A cooperative ensemble method for multistep wind speed probabilistic forecasting

He, Yaoyao; Wang, Yun; Wang, Shuo; Yao, Xin

DOI:
10.1016/j.chaos.2022.112416

Document Version
Peer reviewed version

Citation for published version (Harvard):
He, Y, Wang, Y, Wang, S & Yao, X 2022, 'A cooperative ensemble method for multistep wind speed probabilistic forecasting', *Chaos, Solitons and Fractals*, vol. 162, 112416. https://doi.org/10.1016/j.chaos.2022.112416

A Cooperative Ensemble Method for Multistep Wind Speed Probabilistic Forecasting

Yaoyao He [a,b,*], Yun Wang [a,b], Shuo Wang [c], Xin Yao [c,d]

[a] School of Management, Hefei University of Technology,Hefei 230009, China

[b] Key Laboratory of Process Optimization and Intelligent Decision-Making (Hefei University of Technology), Ministry of Education, Hefei 230009, China

[c] School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[d] School of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Abstract

Accurate wind speed forecasting is of great significance to ensure the safe utilization of wind power. However, the randomness and volatility nature of wind speed give rise to an enormous challenge to the precision of wind speed forecasting. Combining the data preprocess technology, feature selection method, forecasting model, optimization algorithm and data postprocessing technology, the complete ensemble empirical mode decomposition with adaptive noise-least absolute shrinkage and selection operator-quantile regression neural network (CEEMDAN-LASSO-QRNN) model is developed to preform multistep wind speed probabilistic forecasting. Within the proposed model, CEEMDAN technology is firstly employed to decompose original wind speed timeseries into several subsequences. For each subsequence, the explanatory variables constructed by a hybrid multistep forecasting strategy are selected by LASSO regression. Subsequently, QRNN forecasting models are established to obtain multistep conditional quantiles predictions for entire subsequences. Ultimately, the aggregated quantiles are served as the samples to fit approximate distribution through kernel density estimation (KDE), thus obtaining the probability density function, further achieving probabilistic predictions, interval predictions and point predictions. The case studies including four real datasets are provided to validate the dependability and feasibility of the proposed model. Experimental results indicate higher accuracy and robustness of the proposed model occur in multistep wind speed probabilistic forecasting.

Keywords: wind speed forecasting, cooperative ensemble method, CEEMDAN decomposition, multistep probabilistic forecasting

1. Introduction

As the most key aspect in wind research, wind speed prediction plays an important role in meteorological forecast, meteorological disaster prevention, air pollution prediction, renewable energy utilization and so on[1, 2]. The stochastic and fluctuant nature of wind speed result in a great uncertainty in prediction process that is

*Corresponding author. School of Management, Hefei University of Technology, Hefei 230009, China

Email addresses: hy-342501y@163.com (Yaoyao He [a,b,*]), wendywangy1220@163.com (Yun Wang [a,b]), s.wang.2@bham.ac.uk (Shuo Wang [c]), X.Yao@cs.bham.ac.uk (Xin Yao [c,d])

challenging to quantify through traditional deterministic forecasting methods [3]. Therefore, an increasing amount of work has started to focus on probabilistic prediction for more comprehensive forecasting information [2–5]. The probabilistic forecasting methods can better express the uncertainty, in the forms of probability density functions (PDFs), cumulative distribution functions (CDFs), intervals and quantiles, discrete probabilities and moments of probability distribution such as mean, variance and skewness [6]. More specifically, wind speed can be regarded as a random variable to fit PDFs or CDFs, which is preferred most by decision maker in probabilistic forecasting. The intervals and some noteworthy moments like mean, median, mode, can be derived from the PDFs [7].

Existing wind speed probabilistic prediction methods can be roughly divided into parametric methods and nonparametric methods [8]. Relying on rational hypothesis and prior knowledge, parametric methods assume the probability distribution of the future point in advance [9]. By estimating several parameters, the parametric methods are simple and efficient. However, the commonly used hypothesized distributions such as Gaussian distribution [10] and Beta distribution [11] tend to deviate greatly from the real distribution of wind speed, thereby leading to low prediction accuracy. Without a pre-defined distribution, the nonparametric methods make no assumptions [8]. Nonparametric methods relatively complicate the estimating process, but greatly improve accuracy and adaptability in majority scenarios. Consequently, nonparametric methods are dominating current probabilistic forecasting methods. Prediction intervals (PIs) can be estimated by Bootstrap [12], Bayesian [13] or other nonparametric methods. Compared with interval forecasting, the quantile form is capable of providing more insightful information. Besides, as a kind of regression idea, quantiles regression (QR) can integrate with diverse forecasting methods to obtian quantiles predictions. Thereinto, the integration of QR and diverse artificial intelligence (AI) methods have become the mainstream in recent years[14, 15]. When it comes to probability density forecasting, the kernel density estimation (KDE) is a critical role. Some predicted values of one point are treated as sample to fit its PDF by KDE. For example, the quantiles predictions can be converted into probability density functions [16]. With aforementioned superiorities over intervals and quantiles, probability density forecasting has great advantages in revealing predictive information, thereby mitigating the uncertainty, enhancing reliability and guaranteeing robustness.

Despite individual models have potential of exerting own value to an extreme in some particular scenarios and conditions. Restricted by multifarious factors, these single models embrace poor adaptability and robustness. Furthermore, prediction accuracy fails to be satisfied and amelioration is also limited to large degree in some situations. Therefore, the attention paid on the ensemble forecasting methods is rising constantly for seeking precision and stability. Ren et.al [17] reviewed the ensemble forecasting methods in wind forecasting field, and classified them into two types: competitive ensemble methods and cooperative ensemble methods. The competitive ensemble methods create ensemble predictions constructed by different initial conditions, such as various predictors and different parameters [18]. The forecasting results of all the models or the selected models after pruning are integrated eventually. The ways to generate ensemble predictions are various. Bootstrap method was exploited to generate different training sample trained independently with back propagation neural network (BPNN), thereby obtaining different forecasting results [19]. Ma et.al [20] proposed a model-free framework named randomly distributed embedding (RDE), which achieves accurate predictions of short-term high-dimensional data. By randomly generating explanatory variables multiple times, the RDE framework can obtain ensemble predictions to fit probability distribution.

Differently, the cooperative ensemble methods divide the prediction task into two subtasks: one is mast task to complete forecasting and the other is auxiliary task in order to improve the prediction accuracy and consolidate the prosperity ability [18]. Auxiliary task contains data pre-processing, feature selection, parameter optimization, error correction, data post-processing and so on. The combination of various algorithms gives this method more room for optimization and improvement. Especially, the application of decomposition in cooperative ensemble methods has been greatly developed. The empirical wavelet transform (EWT) decomposition combined with cuckoo search (CS) algorithm is developed to assist core forecasting model in [21], which is a kind of cooperative ensemble methods. A novel hybrid architecture for multi-step electricity load forecasting is proposed in [22]. The complementary ensemble empirical mode decomposition (CEEMD) collaborates with backpropagation neural network improved by particle swarm optimization (PSO-BP) for forecasting, and the variational mode decomposition (VMD) and PSO-BP is established for prediction of the subsequent error. In [23], a new hybrid wind speed forecasting model is developed, which integrates long short-term memory neural network (LSTM), decomposition methods and grey wolf optimizer (GWO).

In view of the rapid development of cooperative ensemble methods based on decomposition and the preponderance of probabilistic forecasting, this paper focuses on the application of cooperative ensemble probabilistic methods based on decomposition for wind speed forecasting. Like this method, quite a few have been proposed in literature. Jiang et.al [2] proposed an innovative hybrid method consisting of EMD, VMD and conditional KDE. Cheng et.al [24] developed a an ensemble model for probabilistic wind speed forecasting based on wavelet threshold denoising (WTD), recurrent neural network (RNN) and adaptive neuro fuzzy inference system (ANFIS). More applications of cooperative ensemble were summarized in [5]. The ideal of cooperative ensemble methods based on decomposition was shown to have a superior probabilistic forecasting performance in rich literature. However, it inevitably brings about some problems, summarized as follows.

(1) Traditional ensemble methods only utilize same algorithm to predict all the subsequences but ignore the fact that different subsequences are comprised of different signal characteristics. How to select the optimal forecasting model for specific subsequence is utterly important.

(2) The large fluctuations hidden in partial subsequences may negatively affect on overall prediction accuracy, which is neglected by majority of ensemble forecasting methods. This issue can be avoided by reprocessing decomposed data sequences as a secondary processing step. The reprocessing could significantly improve accuracy.

(3) Considering that the traditional ensemble model only employs single optimization objective, the dilemma of trapping into local optimization easily occurs. Exploiting appropriate objective functions for corresponding forecasting forms can be a pivotal task in wind speed forecasting.

With the above issues in mind, a novel cooperative ensemble method for wind speed probabilistic forecasting is presented in this paper. More specifically, the initial complicated wind speed time series are firstly decomposed by CEEMDAN method to several subsequences with relative terseness and stabilization. Secondly, the selection of vital variables by LASSO regression and the establishment of QRNN forecasting models are implemented respectively for every subsequence. Then, the parameters of QRNN involved in each subsequence are optimized with different objective functions under different prediction forms. Next, the predicted quantiles of each subsequence

are aggravated at same quantile points. Finally, the integrated quantiles are treated as the samples to output an approximate probability density function. The main contributions of this paper can be summarized as follows.

(1) Motivated by the advantages of cooperative ensemble probabilistic forecasting methods based on decomposition, this paper ingeniously integrates master probabilistic forecasting model based on artificial intelligence and several auxiliary methods, which contain a data decomposition technique, a feature selection method, parameter optimization algorithms and a data postprocessing approach. All together form our novel cooperative ensemble method CEEMDAN-LASSO-QRNN model for wind speed multistep probabilistic prediction aiming at better accuracy and robustness.

(2)The massive and complex characteristics inherent in wind speed render an unsatisfactory prediction performance. Data decomposition is an effective tool to dig up more valuable information. However, there are still some subsequences with large randomness and volatility after the decomposition. Taking the historical wind speed data into consideration as much as possible can help us tackle this issue. However, the high dimensional explanatory variables may result in over-fitting and extra time consumption. Therefore, we make use of LASSO regression to eliminate redundant information, so that the prediction accuracy and operational efficiency can be improved. The QRNN can interpret intricate nonlinear relationship between wind speed data, and further combining with KDE, it can fit the probability density function to obtain more comprehensive prediction information.

(3) Due to the large distinctions that the subseries may embrace from each other, the features selection and forecasting procedures are established severally aiming at every subsequence. On one hand, the parameters in LASSO regression for feature selection are optimized by GCV method. Different parameters for each subsequence further lead to the disparities in the chosen variables. On the other hand, the parameters in QRNN forecasting model for each subsequenceare are adjusted by the gird search method. By virtue of these independent modeling processes, the prediction accuracy of all subsequences can be guaranteed, thus maximizing predictive performance.

(4) More in line with practical applications, a three-stage model is adopted in this paper, consisting of model training, model validation as well as model testing. Compared with traditional two-stage model, the inclusion of validation aims to tune parameters more sensibly. Meanwhile, for preventing local optimization, the objective function changes with the prediction forms in the validation stage. The optimal parameters can thus be achieved for each forecasting form.

(5) Case studies are carried out to examine the effectiveness of our proposed model including four real wind speed datasets. For more convincing results, a multistep probabilistic wind speed forecasting is designed. In addition, the interval and point predictions derived from the probability density function are also provided in this paper for full verification.

The remaining parts of this paper are organized as follows. Section 2 introduces the relative methodologies applied in the proposed model. The model training process is elaborated in Section 3. Section 4 demonstrates the improvement of the proposed model in prediction accuracy and model robustness through case studies including four datasets. Finally, further discussions and conclusions are given in Section 5.

2. Methodologies

2.1. CEEMDAN Decomposition

Data preprocessing methods have been popular in wind speed forecasting in recent years, such as data reconstruction, data decomposition, data transformation and other forms, which can effectively facilitate the prediction. In particular, decomposition techniques have been the most widely developed and applied[25–27] Empirical mode decomposition (EMD)[28], one of the most popular decomposition techniques, plays a crucial role in disposing intricate timeseries. As the improvement of EMD, ensemble empirical mode decomposition (EEMD) [29] and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [30] were proposed to reduce the mode mixing effect by adding Gaussian white noise. Compared to EEMD that has time consuming and residual noise issues. CEEMDAN added Gaussian white noise adaptively based on EEMD, further cubing mode mixing phenomenon and enhancing the algorithm convergency. A complete algorithm procedure for decomposing a timeseries $Y = \{y_i\}_{i=1,2,\ldots,I}$ using CEEMDAN is described as follows.

Step 1: Add a set of Gaussian white noise $\{\omega^s\}_{1,2,\ldots,S}$, where ω^s denoted the $s-th$ Gaussian white noise series with zero mean and unit variance. S new series are obtained and listed in Eq (1)

$$Y^s = Y + \omega^s, s = 1, 2, \ldots, S \tag{1}$$

Step 2: Utilize EMD to decompose Y^s to obtain the first subsequence c_1^s, and repeat it S times for each new series to produce $\{c_1^s\}_{s=1,2,\ldots,S}$. As a result, the first intrinsic mode functions (IMF) $c_1 = \frac{1}{S}\sum_{s=1}^{S} c_1^s$ and residue $r_1 = Y - c_1$ can be calculated.

Step 3: Add ω^s to the residue for S times. The new sequences $r_1^s = r_1 + \omega^s, s = 1, 2, \ldots, S$ are decomposed by EMD to obtain the first subsequence $\{c_2^s\}_{s=1,2,\ldots S}$.

