Energy Conversion and Management 138 (2017) 587–602

Contents lists available at ScienceDirect

Energy Conversion and Management

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

A novel probabilistic wind speed forecasting based on combination of the adaptive ensemble of on-line sequential ORELM (Outlier Robust Extreme Learning Machine) and TVMCF (time-varying mixture copula function)

Xiangang Peng [a], Weiqin Zheng [a,*], Dan Zhang [a], Yi Liu [a], Di Lu [a], Lixiang Lin [b]

[a] School of Automation, Guangdong University of Technology, 510006 Guangzhou, China
[b] Guangzhou Power Supply Bureau Co., Ltd., 510000 Guangzhou, China

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form 12 January 2017
Accepted 3 February 2017
Available online 24 February 2017

Keywords:
Probabilistic wind speed forecasting
Time-varying mixture copula function
On-line sequential Outlier Robust Extreme Learning Machine
On-line ensemble using ordered aggregation
Bernaola Galvan Algorithm

ABSTRACT

The uncertainty and nonstationary of wind speed have compelled the power system operators and researchers to search for more accurate and reliable techniques to implement wind speed forecasting (WSF). In allusion to this phenomenon, this paper presents an adaptive ensemble of model for the probabilistic WSF, which is based on combination of the adaptive ensemble of on-line sequential ORELM (OS-ORELM) and the time-varying mixture copula function (TVMCF) to perform multi-step WSF. An OS-ORELM with forgetting mechanism based on Cook's distance (λ_{CDFF} OS-ORELM) serves as a basic WSF model and an on-line ensemble using ordered aggregation (OEOA) technique is employed to improve the prediction performance. In the data pre-processing period, the Bernaola Galvan algorithm (BGA) is employed to partition the raw wind speed series into segments and the adaptive variational mode decomposition (AVMD) is used to decompose each segment into sub-series with different sub-band. Each transformed sub-series is well-modeled with the application of λ_{CDFF} OS-ORELM-OEOA, which is optimized by modified crisscross optimization algorithm (CSO). Eventual forecast results are obtained through aggregate calculation. Then the probabilistic prediction intervals (PIs) of wind speed are established in a TVMCF framework by modeling the conditional forecasting error. Case studies using the real wind speed data from the National Renewable Energy Laboratory (NREL) demonstrate that the proposed model can not only improves point forecasts compared with benchmark methods, but also constructs higher quality of probabilistic PIs.

1. Introduction

WSF is of great significance for wind farm management and power system operation. It is well-known that the behavior of wind power output depends on the quality and variation of wind speed and its prediction can originate from the wind speed forecast [1,2]. However, the randomness, volatility and intermittence of wind speed, considered as 'nondispatchable', have not only restricted the high penetration of wind power integrating to the power system but also impacted the forecast results significantly. For example, it will bring about a great impact on power system operation and scheduling in various aspects, e.g. power system stability, ancillary service and power quality [3,4]. For this reason, an accurate forecast of wind speed (or wind power) is an essential and effective way to tackle these problems, which can provide a prospect to the wind power output and technique support for wind power trading in electricity, thus bringing significant economic benefit [5].

Under this context, a lot of scholars have emphasized the importance of wind speed and wind power forecast study in the past few decades. Various approaches have been proposed in the literatures. The state-of-the-art methods can be classified into different categories: physical modeling methods, statistical models, intelligent approaches and hybrid (or combined) methods. Each forecasting method has its advantages and disadvantages. The first one strongly depends on the meteorological and geographical information for modeling, which has been proved with poor forecast accuracy and in most cases is used as an auxiliary input for other forecasting methods [6,7]. The popular statistical

* Corresponding author at: Room 315, No. 2, Laboratory Building, Guangdong University of Technology, No. 100, West Waihuan Road, Higher Education Mega Center, Panyu District, Guangzhou City, Guangdong Province 510006, China.
E-mail address: zhengweiqin969@yeah.net (W. Zheng).

http://dx.doi.org/10.1016/j.enconman.2017.02.004
0196-8904/

Nomenclature

φ	scaling exponent
θ	threshold region
K	number of modes
i, j, t	common indices
T_0	size of initial training data
$\mathbf{D} = \{(x_t, y_t)\}_{t=1}^{T}$	data set with T sample points
$\mathbf{D}_{online} = \{(x_t, y_t)\}_{t=T_0+1}^{T}$	data set for online learning
$\mathbf{D} = \{(x_t, y_t)\}_{t=1}^{T_0}$	data set for initial the ORELM
$g(\cdot)$	activation function
L	number of hidden nodes
L_{opt}	optimal number of hidden nodes
w_{in}	weight of connection from input neural i to hidden neuron n
$\boldsymbol{\beta}, \boldsymbol{\beta}_0, \hat{\boldsymbol{\beta}}$	output weights, optimal output weights and initial output weights of ORELM
b_i	bias values of the ith hidden nodes
$\mathbf{H}, \mathbf{H}_0$	hidden layer output matrix and initial hidden layer output matrix of ORELM
$\mathbf{H}^{\dagger}$	Moore-Penrose generalized inverse of ORELM's hidden layer output matrix
υ	Lagrange multiplier
k	denote the kth sample
$\mathbf{P}_0$	initial covariance matrix
$\mathbf{y}_0$	initial output vector
m, n	dimension of the input vector and out vector
$\boldsymbol{\Gamma}_k$	gain matrix of the kth sample
D_k	Cook's distance of the kth sample
λ_{CDFF}	Cook's distance based forgetting factor
C_k	Statistic of Cook's distance of the kth sample
e_k	prediction error of the kth sample
σ^2	variance of the forecast error
$S_k(\cdot)$	survivor function
M_k	effective sample size of the k sample
P_{vc}, P_{hc}	vertical and horizontal crossover probability
$P_{vc\max}, P_{vc\min}$	maximum and minimum value of P_{vc}
o_i	fitness value of particle i
o_{avg}	average value of the current fitness value
n_{pop}	number of population
p	index of the training samples
N	total number of testing samples
$c_i(\cdot)$	copula function i
ϕ_i, η_i	parameters and weight of copula function i
γ	regularization parameter
ρ	controlling factor
p_{real}^t, p_{real}^t	measured value and predictive value of wind speed at time t
$1-\alpha$	nominal coverage probability of predication intervals
$\lambda_{\min}$	minimum value of forgetting factor
P_{BG}	significance level
l_0	segment length threshold
a_t, b_t	lower and upper bands of the tth PI
$I_t, \zeta(\cdot)$	0–1 variables
ς, ξ	hyper-parameter
R	interval length of the entire forecast results

models include time series models, grey models and Kalman filter (KF), etc. Time series models, such as autoregressive moving average (ARMA) [8] and autoregressive integrated moving average (ARIMA) [9], are extensively used in practice for short-term wind speed forecast. However, these conventional statistical methods are based on the assumptions: (1) wind speed data obey the norm distribution and (2) a linear correlation structure exists among time series wind speed. Unfortunately, it is well known that wind speed series are not normally distributed and need more complex functions to characterize the nonlinear relations of wind speed. With the development of artificial intelligence (AI), various models for time series-based wind speed prediction are mushrooming and have been employed to overcome the limitation of statistical models mentioned above. The most popular intelligent ones are the artificial intelligent approaches, such as artificial neural networks (ANNs) [10–12] and support vector machines (SVMs) [13–15]. Generally, AI based methods always provide a more competitive performance than the former methodologies because of their potential abilities for data-mining and feature-extracting [16].

Due to the volatile and chaotic nature of weather systems, individual methods mentioned above may not comprehensively represent the inner stochastic nature of wind speed series. As a remedy, many hybrid methods in combination with various models have been proposed for deterministic point WSF. In the literature available, the hybrid methods can be divided into four categories [1]: (1) weighting-based combined approaches [17,18]; (2) combined approaches including data pre-processing techniques [19–23]; (3) combined approaches including parameter selection [14,24]; (4) optimization techniques based combined approaches [14,19,20,22,24] and (5) combined approaches including error processing techniques [25,26]. Liu et al. [23] developed another four different hybrid models by combining four signal decomposing algorithms (e.g., WD (wavelet decomposition)/WPD (wavelet packet decomposition)/EMD (empirical mode decomposition)/EEMD (ensemble empirical mode decomposition)) and ELMs. The authors investigated the promoted percentages of the ELMs after applying these mainstream signal decomposing algorithms in the multiple step WSF. The obtained results showed that all the proposed hybrid algorithms outperform the single ELMs by utilizing the decomposing algorithms. Usually, ELM is applied as a forecasting engine in the combined models. Salcedo-Sanz et al. presented a novel approach for short-term wind speed prediction based on a Coral Reefs Optimization algorithm (CRO) and ELM network [24]. They found that the ELM network was able to provide excellent results within a very short computation time. Wang and Hu [27] proposed a robust short-term wind speed forecasting approach with the combination of ARIMA, ELM, SVM and Least Square SVM (LSSVM) using a Gaussian Process Regression (GPR) model for more precisely appraising wind speed. An overview of hybrid methods based WSF and wind power prediction was presented in [1,28].

Unlike ANN and SVM, which have more complex network structures and more parameters to be tuned, ELM is a single-hidden layer feedforward neural networks with fast learning speed [29]. However, ELM is based on empirical risk minimization principle that easily causes over-fitting phenomenon. In consideration of the outliers robustness problem and sparsity, in this paper, ORELM [30] are selected as the regression application. Furthermore, most of the above models are trained in an offline mode, which means that those models fail to cope with the underlying change of wind speed. Consequently, an algorithm for WSF shall ideally have as requisites: (i) a high flexibility to represent wind speed uncertainty, and (ii) time-adaptive characteristics. Illuminated by Soares and Rui [31], OS-ORELM with forgetting mechanism which reflected the timeliness of training data in the process of learning was presented. The time-varying forgetting factor (FF) is calculated

using Cook's distance (CD) [32] and updated by recursive least square (RLS) algorithm. This online model is termed as λ_{CDFF}OS − ORELM. To further improve the prediction accuracy, a new online learning ensemble of regressor models using an ordered aggregation (referred to as OEOA) technique made it available to provide on-line predictions of wind speed in changing environments [31]. In the data pre-processing phase: (1) BGA [33] is used to cut the raw data into several segments and therefore, the training and testing data is selected; (2) EMD is used to decompose the original signal into a number of modes to determine the number K of variational mode decomposition (VMD). In addition, detrended fluctuation analysis (DFA) is also developed to define the relevant modes to construct the filtered signal [34].

Most of the aforementioned models are useful for point or deterministic WSF in an offline mode. Although the deterministic forecasting method can achieve high accuracy, it fails to convey the uncertain characteristics of wind speed. Unlike the deterministic one, the probabilistic models contain more benefit information and can provide an effective range to estimate the probability of the point forecast result and further to implement power system operation. In the probabilistic framework, GPR [27,35] and quantile regression (QR) [36,37] are the most used models, whereas other models such as copula model, are under-explored in the field of wind speed and wind power probabilistic forecasting. Bessa et al. [38] presented a novel time-adaptive quantile-copula for wind power probabilistic forecasting, which is based on selecting adequate kernels for modeling the different variable types of the wind power problem. The authors stressed that the probabilistic model should represent the uncertainty of wind power and had time-adaptive characteristic. Several studies of time-adaptive models in the field of wind power forecasting have also been reported recently [39,40], yet relatively little research refers to wind speeds. This paper presents a novel hybrid intelligent algorithm for deterministic WSF, the results of which are further evaluated by performing probabilistic forecasts using TVMCF model. TVMCF enables to capture the uncertainty of wind speed, which is constructed by Gaussian copula function (GCF), Rotated Gumbel copula function (RGCF) and Symmetrised Joe-Clayton copula function (SJCCF). The value of parameters in TVMCF model is estimated using EM (expectation-maximization) algorithm [41] combined with WAQR (EM-WAQR) [42] approach. In addition, a modified CSO algorithm [19], called CSO with Self-Adaptive Mutation (CSO-SAM) is applied to optimize the parameters of ORELM networks.

