, 3.43%, 4.06% and 3.62% respectively.	Preprocessing methods, several single models and optimization algorithms are used. Different denoising strategies and optimization methods are compared.	Uncertain predictions and the influence of parameters on model performance are not discussed. The operation time of models is ignored.
Shao et al. [29]	2021	1. CEEMDAN 2. BPNN+GRNN+RBF+ ELM 3. MOGWO	Several experiments verify the effectiveness of the proposed model with the smallest MAPE value of 2.03%.	The same optimization algorithm is used. Different denoising strategies and optimization methods are compared.	Uncertain predictions and the influence of model parameters are not discussed. The operation time of models is ignored. Lack comparison with other literatures.
Liu et al. [23]	2019	1. ICEEMDAN 2. BPNN+GRNN+ENN+ELM +ARIMA 3. MMODA	The proposed model can improve the prediction accuracy. MAPE values of multi-step prediction are 3.15%, 4.41% and 5.02% respectively.	The same denoising strategy is applied. Multiple single models are applied. Different denoising strategies and optimization methods are compared.	The optimization algorithm is improved. Wind speed forecasts for different seasons are discussed. The result of parameter adjustment is not discussed in detail.

SET	Model	MAPE	RMSE	MAE	R^2	IA	PE ±5%	PE ±10%	PE ±15%
Site1:10-min	**CFM**	**2.8645**	**0.4114**	**0.3217**	**0.9749**	**0.9754**	**80.00**	**98.89**	**99.00**
	Niu's model	3.4181	0.4898	0.4322	0.7773	0.9541	32.44	58.42	75.28
	Shao's model	4.0200	0.6628	0.5380	0.7222	0.9149	33.58	61.06	78.36
	Liu's model	3.1835	0.4345	0.4064	0.8247	0.9639	30.97	56.37	73.01
Site2:10-min	**CFM**	**2.1843**	**0.3991**	**0.3111**	**0.9639**	**0.9942**	**92.22**	**94.23**	**98.20**
	Niu's model	2.3214	0.4077	0.3372	0.8267	0.9696	32.29	59.41	77.06
	Shao's model	3.0082	0.5493	0.4388	0.7853	0.9445	37.63	66.10	82.22
	Liu's model	1.8987	0.2917	0.2703	0.9113	0.9845	90.93	93.11	98.94
Site3:10-min	**CFM**	**2.8727**	**0.4299**	**0.3598**	**0.9751**	**0.9582**	**84.44**	**98.89**	**99.00**
	Niu's model	3.1730	0.5150	0.4480	0.8186	0.9515	38.32	51.23	79.02
	Shao's model	3.8116	0.6726	0.5498	0.7906	0.9163	39.57	54.24	72.78
	Liu's model	3.1508	0.4774	0.4442	0.8441	0.9584	40.91	60.58	88.60

Data availability

Because of the sensitive issues involved in the survey, the data would not be shared.

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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 71671029).

Appendix

Table A List of terminologies

NWP	numerical weather prediction	**GPR**	gaussian process regression
ARIMA	autoregressive integrated moving average	**ARMA**	autoregressive moving average
AI	artificial intelligence	**PSO**	particle swarm optimization
ANFIS	adaptive neuro-fuzzy inference system	**EEMD**	ensemble empirical mode decomposition
SSA	singular spectrum analysis	**WA**	wavelet transform
GM	gray prediction model	**CFM**	combined forecasting model
CEE	complete ensemble empirical mode decomposition with adaptive noise	**ICE**	improved complete ensemble empirical mode decomposition with adaptive noise
MOGWO	multi objective grey wolf optimizer	**PP**	point prediction
IP	interval prediction	**AR**	autoregressive model
ANN	artificial neural networks	**SVM**	support vector machines
LSTM	long short-term memory	**CNN**	convolutional neural network
BPNN	back propagation neural network	**ELM**	extreme learning machine
GRNN	general regression neural network	**ENN**	elman neural network
SEM	selection of excellent sub models	**AWD**	cumulative breadth error
RMSE	root mean square error	**MAE**	mean absolute error
MAPE	mean absolute percentage error	**R^2**	goodness of fit
IA	index of agreement of predictive results	**RE**	relative error
FICP	forecasting interval coverage probability	**FINAW**	prediction interval standardized mean breadth
MOGOA	multi-objective grasshopper optimization algorithm	**MOALO**	multi-objective antlion algorithm
MODA	multi-objective dragonfly algorithm	**EMD**	empirical mode decomposition
GRU	gated recurrent unit	**TCN**	temporal convolutional networks
QRNN	quasi-recurrent neural networks	**RNN**	recurrent neural network
WNN	wavelet neural network	**VMD**	variational mode decomposition
RBF	radial basis function	**PSR**	phase-space reconstruction
MOMVO	multi objective multi verse optimization	**RVFL**	random vector functional link network
WOA	whale optimization algorithm	**EPT**	ensemble patch transformation
SVR	support vector regression	**MSSA**	multi objective salp swarm algorithm