Step 4: Repeat Step 3 until the residue becomes a monotonic function, at which time the residue cannot be decomposed by EMD. Supposing that there are L IMFs $\{c_l\}_{l=1,2,\ldots,L}$ and a residue r_L obtained, the original timeseries is formulated as Eq (2).

$$Y = \sum_{l=1}^{L} c_l + r_L \tag{2}$$

2.2. LASSO regression

In order to extract pivotal information, more explanatory variables tend to be kept in modeling, which undoubtedly results in low accuracy and high time consumption. LASSO regression is a variant of ridge regression by substituting L2 regularization term with L1 [31], which explores the linear relationship between explanatory variables X and response variables Y. In addition to alleviating over-fitting and multicollinearity problems, LASSO regression is able to reduce the number of explanatory variables. The learning objective of LASSO can be formulated as Eq (3).

$$\min_{\beta} \left\{ \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{3}$$

in which β represents the coefficient set of explanatory variables and λ is the nonnegative regularization parameter to tune the shrinkage. The larger the value is, the greater the punishment intensity is.

LASSO regression compresses the regression coefficients to exactly zero. Then, the variables with zero coefficient are viewed as nonsignificant, which will be eliminated. Eq (3) can be transformed into an objective function along with a constraint condition, presented as follows.

$$\begin{aligned} &\widehat{\beta}^{LASSO} = \arg\min\left\{\|Y - X\beta\|_2^2\right\} \\ &s.t. \|\beta\|_1 \leq \eta \end{aligned} \tag{4}$$

where $\widehat{\beta}^{LASSO}$ represents the estimator of independent variables in LASSO regression and the constraint parameter η is negatively correlated with λ, which can be preset or optimized. As a quadratic programming problem, Eq (4) cannot be solved by calculus. Therefore, least angle regression (LARS) is used instead.

2.3. Quantile regression neural network (QRNN)

QRNN [32] combines the advantages of quantile regression (QR) [33] and artificial neural networks (ANN) [34]. On the basis of QR theory and the standard multi-layer perceptron ANN, an estimate of the conditional τ quantile of Y is expressed as the following formula.

$$Q_Y\left(\tau \left| X \right.\right) = F\left[X, H\left(\tau\right), O\left(\tau\right)\right] \tag{5}$$

in which $Y = \{y_i\}_{i=1,2,\ldots,I}$ is the response variable, $X = \{x_i\}_{i=1,2,\ldots,I}$ represents the explanatory variables in which $x_i = \{x_{i,j}\}_{j=1,2,\ldots,J}$, $H\left(\tau\right) = \{h_{j,k}\left(\tau\right)\}_{j=1,2,\ldots,J;k=1,2,\ldots,K}$ is assigned as the connection weight of input layer and hidden layer and $O\left(\tau\right) = \{o_k\left(\tau\right)\}_{k=1,2,\ldots,K}$ denotes the weight to connect hidden layer with output layer. For a particular sample $[x_i, y_i]$, this nonlinear function $F\left(\cdot\right)$ can be expressed as follows.

$$F\left[x_i, H\left(\tau\right), O\left(\tau\right)\right] = f\left\{\sum_{k=1}^{K} o_k\left(\tau\right) g\left[\sum_{j=1}^{J} h_{j,k}\left(\tau\right) x_{i,j} + b_{i,j}^h\right] + b_i^o\right\} \tag{6}$$

where $\left\{b_{i,j}^h\right\}_{i=1,2,\ldots,I;j=1,2,\ldots,J}$ is on behalf of the bias of hidden layer and $\{b_i^o\}_{i=1,2,\ldots,I}$ is the bias corresponding to output layer. The two activation functions $g\left(\cdot\right)$ and $f\left(\cdot\right)$ are usually designated as a sigmoid function and a linear model. The weight parameters $H\left(\tau\right)$ and $O\left(\tau\right)$ can be estimated by minimizing the following objective function.

$$\min_{[H(\tau),O(\tau)]} \sum_{i=1}^{I} \rho_\tau\left\{y_i - F\left[x_i, H\left(\tau\right), O\left(\tau\right)\right]\right\} + \xi \sum_{j=1}^{J}\sum_{k=1}^{K}\left[h_{j,k}\left(\tau\right)\right]^2 \tag{7}$$

in which ξ is positive constant that regulates the relative contribution of the weight decay term for avoiding over-fitting and $\rho_\tau\left(v\right)$ represents an asymmetric loss function expressed in Eq (7). The confirmation of ξ and the number of hidden layer nodes need to be predetermined or optimized. Suppose that there are T quantiles denoted as $\{\tau_t\}_{t=1,2,\ldots,T}$, a set of conditional quantiles $\left\{\widehat{Q}_Y\left(\tau_t \left| X \right.\right)\right\}_{i=1,2,\ldots,T}$ estimators are obtained eventually.

$$\rho_\tau\left(v\right) = \begin{cases} \tau v & v \geq 0 \\ \left(\tau - 1\right) v & v < 0 \end{cases} \tag{8}$$

2.4. Kernel density estimation

KDE is a non-parametric method that bridges quantiles forecasting and probability density forecasting without assuming the form of the probability distribution [34]. The conditional quantiles $\left\{\widehat{Q}_Y\left(\tau_t\left|X\right.\right)\right\}_{i=1,2,\ldots,T}$ are treated as samples to fit an optimal estimator of actual distribution. The KDE method is defined as the following formula.

$$\widehat{f}_d\left(q\right)=\frac{1}{TB}\sum_{t=1}^{T}K\left(\frac{q-\widehat{Q}_Y\left(\tau_t\left|X\right.\right)}{B}\right) \tag{9}$$

in which B denotes the bandwidth. $K(\cdot)$ represents the Epanechnikov kernel function, which is defined as:

$$K\left(\alpha\right)=\begin{cases}\frac{3}{4}\left(1-\alpha^2\right) & \alpha\in[-1,1]\\ 0 & \alpha\notin[-1,1]\end{cases} \tag{10}$$

2.5. Direct-Recursive (DR) multistep forecasting strategy

Typically, the further the horizon we attempt to forecast, the more difficult the forecasting is, because of the larger uncertainty. Therefore, several multistep forecasting strategies are introduced based on the recursive and direct strategies in [35]. The recursive strategy is the most traditional multi-step forecasting method, in which the first step needs to be trained at the beginning. Then, after which the first predicted value is treated as known information for predicting the next one. When using the recursive training strategy, inaccuracy may accumulate as the forecasting step moves on. The direct training strategy works differently, without intermediate forecasting. It indicates that H-step ahead forecasting signifies H models to be established. Nevertheless, the independent learning process neglects complicated correlations between variables, which may lead to a negative impact on prediction.

To combine the advantages of both strategies, a hybrid multi-step forecasting method is introduced denoted as DR strategy. Concretely speaking, for H-step in a timeseries $Y=\{y_i\}_{i=1,2,\ldots,I}$, an independent model f_h is built as follows.

$$y_{i+h}=f_h\left(y_{i+h-1},y_{i+h-2},\ldots,y_{i-J+1}\right)+\theta \tag{11}$$

in which $i=J,J+1,\ldots,I-h$, θ represents the error term. Therefore, the H-step ahead predictions are given by:

$$\widehat{y}_{I+h}=\begin{cases}\widehat{f}_h\left(y_I,y_{I-1},\ldots,y_{I-J+1}\right) & \text{if } h=1,\\ \widehat{f}_h\left(\widehat{y}_{I+h-1},\ldots,\widehat{y}_{I+1},y_I,\cdots,y_{I-J+1}\right) & \text{if } h\in\{2,\ldots,H\}\end{cases} \tag{12}$$

in which $\left\{\widehat{f}_h\right\}_{h=1,2,\ldots,H}$ denotes the multistep forecasting models.

3. The cooperative ensemble model based on CEEMDAN-LASSO-QRNN model

The cooperative ensemble model proposed in this paper, intergates CEEMDAN for data processing, LASSO regression with GCV for feature selection, QRNN optimized by gird search for forecasting and KDE for postprocessing. This ensemble model is named as CEEMDAN-LASSO-QRNN. In addition, a hybrid multistep forecasting strategy is introduced for wind speed multistep probabilistic forecasting. For evaluating this model more comprehensively, the interval and point forecasting are provided except for the probabilistic forecasting. The flowchart of the proposed model is depicted in Fig. 1. The details on how this ensemble model works are given in this section.

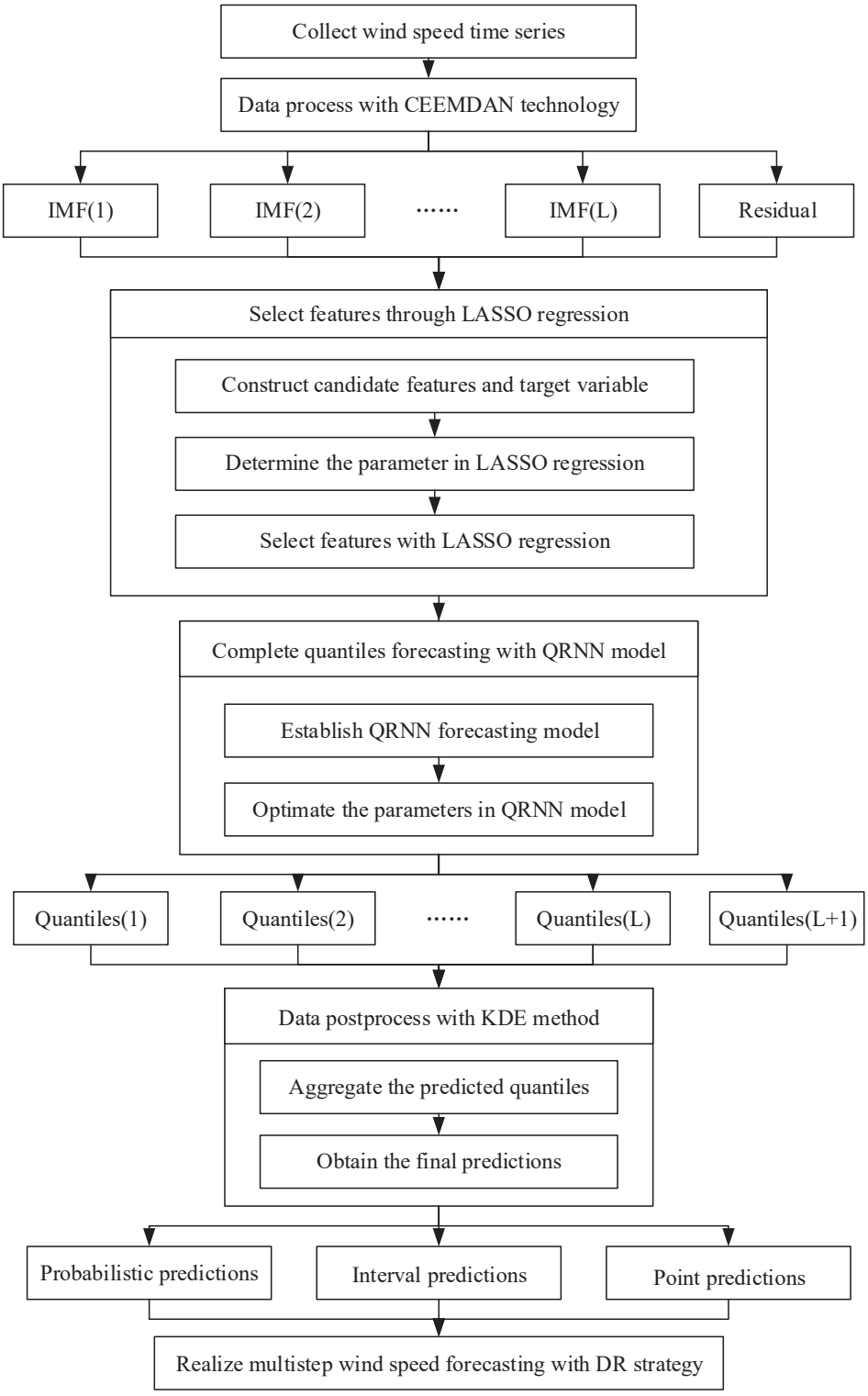

Fig. 1: The flowchart of CEEMDAN-LASSO-QRNN model

3.1. Data preprocess with CEEMDAN

Wind speed time series often embrace complexity and irregularity. Data preprocessing is thus necessary for accurate wind speed forecasting. For a given time series $Y = \{y_i\}_{i=1,2,\ldots,I}$, CEEMDAN is employed to decompose it into a cluster of IMF subsequences $Y = \{y_i\}_{i=1,2,\ldots,I}$ and a residue series r_L. The samples of certain subsequences c_l are indicated as $\left\{y_i^l\right\}_{i=1,2,\ldots,I}$ in which $l = 1, 2, \ldots, L$, as $\left\{y_i^{L+1}\right\}_{i=1,2,\ldots,I}$ for r_L correspondingly. They all satisfy the formula shown below.

$$y_i = \sum_{l=1}^{L} y_i^l + y_i^{L+1}, i = 1, 2, \ldots, I \tag{13}$$

3.2. Select features through LASSO regression

Owing to the stochastic and irregularity nature in wind speed time series, it is very difficult to extract useful information. Too many explanatory variables are constructed aiming to capture more comprehensive information. However, it leads to high-dimensional data structure that makes modelling for wind speed modelling extremely difficult. To solve the dilemma, the LASSO regression is applied for selecting significant features. For every subsequence, LASSO regression is executed independently as follows.

3.2.1. Construct candidate features and target variables

For the $l-th$ subsequence $\left\{y_i^l\right\}_{i=1,2,\ldots,I}$, in order to extract as much valid information from historical data as possible, the previous J historical points $\left\{y_{i-j}^l\right\}_{j=1,2,\ldots,J}$ are chosen as candidate features and the next points y_i^l are treated as the target variable through rolling one-step forecasting method.