On the whole, a hybrid WSF forecasting model combining BGA, AVMD, λ_{CDFF} OS-ORELM-OEOA, CSO-SAM and TVMCF is developed to estimate the deterministic prediction and probabilistic PIs associated with short-term wind speeds. The major contributions of this work are as follows:

- A hybrid model with on-line learning is initially developed to partition the raw wind speed series into sub-series by employing BGA and each sub-series is decomposed by AVMD.
- CSO-SAM algorithm, with a strong searching ability and low computational burden, is integrated into ORELM networks.
- An adaptive technique of λ_{CDFF} OS-ORELM-OEOA serves as an instrument to multi-step wind speed forecast and provides the deterministic prediction results.
- Probabilistic PIs are implemented based on the TVMCF model that, the conditional distribution of forecast error is formulated to estimate the lower and upper bound of the forecasting intervals.

The rest of the paper is organized as follows. Section 2 gives a brief description of the principle of BGA, AVMD, λ_{CDFF} OS-ORELM-OEOA model and TVMCF model. Modeling the approaches of the proposed hybrid technique and experimental setup is discussed in Section 3 while Section 4 presents the forecasting results of the proposed hybrid model and comparison analysis. Followed that conclusions are presented.

2. Methodology

2.1. BGA

With the characteristics of uncertainty and intermittence, extreme winds exhibit transient nonstationary features which may remarkably influence wind structure interaction. Inspired by the ideas of Bernaola-Galván, wind speed series can be associated with segments of different statistical properties (means and standard deviation). The detailed steps of BGA can refer to [33]. After the partition, the nonstationary and fluctuating wind speed series is composed of several segments with different mean value, in such way as to maximize the differences in the mean values between adjacent segments. Under this circumstance, there is no larger fluctuation in the subsequences and high forecasting accuracy will be obtained when applying the proposed forecasting model to each segment individually.

2.2. AVMD

Original VMD [43] aims to decompose a time series signal into a number of discrete sub-signals with certain sparsity properties and all sub-signals are mostly compact around a center frequency. In the application of VMD, the number K is required to predefine, and too few or too many modes K (overbinning or underbinning) has predictable impact on the efficiency of filtering. Motivated by the self-adaptability to the stochastic signal and the finite of decomposition (called intrinsic mode functions (IMFs)), EMD is used as an auxiliary method to help to preset the initial number of mode in VMD. Then a reliable criterion based on detrended fluctuation analysis (DFA) [36] is designed to select the optimal number of mode (i.e. K). DFA method is a scaling analysis method to measure long-range dependency for the non-stationary time series. The scaling exponent φ can be considered as an indicator of the series roughness: the larger the value, the smoother the time series or slower fluctuations. Based on the scaling exponent φ, the threshold region of $\theta = \varphi + 0.25$ is determined, where the IMFs with higher slope than the threshold are treated as denoising signal and used for reconstruction. Detailed steps of wind speed series decomposition using AVMD approach can be illustrated in Fig. 1.

2.3. λ_{CDFF} OS-ORELM-OEOA model

2.3.1. Basic principle of ORELM

ELM, proposed by Huang et al. [29], is a single-hidden layer feedforward networks (SLFNs) with the weights and bias in hidden nodes randomly assigned and the output weights decided by Moore-Penrose generalized inverse. The ELM training method can be defined in the following way: Given N arbitrary distinct samples (x_i, t_i), where $x_i = [x_{i1}, \ldots, x_{in}]^T \in \mathrm{R}^n$ and $y_i = [y_{i1}, \cdots, y_{im}]^T \in \mathrm{R}^m$, ELM with L ($L < N$) hidden nodes and the activation function $g(x)$ can be mathematically modeled as

$$\sum_{i=1}^{L} \beta_i g(w_i \cdot x_j + b_i) = y_j,\ j = 1, 2, \ldots, N \tag{1}$$

where $\mathbf{w}_i = [w_{i1}, \ldots, w_{in}]^T$ is the randomly selected input weight vector connecting the ith hidden neuron and the nth input neurons, b_i is the randomly selected bias of the ith hidden nodes, and β_i is the

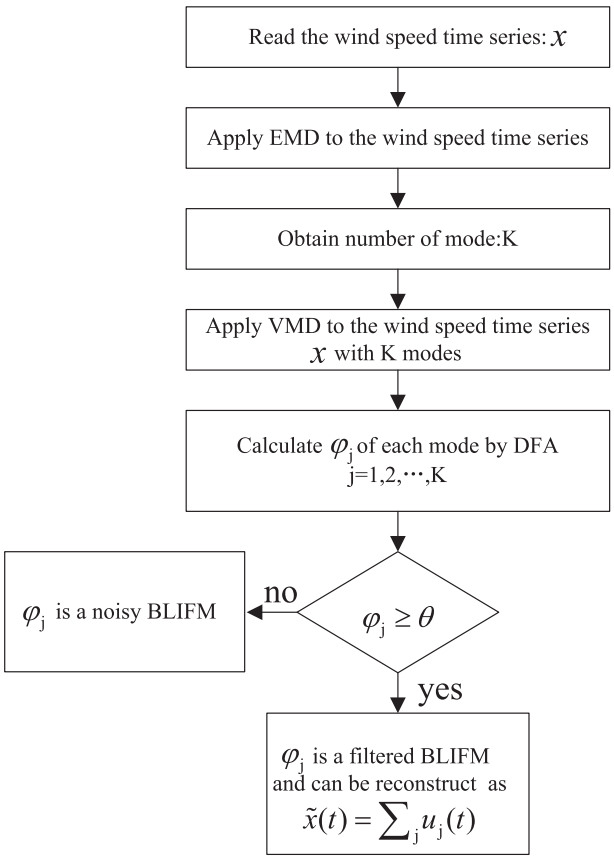

Fig. 1. The diagram of AVMD denoising.

weight that connects to the ith hidden neuron and the output neuron. The above N equations can be written compactly as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{y} \tag{2}$$

where

$$\begin{aligned}\mathbf{H} &= (w_1,\cdots,w_L,x_1,\cdots,x_N,b_1,\cdots,b_L)\\ &= \begin{bmatrix} g(w_1\cdot x_1+b_1) & \cdots & g(w_L\cdot x_1+b_L)\\ \vdots & \cdots & \vdots\\ g(w_1\cdot x_N+b_1) & \cdots & g(w_L\cdot x_N+b_L)\end{bmatrix}_{N\times L}\end{aligned} \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix}\beta_1^{\mathrm{T}}\\ \vdots\\ \beta_L^{\mathrm{T}}\end{bmatrix}_{L\times m} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix}y_1^{\mathrm{T}}\\ \vdots\\ y_L^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{4}$$

The optimal output weight vector $\hat{\boldsymbol{\beta}}$ can be calculated through

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{y} \tag{5}$$

by minimizing the associated loss function

$$\min\|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\| \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of the hidden layer output matrix $\mathbf{H}$.

An essential part in ELM is to minimize the training error, however, this may give rise to overfitting which will depress the predicting accuracy. In order to maintain the best tradeoff between train error and norm of output weight, a regularization parameter γ is introduced, which will make the ELM model with good generalization ability. Furthermore, the outliers occupy little of the wind speed series, which can be described with sparsity. As indicated in [30], ℓ_1-norm can not only guarantee the sparsity characteristic but also leads to the overall minimization convex. We rewrite the Eq. (6) as follow

$$\begin{cases}\min_{\beta}\|\mathbf{e}\|_1 + \frac{1}{\gamma}\|\boldsymbol{\beta}\|_2^2\\ \text{s.t.}\quad \mathbf{e} = \mathbf{y} - \mathbf{H}\boldsymbol{\beta}\end{cases} \tag{7}$$

This constrained convex optimization problem can be solved by the augmented Lagrange multiplier (ALM) method. As a result, a new iteration is generated as follow

$$\begin{cases}\beta_{k+1} = (\mathbf{H}^T\mathbf{H} + 2/\gamma\mu\mathbf{I})^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{e}_k + \upsilon_k/\mu)\\ \upsilon_{k+1} = \upsilon_k + \mu(\mathbf{y} - \mathbf{H}\beta_{k+1} - \mathbf{e}_{k+1})\end{cases} \tag{8}$$

and e_{k+1} is given explicitly by

$$\begin{aligned}e_{k+1} &= \text{shrink}(\mathbf{y} - \mathbf{H}\beta_{k+1} + \upsilon_k/\mu, 1/\mu)\\ &\triangleq \max\{|\mathbf{y} - \mathbf{H}\beta_{k+1} + \upsilon_k/\mu| - 1/\mu, 0\}\circ\text{sign}(\mathbf{y} - \mathbf{H}\beta_{k+1} + \upsilon_k/\mu)\end{aligned} \tag{9}$$

where μ is a penalty parameter and υ_k is the Lagrange multiplier; "$\circ$" represents the element-wise multiplication.

2.3.2. On-line sequential and update of ORELM

ELM and its improved algorithms (e.g. RELM and ORELM) are batch learning algorithms, which cannot reflect the timeliness of sequential training data well. Fortunately, online sequential extreme learning machine (OS-ELM) has been proposed to tackle this problem, learning data one-by-one or chunk-by-chunk (a block of data) with fixed or varying chunk size [31]. In this paper, the forgetting technique and recursive least square (RLS) method are used to modify the OS-ELM method. Variable FF using the Cook's distance enable to gradually expel the outdated data that could become a possible source of misleading information. The method is termed as λ_{CDFF} OS-ORELM.

The λ_{CDFF} OS-ORELM model has two phases: initialization phase and on-line learning phase. In the initialization phase, a wind speed series training data set of size T_0, $\mathbf{D}_0 = \{(x_t, y_t)\}_{t=1}^{T_0}$ from a total wind speed series data set $\mathbf{D} = \{(x_t, y_t)\}_{t=1}^{T}$ (with $T_0 < T$), is applied to initiate the model. In the on-line sequential learning phase, on-line samples from a wind speed series data set $\mathbf{D}_{online} = \{(x_t, y_t)\}_{t=T_0+1}^{T}$ are given incrementally one-by-one to retrain the forecast model online.

Step 1. Initialization phrase:

(1) Randomly pre-set the hidden nodes $L(L > 1)$ and assign input weights $\mathbf{w}$ and bias b ; $k = 0$;
(2) CSO (describe in Section 2.3.4) method is applied to search for optimal input weights $\mathbf{w}$ and bias b. The number of hidden nodes L_{opt} is selected based on the best performance on a 10-fold cross validation using the training data set in 1 trial;
(3) Calculate the output weight vector β_0 [31]

$$\boldsymbol{\beta}_0 = \mathbf{P}_0\mathbf{H}_0^T\mathbf{y}_0 \tag{10}$$

where $\mathbf{P}_0 = (\mathbf{H}_0^T\mathbf{H}_0)^{-1}$ and $\mathbf{y}_0 = [y_1,\cdots,y_{T_0}]^{\mathrm{T}}$; $\mathbf{H}_0$ is the initial hidden layer output matrix and is obtained using Eq. (3).