References

[1] Zhang Z, Qin H, Liu Y, Wang Y, Yao L, Li Q, et al. Long Short-Term Memory Network based on Neighborhood Gates for processing complex causality in wind speed prediction. Energy Conversion and Management. 2019;192:37-51. https://doi.org/10.1016/j.enconman.2019.04.006

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

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

[4] Allen DJ, Tomlin AS, Bale CSE, Skea A, Vosper S, Gallani ML. A boundary layer scaling technique for estimating near-surface wind energy using numerical weather prediction and wind map data. Applied Energy. 2017;208:1246-57. https://doi.org/10.1016/j.apenergy.2017.09.029

[5] Chu J, Yuan L, Pan L, Liu Q, Yan J, Liu Y. NWP Combination Correction Model Based on Variable-weight Stacking Algorithm. Energy Procedia. 2019;158:6309-14. https://doi.org/10.1016/j.egypro.2019.01.408

[6] Li L-L, Chang Y-B, Tseng M-L, Liu J-Q, Lim MK. Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm. Journal of Cleaner Production. 2020;270:121817. https://doi.org/10.1016/j.jclepro.2020.121817

[7] Hoolohan V, Tomlin AS, Cockerill T. Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. Renewable Energy. 2018;126:1043-54. https://doi.org/10.1016/j.renene.2018.04.019

[8] Dong L, Wang L, Khahro SF, Gao S, Liao X. Wind power day-ahead prediction with cluster analysis of NWP. Renewable and Sustainable Energy Reviews. 2016;60:1206-12. https://doi.org/10.1016/j.rser.2016.01.106

[9] Liu Z, Jiang P, Wang J, Zhang L. Ensemble forecasting system for short-term wind speed forecasting based on optimal sub-model selection and multi-objective version of mayfly optimization algorithm. Expert Systems with Applications. 2021;177:114974. https://doi.org/10.1016/j.eswa.2021.114974

[10] Grivel E, Diversi R, Merchan F. Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. Digital Signal Processing. 2021;116:103089. https://doi.org/10.1016/j.dsp.2021.103089

[11] Wang L, Tao R, Hu H, Zeng Y-R. Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. Renewable Energy. 2021;164:642-55. https://doi.org/10.1016/j.renene.2020.09.108

[12] Liu H, Tian H-q, Li Y-f. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Applied Energy. 2012;98:415-24. https://doi.org/10.1016/j.apenergy.2012.04.001

[13] Modaresi Movahed T, Jalaly Bidgoly H, Khoshgoftar Manesh MH, Mirzaei HR. Predicting cancer cells progression via entropy generation based on AR and ARMA models. International Communications in Heat and Mass Transfer. 2021;127:105565. https://doi.org/10.1016/j.icheatmasstransfer.2021.105565