3.2.2. Select features with LASSO regression

The linear relationship between the candidate features $\left\{y_{i-j}^l\right\}_{j=1,2,\ldots,J}$ and the target variables y_i^l is explored by LASSO regression firstly. A multiple regression model estimated by LASSO method can be expressed as below.

$$y_i^l = \widehat{\beta}_1^{LASSO,l} y_{i-1}^l + \widehat{\beta}_2^{LASSO,l} y_{i-2}^l + \cdots + \widehat{\beta}_J^{LASSO,l} y_{i-J}^l \tag{14}$$

where $\widehat{\beta}^{LASSO,l} = \left\{\widehat{\beta}_j^{LASSO,l}\right\}_{j=1,2,\ldots,J}$ are the estimated coefficients of independent variables. Through LASSO regression, the coefficients of insignificant variables are compressed to zero. The features with zero coefficient will be eliminated, and the remaining ones are treated as explanatory variables. Supposing that the dimension of features vector is shrunk from J into J^l, the explanatory variables are expressed as $x_i^l = \left\{x_{i,j}^l\right\}_{j=1,2,\ldots,J^l}$ to facilitate modeling. Remarkably, the shrinkage is adjusted by the parameter η^l, which is determined by the GCV method in this paper. The parameters in all the subsequences varies from each other, thus leading to the difference between the selected variables of each subsequence.

3.3. Complete quantiles forecasting with QRNN model

Because of the volatility and randomness in wind speed data, traditional linear probability density forecasting methods cannot describe the intricate correlation between data. Therefore, based on the explanatory variables $x_i^l = \left\{x_{i,j}^l\right\}_{j=1,2,\ldots,J^l}$ and response variables y_i^l for the $l-th$ subsequence, the QRNN model are utilized to establish the nonlinear relationship, further quantifying the nondeterminacy in forecasting process with the assistance of KDE method.

3.3.1. Establish QRNN forecasting model

Supposing that there are T conditional quantiles at quantile points $\{\tau_t\}_{t=1,2,\ldots,T}$, the estimator of the conditional τ_t-quantile of y_i^l is expressed as follows.

$$Q_{y_i^l}^l\left(\tau_t \left| x_i^l\right.\right) = f\left\{\sum_{k=1}^{K^l} o_k^l\left(\tau_t\right) g\left[\sum_{j=1}^{J^l} h_{j,k}^l\left(\tau_t\right) x_{i,j}^l\right]\right\} \tag{15}$$

where K^l represents the number of hidden layers of QRNN model in $l-th$ subsequence. The estimators $\widehat{H}^l\left(\tau_t\right) = \left\{\widehat{h}_{j,k}^l\left(\tau_t\right)\right\}_{j=1,2,\ldots,J^l;k=1,2,\ldots,K^l}$ and $\widehat{O}^l\left(\tau_t\right) = \left\{\widehat{o}_k^l\left(\tau_t\right)\right\}_{k=1,2,\ldots,K^l}$ can be calculated by solving the following objective function and inequality constraint.

$$\min_{\left[H^l(\tau_t),O^l(\tau_t)\right]} \sum_{i=1}^{I} \rho_{\tau_t}\left\{y_i^l\left(\tau_t\right) - Q_{y_i^l}^l\left(\tau_t \left| x_i^l\right.\right)\right\} + \xi^l \sum_{j=1}^{J^l}\sum_{k=1}^{K^l}\left[h_{j,k}^l\left(\tau_t\right)\right]^2 \tag{16}$$

where ξ^l represents the penalty parameter that corresponds to the $l-th$ subsequence. Finally, the quantiles forecasting $\widehat{Q}_{y_i^l}^l\left(\tau_t \left| x_i^l\right.\right)$ can be obtained.

3.3.2. Optimate the parameters in QRNN model

The number of hidden layer nodes and the penalty parameter are optimized by gird search in this paper. For each subsequence, the two pivotal parameters K^l and ξ^l are tuned respectively at given ranges for the best forecasting performance. Furthermore, the optimization objectives are derived from the evaluation indices corresponding to the forecasting forms, which is designed for remaining optimal in their respective prediction forms. Within the optimal parameters, the outputs of QRNN model are expressed as $\left\{\widehat{q}_{i,t}^l\right\}_{t=1,2,\ldots,T}$ for $l-th$ each subsequence.

3.4. Data postprocess with KDE

3.4.1. Aggregate the predicted quantiles

Based on the optimal conditional quantiles obtained for all the subsequences, denoted as $\left\{\widehat{q}_{i,t}^l\right\}_{t=1,2,\ldots,T;l=1,2,\ldots,L+1}$, the final conditional quantiles $\{\widehat{q}_{i,t}\}_{t=1,2,\ldots,T}$ can be calculated by Eq (17).

$$\{\widehat{q}_{i,t}\}_{t=1,2,\ldots,T} = \sum_{l=1}^{L+1} \widehat{q}_{i,t}^l, t = 1, 2, \ldots, T \tag{17}$$

3.4.2. Obtain the final predictions

To bridge quantiles and probability density curves, KDE is used to convert the forecasting results from the QRNN model into probability density functions. Specifically, the conditional quantiles $\{\widehat{q}_{i,t}\}_{t=1,2,\ldots,T}$ of the $i-th$ targeted prediction point are viewed as the samples of an unknown distribution. With an appropriate bandwidth and a predetermined kernel function, the KDE method can fit approximate probability density functions for the wind speed targeted points. The probabilistic forecasting is therefore completed, from which the interval and point predictions can be obtained.

3.5. Multistep wind speed forecasting with DR strategy

For a complete assessment of our proposed model, a hybrid multistep forecasting methods, DR strategy, is applied in this paper. With regard to each subsequence, the H-step ahead forecasting model need to be established. Notably, the DR strategy is embodied in the LASSO regression and QRNN model. Different forecasting horizons possess different feature vectors constructed through the DR strategy. In the QRNN model, the predicted probability means of subsequences will be served as the feature that will be utilized in the next forecasting step. Therefore, the LASSO regression and QRNN model cooperate with DR strategy to complete the multistep forecasting for each subsequence. Subsequently, the same forecasting procedure is repeated in all the subsequences, and the final multistep wind speed probabilistic predictions can be obtained.

3.6. Evaluation indicators for different forecasting forms

3.6.1. Probabilistic predictions

Assuming that P points $\{w_{i,p}\}_{p=1,2,\ldots,P}$ that meet the inequation of $w_{i,1} \leq w_{i,2} \leq \cdots \leq w_{i,P}$ are extracted from the probability density functions, corresponding probabilities $\{v_{i,p}\}_{z=1,2,\ldots,P}$ can be obtained concurrently. A series of forecasting with corresponding probabilities can be regarded as probabilistic predictions.

The quantile score (QS) [36], a universally accepted assessment criteria for quantiles prediction, is proposed based on pinball loss function defined by Eq (18). Aiming at each quantile and each forecasted point, the pinball loss function need to be calculated, where their average is specified as the value of QS. A lower QS indicates a more accurate quantiles prediction, which can be calculated by Eq (19).

$$\varphi\left(y_i, \widehat{q}_{i,t}, \tau_t\right) = \begin{cases} \left(1-\tau_t\right)\left(\widehat{q}_{i,t} - y_i\right) y_i < \widehat{q}_{i,t} \\ \tau_t\left(\widehat{q}_{i,t} - y_i\right) y_i \geq \widehat{q}_{i,t} \end{cases} \tag{18}$$

$$QS = \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=1}^{T}\varphi\left(y_i, \widehat{q}_{i,t}, \tau_t\right) \tag{19}$$

The continuous rank probability score (CRPS) is utilized to verify the capacity of the proposed model in probability density prediction, which can be calculated by Eq (20) [37].

$$CRPS = \frac{1}{I}\sum_{i=1}^{I}\left[\frac{1}{P}\sum_{p=1}^{P}\left|w_{i,p} - y_i\right| - \frac{1}{2P^2}\sum_{p=1}^{P}\sum_{q=1}^{P}\left|w_{i,p} - w_{i,q}\right|\right] \tag{20}$$

which rewards the predicted values clustered around actual values and punishes the deviations from it. Therefore, the value of CRPS is negatively related with the precision of probability density forecasting.

3.6.2. Interval predictions

Taking the minimum and the maximum of the predictions $\{w_{i,p}\}_{p=1,2,\ldots,P}$ as the lower and upper bounds of forecasting interval, the PIs $[w_{i,1}, w_{i,P}]$ can be obtained for the $i-th$ target point. The assessment of interval forecasting requires two indicators simultaneously, one for reliability and the other for sharpness. In order to satisfy the two requirements, the prediction interval coverage probability (PICP) collaborates with the prediction interval

normalized averaged width (PIANW) to demonstrate the plausibility and validity of interval predictions, which the formulas can be formulated as below [38].

$$PICP = \frac{1}{I}\sum_{i=1}^{I}\omega_i \tag{21}$$

$$\omega_i = \begin{cases} 1, \ y_i \in [W_{i,1}, W_{i,P}] \\ 0, \ y_i \notin [W_{i,1}, W_{i,P}] \end{cases} \tag{22}$$

$$PINAW = \sum_{i=1}^{I}\frac{W_{i,P} - W_{i,1}}{ID} \tag{23}$$

where D is the difference between the maximum and minimum of target forecasting value. A larger PICP reveals more predicted values falling into the PIs, and a smaller PINAW reflects lower uncertainty in the PIs. From a comprehensive point of view, coverage width-based criterion (CWC) proposed in [39] takes reliability and sharpness into consideration for evaluating the interval predictions. It is defined as below.

$$CWC = PINAW\left\{1 + \gamma\left(PICP\right)\exp\left[-\phi\left(PICP - \mu\right)\right]\right\} \tag{24}$$

in which ϕ and μ are two controlling parameters. They indicate the preset confidence level. ϕ is adopted to enlarge any slight difference between the confidence level and PICP. As an indicator function, $\gamma\left(PICP\right)$ is equal to 0 when $PICP \geq \mu$, else is 1. From a different point of view, PINAW plays a pivotal role in assessing PIs when the PICP satisfies the preset confidence level. Accordingly, a lower CWC is preferred in the evaluation of interval forecasting.

3.6.3. Point predictions

Based on the probabilistic predictions $\{w_{i,p}\}_{p=1,2,\ldots,P}$ and corresponding probabilities $\{v_{i,p}\}_{p=1,2,\ldots,P}$, the probability means served as point predictions can be calculated as the Eq (25) shows.

$$\widehat{y}_i^{mean} = \frac{\sum_{p=1}^{P} w_{i,p}v_{i,p}}{\sum_{p=1}^{P} v_{i,p}} \tag{25}$$

The point predictions can be validated by two error criteria: mean absolute percentage errors (MAPE) and root mean square errors (RMSE), defined as below [40].

$$MAPE = \frac{1}{I}\sum_{i=1}^{I}\left|\frac{y_i - \widehat{y}_i^{mean}}{y_i}\right| \tag{26}$$

$$RMSE = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(y_i - \widehat{y}_i^{mean}\right)^2} \tag{27}$$

MAPE and RMSE are negatively correlated with the prediction performance, which means a smaller value indicates a more superior point forecasting performance.

4. Case studies and result analysis

Case studies are designed to examine the effectiveness of the proposed multi-step forecasting model in practice. Four datasets collected from actual wind farms and seven evaluation indicators are employed to adequately show the superiorities of the proposed model over other five comparative models comprising single or hybrid models. Multistep probabilistic wind speed forecasting is implemented in six models, in which probabilistic predictions, interval predictions and point predictions are analyzed and compared in terms of the seven assessment criteria, respectively.

4.1. Data collection

The wind speed data recorded in 1-h interval are collected from Sotavento wind farm in Spain [41], depicted as Fig. 2. The wind speed has different characteristics in different seasons and months: data in March, July, October and January are selected to form four datasets for verification, each of which represents a season. Each dataset contains 744 samples, further partitioned into three datasets in the partition ratio at 20:4:7: the first 480 samples are used as the training set, the subsequent 96 samples form the validation set and the last 168 samples form the testing set.

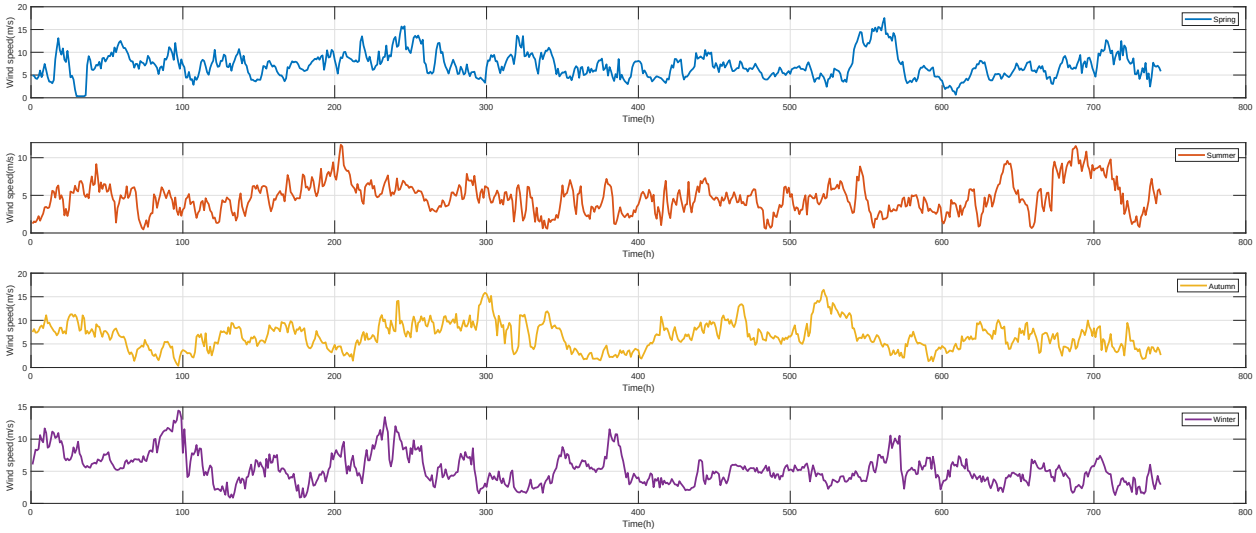

Fig. 2: Original hourly wind speed data of four seasons

4.2. The establishment of the proposed model

The decomposition technology CEEMDAN is first applied to the raw wind speed timeseries. 8 IMFs and a residual are obtained as shown in Fig. 3 and Fig. 4, in which the 9-th subseries represents the residual. All the IMF subsequences have a smaller magnitude and slighter variation than the original wind speed. These subsequences will be used to model severally by the procedures elaborated in the above section.