Step 2. On-line sequential learning: when a new sample (x_t, y_t) from $\mathbf{D}_{online}$ is available, it is employed to obtain a new output weight vector β_{k+1} using the RLS with CDFF as follows:

(1) Calculate the $(k+1)$-th hidden layer output matrix

$$h_{k+1} = [g(\mathbf{a}_1, b_1, \mathbf{x}_t), \ldots, g(\mathbf{a}_{L_{opt}}, b_{L_{opt}}, \mathbf{x}_t)] \tag{11}$$

(2) Update β_{k+1} using RLS

$$\beta_{k+1} = \beta_k + \mathbf{\Gamma}_{k+1} h_{k+1}^T e_{k+1} \tag{12}$$

with $e_{k+1} = y_{k+1} - h_{k+1}^T \beta_k$ being the one-step-ahead prediction error. The gain matrix $\mathbf{\Gamma}_{k+1}$ is a measure of the dispersion of the estimate β_{k+1} that holds

$$\mathbf{\Gamma}_{k+1} = \frac{1}{\lambda_{k+1}} \left(\mathbf{\Gamma}_k - \frac{\mathbf{\Gamma}_k \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{\Gamma}_k}{\lambda_{k+1} + \mathbf{h}_{k+1}^T \mathbf{\Gamma}_k \mathbf{h}_{k+1}} \right) \tag{13}$$

(3) Calculate the time-varying Cook's distance of new observation [32]

$$D_{k+1} = \frac{\mathbf{h}_{k+1}^T \mathbf{\Gamma}_k \mathbf{h}_{k+1} e_{k+1}^2}{m \sigma_k^2 (1 + \mathbf{h}_{k+1}^T \mathbf{\Gamma}_k \mathbf{h}_{k+1})} \tag{14}$$

where m is the length of β and σ_k^2 is a consistent estimate of σ^2, with

$$\sigma_k^2 = M_k^{-1} \sum_{j=1}^{k} \left(\prod_{i=j+1}^{k} \lambda_i \right) (y_j - \mathbf{h}_j^T \beta_j)^2 \tag{15}$$

$$M_k = \sum_{j=1}^{k+1} \left(\prod_{i=j+1}^{k+1} \lambda_i \right) = 1 + \lambda_{k+1} M_k \tag{16}$$

where M_k is the so-called effective sample size with $M_0 = 0$.

(4) Update adaptive FF based on Cook's distance with the statistics

$$C_k = v D_k \tag{17}$$

where C_k obey the chi-squared distribution with v degrees of freedom. Denote

$$S_k = P(\chi_v > C_k)(0 < S_k < 1) \tag{18}$$

Therefore, Cook's distance is translated to FF as

$$\lambda_k^{CDFF} = \lambda_{\min} + (1 - \lambda_{\min}) S_k \tag{19}$$

where $S_k \equiv S(C_t)$ is the survivor function and $\lambda_{\min}$ is the lower value of threshold.

2.3.3. Multiple λ_{CDFF} OS-ORELM selection using OEOA

Alternatively, the ensemble learning technique has been proven itself as an effective system to dispose the concept drift problem. This paper proposes an adaptive and online ensemble based on λ_{CDFF} OS-ORELM using ordered aggregation (OA). OEOA designs an ensemble model by employing a slide window (SW) approach, which assumes that the most recent data provides the best and most relevant representation of the current state of the process and of the near-future state [31]. It starts from a fixed data window and when a new sample is available, the oldest sample will be removed from the window. New model trained by data window will be launched to the ensemble while models that do not contribute to the ensemble are excluded. The basic procedure of OEOA is listed as follows [31]:

- Creat a pool of multiple λ_{CDFF} OS-ORELM model
- On-line learning phase
  - Multiple model dynamical selection from the pool of models
  - Output prediction based on the ordered aggregation of the best models where weights of the subset of models are assigned dynamically
  - Update the pool of models using the recent samples

2.3.4. λ_{CDFF} OS-ORELM optimized by CSO-SAM

The CSO is a recent evolutionary algorithm which has a huge advantage in solution accuracy and convergence speed when addressing complex optimization problems. However, as stated in [19], there exists the probability both in the Horizontal crossover (HC) and Vertical crossover (VC) operation, which have great influence on the performance of these two operations (i.e. escaping from local optimum). While the VC probability (i.e. P_{vc}) is difficult to predefine due to the fact that fixed (overestimating or underestimating) P_{vc} value sometimes limits the global search capability of CSO. Therefore, we design a self-adaptive adjustment mechanism to adjust the P_{vc} value automatically based on the statistics properties of the population fitness value (i.e. variance) during the convergence procedure. The proposed modified CSO with Self-Adaptive Mutation (CSO-SAM) to fit the P_{vc} value is based on the variance τ^2 of the population fitness values. The expression of τ^2 can be shown in Eq. (20)

$$\begin{cases} \tau^2 = \sum_{i=1}^{n} [(o_i - o_{avg})/o]^2 \\ o = \max\left\{ \max_{1<i<n} |o_i - o_{avg}|, 1 \right\} \end{cases} \tag{20}$$

where o_i is the fitness value of particle i; o_{avg} is the average value of the current fitness value; n_{pop} is the number of population.

Finally, the P_{vc} value is determined by adopting a linearly decreasing strategy, as shown in (21)

$$P_{vc}(t) = (P_{vc\max} - P_{vc\min})(1 - \tau_t^2/n) + P_{vc\min} \tag{21}$$

where $P_{vc\min}$ and $P_{vc\max}$ are the preset maximum and minimum value of P_{vc}, respectively. Referring to the empirical values given by Meng et al. [19], we set $P_{vc\min} = 0.2$ and $P_{vc\max} = 0.8$.

The MSE of the wind speed prediction modeled by OS-ORELM will be used as the object function in the CSO-SAM. The fitness of every convergence solution can be measured as follow

$$E_{rr} = \frac{1}{N} \sum_{p=1}^{N} \sum_{t=1}^{m} \left(y_t^{(p)} - \hat{y}_t^{(p)} \right)^2 \tag{22}$$

where $y_t^{(p)}$ and $\hat{y}_t^{(p)}$ are the actual wind speed and forecast value of the tth neuron, respectively, N is the number of training samples and p is the index of training sample.

2.4. TVMCF model

Stochastic dependence refers to the behavior of a random variable that is affected by others (e.g., wind speed prediction and forecasting errors are treated as two dependent probabilistic variables). Modeling the stochastic dependence is challenging because it is hard to find multivariate analytical formulas for variables with complex marginal distributions and in most cases, the distribution of wind speed (or forecast errors) is designed to a special distribution, such as Beta distribution, Weibull distribution, norm distribution et al. However, as pointed out in [44], the forecast error may not have a special distribution for all the point forecast levels.

The copula technique can provide an effective way of modeling the stochastic dependence of wind speed forecast and its corresponding forecast errors. There are several families of copulas, for example Gaussians, Elliptical or Archimedian copulas. Each single copula has its advantage to reflect the correlation of the bivariate. However, single copula is unable to convey the probabilistic wind speed comprehensively (especially the tail dependence). Additionally, a constant model cannot adapt to the time-varying wind speed series. For this reason, a time-varying mixture copula function (TVMCF) is proposed to fit a good joint distribution of point forecast and corresponding error. The time-varying Gaussian copula function (TVGCF), time-varying Rotated Gumbel Copula func-

tion (TVRGCF) and time-varying Symmetrized Joe-Clayton Copula function (TVSJCCF) are chosen to construct the mixture copula (MC). The TVMCF is expressed as follow

$$c(F_X(x), F_Y(y)) = \sum_{i=1}^{3} \eta_i c_i(F_X(x), F_Y(y), \phi_i) \quad (23)$$

where $c_i(\bullet)$ denotes to TVGCF, TVRGCF and TVSJCF; ϕ_i is the corresponding parameters and η_i is the weighting factor of each single copula, with $\sum_{i=1}^{3} \eta_i = 1$.

The core problem of TVMCF model is parameters estimation. General strategy of copula parameter estimation is not available for the proposed model due to its mixed nature. Therefore, the EM Algorithm combined the weighted average quantile regression (EM-WAQR) method are applied to estimate the parameters of TVMCF. The EM algorithm [41] includes two mainly steps: the EM iteration alternates between performing an expectation step (E step), which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization step (M step), which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. During each iteration of the M step, the WAQR [42] approach is used to estimate the parameters of single copula.

3. Hybrid WSF framework

3.1. General framework of proposed short-term WSF model

The proposed deterministic and probabilistic WSF approach is a combination of AVMD, λ_{CDFF} OS-ORELM and TVMCF. While BGA is used to select the training and testing data and just implements one time during the WSF simulation. The structure of the proposed forecast model is shown in Fig. 2. As illustrated, the deterministic prediction includes three parts. In part I, data preprocessing is considered which the available inputs are linearly normalized to the range [0, 1] to overcome the saturation phenomenon. Then BGA is used to partition the non-stationary wind speed time series into segments (i.e. stationary sub-series), this will help to escape from the extreme circumstance, especially the abrupt changing of the wind speed. In part II, every segment time series is decomposed applying AVMD. We adopt the CSO-SAM algorithm to optimize the λ_{CDFF} OS-ORELM-OEOA forecast model of each subsequence in the part III. Each segment recombines its subseries in the last step of deterministic prediction to obtain the final deterministic prediction results. In this paper, we do the one-step, three-step and five-step prediction for each subseries.

During the probabilistic WSF, training and retraining results of the forecast model are obtained to conduct a statistics mathematically. We focus on the post-processing technique for the training and retraining results, and then model the conditional probability density function (PDF) of forecast error using conditional TVMCF. Based on this approach, the probabilistic PIs at a confidence level are constructed.

3.2. Parameters selection

BGA: The significance level P_{BG} and segment length threshold l_0 are set to be 95% and 700, respectively.

Input selection: As proved in [19], the 6 inputs of the ANN has performed a good prediction accuracy, so in this study, for ORELM, ELM and BPNN (back propagation neural network) applied to all the sub-sequences in every season, the input number of input neuron is set to be 6.

CSO-SAM algorithm:

(1) Horizontal crossover probability $P_{hc} = 1.0$.
(2) Maximum and minimum vertical crossover probability $P_{v\text{cmax}} = 0.8$ and $P_{v\text{cmin}} = 0.2$.
(3) Population size: 10.
(4) Maximum iterative times: 100.

λ_{CDFF} OS-ORELM-OEOA algorithm: In initial phrase of λ_{CDFF} OS-ORELM, the initial training dataset size T_0 is set to be 20 according to the empirical suggestion by Soares et al. [31]. Since the number of neuron in the hidden layer L should not be greater than T_0 in order to comply with the assumptions in the ELM algorithm. Therefore, L is selected by varying it in the interval of [1,20]. Another parameter called regularization parameter γ should be estimated with L together. Let $\gamma = [2^{-50}, 2^{-49}, \cdots, 2^{49}, 2^{50}]$, then each pair of (L, γ) are obtained based on the best performance on a 10-fold cross-validation method using the initial training dataset. The minimum forgetting factor $\lambda_{\min} = 0.6$. The controlling factor of OEOA is set to be $\rho = 0.04$.