[14] Fu W, Fang P, Wang K, Li Z, Xiong D, Zhang K. Multi-step ahead short-term wind speed

forecasting approach coupling variational mode decomposition, improved beetle antennae search algorithm-based synchronous optimization and Volterra series model. Renewable Energy. 2021;179:1122-39. https://doi.org/10.1016/j.renene.2021.07.119
[15] Wang Y, Wang J, Li Z, Yang H, Li H. Design of a combined system based on two-stage data preprocessing and multi-objective optimization for wind speed prediction. Energy. 2021;231:121125. https://doi.org/10.1016/j.energy.2021.121125
[16] Pousinho HMI, Mendes VMF, Catalão JPS. A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal. Energy Conversion and Management. 2011;52:397-402. https://doi.org/10.1016/j.enconman.2010.07.015
[17] Jiang P, Liu Z, Wang J, Zhang L. Decomposition-selection-ensemble forecasting system for energy futures price forecasting based on multi-objective version of chaos game optimization algorithm. Resources Policy. 2021;73:102234. https://doi.org/10.1016/j.resourpol.2021.102234
[18] Huang W, Liu H, Zhang Y, Mi R, Tong C, Xiao W, et al. Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. Applied Soft Computing. 2021;109:107541. https://doi.org/10.1016/j.asoc.2021.107541
[19] An F-P, Ma X-m, Bai L. Image fusion algorithm based on unsupervised deep learning-optimized sparse representation. Biomedical Signal Processing and Control. 2022;71:103140. https://doi.org/10.1016/j.bspc.2021.103140
[20] Gnatowski M, Buchaniec S, Brus G. The prediction of the polarization curves of a solid oxide fuel cell anode with an artificial neural network supported numerical simulation. International Journal of Hydrogen Energy. 2021. https://doi.org/10.1016/j.ijhydene.2021.09.100
[21] Du P, Wang J, Yang W, Niu T. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renewable Energy. 2018;122:533-50. https://doi.org/10.1016/j.renene.2018.01.113
[22] Wang J, Zhang L, Wang C, Liu Z. A regional pretraining-classification-selection forecasting system for wind power point forecasting and interval forecasting. Applied Soft Computing. 2021:107941. https://doi.org/10.1016/j.asoc.2021.107941
[23] Liu Z, Jiang P, Zhang L, Niu X. A combined forecasting model for time series: Application to short-term wind speed forecasting. Applied Energy. 2020;259:114137. https://doi.org/10.1016/j.apenergy.2019.114137
[24] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. Renewable Energy. 2021;165:573-94. https://doi.org/10.1016/j.renene.2020.11.002
[25] Li H, Liu T, Wu X, Li S. Research on test bench bearing fault diagnosis of improved EEMD based on improved adaptive resonance technology. Measurement. 2021;185:109986. https://doi.org/10.1016/j.measurement.2021.109986
[26] Li D, Jiang F, Chen M, Qian T. Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks. Energy. 2022;238:121981. https://doi.org/10.1016/j.energy.2021.121981
[27] Zhang X, Wang J, Gao Y. A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. Energy Economics. 2019;81:899-913. https://doi.org/10.1016/j.eneco.2019.05.026
[28] Wang D, Luo H, Grunder O, Lin Y. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renewable Energy. 2017;113:1345-58. https://doi.org/10.1016/j.renene.2017.06.095