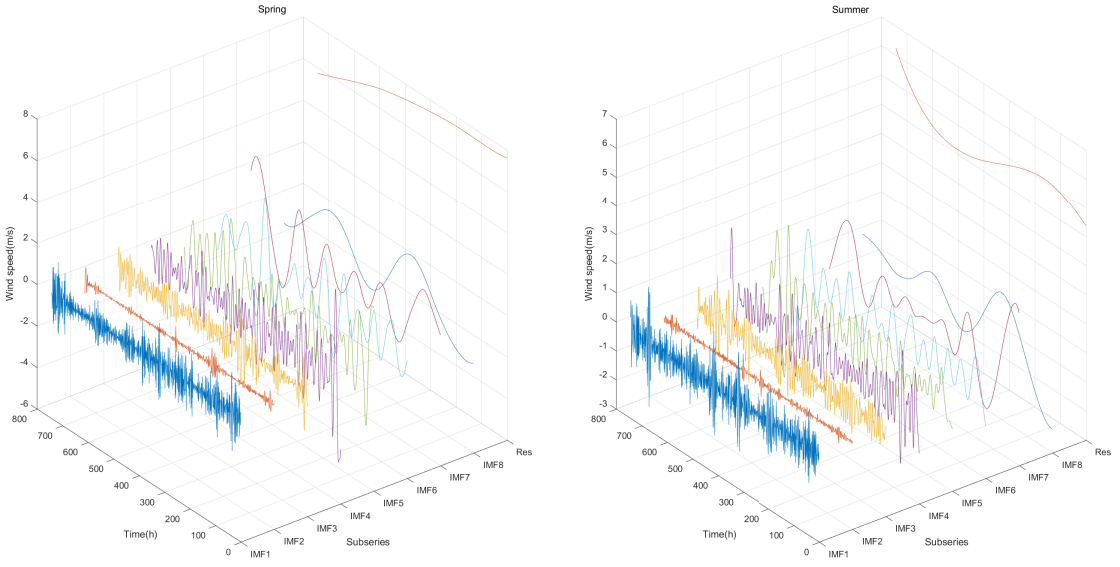

Fig. 3: The decomposition results with CEEMDAN in spring and summer

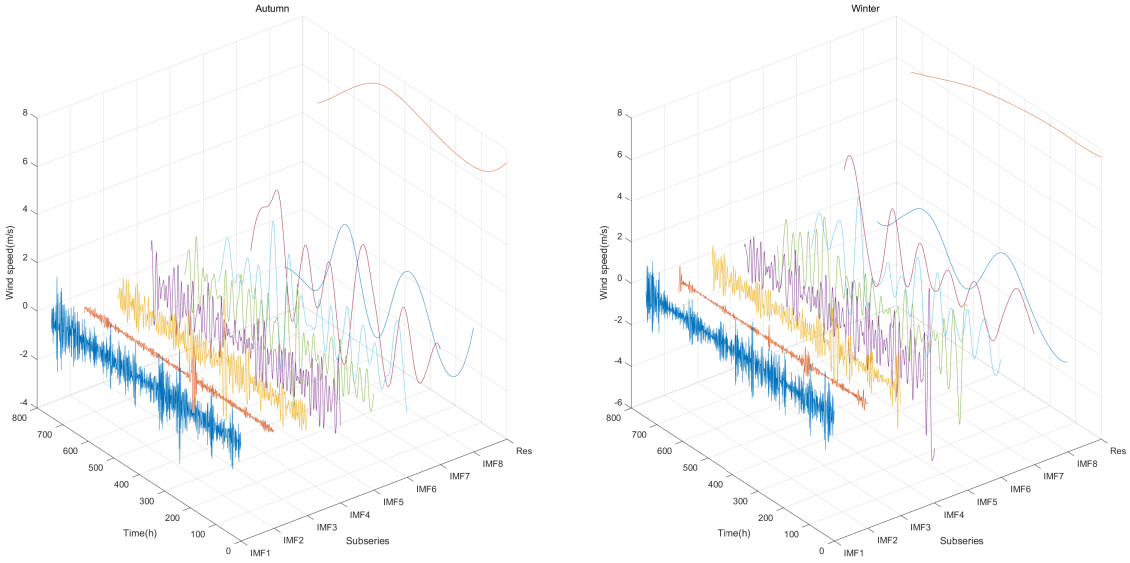

Fig. 4: The decomposition results with CEEMDAN in autumn and winter

In order to extract as much information as possible from the historical data, the previous three days of wind speed data are taken into consideration, and 72 explanatory variables are constructed. There are three hyper parameters in LASSO regression and QRNN forecasting model. The constraint parameter η^l in LASSO regression is optimized by GCV method in each subsequence. The optimal parameter values and the compressed dimension are listed in Table. 1, from which it can be observed that the optimized results vary significantly with the subsequences. The original dimensions of the explanatory variables are reduced from 72 to 8-51, improving the predictive efficiency to some extent. It is worth noting that the number of hidden layer nodes and the penalty parameter are determined by the grid search method. A validation dataset is established to search for the optimal. More detailed descriptions about the optimal values of parameters and the corresponding objective functions are shown in the next section. For each subsequence, the LASSO regression and QRNN model with the optimal parameters are implemented in the testing set to verify the forecasting capacity of the proposed models.

The ultimate predictions aggregated by all forecasted conditional quantiles corresponding to each subsequence serve as the samples to fit the approximate distributions through KDE. Consequently, probabilistic forecasting, interval forecasting and point forecasting of one-step, two-step along with three-step for wind speed are accomplished. Their specific predictions are presented by graphs in the next section.

4.3. Result comparison and analysis

There are six models established for case studies, consisting of five comparative models: QRNN, LASSO-QRNN,CEEMDAN-QRNN, EMD-LASSO-QRNN and EEMD-LASSO-QRNN along with our proposed CEEMDAN-LASSO-QRNN model. All the six models are denoted as Model 1-Model 6. All the models are preformed through the same prediction pattern Given a subsequence, all the models are firstly trained on the training set. The trained models are then executed on validation dataset in order to find the optimal parameters that corresponds to the optimal solution of the objective functions in QRNN. The objective functions differ in three predictive forms: probability density predictions, interval predictions and point predictions. The models that possess the optimal solution are then tested on the testing dataset as the final forecasting model. All the models are preformed in same operating environment. The operating system is Windows 10 version, the CPU consists of Intel (R) Core (TM) i9-10900k CPU @ 3.70 GHz, the experiment platform is R x64 4.0.3 and Rstudio

4.3.1. Probabilistic predictions

In probabilistic predictions, the CRPS is used to serve as the objective function. For the *l-th* subsequence, the objective formula is expressed as below.

Objective: seek for the optimal solution set of K^l and ξ^l such to:

$$\begin{aligned} &\min \left\{ CRPS\left(K^l, \xi^l\right) \right\} \\ &s.t. \quad 1 \leq K^l \leq 5 \\ &\qquad 1 \leq \xi^l \leq 5 \end{aligned} \tag{28}$$

The mutil-step ahead probabilistic predictions of the six models in spring dataset with the optimal parameters are displayed in Figs. 5- 7. The probability density curves of the six models in the 3rd, 27th, 51st, 75th, 99th

Table 1: The optimized results in LASSO regression for each subsequence

Datasets	Horizons	Parameters	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	Res
Spring	1-step	η^l	0.26	0.42	0.33	0.28	0.4	0.46	1	1	1
		J^l	19	35	29	43	42	50	13	14	23
	2-step	η^l	0.26	0.43	0.33	0.32	0.36	0.39	1	1	1
		J^l	21	37	30	44	41	45	14	15	24
	3-step	η^l	0.25	0.41	0.33	0.28	0.36	0.39	1	1	1
		J^l	20	37	28	41	44	45	14	14	19
Summer	1-step	η^l	0.17	0.37	0.34	0.31	0.3	1	1	1	1
		J^l	15	32	29	39	32	36	11	17	19
	2-step	η^l	0.15	0.35	0.34	0.32	0.32	1	1	1	1
		J^l	13	31	28	38	34	33	11	17	15
	3-step	η^l	0.16	0.35	0.34	0.32	0.38	1	1	1	1
		J^l	14	32	28	38	46	36	10	16	15
Autumn	1-step	η^l	0.31	0.2	0.36	0.31	0.43	1	1	1	1
		J^l	22	17	29	44	42	34	16	**8**	19
	2-step	η^l	0.3	0.27	0.33	0.3	0.34	1	1	1	1
		J^l	22	33	30	41	33	31	15	9	17
	3-step	η^l	0.3	0.27	0.37	0.27	0.48	1	1	1	1
		J^l	22	35	34	36	48	34	16	10	17
Winter	1-step	η^l	0.33	0.35	0.32	0.39	0.4	1	1	1	1
		J^l	28	34	24	38	41	43	9	17	19
	2-step	η^l	0.35	0.37	0.31	0.5	0.34	1	1	1	1
		J^l	30	38	22	**51**	36	43	9	19	10
	3-step	η^l	0.29	0.4	0.31	0.38	0.43	1	1	1	1
		J^l	23	40	24	38	46	42	11	17	12

and 123rd future points are depicted in the same figure, from which it can be observed that the actual values represented by the dotted lines all fall on the probability density curve. In general, the proposed model outperforms the other models. The models without decomposition show a flatter probability density curves, while the others show more concentrated around the actual value. It can be drawn that the decomposition technique has the capacity of improving the accuracy of probabilistic predictions.

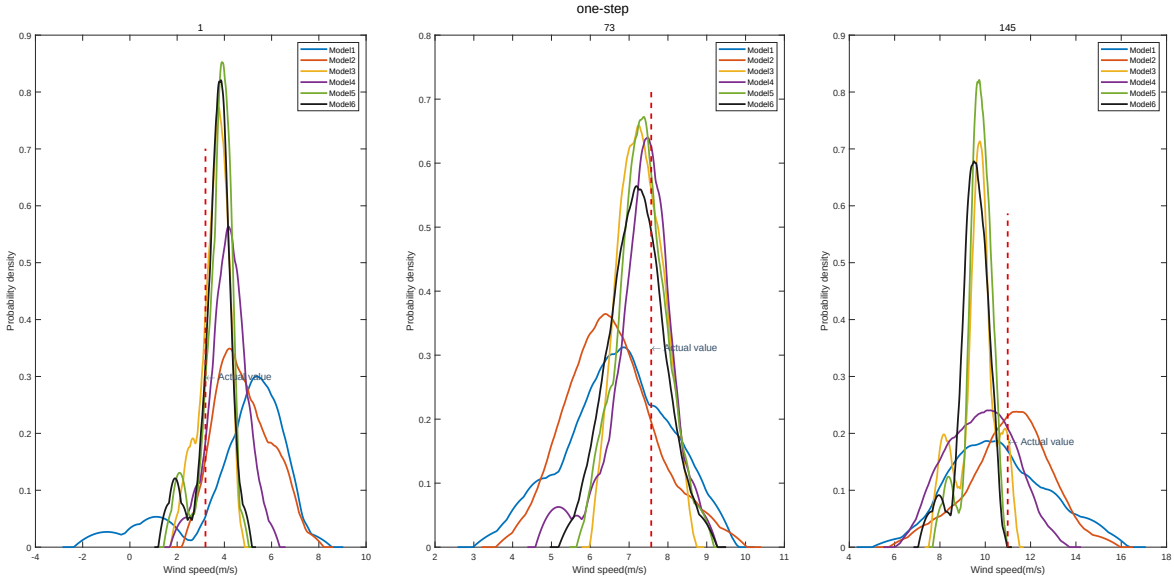

Fig. 5: Partial one-step probability density curves of all the models in spring testing dataset

A more detailed comparison are shown in Table. 2. According to the two index values of the probabilistic prediction, the models using CEEMDAN preform the best synthetically. The proposed model does not always outperform the other models for all the horizons in the four datasets, but its performance is very close to the best. Moreover, the average QS values of one-step, two-step and three-step occupy the minimum in the four datasets. The average CRPS values are the lowest in autumn and winter datasets. Compared with model1, the QS values of our proposed model reduce by 75.34%, 82.35%, 64.00% and 63.16%, respectively for the four datasets. Likewise, the CRPS values of our proposed model in the autumn and winter datasets decrease by 56.44% and 46.58% than Model1. Overall speaking, the proposed model presents the best probabilistic prediction performance.

4.3.2. Interval predictions

From the probability density function, the minimum and the maximum are extracted to be the upper and lower bounds of the interval prediction. As a comprehensive evaluation index for interval predictions, CWC are selected to be the objective function. The optimal solution can be calculated by using the following formula.