EEMD algorithm [22]: In this case study, the amplitude (standard deviation) of the white noise is 0.5 and the ensemble number is set as 200.

4. Forecast model validation and results

4.1. Experimental data description and setup

In this paper, the proposed hybrid WSF approach has been comprehensively tested and benchmarked on wind speed datasets at a wind farm located in the Texas, USA (Latitude: 31.19N, Longitude: 102.24W). Those wind speed data are measured at the elevation of about 850 m and are available in [45]. The historical wind speed data are collected in 2004, 2005 and 2006 with a 10 min resolution. To reduce the influence of seasonal patterns on wind speed prediction, the following months are randomly selected in each year: April, July, October, and January, corresponding to the four seasons of each year (as shown in Fig. 3). The total sampling points of the four representative months are 4320, 4464, 4464 and 4464, respectively. Accordingly, four different prediction models with the same architecture are adopted to fulfill the seasonal forecast task, because the weather conditions and the wind speed data vary significantly in different seasons. Unlike most paper that randomly selected the wind speed data set, in this paper, the data set is available after using BGA to partition the monthly wind speed series into several segments, as detailed in Section 2.1. The descriptive statistics of the partitioned data sets, including the mean and the standard deviation are summarized in Table 1. Among each segment, the last 100 points are chosen as the validation data and the remaining samples are used as training data. For the online models (i.e. λ_{CDFF} OS-ORELM and λ_{CDFF} OS-ORELM-OEOA), the first 20 samples (i.e. $T_0 = 20$) of the training data are applied to initiate the model according to the empirical value in Ref. [31]. To fully validate the effectiveness of the proposed algorithm, the results are compared with the λ_{CDFF} OS-ORELM, Offline-ORELM, persistence model BPNN and Offline-ELM. To further validate the effectiveness of the proposed method, EEMD based above approaches are used as contrasts. In our study, the 5th segment in Table 1(i.e. S5) is chosen to implement the multi-step predictions, the 1th to 551th points are used as training data and the remaining (552th to 651th points) are used as testing data. All of the prediction simulations are implemented in MATLAB R2014a and carried out on a personal computer with a Pentium(R) Dual-Core E6600 3.2-GHz CPU and 4.00 GB of RAM. In order to get a fair prediction results as well as reduce the fore-

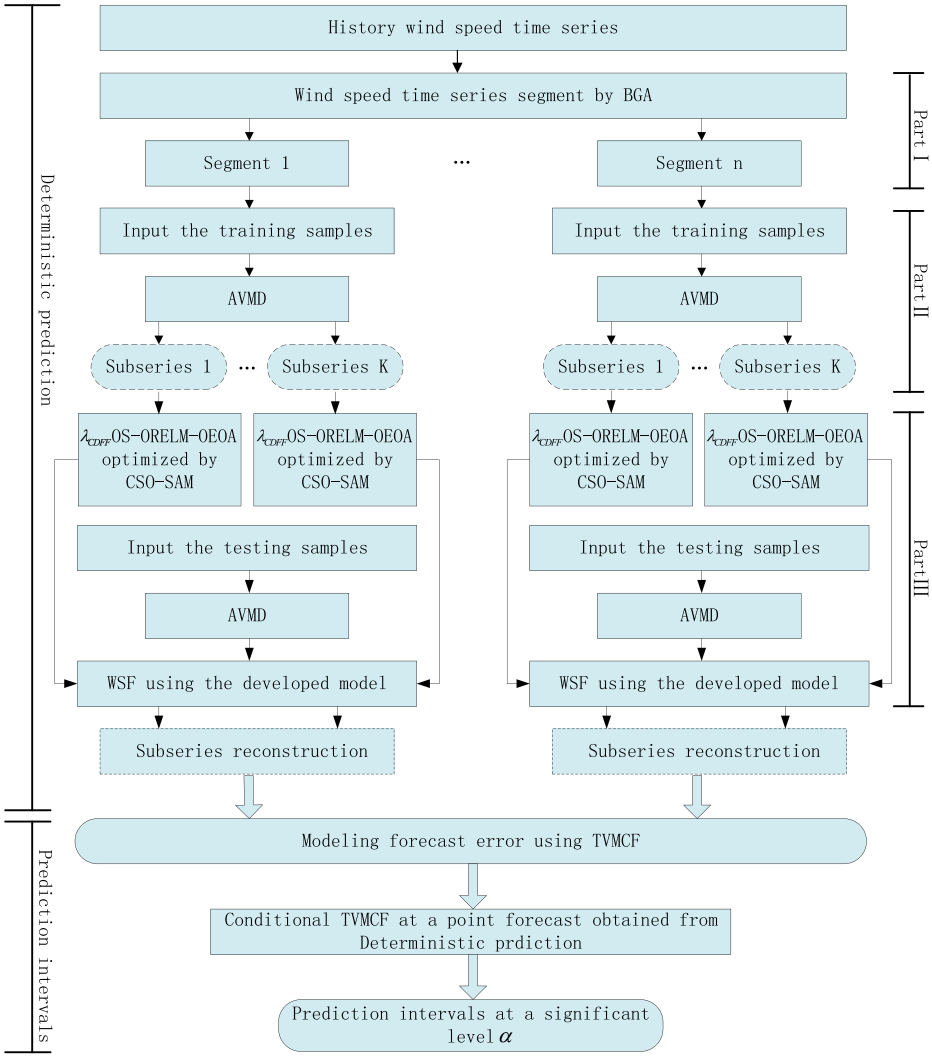

Fig. 2. The structure of the proposed forecast model.

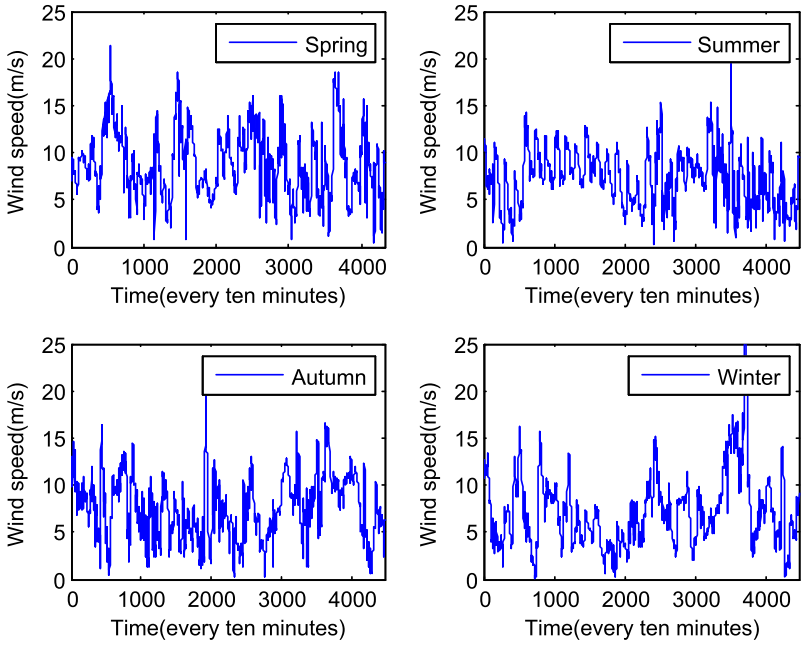

Fig. 3. Seasonal characteristic of wind speed time series.

Table 1
Partitioned results of wind speed time series in spring, 2004.

Segment	Start point	Terminal point	Cut point	Mean	Standard deviation
S1	1	434	435(1)	8.324	1.697
S2	436	767	768(1)	13.111	2.363
S3	769	1395	1396(1)	6.821	2.348
S4	1397	2012	2013(1)	9.337	3.662
S5	2014	2663	2664(1)	10.768	2.348
S6	2665	2972	2973(1)	8.653	3.227
S7	2974	3602	3603(1)	6.436	2.625
S8	3604	3757	3758(1)	14.354	2.929
S9	3759	4320	–	7.365	2.779
All	1	4320	–	8.827	3.487

'–' denotes a null value and the value 1 in the bracket means cutting the point with probability 1.

cast error, the forecast model corresponding to the sub-series will be executed 50 times independently.

4.2. Statistics metric of the forecast model

In order to quantitatively determine the performance of the predictive model, three criteria are used [19,20]. They are the mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean-squared error (RMSE), which are calculated as follows, respetively.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{p_{real}^{t} - p_{pre}^{t}}{p_{real}^{t}}\right| \times 100\% \tag{24}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|p_{real}^{t} - p_{pre}^{t}\right| \tag{25}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(p_{real}^{t} - p_{pre}^{t}\right)^{2}} \tag{26}$$

where p_{real}^{t} and p_{pre}^{t} denote measured value and predictive value of wind speed. N is the total number of samples used for performance evaluation.

Practically, narrow width of probabilistic PIs is more informative than wide ones. Therefore, for the probabilistic PIs at a confidence level $1-\alpha$, prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) [18] are used as the assessment indexes, which can be expressed mathematically as

$$\mathrm{PICP} = \frac{1}{N}\sum_{t=1}^{N} I_t \tag{27}$$

where

$$I_t = \begin{cases} 1, & \text{if } p_{pre}^{t} \in [a_t, b_t] \\ 0, & \text{if } p_{pre}^{t} \notin [a_t, b_t] \end{cases}$$

$$\mathrm{PINAW} = \frac{1}{RN}\sum_{t=1}^{N}(a_i - b_i) \tag{28}$$

where N is the number of test samples, p_{pre}^{t} represents the point forecast, and μ and b_t are the lower and upper bands of the tth PI, respectively. R is the interval length of the entire forecast results (i.e. the subtraction between the maximum and the minimum forecast value).

A combinational coverage width-based criterion (CWC), introduced by Khosravi et al. [46], is used for simultaneous evaluation of the quality of constructed probabilistic PIs from both perspectives (i.e. PICP and PINAW), namely

$$\mathrm{CWC} = \mathrm{PINAW}(1 + \zeta(\mathrm{PICP})e^{-\varsigma(\mathrm{PICP}-\xi)}) \tag{29}$$

where $\zeta(\mathrm{PICP})$ is defined by the following step function

$$\zeta(\mathrm{PICP}) = \begin{cases} 1, & \text{if PICP} < \xi \\ 0, & \text{if PICP} \geqslant \xi \end{cases} \tag{30}$$

t and ξ in (29) are two hyper-parameters controlling the magnitude of CWC index. ξ is the nominal confidence level (i.e. $1-\alpha$) for which the intervals are constructed and ς regulates the quality of probabilistic PIs by imposing penalty on invalid PIs. In practice, ξ equals to confidence level $1-\alpha$ and ς is predefined to be a value between 50 and 100 to highly penalize invalid PIs. In fact, the CWC measure tries to compromise between informativeness (measured by PINAW) and correctness (measured by PICP) of probabilistic PIs.

