Energy 93 (2015) 41–56

Contents lists available at ScienceDirect

Energy

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

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

Jianzhou Wang [b], Jianming Hu [a, *]

[a] School of Mathematics & Statistics, Lanzhou University, Lanzhou 730000, China
[b] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form
17 July 2015
Accepted 11 August 2015
Available online 29 September 2015

Keywords:
Gaussian Process Regression
Wind speed forecasting
Empirical Wavelet Transform
Extreme Learning Machine
Support Vector Machine

ABSTRACT

With the increasing importance of wind power as a component of power systems, the problems induced by the stochastic and intermittent nature of wind speed have compelled system operators and researchers to search for more reliable techniques to forecast wind speed. This paper proposes a combination model for probabilistic short-term wind speed forecasting. In this proposed hybrid approach, EWT (Empirical Wavelet Transform) is employed to extract meaningful information from a wind speed series by designing an appropriate wavelet filter bank. The GPR (Gaussian Process Regression) model is utilized to combine independent forecasts generated by various forecasting engines (ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM)) in a nonlinear way rather than the commonly used linear way. The proposed approach provides more probabilistic information for wind speed predictions besides improving the forecasting accuracy for single-value predictions. The effectiveness of the proposed approach is demonstrated with wind speed data from two wind farms in China. The results indicate that the individual forecasting engines do not consistently forecast short-term wind speed for the two sites, and the proposed combination method can generate a more reliable and accurate forecast.

Definition

Multistep ahead forecast: suppose that we are at the time index *h*and are interested in forecasting $\hat{r}_{h+l}$, where $l \geq 1$. The time index h is called the forecast origin and the positive integer l is the forecast horizon. Let $\hat{r}_h(l)$ be the forecast of r_{h+l}, we refer to $\hat{r}_h(l)$ as the l-step ahead forecast of r_t at the forecast origin h. When $l = 1$, we refer to $\hat{r}_h(1)$ as the **one-step ahead forecast** of r_t at the forecast origin h.

Hyberparameters: In Bayesian statistics, a hyperparameter is a parameter of a prior distribution; the term is used to distinguish them from parameters of the model for the underlying system under analysis.

1. Introduction

Due to increasing energy demand and environmental concerns, wind power, as an environmentally friendly source of renewable energy, has attracted global attention. The rapid development of wind energy gives wind power the potential to support sustainable economic development and environmental protection [1]. However, the stochastic and intermittent nature of wind power complicates the large-scale penetration of wind power into the grid system because this could decrease system reliability and power quality. Wind speed forecasts can effectively reduce the risk to the power system induced by wind-related uncertainties.

Many methods and models have been proposed in recent decades to obtain accurate wind speed predictions, including physical approaches and statistical models. Physical approaches are generally used for long-term wind forecasts and utilize weather data, while statistical models are more commonly used for short-term

* Corresponding author. Tel.: +86 189 93173569; fax: +86 931 8912481.
E-mail address: hujm11@lzu.edu.cn (J. Hu).

http://dx.doi.org/10.1016/j.energy.2015.08.045
0360-5442/

Symbols and notation

EWT

V_n	the nth segment in [0,π]
ω_n	the nth maxima of Fourier spectrum
∀	randomly assign
$\widehat{\phi}_n(\omega)$	the empirical scaling function
$\widehat{\varphi}_n(\omega)$	the empirical wavelets
γ	the real number belong to [0,1]
$\langle\rangle$	inner product
$f(t)$	the signal
$\overline{\varphi_n(\tau-t)}$	the conjugate of the empirical wavelets
$w_f^{\varepsilon}(n,t)$	The detail coefficients

ARIMA model

$\varphi(B)$	the AR (autoregressive) model
$\theta(B)$	the moving average function
q	the lag order of moving average terms
p	the lag order of autoregressive terms
d	lag order of non-seasonal differences

LSSVM model

W	the coefficient vector
$\Phi(x)$	the mapping function
γ	regulation constant
e_i	error variable
α_i	the Lagrange multipliers
$K(x_i,x)$	kernel function
b	bias term

SVM model

w	the coefficient vector
$\phi(x)$	the mapping function
ξ_i	the slack variable (feasible case)
ξ_i^*	slack variable (non-feasible case)
α	the Lagrange multipliers for ξ_i
α^*	the Lagrange multipliers for ξ_i^*
$K(x,y)$	kernel function
ε	the approximation precision
C	regulation constant
b	bias term

ELM model

β	the coefficient
$h(x)$	the mapping function
t_i	the indicator vector
H	the mapping function matrix
C	constant
ξ_i	training error vector

GPR model

$p(\bullet)$	the probability function
ϕ	parameters of observation model
$m(x)$	the mean function
f	the latent variables
θ	parameters of covariance function
$k(x,x'\|\theta)$	the covariance function

prediction and are primarily performed through analysis of historical data. NWP (Numerical Weather Prediction) models are representative of the physical methods most often presented in the literature [2,3]. Statistical models mainly include time series techniques (ARMA (Auto-Regressive Moving Average) [4], ARIMA (Autoregressive Integrated Moving Average) [5], FARIMA (fractional ARIMA) [6], exponential smoothing techniques [7] and gray predictors [8]) and statistical learning methods (ANN (Artificial Neural Networks) [9–14], fuzzy logic [15,16], and SVM (Support Vector Machine) [17–19]).