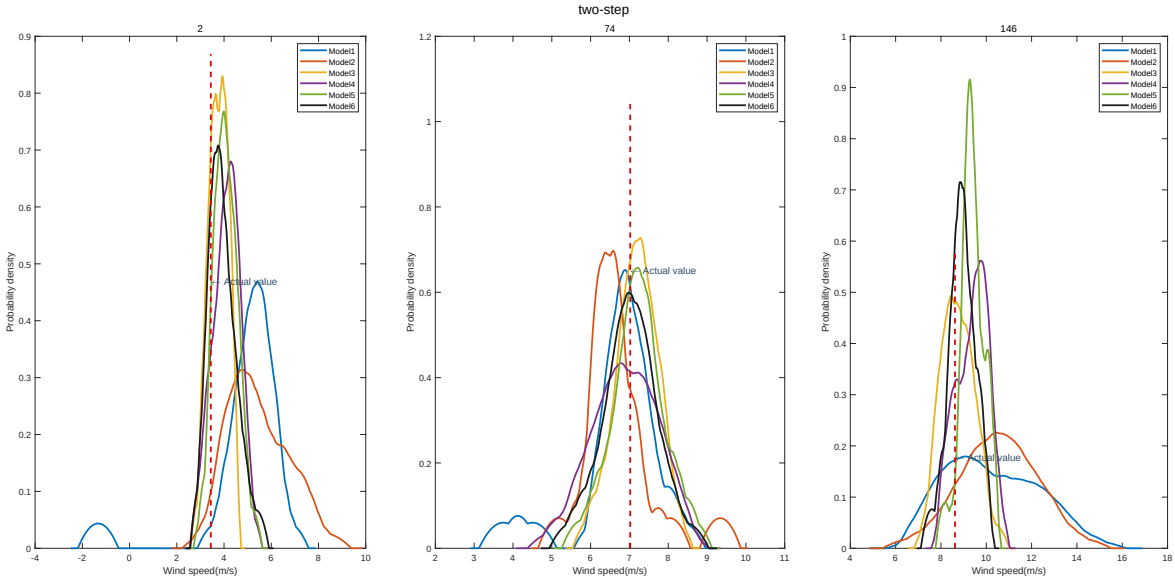

Fig. 6: Partial two-step probability density curves of all the models in spring testing dataset

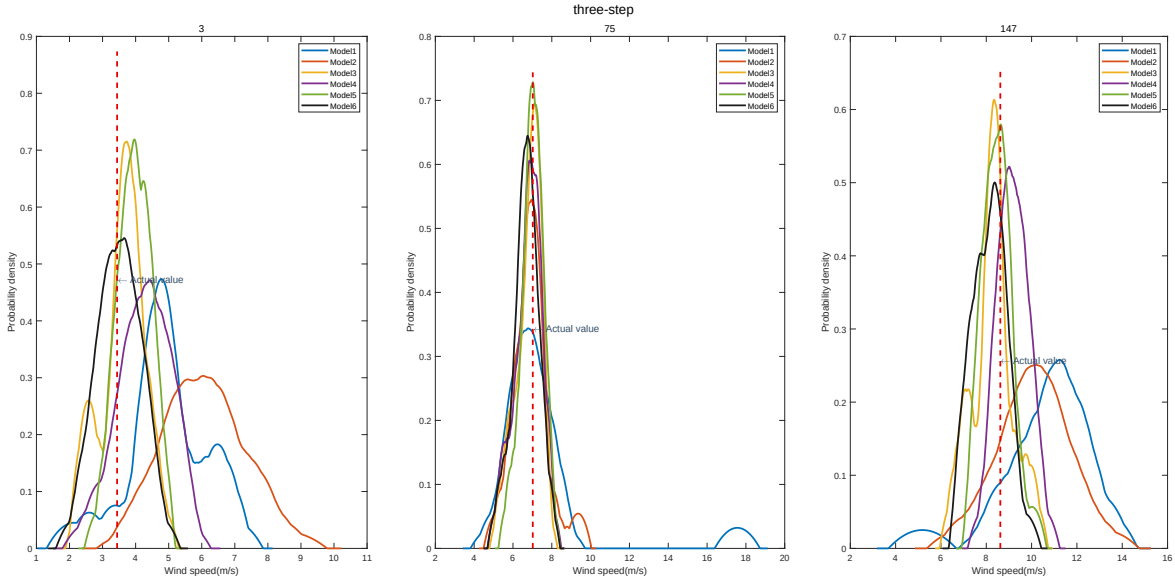

Fig. 7: Partial three-step probability density curves of all the models in spring testing dataset

Table 2: Multistep probabilistic prediction performance evaluations of all the models in four datasets

Datasets	Models	One-step		Two-step		Three-step		Average	
		CRPS	QS	CRPS	QS	CRPS	QS	CRPS	QS
Spring	Model 1	1.53	0.61	1.56	0.76	1.77	0.82	1.62	0.73
	Model 2	0.81	0.31	0.90	0.41	0.99	0.47	0.90	0.40
	Model 3	**0.38**	0.14	**0.44**	0.20	**0.48**	0.22	**0.43**	**0.18**
	Model 4	0.52	0.18	0.52	0.23	0.57	0.25	0.54	0.22
	Model 5	**0.38**	**0.12**	0.47	0.20	0.51	0.23	0.45	**0.18**
	Model 6	0.43	0.13	0.48	**0.19**	0.51	**0.21**	0.47	**0.18**
Summer	Model 1	0.86	0.42	1.89	1.02	2.86	1.61	1.87	1.02
	Model 2	0.78	0.34	0.90	0.47	1.12	0.60	0.93	0.47
	Model 3	0.42	0.17	0.45	0.21	0.48	0.23	0.45	0.21
	Model 4	0.55	0.20	0.58	0.25	0.59	0.29	0.57	0.24
	Model 5	0.39	**0.14**	**0.43**	0.19	0.48	0.23	**0.43**	0.19
	Model 6	0.41	0.15	0.44	**0.18**	**0.46**	**0.21**	0.44	**0.18**
Autumn	Model 1	0.88	0.40	1.02	0.52	1.12	0.58	1.01	0.50
	Model 2	0.84	0.34	1.01	0.48	1.06	0.53	0.97	0.45
	Model 3	0.41	0.15	0.46	0.20	**0.47**	0.22	**0.44**	0.19
	Model 4	0.61	0.24	0.65	0.27	0.68	0.34	0.65	0.28
	Model 5	**0.38**	**0.14**	0.45	0.21	0.53	0.26	0.45	0.20
	Model 6	**0.41**	0.15	**0.44**	**0.19**	0.48	**0.21**	**0.44**	**0.18**
Winter	Model 1	0.63	0.33	0.75	0.38	0.81	0.43	0.73	0.38
	Model 2	0.63	0.24	0.79	0.32	0.83	0.36	0.75	0.30
	Model 3	**0.33**	0.11	**0.38**	**0.15**	**0.40**	0.17	**0.37**	**0.14**
	Model 4	0.44	0.14	0.49	0.19	0.53	0.27	0.49	0.20
	Model 5	**0.33**	**0.10**	0.81	0.16	0.42	0.18	0.52	0.15
	Model 6	0.36	0.11	**0.38**	**0.15**	0.41	**0.16**	0.39	**0.14**

The bolder ones mean the best forecasting performance in corresponding evaluation indicators.

Objective: seek for the optimal solution set of K^l and ξ^l such to:

$$\begin{aligned} &\min \left\{ CWC\left(K^l, \xi^l\right) \right\} \\ &s.t. \quad 1 \leq K^l \leq 5 \\ &\qquad 1 \leq \xi^l \leq 5 \end{aligned} \tag{29}$$

in which the three parameters involved in CWC are set beforehand. ϕ and μ are 50 and 0.95 respectively. The PINAW will dominate the solution when the PICP satisfies the preset confidence. The multi-step interval forecasting results on the spring dataset are depicted in Figs. 8- 10. It can be concluded that the models using the decomposition technique produce more narrower PIs, which reduces the uncertainty of the forecast. Except for the QRNN model (Model 1), the interval width of the other models are close to each other, which indicates these models can eliminate uncertainty to some extent. A more detailed comparison can only be made through the specific values of CWC.

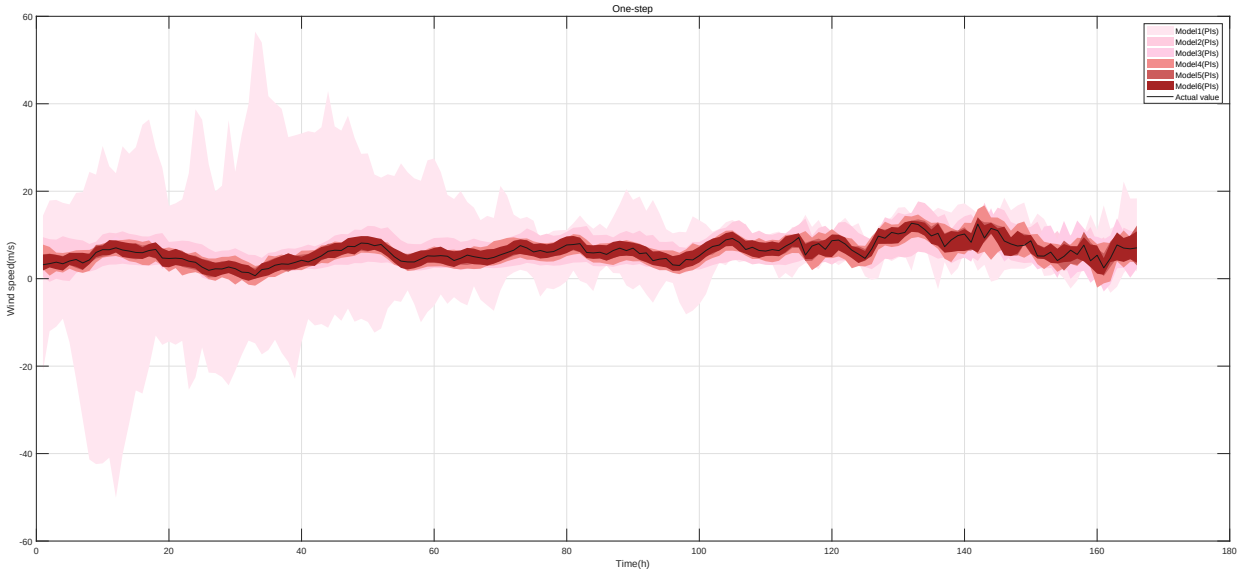

Fig. 8: The one-step interval predictions of all the models in spring testing dataset

When the value of PICP cannot satisfy the confidence level, the interval prediction is considered as a failure. It can be concluded from Table. 3, the situations that PICP value may be inferior to 95% in the five comparative models with the forecasting horizon become longer. However, in our proposed model, the PICP values is always superior to 95%. It fully attests the superiorities of the proposed model in terms of stability and adaptability. Despite failing to perform best in entire assessment criteria, the proposed model embraces the most robust interval forecasting performance. The CEEMDAN-QRNN model (Model3) possesses the lowest CWC value in the majority. While in the spring dataset, it cannot meet the confidence level. The CEEMDAN-LASSO-QRNN (Model 6) overcomes the dilemma, which presents reduction in forecasting uncertainty and contentment for confidence interval. When priority is given to ensuring the success of interval prediction, the proposed model is the preferred most.

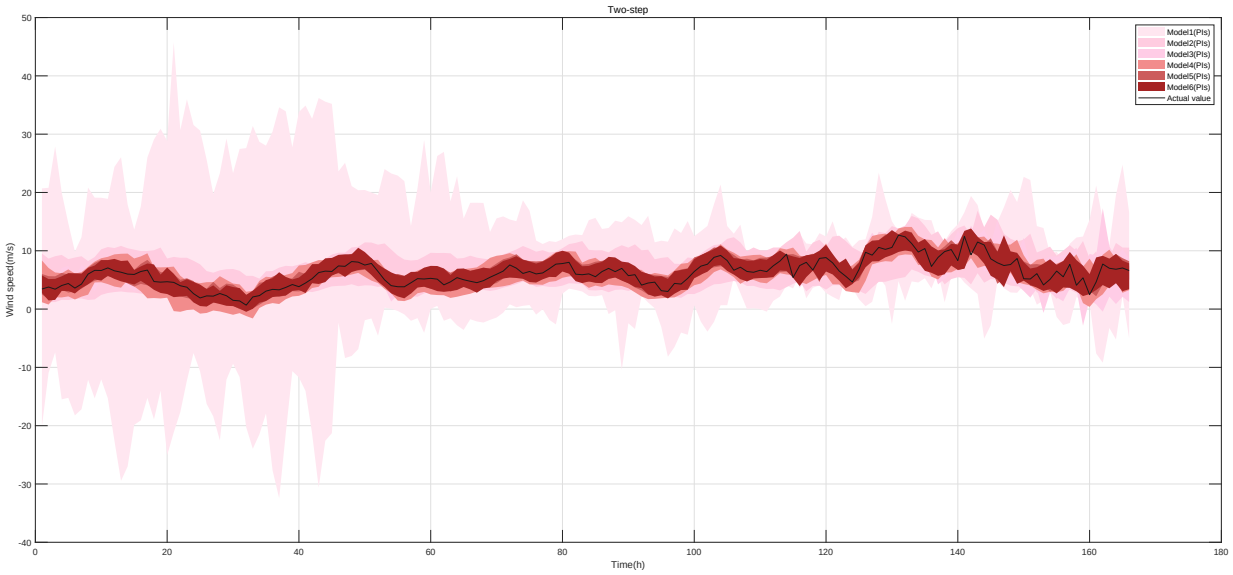

Fig. 9: The two-step interval predictions of all the models in spring testing dataset

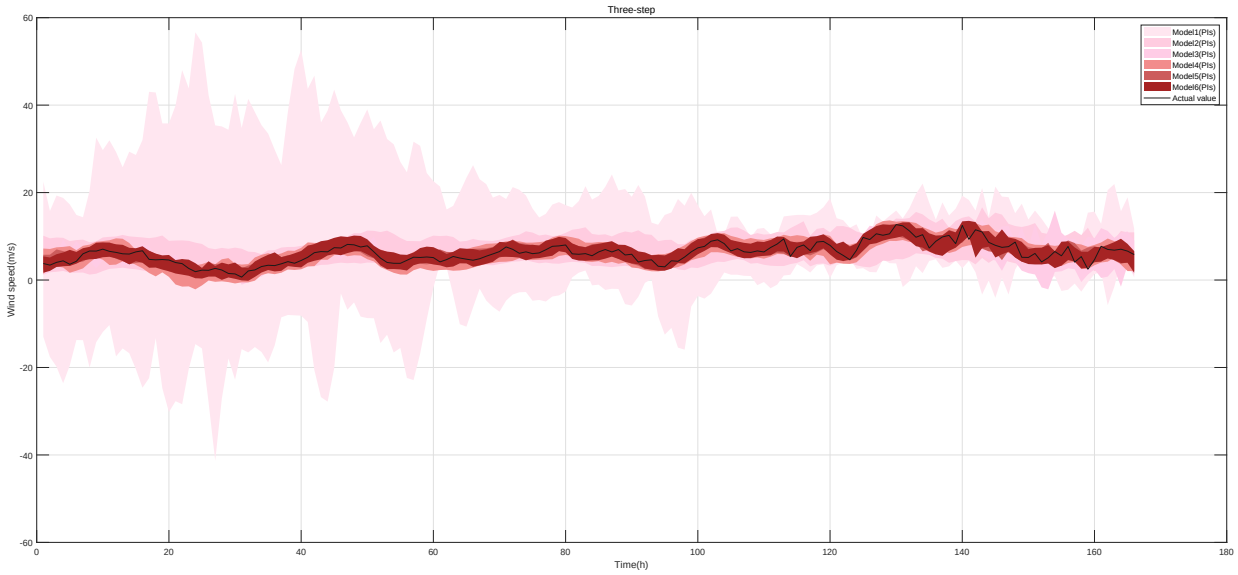

Fig. 10: The three-step interval predictions of all the models in spring testing dataset