4.3. Deterministic WSF results and comparative analysis

4.3.1. Wind speed series signal decomposition and reconstruction

According to Ref. [43], Dragomiretskiy and Zosso indicated that underestimation or overestimating of modes K can be tested empirically by checking the spectral overlap between modes, or by considering the residuals. That means the number of modes by VMD should be assigned in advance. In this study, the initial value of K is determined by EMD. The experimental results have proved that pure band-limited intrinsic mode functions (BLIMFs) can be extracted by VMD with the same number of modes of EMD, using the scaling exponent α as a reliable metric to evaluate the denoised sub-sequence. We take the 5th segment (see in Fig. 4) obtained by BGA as an example to demonstrate the AVMD to select the K modes. The initial K value is obtained from EMD, and the threshold region is set to be $\theta = 0.5 + 0.25$ (i.e. $\theta = 0.75$) according to the empirical value in Ref. [34]. Fig. 4 depicts the decomposition of the wind speed time sub-series of the 5th segment. The scaling exponent values for all the modes are shown in Fig. 5, and the red line represents the threshold θ according to the Ref. [34]. From Fig. 5, we can see that the scaling exponent of the first three BLIMFs are larger than the threshold θ, which are considered as noise free sub-series. Consequently, the 1th to 3th modes are reconstructed into one subseries. While the rest sub-series will be treated as impure sub-signals and should be an individual input to the well-developed forecast models. In this way, the final number of modes obtained by AVMD is 5.

4.3.2. Results and comparative analysis

To determine quantitatively the best model, the corresponding evaluation results of the 5th segment in Table 1 using different forecasting engines without signal decomposition technique (i.e. AVMD and EEMD) are tabulated in Table 2. Additionally, in combination with AVM and EEMD, the estimated error results are indicated in Tables 3 and 4, respectively.

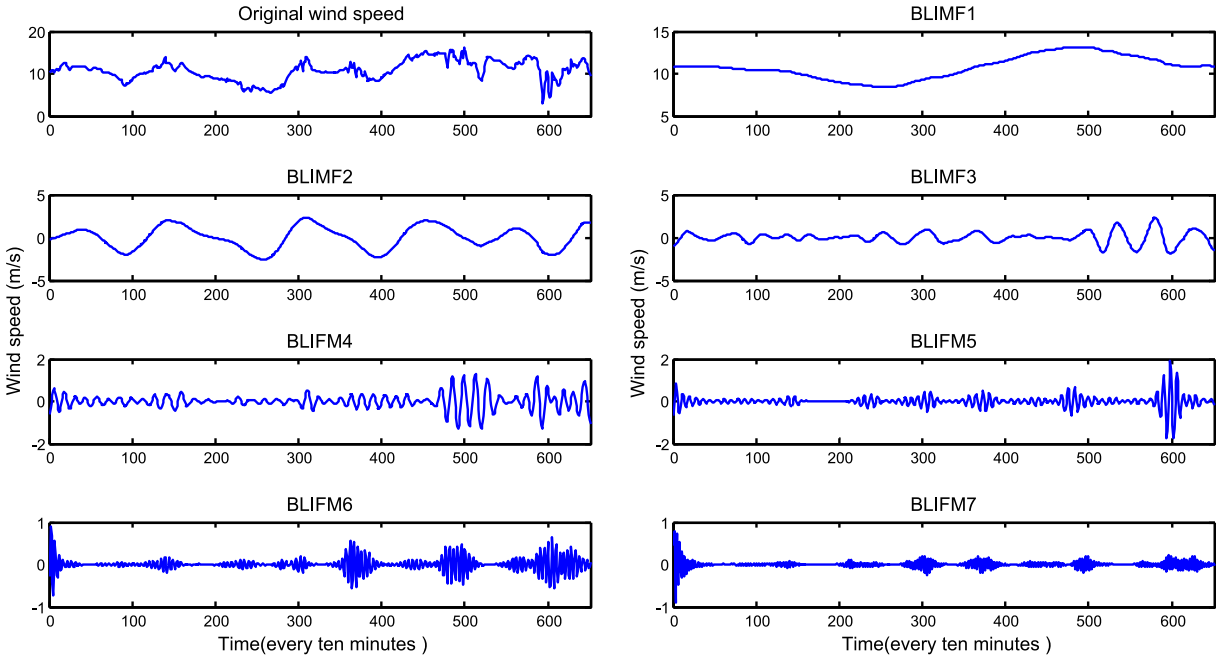

Fig. 4. Decomposition of original wind speed by EMD-VMD with its BLIMFs 1–7.

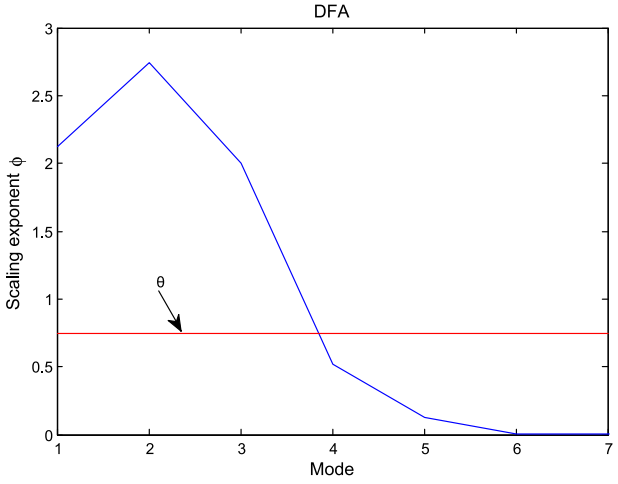

Fig. 5. DFA of wind speed sub-series.

From Table 2, it can be analyzed that the proposed λ_{CDFF} OS-ORELM-OEOA model performs the best among all the forecasting models in one-step, three-step and five-step predictions, with the MAPE index of 0.384%, 0.851% and 0971%, respectively. In terms of one-step prediction, the persistence model presents a worse performance. While in terms of three-step and five-step predictions, the poor results come from Offline-ORELM and Offline-ELM, respectively. This means that, though single ORELM and ELM has the fast learning speed due to their better generalization capacity, they are not always suitable for all the forecasting horizons.

From Tables 2 and 4, it can be concluded that: (a) all the AVMD and EEMD based forecasting approaches have much better forecasting performance than the individual model. For example, the MAPE index decline to 0.324% and 0.341% respectively from 0.384%. This is mainly because after the AVMD and EEMD auxiliary, the wind speed time series is decomposed into an ensemble of band-limited intrinsic mode functions and more pure sub-signals are obtained, which meet the stationary forecast model mathematically; (b) compared to λ_{CDFF} OS-ORELM, Offline-ORELM, Offline-

Table 2
Multi-step forecasting evaluation indexes of different approaches and their CPU time (in second).

Index	One-step	Three-step	Five-step	One-step	Three-step	Five-step
	λ_{CDFF} OS-ORELM-OEOA			λ_{CDFF} OS-ORELM		
MAE (m/s)	0.041	0.096	0.112	0.094	0.146	0.329
MAPE (%)	0.384	0.851	0.917	0.134	2.512	4.077
RMSE (m/s)	0.052	0.136	0.159	0.078	0.368	0.571
CPU time (s)	58.375	69.347	89.048	31.122	35.879	42.763
	Offline-ORELM			Persistence model		
MAE (m/s)	0.211	0.566	0.910	0.306	0.506	0.634
MAPE (%)	3.750	6.662	11.557	3.136	5.255	6.602
RMSE (m/s)	0.550	1.056	1.661	0.528	0.875	1.065
CPU time (s)	5.064	5.231	5.527	4.352	4.696	4.542
	BPNN			Offline-ELM		
MAE (m/s)	0.157	0.623	0.928	0.245	0.536	0.959
MAPE (%)	2.831	8.108	12.291	3.971	6.611	12.524
RMSE (m/s)	0.409	1.187	1.631	0.535	1.019	1.712
CPU time (s)	68.440	72.839	105.761	5.726	5.841	5.924

Table 3
Multi-step forecasting evaluation indexes of AVMD based approaches and their CPU time (in second).

Index	One-step	Three-step	Five-step	One-step	Three-step	Five-step
	λ_{CDFF} OS-ORELM-OEOA			λ_{CDFF} OS-ORELM		
MAE (m/s)	0.034	0.083	0.091	0.061	0.126	0.259
MAPE (%)	0.324	0.825	0.892	0.908	2.086	3.877
RMSE (m/s)	0.049	0.121	0.133	0.068	0.318	0.441
CPU time (s)	338.235	464.887	579.078	131.655	135.451	143.842
	Offline-ORELM			Offline-ELM		
MAE (m/s)	0.110	0.277	0.333	0.101	0.279	0.350
MAPE (%)	1.245	3.232	4.237	1.323	3.309	4.328
RMSE (m/s)	0.165	0.438	0.532	0.165	0.406	0.561
CPU time (s)	31.964	32.201	32.925	11.726	11.841	11.924
	BPNN			Persistence model		
MAE (m/s)	0.067	0.260	0.495	0.300	0.474	0.597
MAPE (%)	0.890	3.833	6.507	3.076	4.930	6.223
RMSE (m/s)	0.087	0.308	0.795	0.498	0.801	0.979
CPU time (s)	410.543	594.419	717.670	25.388	26.029	29.091

Table 4
Multi-step forecasting evaluation indexes of EEMD based approaches and their CPU time (in second).

Index	One-step	Three-step	Five-step	One-step	Three-step	Five-step
	λ_{CDFF} OS-ORELM-OEOA			λ_{CDFF} OS-ORELM		
MAE (m/s)	0.041	0.091	0.103	0.088	0.139	0.303
MAPE (%)	0.341	0.894	0.928	0.125	2.487	3.966
RMSE (m/s)	0.079	0.157	0.289	0.074	0.349	0.514
CPU time (s)	457.656	552.263	657.273	151.340	155.278	163.641
	Offline-ORELM			Offline-ELM		
MAE (m/s)	0.170	0.304	0.455	0.156	0.387	0.525
MAPE (%)	2.018	3.887	5.395	1.952	4.564	6.293
RMSE (m/s)	0.277	0.574	0.782	0.270	0.604	0.758
CPU time (s)	42.003	43.778	44.782	44.179	44.661	45.483
	BPNN			Persistence model		
MAE (m/s)	0.073	0.282	0.554	0.316	0.512	0.640
MAPE (%)	0.914	4.168	7.651	3.242	5.320	6.654
RMSE (m/s)	0.091	0.341	0.979	0.532	0.877	1.067
CPU time (s)	607.878	675.487	773.890	106.715	108.439	110.165

ELM, BPNN and persistence model, the proposed approach has the best performance in the multi-step predictions regardless of the combination of AVMD and EEMD technique. For example, in terms of AVMD based model, the MAE index has been averagely improved by 44.262%, 69.091%, 88.889%, 49.254% and 69.307% respectively. The results confirm the conclusion that: (i) the prediction accuracy can be improved by AVMD when applied to the ORELM networks trained by CSO-SAM algorithms; (ii) the adaptive ensemble of online ORELM develops a robust forecast model to adjust the changing condition and enhances the prediction precision; (c) in terms of both AVMD and EEMD based models, persistence model presents a poor performance; (d) when comparing the proposed approach with EEMD-λ_{CDFF} OS-ORELM-OEOA, the promoted percentages of MAPE by the proposed approach from one-step to five-step are 4.985%, 7.718% and 3.879%, respectively. Furthermore, there is an obvious phenomenon that in the multi-step prediction results the online models (i.e. the first two model) are superior to the offline models (i.e. the remaining four models). The large dispersion of the error indicates that due to its characteristic of offline prediction, which disables to consider the influence of old data on the new observers and keeps the model parameters to every new observer. Conversely, the forgetting mechanism based on Cook's distance helps progressively reduce the importance of old data and is well-developed in the online sequence forecast model to estimate the model parameters adaptively and greatly enhance the forecast accuracy; (e) in terms of CPU times, however, the high prediction accuracy of the proposed hybrid model is at the cost of CPU time, yet it is still slightly fast than the BPNN model. Instead, the persistence model has the least CPU time both adopting the AVMD and EEMD technique. In brief, the results given in Tables 2 and 4 further substantiate the effec-