[29] Shao Y, Wang J, Zhang H, Zhao W. An advanced weighted system based on swarm intelligence optimization for wind speed prediction. Applied Mathematical Modelling. 2021;100:780-804. https://doi.org/10.1016/j.apm.2021.07.024
[30] Xu Y, Yang W, Wang J. Air quality early-warning system for cities in China. Atmospheric Environment. 2017;148:239-57. https://doi.org/10.1016/j.atmosenv.2016.10.046
[31] Wang J, Cheng Z. Wind speed interval prediction model based on variational mode decomposition and multi-objective optimization. Applied Soft Computing. 2021;113:107848. https://doi.org/10.1016/j.asoc.2021.107848
[32] Niu X, Wang J. A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. Applied Energy. 2019;241:519-39. https://doi.org/10.1016/j.apenergy.2019.03.097
[33] He Y, Tsang KF. Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. Energy Reports. 2021;7:6473-88. https://doi.org/10.1016/j.egyr.2021.09.115
[34] Mirjalili S, Saremi S, Mirjalili SM, Coelho LdS. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. Expert Systems with Applications. 2016;47:106-19. https://doi.org/10.1016/j.eswa.2015.10.039
[35] Onyelowe KC, Shakeri J, Amini-Khoshalann H, Salahudeen AB, Arinze EE, Ugwu HU. Application of ANFIS hybrids to predict coefficients of curvature and uniformity of treated unsaturated lateritic soil for sustainable earthworks. Cleaner Materials. 2021;1:100005. https://doi.org/10.1016/j.clema.2021.100005
[36] Mahdevari S, Khodabakhshi MB. A hybrid PSO-ANFIS model for predicting unstable zones in underground roadways. Tunnelling and Underground Space Technology. 2021;117:104167. https://doi.org/10.1016/j.tust.2021.104167
[37] Ozkok FO, Celik M. A hybrid CNN-LSTM model for high resolution melting curve classification. Biomedical Signal Processing and Control. 2022;71:103168. https://doi.org/10.1016/j.bspc.2021.103168
[38] Shi X, Li Y, Yang Y, Sun B, Qi F. Multi-models and dual-sampling periods quality prediction with time-dimensional K-means and state transition-LSTM network. Information Sciences. 2021;580:917-33. https://doi.org/10.1016/j.ins.2021.09.056
[39] Elmaz F, Eyckerman R, Casteels W, Latré S, Hellinckx P. CNN-LSTM architecture for predictive indoor temperature modeling. Building and Environment. 2021;206:108327. https://doi.org/10.1016/j.buildenv.2021.108327
[40] İnik Ö, Altıok M, Ülker E, Koçer B. MODE-CNN: A fast converging multi-objective optimization algorithm for CNN-based models. Applied Soft Computing. 2021;109:107582. https://doi.org/10.1016/j.asoc.2021.107582
[41] Fetimi A, Dâas A, Benguerba Y, Merouani S, Hamachi M, Kebiche-Senhadji O, et al. Optimization and prediction of safranin-O cationic dye removal from aqueous solution by emulsion liquid membrane (ELM) using artificial neural network-particle swarm optimization (ANN-PSO) hybrid model and response surface methodology (RSM). Journal of Environmental Chemical Engineering. 2021;9:105837. https://doi.org/10.1016/j.jece.2021.105837
[42] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM)

forecasts using a GPR (Gaussian Process Regression) model. Energy. 2015;93:41-56. https://doi.org/10.1016/j.energy.2015.08.045

[43] Bardhan A, Samui P, Ghosh K, Gandomi AH, Bhattacharyya S. ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. Applied Soft Computing. 2021;110:107595. https://doi.org/10.1016/j.asoc.2021.107595

[44] Zhang Y, Aslani F, Lehane B. Compressive strength of rubberized concrete: Regression and GA-BPNN approaches using ultrasonic pulse velocity. Construction and Building Materials. 2021;307:124951. https://doi.org/10.1016/j.conbuildmat.2021.124951

[45] Izonin I, Tkachenko R, Gregus ml M, Zub K, Tkachenko P. A GRNN-based Approach towards Prediction from Small Datasets in Medical Application. Procedia Computer Science. 2021;184:242-9. https://doi.org/10.1016/j.procs.2021.03.033

[46] Wang J, Zhou Y, Li Z. Hour-ahead photovoltaic generation forecasting method based on machine learning and multi objective optimization algorithm. Applied Energy. 2022;312:118725. https://doi.org/10.1016/j.apenergy.2022.118725

[47] Liu J, Meng X, Ma Y, Liu X. Introduce canopy temperature to evaluate actual evapotranspiration of green peppers using optimized ENN models. Journal of Hydrology. 2020;590:125437. https://doi.org/10.1016/j.jhydrol.2020.125437

[48] Jiang P, Liu Z, Niu X, Zhang L. A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. Energy. 2021;217:119361. https://doi.org/10.1016/j.energy.2020.119361

[49] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Transactions on Power Systems. 1989;4:1484-91. https://doi.org/10.1109/59.41700