Table 3: Multistep interval prediction performance evaluations of all the models in four datasets

Datasets	Models	One-step			Two-step			Three-step			Average		
		PINAW (%)	PICP (%)	CWC (%)	PINAW (%)	PICP (%)	CWC (%)	PINAW (%)	PICP (%)	CWC (%)	PINAW (%)	PICP (%)	CWC (%)
Spring	Model 1	23.86	100.00	23.86	29.85	99.40	29.85	29.94	98.80	29.94	27.88	99.40	27.88
	Model 2	39.99	99.40	39.99	41.44	97.59	41.44	44.98	98.19	44.98	42.14	98.39	42.14
	Model 3	19.04	100.00	**19.04**	17.62	96.99	**17.62**	20.97	93.98	187.84	19.21	96.99	74.83
	Model 4	26.77	100.00	26.77	29.89	98.8	29.89	27.99	95.78	27.99	28.21	98.19	28.21
	Model 5	24.91	99.40	24.91	23.53	95.78	23.53	22.87	92.77	327.66	23.77	95.98	125.37
	Model 6	24.14	99.40	24.14	25.40	98.19	25.40	26.26	96.39	**26.26**	25.27	97.99	**25.27**
Summer	Model 1	19.12	100.00	19.12	26.46	99.40	26.46	22.55	99.40	22.55	22.71	99.60	22.71
	Model 2	36.75	98.80	36.75	40.70	92.17	452.62	32.98	86.14	8406.93	36.81	92.37	2965.44
	Model 3	14.80	99.40	**14.80**	19.86	98.8	**19.86**	22.74	99.40	22.74	19.13	99.20	19.13
	Model 4	25.57	99.40	25.57	25.03	98.19	25.03	21.87	97.59	21.87	24.16	98.39	24.16
	Model 5	15.28	99.40	15.28	19.94	96.39	19.94	20.29	93.37	245.82	18.50	96.39	93.68
	Model 6	17.27	98.80	17.27	20.83	96.99	20.83	19.18	96.99	**19.18**	19.10	97.59	**19.10**
Autumn	Model 1	25.26	99.40	25.26	40.08	98.19	40.08	28.60	95.78	28.60	31.31	97.79	31.31
	Model 2	46.31	98.80	46.31	48.20	97.59	48.20	51.13	92.17	463.05	48.55	96.18	185.86
	Model 3	21.83	98.80	**21.83**	20.87	98.19	**20.87**	23.21	96.38	**23.21**	21.97	97.79	**21.97**
	Model 4	31.06	98.80	31.06	36.52	98.80	36.52	32.38	94.58	155.85	33.32	97.39	74.48
	Model 5	24.22	100.00	24.22	22.32	93.37	247.84	25.06	90.36	1041.90	23.87	94.58	437.99
	Model 6	27.60	99.40	27.60	30.59	98.80	30.59	24.34	96.39	24.34	27.51	98.19	27.51
Winter	Model 1	41.12	98.80	41.12	36.99	93.37	262.52	36.72	97.59	36.72	38.28	96.59	113.45
	Model 2	51.88	100.00	51.88	55.81	99.40	55.81	58.12	97.59	58.12	55.27	99.00	55.27
	Model 3	28.22	100.00	28.22	31.33	98.19	31.33	29.82	99.40	29.82	29.79	99.20	**29.79**
	Model 4	37.12	100.00	37.12	40.85	99.40	40.85	33.30	97.59	33.30	37.09	99.00	37.09
	Model 5	26.94	99.40	**26.94**	29.20	98.19	**29.20**	29.86	92.77	334.65	28.67	96.79	130.26
	Model 6	27.34	100.00	27.34	33.48	99.40	33.48	31.69	98.19	**31.69**	30.84	99.20	30.84

The bolder ones mean the best forecasting performance corresponding to different datasets in CWC evaluation indicator.

4.3.3. Point predictions

With regard to point predictions, MAPE is the most representative evaluation index to assess accuracy. Therefore, through minimizing the value of MAPE, the optimal solution can be obtained.

Objective: seek for the optimal solution set of K^l and ξ^l such to:

$$\begin{aligned} &\min \left\{ MAPE\left(K^l, \xi^l\right) \right\} \\ &s.t. \quad 1 \leq K^l \leq 5 \\ &\qquad 1 \leq \xi^l \leq 5 \end{aligned} \tag{30}$$

The probability mean can be calculated from the probability density functions, which is viewed as the point prediction. We can observe from Fig. 11- 13 that he point predictions of all models in one-step forecasting distinguish inconspicuously from each other. While in three-step forecasting, parts of comparative models such as Model1 and Model2 diverge from the actual value to a large extent. In conclusion, the increasing horizons generate more indeterminacy, thereby bringing in more challenges in forecasting accuracy. However, the proposed model is performs very well in coping with this challenge, together with the robustness.

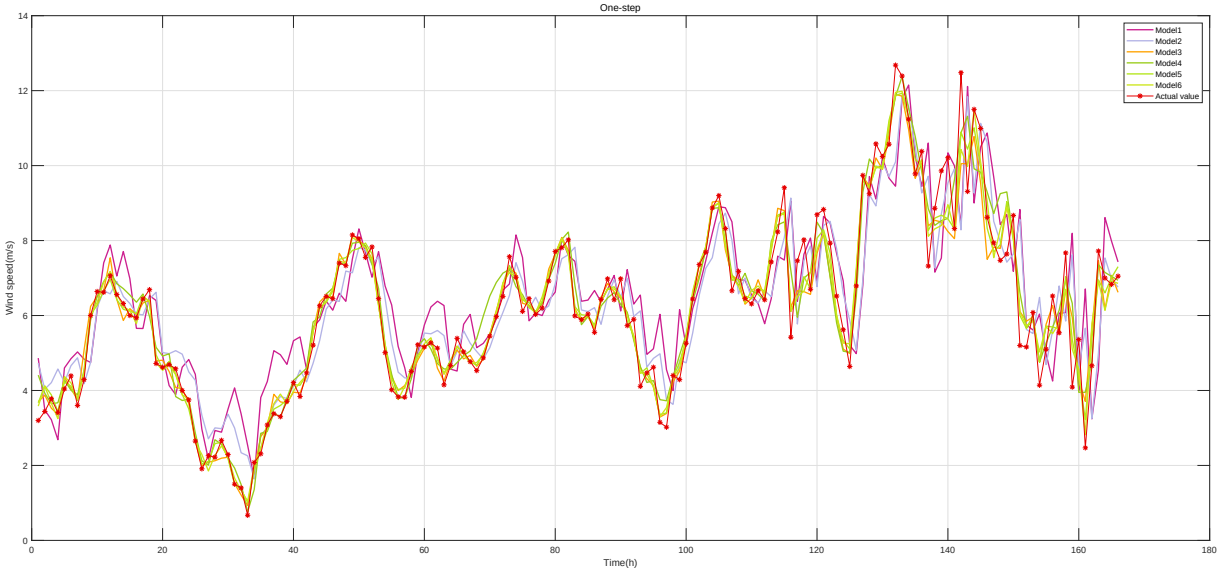

Fig. 11: The one-step point predictions of all the models in spring testing dataset

Which can be seen from Table. 4, the proposed model presents more superiorities over the other five models in most cases. As the forecast period increases, its advantages in terms of robustness and reliability become more significant. The average values of three steps assessment in the two indicators are minimum in all datasets, which indicates a satisfactory forecasting performance. With regard to the spring dataset, the indices of RMSE and MAPE of multistep average of our proposed model CEEMDAN-LASSO-QRNN are 0.65m/s and 8.99%, which is lessened by 64.09%, 72.79% than that of Model 1. Correspondingly for the remaining datasets, the rate of reduction all exceed 50%. The enhancement in multistep point forecasting effectively illustrates that the randomness and volatility can be partly mitigated through the proposed model.

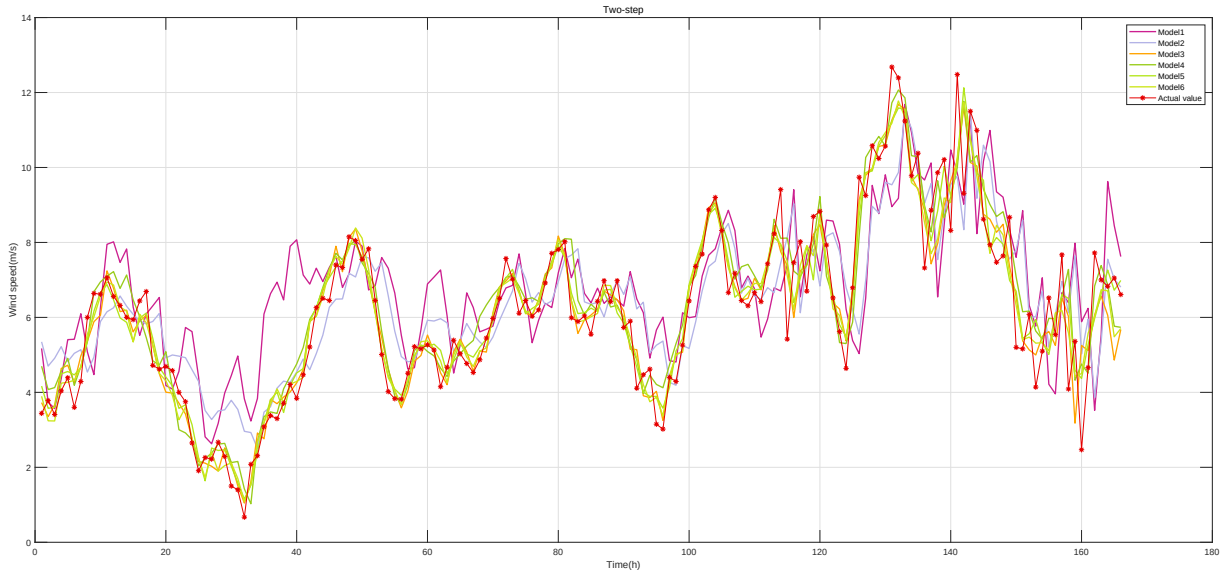

Fig. 12: The two-step point predictions of all the models in spring testing dataset

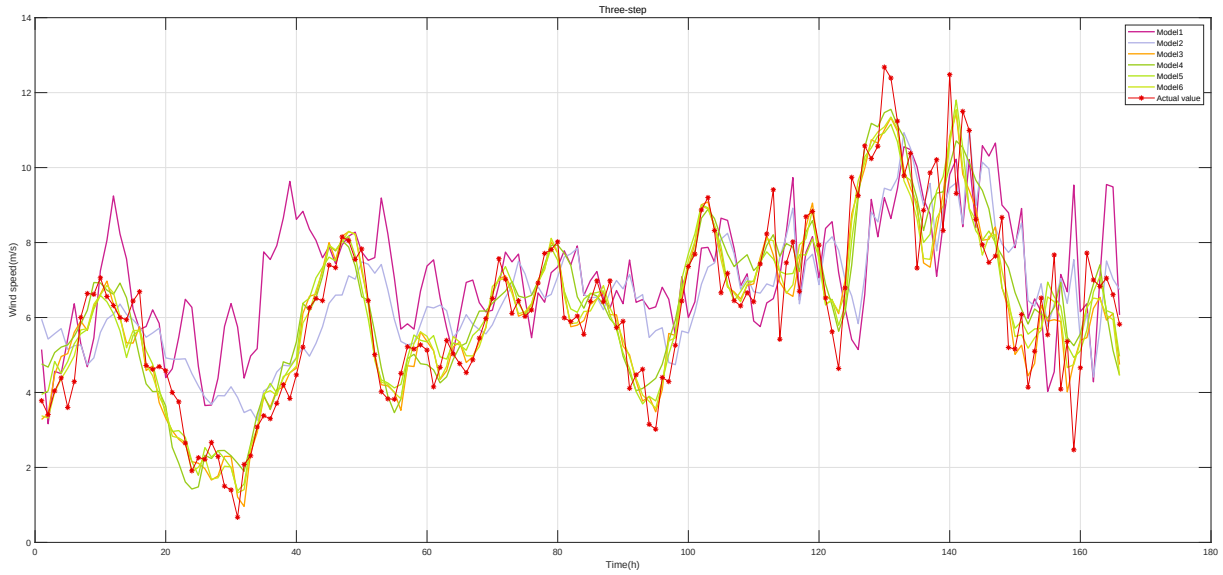

Fig. 13: The three-step point predictions of all the models in spring testing dataset

Table 4: Multistep point prediction performance evaluations of all the models in four datasets

Datasets	Models	One-step		Two-step		Three-step		Average	
		RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Spring	Model 1	1.39	23.04	1.79	32.73	2.24	43.35	1.81	33.04
	Model 2	1.16	17.81	1.40	24.21	1.61	29.71	1.39	23.91
	Model 3	0.53	6.48	0.76	10.40	0.82	11.81	0.70	9.56
	Model 4	0.66	8.69	0.82	12.03	0.92	15.26	0.80	11.99
	Model 5	**0.44**	**5.61**	0.72	9.92	0.81	12.04	0.66	9.19
	Model 6	0.47	6.07	**0.69**	**9.64**	**0.77**	**11.27**	**0.65**	**8.99**
Summer	Model 1	1.68	35.22	1.92	46.36	2.08	54.00	1.89	45.19
	Model 2	1.15	25.24	1.55	37.37	1.88	47.57	1.53	36.73
	Model 3	0.63	11.58	0.79	15.89	0.85	18.60	0.76	15.36
	Model 4	0.68	12.47	0.90	18.41	1.08	24.93	0.89	18.61
	Model 5	**0.54**	10.57	0.74	15.37	0.92	18.19	0.73	14.71
	Model 6	**0.54**	**9.82**	**0.65**	**12.78**	**0.84**	**16.25**	**0.68**	**12.95**
Autumn	Model 1	1.31	21.02	1.69	29.24	1.87	33.22	1.62	27.83
	Model 2	1.26	20.14	1.65	28.45	1.80	32.26	1.57	26.95
	Model 3	0.58	8.89	0.83	13.37	0.83	13.72	0.75	11.99
	Model 4	0.86	13.07	0.99	15.19	1.17	18.79	1.01	15.68
	Model 5	**0.54**	8.76	0.82	13.26	0.91	15.44	0.76	12.48
	Model 6	**0.54**	**8.19**	**0.74**	**11.97**	**0.75**	**12.62**	**0.68**	**10.93**
Winter	Model 1	0.94	19.83	1.24	26.91	1.37	30.93	1.18	25.89
	Model 2	0.84	18.16	1.11	25.62	1.26	30.04	1.07	24.61
	Model 3	0.40	8.83	0.52	10.94	**0.56**	**12.17**	**0.49**	10.65
	Model 4	0.49	9.73	0.72	13.31	0.88	17.98	0.69	13.67
	Model 5	**0.37**	**7.66**	0.56	11.47	0.62	13.38	**0.52**	10.84
	Model 6	0.39	7.99	**0.52**	**10.79**	0.58	12.49	0.50	**10.42**

The bolder ones mean the best forecasting performance in corresponding evaluation indicators.