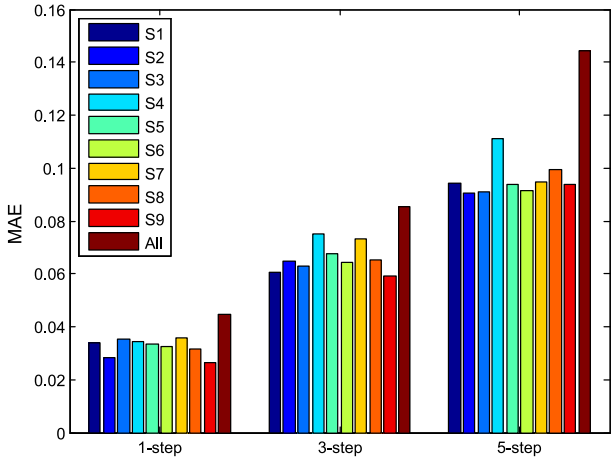

Fig. 6. Multi-step forecasting results of MAE between different segments in Table 1.

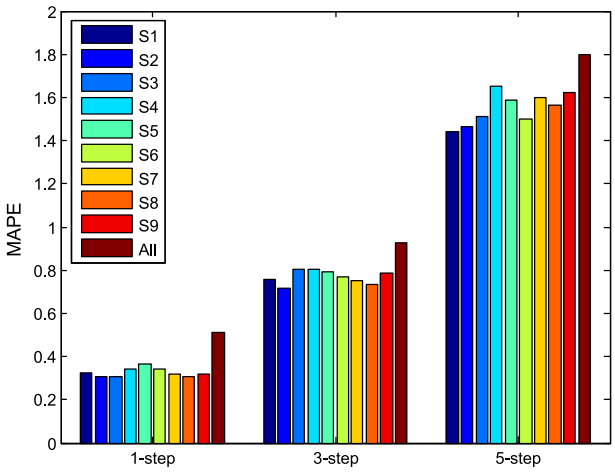

Fig. 7. Multi-step forecasting results of MAPE between different segments in Table 1.

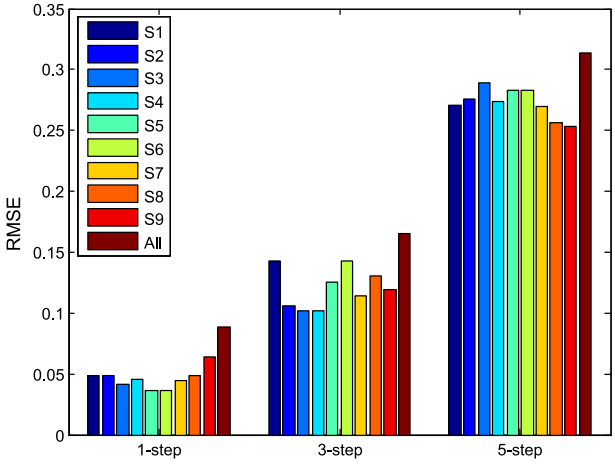

Fig. 8. Multi-step forecasting results of RMSE between different segments in Table 1.

tiveness of the proposed hybrid approach for short-term wind speed prediction.

Similarly, the other segments cut by BGA can adopt the proposed model to execute WSF. The full prediction results of the wind speed time series in Fig. 1 are shown in Figs. 6–8. Meanwhile, it can be observed that the prediction error results of different segments do rise to certain extent when the prediction horizon increases. It indicates that longer prediction horizon implies the stronger stochastic uncertainty for forecasting wind speed. What's more, after cut the wind speed into segments using BGA, the forecast error results degrade outstandingly. This further indicates the effectiveness of our proposed method. Fig. 9 illustrates the histogram of the one-step forecasting results in 2006. It is clearly found that the distribution of the forecasting results and the actual value is quite similar and the forecasting error is close to zero.

4.3.3. Additional case study

To future investigate the performance of the proposed model, an additional site which located offshore (Latitude: 47.11N, Longitude: −90.45W), is applied to do the multi-step ahead predictions in our paper. As illustrated in Fig. 10, the first 587 samples are used for training and the remaining 100 samples are applied to test the well-developed model. The forecasting results of the additional case are presented in Table 5 and the conclusions are similar to the ones which made in Section 4.3.2. The results manifest that the proposed model is effective in forecasting the short-term WSF.

4.4. Probabilistic WSF results and comparative analysis

In this experiment, the wind speed time series collected in 1 January 2004 to 31 December 2005 are selected as the training datasets and the test datasets is from 1 January 2006 to 31 December 2006. Based on the deterministic forecast results of wind speed in 2004 and 2005, the TVMCF model discussed in Section 3 is applied to execute the probabilistic PIs of wind speed in 2006. In our study, the wind speed time series in January 2006 is selected as an example to implement the probabilistic PIs using the one-step, three-step and five-step wind speed forecasting results. The probabilistic PIs model based on GPR is used for forecasting performance comparison with our proposed approach.

The series-based wind speed is first partitioned into 12 subsequences using BGA, as illustrated in Fig. 11, and each subsequence implements the probabilistic PIs individually. Similarly, the maximum length of the sub-sequences (i.e. the 4th subsequence with 677 points) is selected as an example to present the results and comparison analysis. Similarly, in this subsection, we assess the performance of TVMCF with three different forecasting horizons, i.e., one-step, three-step and five-step ahead.

The analysis in the Section 4.3 focuses on the forecasting accuracy of deterministic prediction. The TVMCF model, by modeling conditional forecast error, provides an effective way of modeling stochastic dependence and probabilistic predictive distribution beyond the special distribution, such as the normal distribution. Fig. 12 shows the predictive PDF and the 95% confidence interval of the wind speed prediction, which mirror the uncertainty and volatility in the predicted wind speed. We can see that the actual wind speed values all fall within the regions predicted from the forecast distributions. In addition, all the actual wind speed observations fall within the 95%-level PIs. In Ref. [27], the uncertainties of wind speed have been characterized by Gaussian PDF. The model comparison between TVMCF and that in Ref. [27] demonstrates that:

(1) The distribution parameters of Gaussian PDF for some particular wind speed levels are difficult to determine using historical measurements because of the low occurrence probability of these output levels. In contrast, the functions in the proposed model are not specific to any wind speed level. Thus, all of the historical data can be used to give the best estimate of the parameters.
(2) From Fig. 12, it can be found that the distributions of the predictive wind speed cannot always obey the Gaussian distribution for all the point forecast levels, in other word, the point forecast does not always provide an unbiased estimation. The predictive PDF of the mean 10-min wind speed for the 81th test of data can illustrates this phenomenon obviously.
(3) The TVMCF based probabilistic PIs formulate the predictive PDF by considering the forecast error, which can provide a comprehensive correlation between forecast error and the predicted value. The proposed model can also capture the tail dependency of wind speed, though with low occurrence probability, which is important to the system uncertainty and operational risk. This quite different from that using the GPR in Ref. [27].

Figs. 13–15 visually illustrate the forecast values obtained from the suggested model and the corresponding 95% confidence intervals from one-step to five-step, employing the TVMCF and GPR, respectively. Khosravi et al. [46] demonstrated that the con-

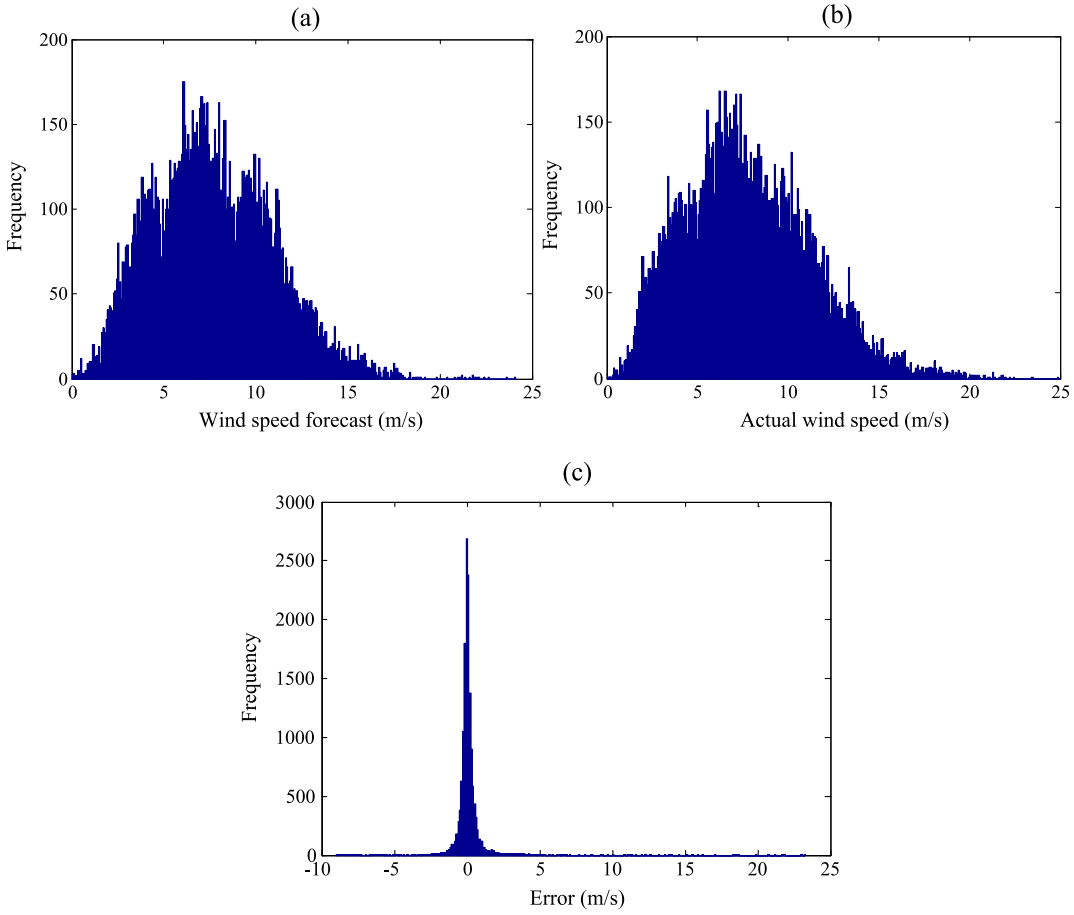

Fig. 9. Histogram of the one-step wind speed forecasting results during the year of 2006: (a) the forecasting value; (b) the actual value and (c) the corresponding forecast error.

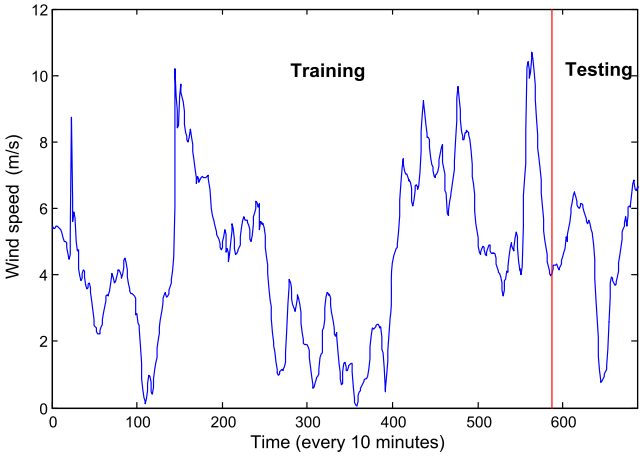

Fig. 10. Offshore wind speed time series.

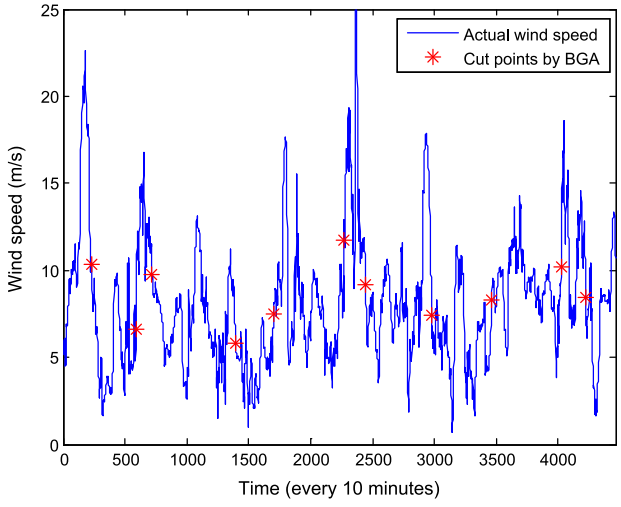

Fig. 11. Series-base wind speed in January 2006.

Table 5
Multi-step forecasting results of the additional case.

Index	AVMD-λ_{CDFF} OS-ORELM-OEOA		
	1-step	3-step	5-step
MAE (m/s)	0.013	0.033	0.115
MAPE (%)	0.622	1.561	1.769
RMSE (m/s)	0.020	0.048	0.064
CPU time (s)	331.866	436.328	542.705

structed probabilistic PIs are theoretically valid if their coverage probability is greater than or equal to the corresponding nominal confidence level. It can be discovered from Figs. 13–15 that all of the actual observations fall within the probabilistic PIs when using the TVMCF. For the GPR, however, except the one-step horizon, the other two horizons cannot make sure that all the actual wind speed observations fall within the 95%-level PIs and the width of PIs

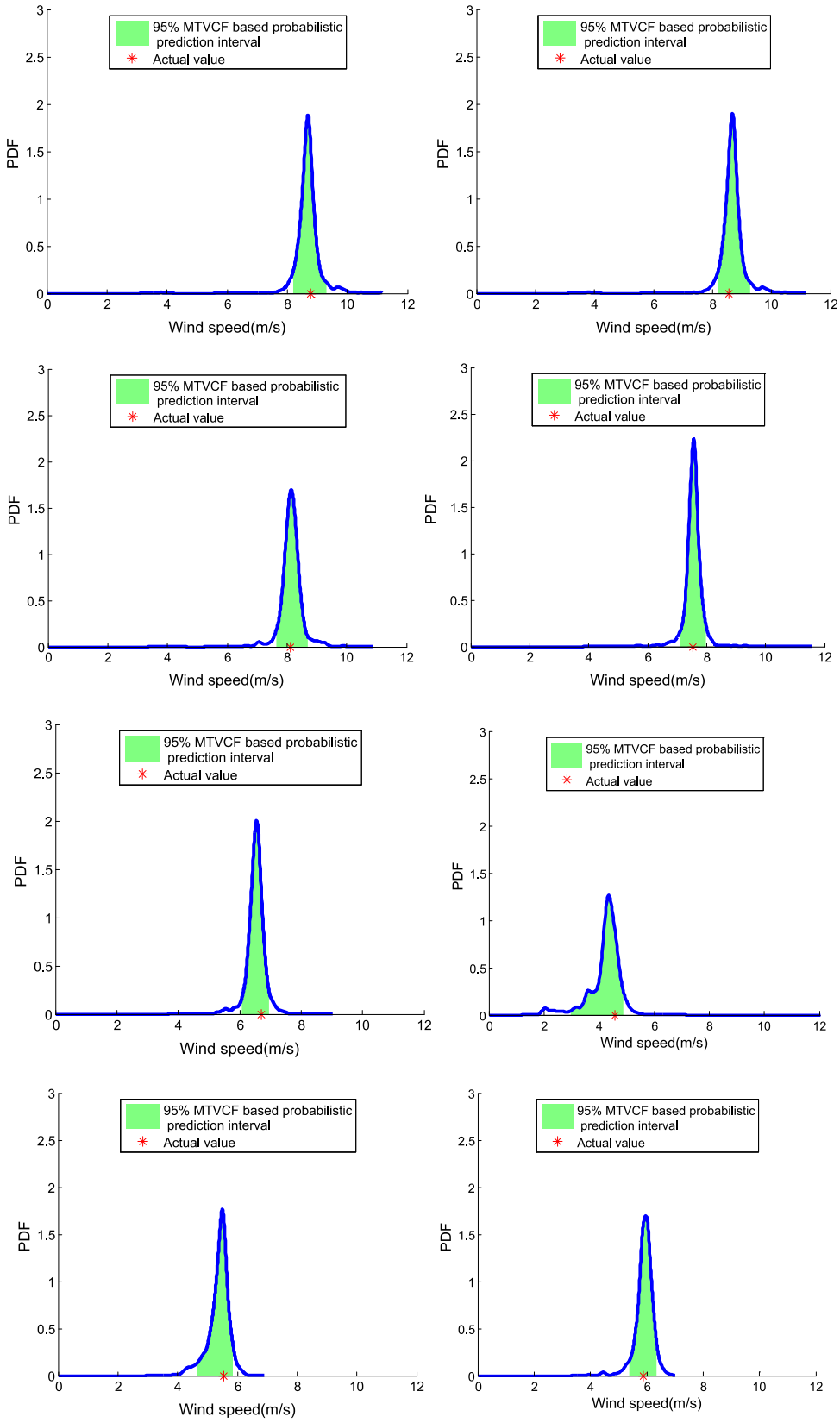

Fig. 12. The predictive PDF of the mean 10-min wind speed for the 1st, 17th, 33rd, 49th, 65th, 81th, 97th and 144th tests of data in January 1.

become more and more border from one-step ahead to five-step ahead. It can also be seen that the proposed method can keep a reliable width of PIs to weaken the influence of wind speed's volatility. This is due to: (1) the point forecast results are obtained using an on-line mode which has a high forecasting accuracy and (2) TVMCF method is data driven which contains more information about the wind speed and enables to capture the fluctuation scale of wind speed to build the probabilistic PIs adaptively. Such inter-

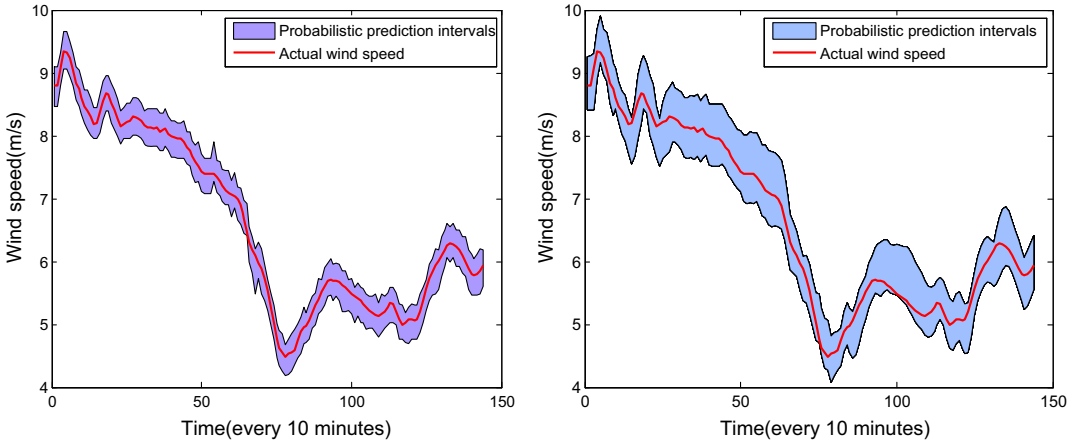

Fig. 13. The 95% probabilistic PIs of the one-step WSF given by TVMCF and GPR, against the observations.

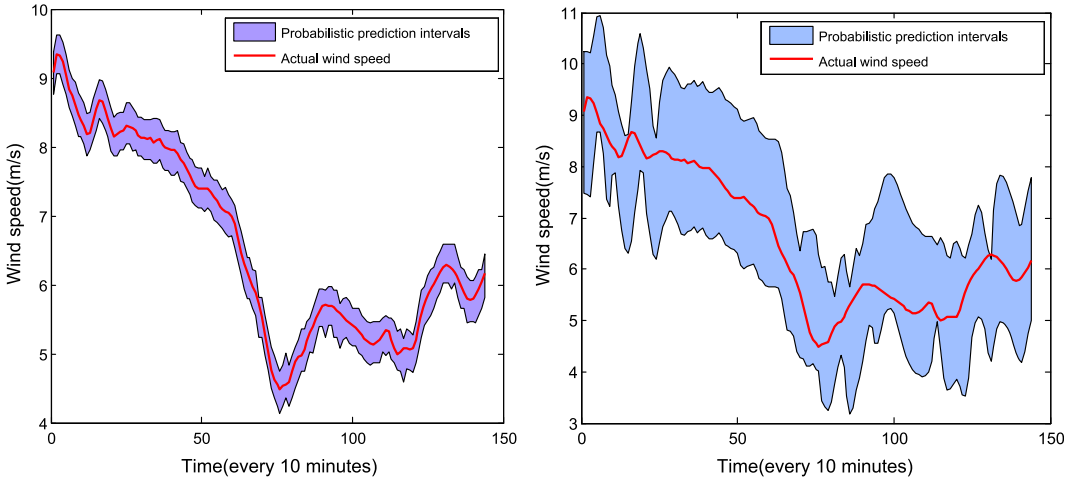

Fig. 14. The 95% probabilistic PIs of the three-step WSF given by TVMCF and GPR, against the observations.