Each step ahead forecasting is independently optimized, and all the models mentioned above are optimized by the same procedure for all the datasets. Because of the space limit, only the best outcome of the proposed model is listed in Table. 5. Based on the analysis and discussion in three forecasting forms, all the six models are demonstrated to be feasible and effective. The experimental studies of one-step, two-step and three-step wind speed predictions show our proposed model performs the best. As a whole, CEEMDAN-LASSO-QRNN is effective in forecasting wind speed series that fluctuates irregularly and variationally.

4.3.4. The comparison with advanced methods

The ensemble methods are discussed by the two categories of competitive ensemble methods and cooperative ensemble methods in this paper. The proposed CEEMDAN-LASSO-QRNN multi-step wind speed probabilistic forecasting belongs to the latter. It is proved to be effective and steady in predicting complex univariate time series. There are many advanced ensemble forecasting frameworks, most of which can be classified as the former. Randomly distributed embedding (RDE) was proposed in [20], which is a model-free ensemble forecasting framework. By randomly generating explanatory variables multiple times to achieves accurate predictions of short-term high-dimensional data. Multiple repeated experiments can generate the ensemble predictions, which are utilized to fit probability distribution by KDE. For addressing complexity in dynamic interconnected systems, Ye et.al [42] proposed a general approach called multiview embedding (MVE) that utilizes complexity to amplify information, thus overcoming the limitations of short and multidimensional noisy time series. Compared with the above two frameworks to forecast short-term high-dimensional data, the reservoir computing (RC) approach proposed in [43] aims to overcome the forecasting of large spatial extent and attractor dimension.

Since univariate time series can be transformed into multi-dimensional time series through reconstruction, the model proposed in this paper is compared with these three advanced prediction frameworks through experiments to further illustrate the effectiveness of the model proposed in this paper on univariate complex time series. The three comparative models: RDE, MVE and RC are performed in the same system environment with the proposed. Except the running platform of RDE is MATLAB, the rest are R. The lag order is also set to 72. All the conditions are consistent with the model CEEMDAN-LASSO-QRNN. More detailed parameters involved in the three frameworks are represented in the Table. 6. Note that the RDE framework combined with KDE to obtain probability density function, which can compare with the proposed model in three forecasting forms. The comparison results of evaluation indicators are listed in the Table. 7. While the MVE and RC frameworks only compare in the point forecasting, and the results are displayed in Table. 8.

The evaluation results of three-step wind speed forecasting show that the proposed model is demonstrated a significant advantage at the univariate time series. Each forecaasting framework is suitable for different conditions and scenarios. The RDE framework is appropriate for short-term and high-dimensional data; the MVE framework is effective in complex interconnected systems with short and high-dimensional data; the RC framework can be applied in the long and high-dimensional data. While, refer to the wind speed univariate time series, the proposed model demonstrates its superiorities over the three advanced framework.

Table 5: The optimal structure parameters of CEEMDAN-LASSO-QRNN model

Datasets	Objectives	Horizons	subsequence								
			IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	Res
Spring	CPRS	1-step	(1,5)	(1,4)	(1,5)	(1,5)	(1,1)	(1,2)	(1,3)	(1,3)	(3,5)
		2-step	(1,4)	(2,3)	(2,5)	(2,3)	(1,1)	(1,2)	(1,3)	(1,3)	(3,4)
		3-step	(1,3)	(1,2)	(2,2)	(2,3)	(1,1)	(1,2)	(1,3)	(1,4)	(2,3)
	CWC	1-step	(5,2)	(2,1)	(5,5)	(2,5)	(2,1)	(3,4)	(1,2)	(1,3)	(1,4)
		2-step	(5,4)	(5,3)	(2,5)	(2,5)	(2,1)	(3,4)	(2,1)	(1,4)	(1,5)
		3-step	(4,1)	(1,2)	(3,3)	(2,5)	(2,1)	(3,4)	(4,2)	(2,1)	(1,2)
	MAPE	1-step	(5,5)	(5,4)	(5,4)	(5,3)	(1,1)	(1,1)	(1,5)	(1,4)	(2,2)
		2-step	(3,2)	(2,2)	(4,1)	(2,5)	(1,3)	(1,1)	(1,5)	(1,3)	(2,2)
		3-step	(1,1)	(2,2)	(5,3)	(5,3)	(1,3)	(1,1)	(1,5)	(1,3)	(2,1)
Summer	CPRS	1-step	(1,1)	(1,1)	(1,1)	(1,1)	(1,4)	(1,3)	(1,4)	(1,4)	(1,3)
		2-step	(1,1)	(1,2)	(1,5)	(1,1)	(1,5)	(1,1)	(1,4)	(1,4)	(1,5)
		3-step	(2,3)	(1,4)	(1,3)	(1,1)	(1,4)	(1,3)	(1,4)	(1,1)	(1,5)
	CWC	1-step	(5,5)	(5,1)	(2,3)	(1,5)	(1,2)	(1,1)	(1,1)	(1,3)	(1,1)
		2-step	(5,5)	(5,3)	(2,2)	(1,5)	(1,2)	(1,1)	(4,1)	(1,2)	(1,2)
		3-step	(5,1)	(5,1)	(2,2)	(1,5)	(1,2)	(1,1)	(4,1)	(4,5)	(3,4)
	MAPE	1-step	(5,1)	(2,2)	(3,5)	(3,1)	(1,5)	(1,4)	(1,4)	(1,2)	(1,3)
		2-step	(2,5)	(4,1)	(4,2)	(2,1)	(2,4)	(1,1)	(1,4)	(1,1)	(1,5)
		3-step	(5,1)	(5,2)	(2,5)	(1,1)	(3,4)	(1,4)	(1,4)	(1,1)	(1,5)
Autumn	CPRS	1-step	(1,5)	(4,5)	(1,1)	(1,1)	(2,2)	(1,4)	(1,4)	(1,5)	(1,2)
		2-step	(1,1)	(4,3)	(2,1)	(1,3)	(3,1)	(1,4)	(1,4)	(1,5)	(1,4)
		3-step	(1,1)	(1,1)	(2,2)	(1,1)	(3,1)	(1,4)	(1,5)	(1,5)	(1,2)
	CWC	1-step	(4,1)	(4,1)	(5,1)	(2,1)	(1,1)	(2,5)	(2,4)	(1,1)	(1,3)
		2-step	(3,5)	(3,4)	(3,4)	(2,1)	(1,1)	(2,5)	(2,1)	(1,2)	(1,1)
		3-step	(4,1)	(5,3)	(2,3)	(2,1)	(1,1)	(2,5)	(2,1)	(2,2)	(3,4)
	MAPE	1-step	(3,4)	(2,1)	(5,3)	(3,4)	(3,2)	(1,1)	(1,2)	(1,1)	(1,1)
		2-step	(5,5)	(5,2)	(2,4)	(1,2)	(4,1)	(1,1)	(1,4)	(1,1)	(1,1)
		3-step	(3,2)	(4,2)	(5,3)	(3,2)	(3,4)	(1,1)	(1,3)	(1,1)	(1,2)
Winter	CPRS	1-step	(1,1)	(1,1)	(1,5)	(1,4)	(1,2)	(1,3)	(1,1)	(1,1)	(1,5)
		2-step	(1,1)	(1,1)	(2,2)	(4,5)	(2,1)	(1,3)	(1,3)	(1,3)	(1,4)
		3-step	(1,1)	(1,1)	(4,3)	(4,5)	(2,1)	(1,3)	(1,2)	(1,3)	(1,4)
	CWC	1-step	(4,1)	(4,1)	(5,1)	(2,1)	(1,1)	(2,5)	(2,4)	(1,1)	(1,3)
		2-step	(3,5)	(3,4)	(3,4)	(2,1)	(1,1)	(2,5)	(2,1)	(1,2)	(1,1)
		3-step	(4,1)	(5,3)	(2,3)	(2,1)	(1,1)	(2,5)	(2,1)	(2,2)	(3,4)
	MAPE	1-step	(4,3)	(5,2)	(3,5)	(3,3)	(1,2)	(1,2)	(1,1)	(1,3)	(1,5)
		2-step	(3,5)	(4,1)	(5,1)	(2,1)	(4,3)	(1,3)	(1,1)	(1,2)	(1,5)
		3-step	(4,4)	(3,4)	(4,5)	(1,4)	(1,3)	(1,3)	(1,1)	(1,3)	(1,4)

Table 6: The parameters of RDE, MVE and RC

Models	Parameters	Definition	Vuale
RDE	S	the number of non-delay embedding	200
	L	embedding dimension	6
MVE	E	the embedding dimensions to use for time delay embedding	3
	tau	the lag to use for time delay embedding	1
RC	n.neurons	the number of neurons in the reservoir	1000
	density	the density of connections in the reservoir	0.02
	lambda	the tuning parameter for ridge regression	10

5. Conclusion and discussion

With the increasing importance of wind energy worldwide, reliable and stable wind speed prediction is paid more and more attention. In view of the randomness and uncertainty of wind speed prediction, this paper develops a cooperative ensemble method based on CEEMDAN-LASSO-QRNN for multi-step wind speed probabilistic forecasting. In this method, CEEMDAN is utilized to decompose original wind speed time series; high dimensional features are compressed by LASSO regression optimized by GCV method; QRNN is used to perform wind speed quantiles forecasting, whose parameters are tuned by gird search; for more future information, the KDE method is employed to estimate probability density functions, further producing probabilistic predictions, interval predictions as well as point predictions. The experimental results show that the proposed model is superior to the other eight comparative models in most cases. To be specific, compared with the model QRNN and LASSO-QRNN, the models based on decomposition method have displayed more reliability and stability in multi-step wind speed forecasting. Further comparison shows that the CEEMDAN-LASSO-QRNN model owns more robust performance as the number of predicted steps increases. Therefore, for wind speed time series with strong fluctuations, the proposed method can provide a reliable and stable multi-step probabilistic forecasting model.

The validity and practicability of the cooperative ensemble method for multistep wind speed probabilistic forecasting are exemplified clearly in case analysis. There are inevitably some shortcomings and problems in our method that is worth further investigations. The most crucial issue is time consumption. In particular, the decomposition technique increases the training time. In the future, we aim at developing more efficient models.

Acknowledgment

The authors would like to thank the National Natural Science Foundation of China (Nos. 71771073 and 72171068) and the Anhui Provincial Natural Science Foundation for Distinguished Young Scholars (No.2108085J36).