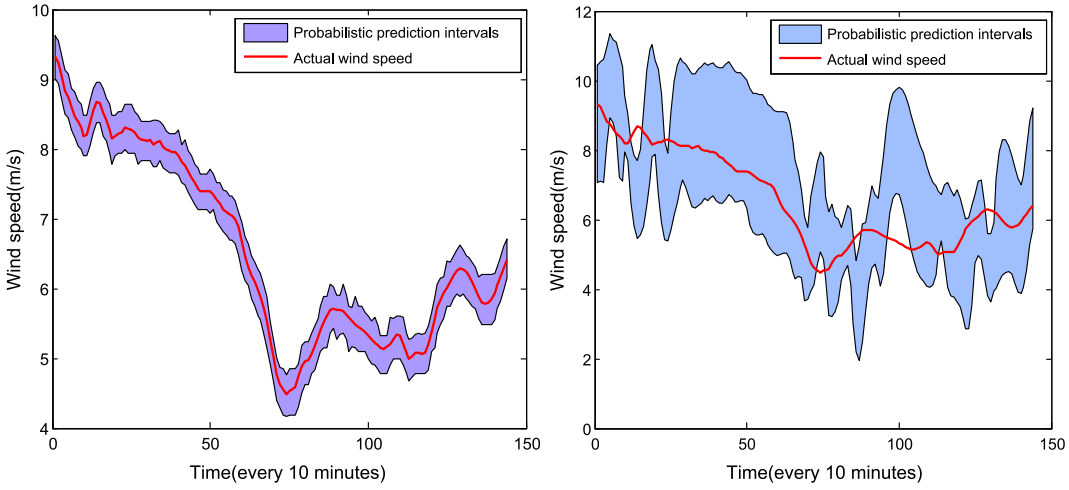

Fig. 15. The 95% probabilistic PIs of the five-step WSF given by TVMCF and GPR, against the observations.

val estimation can assist in the decision-making process for power system scheduling and control. It could also reduce the opportunity costs of overly conservative bidding in the forward market that can result from uncertain availability.

Table 6 summarizes the values of PICP, PINAW and CWC for experiments conducted using 2006 monthly test data sets. Firstly, the maximum length of sub-series cut by BGA in each month is chosen as an example. Then the first 144 sampling points in every

Table 6
Probabilistic PIs evaluation indexes for wind speed in 2006.

Month	One-step			Three-step			Five-step		
	PICP (%)	PINAW (%)	CWC (%)	PICP (%)	PINAW (%)	CWC (%)	PICP (%)	PINAW (%)	CWC (%)
January	100	11.57	11.57	100	12.02	12.02	100	12.80	12.80
February	100	11.57	11.57	100	12.38	12.38	100	13.50	13.50
March	100	6.22	6.22	100	6.47	6.47	100	6.90	6.90
April	100	10.13	10.13	100	10.62	10.62	100	11.06	11.06
May	100	5.24	5.24	100	5.44	5.44	100	6.90	6.90
June	100	5.96	5.96	100	6.25	6.25	99.31	7.08	7.08
July	100	8.83	8.83	100	9.32	9.32	100	9.68	9.68
August	100	7.28	7.28	100	7.58	7.58	100	7.91	7.91
September	100	5.57	5.57	100	6.18	6.18	100	6.30	6.30
October	100	5.05	5.05	100	5.20	5.20	98.61	5.24	5.24
November	100	7.96	7.96	100	8.10	8.10	100	8.52	8.52
December	100	16.75	16.75	100	17.67	17.67	100	18.24	18.24

sub-series are employed to calculate the values of PICP, PINAW and CWC. The reason for our experiment is that the online operation of our proposed wind speed forecast model is available and each sub-series can perform the probabilistic PIs individually. It is important to note that most of PICPs are equal to 1 except the five-step ahead forecasting in June and October and all the PICPs are always greater than the nominal confidence level (i.e. 95%). From a practical point of view, this means that scheduling operation of this wind farm suffers from fewer problems in longer terms and is not easy to prone to catastrophic mistakes. In addition, comparing CWCs in Tables 6 reveals that effects of seasons have a significant impact on the quality of PIs. May and October are the best months from the PINAW and CWC perspective, as the values of PINAW and CWC are smallest. In contrast, January, February and December are the worst months with larger values of PINAW and CWC than other months. These three months correspond to the end of winter and beginning of spring in America, in which the wind speed have high level of uncertainty and volatility.

5. Conclusion

Accurate and reliable technique for short-term wind speed forecasting is vital to power system scheduling and operation due to the greater incorporation of wind energy into power system. Furthermore, high fluctuation of wind speed will have great impact on forecasting precision. Aiming to address these problems, a hybrid forecast model with the combination of BGA, AVMD and λ_{CDFF} OS-ORELM-OEOA optimized by CSO-SAM is proposed to predict the short-term wind speed at one-step, three-step and five-step ahead. Based on the deterministic results, the probabilistic PIs is implemented applying the TVMCF which enables to overcome the deficiencies of the deterministic forecast to dispose uncertainties.

For the deterministic forecast, Offline-ORELM, Offline-ELM, persistence model and the BPNN model are selected as benchmark models. For further comparison, the EEMD technology is employed to establish the EEMD-based hybrid models. The simulation results shown in Tables 2 and 4 indicate that: (1) by utilization of AVMD and EEMD, the hybrid models can promote the forecasting performance of the single models significantly; (2) the AVMD-based proposed hybrid models have evidently superior forecasting capacity compared with the EEMD-based hybrid models; (3) among all the hybrid models, AVMD-λ_{CDFF} OS-ORELM-OEOA has the best forecasting performance in the multi-step predictions; (3) high forecasting accuracy of the proposed hybrid model is at the cost of high time consumption.

To obtain comprehensive prediction information, TVMCF is adopted and developed to construct the probabilistic PIs, where the correlation between the forecast errors and point forecast results are built to formulate the TVMCF. As a result, the conditional distribution at a given point forecast is used to determine the probabilistic PIs. The simulation results have demonstrated that the distribution of wind speed cannot always obey the Gaussian distribution and this is quite different from that in Ref. [27]. The probability information provided by the combined model can benefit the industry in the operation of wind turbines and the integration of wind energy into the power system.

On-line machine learning is a developing tendency to wind speed or wind power forecasting, in the future study, we are preparing to use the ensemble of other state-of-the-art AI approaches, such as SVMs, to perform wind speed or wind power prediction. Another direction is to reduce the computational costs by applying sparsification strategy.

Acknowledgments

This work was supported by the Guangdong Nature Science Foundation Project (GNSF 10151009001000045). The authors would like to thank National Renewable Energy Laboratory for providing the original wind speed data of this research.

References

[1] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34:243–54.
[2] Sweeney CP, Lynch P, Nolan P. Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. Meteorol Appl 2013;20:32–40.
[3] Albadi MH, El-Saadany EF. Overview of wind power intermittency impacts on power systems. Elect Power Syst Res 2010;80(June):627–32.
[4] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2013;62(September):592–7.
[5] Zhang Yao, Wang Jianxue, Wang Xifan. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70.
[6] Ren C, An N, Wang J, Li L, Hu B, Shang D. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl-Based Syst 2014;56:226–39.
[7] Xiao L, Wang JZ, Dong Y, Wu J. Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 2015;44:271–88.
[8] Torres JL, García A, de Blas M, de Francisco A. Forecast of hourly averages wind speed with ARMA models in Navarre. Sol Energy 2005;79:65–77.
[9] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using fARIMA models. Renew Energy 2009;34:1388–93.
[10] Chen L, Lai X. Comparison between ARIMA and ANN models used in short-term wind speed forecasting. In: Power and energy engineering conference (APPEEC), 2011 Asia-Pacific. IEEE; 2011. p. 1–4.
[11] Cadenas E, Rivera W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. Renew Energy 2009;34(1):274–8.
[12] Noorollahi Y, Jokar MA, Kalhor A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. Energy Convers Manage 2016;115:17–25.
[13] Santamaria-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. Renew Energy 2016;85:790–809.

[14] Kong X, Liu X, Shi R, Lee K. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[15] Zhou JY, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manage 2011;52:1990e8.
[16] Wang Y, Wang JZ, Wei X. A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: a case study of wind farms in northwest China. Energy 2015;91:556–72.
[17] Sánchez I. Short-term prediction of wind energy production. Int J Forecast 2006;22:43–56.
[18] Zhang Yachao, Liu Kaipei, Qin Liang, An Xueli. Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods. Energy Convers Manage 2016;112:208–19.
[19] Meng Anbo, Ge Jiafei, Yin Hao, Chen Sizhe. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. Energy Convers Manage 2016;114:75–88.
[20] Liu Da, Niu Dongxiao, Wang Hui, Fan Leilei. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62:592–7.
[21] Wang J, Zhang W, Li Y, Wang J, Dang Z. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl Soft Comput 2014;23:452–9.
[22] Wang S, Zhang N, Wu L, et al. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. Renew Energy 2016;94:629–36.
[23] Liu H, Tian H, Li Y. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Convers Manage 2015;100:16–22.
[24] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, et al. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – extreme learning machine approach. Energy Convers Manage 2014;87:10–8.
[25] Wang MD, Qiu QR, Cui BW. Short-term wind speed forecasting combined time series method and arch model. In: Proceedings of the international conference on machine learning and cybernetics (ICMLC); 2012. p. 924–7.
[26] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerodyn 2008;96:2348–62.
[27] Wang Jianzhou, Hu Jianming. 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.
[28] Okumus I, Dinler A. Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Convers Manage 2016;123:362–71.
[29] Huang GB, Zhu QY, Siew CK. Extreme learning machine: a new learning scheme of feedforward neural networks. In: Proceedings of the international joint conference on neural networks, IEEE; 2004. p. 985–90.
[30] Zhang K, Luo M. Outlier-robust extreme learning machine for regression problems. Neurocomputing 2015;151:1519–27.
[31] Soares SG, Rui A. An adaptive ensemble of on-line extreme learning machines with variable forgetting factor for dynamic system prediction. Neurocomputing 2015;171:693–707.
[32] Sánchez I. Recursive estimation of dynamic models using cook's distance, with application to wind energy forecast. Technometrics 2006;48:61–73.
[33] Bernaola-Galván P, Ivanov PC, Nunes Amaral LA, et al. Scale invariance in the nonstationarity of human heart rate. Phys Rev Lett 2001;87:168105.
[34] Liu Y, Yang G, Li M, et al. Variational mode decomposition denoising combined the detrended fluctuation analysis. IEEE Trans on Signal Process 2016;125:349–64.
[35] Zhang C, Wei H, Zhao X, et al. A Gaussian process regression based hybrid approach for short-term wind speed prediction. Energy Convers Manage 2016;126:1084–92.
[36] Bremnes John B. Probabilistic wind power forecasts using local quantile regression. Wind Energy 2004;7:47–54.
[37] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy 2006;9:95–108.
[38] Bessa Ricardo J, Miranda V, Botterud A, Zhou b Z, Wang J. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renew Energy 2012;40:29–39.
[39] Kou P, Liang D, Gao F, et al. Probabilistic wind power forecasting with online model selection and warped gaussian process. Energy Convers Manage 2014;84:649–63.
[40] Kou P, Gao F, Guan X. Sparse online warped Gaussian process for wind power probabilistic forecasting. Appl Energy 2013;108(8):410–28.
[41] Moon TK. The expectation-maximization algorithm. IEEE Signal Process Mag IEEE Signal Proc Mag 1996;13:47–60.
[42] Xie Q. Computation and application of Copula-based weighted average quantile regression. J Compute Appl Math 2015;281:182–95.
[43] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62:531–44.
[44] Zhang N, Kang C, Xia Q, et al. Modeling conditional forecast error for wind power in generation scheduling. IEEE Transactions on Power Systems 2014;29 (3):1316–24.
[45] National Renewable Energy Laboratory (NREL). URL: <http://wind.nrel.gov/wind>.
[46] Khosravi A, Nahavandi S, Creighton D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts. IEEE Trans Sustain Energy 2013;4:602–10.