Appendix

Nomenclature

Indices and parameters

i	Index of samples, $i = 1, 2, \ldots, I$
s	Index of the adding Gaussian white noise, $s = 1, 2, \ldots, S$
l	Index of IMFs, $l = 1, 2, \ldots, L$
j	Index of explanatory variables or input nodes, $j = 1, 2, \ldots, J$
k	Index of hidden nodes, $k = 1, 2, \ldots, K$
t	Index of quantile points, $t = 1, 2, \ldots, T$
p	Index of predicted values, $p = 1, 2, \ldots, P$
h	Index of forecasting horizons, $h = 1, 2, \ldots, H$
λ	The penalty parameter in LASSO regression
J^l	The number of explanatory variables after feature selection in $l-th$ subsequence
η	The constraint parameter corresponding to J^l
η^l	The constraint parameter in $l-th$ subsequence
ξ	The penalty parameter in QRNN
ξ^l	The penalty parameter in $l-th$ subsequence
ϕ	The parameter in CWC
μ	The preset confidence level

Variables

Y	Original wind speed samples set
y_i	The $i-th$ sample in Y
Y^s	The $s-th$ timeseries added Gaussian white noise
ω^s	The $s-th$ Gaussian white noise
c_l^s	The $s-th$ series of the $l-th$ IMF subsequence decomposed by EMD
c_l	The $l-th$ IMF
y_i^l	The $i-th$ value in c_l
r_l	The $l-th$ residue series
r_l^s	The $s-th$ residue series obtained by adding Gaussian white noise to r_l
r_L	The final residue series
y_i^{L+1}	The $i-th$ value in r_L
X	The explanatory variables samples set
x_i	The $i-th$ sample in X
$x_{i,j}$	The $i-th$ sample of the $j-th$ explanatory variable in x_i
β	The coefficient set of explanatory variables
$\widehat{\beta}^{LASSO}$	The estimated coefficient set in LASSO regression
τ	The quantile point set
τ_t	The $t-th$ quantile point in τ

$Q_Y(\tau \mid X)$	The conditional quantile set
$Q_Y(\tau_t \mid X)$	The conditional quantile set at $t-th$ quantile point
$\widehat{Q}_Y(\tau_t \mid X)$	The estimated conditional quantile set at $t-th$ quantile point
$H(\tau)$	The weight between input nodes and hidden nodes at τ quantile point
$h_{j,k}(\tau)$	The weight between $j-th$ input node and the $k-th$ hidden node in $H(\tau)$
$O(\tau)$	The weight between hidden nodes and output node at τ quantile point
$o_k(\tau)$	The weight between $k-th$ hidden node and the output node in $O(\tau)$
$b_{i,j}^h$	The bias of $i-th$ sample in $i-th$ hidden layer
b_i^o	The bias of $i-th$ sample in output layer
θ	The error term in multistep forecasting strategy
B	The bandwidth in KDE method
y_{i-j}^l	The $j-th$ candidate feature of the $l-th$ subsequence
$\widehat{\beta}^{LASSO,l}$	The estimated coefficient set in LASSO regression
$\widehat{\beta}_j^{LASSO,l}$	The estimated coefficient corresponding to the $j-th$ candidate variable in $\widehat{\beta}^{LASSO,l}$
x_i^l	The $i-th$ sample of explanatory variables in the $l-th$ subsequence
$x_{i,j}^l$	The $j-th$ explanatory variables in x_i^l
$\widehat{Q}_{y_i^l}^l\left(\tau_t \left\| x_i^l\right.\right)$	The estimated conditional quantile set at $t-th$ quantile point in $l-th$ subsequence
$\widehat{H}^l(\tau_t)$	The estimated weight of input and hidden nodes at τ_t quantile point in $l-th$ subsequence
$\widehat{h}_{j,k}^l(\tau_t)$	The estimated weight of the $j-th$ input node and the $k-th$ hidden node in $\widehat{H}^l(\tau_t)$
$\widehat{H}^l(\tau_t)$	The estimated weight of hidden and output node at τ_t quantile point in $l-th$ subsequence
$\widehat{q}_{i,t}$	The estimated quantiles at τ_t quantile at $i-th$ sample
$\widehat{q}_{i,t}^l$	The estimated quantiles at τ_t quantile in $l-th$ subsequence at $i-th$ sample
$w_{i,p}$	The $p-th$ predicted value of the $i-th$ time point through KDE
$v_{i,p}$	The predicted probability corresponding to $w_{i,p}$
$\widehat{q}_{i,t}$	The final predicted quantile of $i-th$ time point
$\widehat{y}_i^{mean}$	The final predicted value of the $i-th$ time point
$w_{i,1}$	The lower bound of subinterval
$w_{i,P}$	The upper bound of subinterval
D	The difference between the maximum and minimum of entire datasets

Functions

$\left\|\cdot\right\|_1$	L1 regularization
$\left\|\cdot\right\|_2$	L2 regularization
$g(\cdot)$	The activation function of hidden layer
$f(\cdot)$	The activation function of output layer
$\rho_{\tau_t}(\cdot)$	The check function at quantile point
$K(\cdot)$	The kernel function

$f_h(\cdot)$	The h-step model
$\widehat{f}_h(\cdot)$	The h-step estimated model

References

[1] Zhou Y, Wang J, Lu H, Zhao W. Short-term wind power prediction optimized by multi-objective dragonfly algorithm based on variational mode decomposition. Chaos, Solitons & Fractals 2022;157:111982.

[2] Jiang Y, Huang G, Yang Q, Yan Z, Zhang C. A novel probabilistic wind speed prediction approach using real time refined variational model decomposition and conditional kernel density estimation. Energy Conversion and Management 2019;185:758–73.

[3] Iversen EB, Morales JM, Møller JK, Madsen H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. International Journal of Forecasting 2016;32(3):981–90.

[4] Zhao X, Jiang N, Liu J, Yu D, Chang J. Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework. Energy Conversion and Management 2020;203:112239.

[5] Hu J, Heng J, Wen J, Zhao W. Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm. Renewable Energy 2020;162:1208–26.

[6] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renewable and Sustainable Energy Reviews 2014;32:255–70.

[7] Zhang L, Xie L, Han Q, Wang Z, Huang C. Probability density forecasting of wind speed based on quantile regression and kernel density estimation. Energies 2020;13(22):6125.

[8] Hu T, Guo Q, Li Z, Shen X, Sun H. Distribution-free probability density forecast through deep neural networks. IEEE transactions on neural networks and learning systems 2019;31(2):612–25.

[9] Hu J, Tang J, Lin Y. A novel wind power probabilistic forecasting approach based on joint quantile regression and multi-objective optimization. Renewable Energy 2020;149:141–64.

[10] Wang Y, Hu Q, Meng D, Zhu P. Deterministic and probabilistic wind power forecasting using a variational bayesian-based adaptive robust multi-kernel regression model. Applied energy 2017;208:1097–112.

[11] Bludszuweit H, Domínguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. IEEE Transactions on Power Systems 2008;23(3):983–91.

[12] Hwang E. Prediction intervals of the covid-19 cases by har models with growth rates and vaccination rates in top eight affected countries: Bootstrap improvement. Chaos, Solitons & Fractals 2022;:111789.

[13] Yang X, Fu G, Zhang Y, Kang N, Gao F. A naive bayesian wind power interval prediction approach based on rough set attribute reduction and weight optimization. Energies 2017;10(11):1903.

[14] Wan C, Lin J, Wang J, Song Y, Dong ZY. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. IEEE Transactions on Power Systems 2016;32(4):2767–78.

[15] Zhang Z, Qin H, Liu Y, Yao L, Yu X, Lu J, et al. Wind speed forecasting based on quantile regression minimal gated memory network and kernel density estimation. Energy Conversion and Management 2019;196:1395–409.

[16] Zhang L, Xie L, Han Q, Wang Z, Huang C. Probability density forecasting of wind speed based on quantile regression and kernel density estimation. Energies 2020;13(22):6125.

[17] Ren Y, Suganthan P, Srikanth N. Ensemble methods for wind and solar power forecasting—a state-of-the-art review. Renewable and Sustainable Energy Reviews 2015;50:82–91.

[18] He Y, Wang Y. Short-term wind power prediction based on eemd–lasso–qrnn model. Applied Soft Computing 2021;105:107288.

[19] Wu W, Peng M. A data mining approach combining k-means clustering with bagging neural network for short-term wind power forecasting. IEEE Internet of Things Journal 2017;4(4):979–86.

[20] Ma H, Leng S, Aihara K, Lin W, Chen L. Randomly distributed embedding making short-term high-dimensional data predictable. Proceedings of the National Academy of Sciences 2018;115(43):E9994–10002.

[21] Altan A, Karasu S, Bekiros S. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos, Solitons & Fractals 2019;126:325–36.

Table 7: The evaluations of the proposed model (Model6) and RDE

Datasets	Step	Models	RMSE	MAPE	PINAW	PICP	CRPS
spring	one-step	Model6	**0.47**	**6.07**	**24.14**	**99.40**	**0.43**
		RDE	2.67	42.08	51.70	71.69	1.50
	two-step	Model6	**0.69**	**9.64**	**25.40**	**98.19**	**0.48**
		RDE	2.62	41.46	55.02	80.12	1.49
	three-step	Model6	**0.77**	**11.27**	**26.26**	**96.39**	**0.51**
		RDE	6.39	95.60	55.67	0.00	5.91
	Average	Model6	**0.65**	**8.99**	**25.27**	**97.99**	**0.47**
		RDE	3.89	59.71	54.13	50.60	2.96
summer	one-step	Model6	**0.54**	**9.82**	**17.27**	**98.80**	**0.41**
		RDE	2.86	68.40	51.70	60.24	1.62
	two-step	Model6	**0.65**	**12.78**	**20.83**	**96.99**	**0.44**
		RDE	2.83	69.44	55.02	67.47	1.61
	three-step	Model6	**0.84**	**16.25**	**19.18**	**96.99**	**0.46**
		RDE	5.75	93.72	55.67	0.00	5.01
	Average	Model6	**0.68**	**12.95**	**19.10**	**97.59**	**0.44**
		RDE	3.81	77.19	54.13	42.57	2.74
autumn	one-step	Model6	**0.54**	**8.19**	**27.60**	**99.40**	**0.41**
		RDE	2.34	38.91	51.70	78.92	1.36
	two-step	Model6	**0.74**	**11.97**	**30.59**	**98.80**	**0.44**
		RDE	2.33	39.52	55.02	83.13	1.37
	three-step	Model6	**0.75**	**12.62**	**24.34**	**96.39**	**0.48**
		RDE	5.74	95.36	55.67	0.00	5.31
	Average	Model6	**0.68**	**10.93**	**27.51**	**98.19**	**0.44**
		RDE	3.47	57.93	54.13	54.02	2.68
winter	one-step	Model6	**0.39**	**7.99**	**27.34**	**100.00**	**0.36**
		RDE	1.42	38.05	51.70	94.58	1.01
	two-step	Model6	**0.52**	**10.79**	**33.48**	**99.40**	**0.38**
		RDE	1.52	40.88	55.02	95.18	1.09
	three-step	Model6	**0.58**	**12.49**	**31.69**	**98.19**	**0.41**
		RDE	4.37	94.11	55.67	0.00	4.05
	Average	Model6	**0.50**	**10.42**	**30.84**	**99.20**	**0.39**
		RDE	2.44	57.68	54.13	63.25	2.05

The bolder ones mean the best forecasting performance in corresponding evaluation indicators.

Table 8: The evaluations of the proposed model (Model6), MVE and RC

Datasets	Models	Step_1		Step_2		Step_3		Average	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Spring	Model6	**0.47**	**6.07**	**0.69**	**9.64**	**0.77**	**11.27**	**0.65**	**8.99**
	MVE	0.98	15.71	1.34	21.32	1.53	25.42	1.28	20.82
	RC	1.93	31.91	2.03	31.49	3.08	49.29	2.35	37.56
Summer	Model6	**0.54**	**9.82**	**0.65**	**12.78**	**0.84**	**16.25**	**0.68**	**12.95**
	MVE	0.98	25.58	1.34	37.96	1.58	46.13	1.30	36.56
	RC	2.37	52.73	3.03	56.62	2.6	46.86	2.67	52.07
Autumn	Model6	**0.54**	**8.19**	**0.74**	**11.97**	**0.75**	**12.62**	**0.68**	**10.93**
	MVE	1.28	16.72	1.77	23.19	2.18	27.89	1.74	22.60
	RC	2.1	40.01	3.07	58.65	3.24	63.59	2.81	54.08
Winter	Model6	**0.39**	**7.99**	**0.52**	**10.79**	**0.58**	**12.49**	**0.50**	**10.42**
	MVE	0.97	21.54	1.28	30.82	1.43	34.36	1.23	28.91
	RC	1.75	42.40	2.39	60.68	3.11	77.35	2.42	60.14

The bolder ones mean the best forecasting performance in corresponding evaluation indicators.

[22] Wang D, Yue C, ElAmraoui A. Multi-step-ahead electricity load forecasting using a novel hybrid architecture with decomposition-based error correction strategy. Chaos, Solitons & Fractals 2021;152:111453.

[23] Altan A, Karasu S, Zio E. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. Applied Soft Computing 2021;100:106996.

[24] Cheng L, Zang H, Ding T, Sun R, Wang M, Wei Z, et al. Ensemble recurrent neural network based probabilistic wind speed forecasting approach. Energies 2018;11(8):1958.

[25] Lingyu T, Jun W, Chunyu Z. Mode decomposition method integrating mode reconstruction, feature extraction, and elm for tourist arrival forecasting. Chaos, Solitons & Fractals 2021;143:110423.

[26] Zhou Y, Wang J, Lu H, Zhao W. Short-term wind power prediction optimized by multi-objective dragonfly algorithm based on variational mode decomposition. Chaos, Solitons & Fractals 2022;157:111982.

[27] Saxena BK, Mishra S, Rao KVS. Offshore wind speed forecasting at different heights by using ensemble empirical mode decomposition and deep learning models. Applied Ocean Research 2021;117:102937.

[28] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A: mathematical, physical and engineering sciences 1998;454(1971):903–95.

[29] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in adaptive data analysis 2009;1(01):1–41.

[30] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2011, p. 4144–7.

[31] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2011;73(3):273–82.

[32] Taylor JW. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 2000;19(4):299–311.

[33] Koenker R, Hallock KF. Quantile regression. Journal of economic perspectives 2001;15(4):143–56.

[34] Tenreiro C. Fourier series-based direct plug-in bandwidth selectors for kernel density estimation. Journal of Nonparametric Statistics

2011;23(2):533–45.

[35] Taieb SB, Bontempi G, Atiya AF, Sorjamaa A. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. Expert systems with applications 2012;39(8):7067–83.

[36] Zhang W, Quan H, Srinivasan D. Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. Energy 2018;160:810–9.

[37] Messner JW, Pinson P, Browell J, Bjerregård MB, Schicker I. Evaluation of wind power forecasts—an up-to-date view. Wind Energy 2020;23(6):1461–81.

[38] Pedro HT, Coimbra CF, David M, Lauret P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. Renewable Energy 2018;123:191–203.

[39] Khosravi A, Nahavandi S, Creighton D, Atiya AF. Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE transactions on neural networks 2010;22(3):337–46.

[40] Quan H, Srinivasan D, Khosravi A. Short-term load and wind power forecasting using neural network-based prediction intervals. IEEE transactions on neural networks and learning systems 2013;25(2):303–15.

[41] Sotavento. Website; 2018. `http://www.sotaventogalicia.com/en/`.

[42] Ye H, Sugihara G. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. Science 2016;353(6302):922–5.

[43] Pathak J, Hunt B, Girvan M, Lu Z, Ott E. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. Physical review letters 2018;120(2):024102.