Recent studies have predominantly focused on short-term wind forecasting ranging from minutes to hours because of the importance of these forecasts for power systems. Various attempts have been made to use hybrid methods for short-term wind forecasting. The combined approaches most commonly seen in the literature are data preprocessing-based approaches, parameter-optimization-based approaches and weighting-based approaches.

When employing data preprocessing-based approaches, techniques such as WT (Wavelet Transform) and EMD (Empirical Mode Decomposition) are used as data preprocessors to decompose wind series or eliminate stochastic volatility. The WT method is widely utilized for wind speed series data processing by researchers [17,20–22] due to its adaptive ability of time–frequency analysis. For example, Liu et al. [17] proposed a hybrid model consisting of WT, GA (Genetic Algorithm) and SVM in which WT decomposed the wind speed series into two components for wind speed forecasting. The EMD method, introduced by Huang et al. [23] for the decomposition of non-stationary signals into a series of IMFs (Intrinsic Mode Functions), has also been recently utilized for the purpose of wind data processing [9,19,24–26]. For example, Hu et al. [19] suggested a forecasting approach associated with ensemble EMD and the SVM, with Ensemble EMD used to decompose the wind speed series for the SVM to enhance wind speed prediction precision. Among the parameter-optimization-based approaches, many studies have concentrated on stochastic heuristic optimization algorithms that have the capability of fast convergence to global optima and relatively simple implementation. These stochastic heuristic optimization algorithms include the GA [27,28], DE (Differential Evolution) algorithm [29], PSO (Particle Swarm Optimization) algorithm [30], evolutionary programming [31], MTS (Memory Tabu Search) algorithm [32], and ICA (Imperialist Competitive Algorithm) [33]. The weighting-based approaches presented in the literature [4,33–36] combine several independent forecasting techniques through a weight that determines the relative effectiveness of each model. Most of the independent forecasting techniques involved include time series techniques and statistical learning methods.

The aforementioned models are useful for point or deterministic wind speed or power forecasts. These combined models are mostly parametric models that offer point estimation of wind speed given appropriate inputs. However, due to the stochastic and variable nature of the wind, the accuracy of such forecasts cannot be guaranteed and tends to be fairly low. In addition, the uncertainty in wind energy gives rise to economic risks for wind farm owners, especially in competitive electricity markets [4,5]. In such circumstances, probabilistic forecasting of wind power becomes highly meaningful for both utilities and system operators who manage wind farms. Probabilistic forecasting can effectively reflect the uncertainties associated with prediction results, and help to assess the risk of relying on a forecast [6]. For these reasons, some researchers have examined the interval wind speed or power prediction with respect to uncertainty

information [37–42]. For example, Jiang et al. [37] proposed a forecasting model based on Bayesian theory and structural break modeling to forecast very short-term wind speed and its intervals. Pinson and Kariniotakis [40] categorized forecast situations into several scenarios, and predicted intervals by fuzzy inference and bootstrap sampling from empirical distributions of the forecasting residuals of the each scenario prediction. Kou et al. [41] made use of online model selection and the warped Gaussian process to forecast probabilistic short-term wind power in consideration of the non-Gaussian uncertainties in wind power series.

This paper is organized as follows. Section 2 discusses the drawbacks of existing models and the contributions of the model proposed here, and Section 3 introduces the individual models required and describes the hybrid model that was developed. In Section 4, the wind speed predictions and advantages of this strategy are analyzed and discussed in comparison with other benchmark models. Finally, conclusions from this study are presented in Section 5.

2. Innovations of the proposed methodology

The hybrid approaches detailed above can generate somewhat more accurate and reliable wind speed predictions. These combined parametric models provide point estimation of wind speed for a single wind farm or at the regional level, but they cannot provide information beyond the single-valued predictions. By contrast, probabilistic forecasting can provide more information. Probabilistic forecasting can effectively reflect the uncertainties associated with the prediction results, and help to assess the risk of relying on the forecast [6]. Relatively few existing studies have used probabilistic forecasting, and there are even fewer studies that use deterministic forecasting transformed into probabilistic forecasting.

Data preprocessing-based approaches commonly use the WT or EMD techniques to decompose short-term wind speed series or eliminate the stochastic volatility. However, WT lacks the ability of self-adaptive data processing and requires the specification of wavelet basis and parameters beforehand, while the EMD is sensitive to noise and sampling and lacks a basis in mathematical theory. The EWT (Empirical Wavelet Transform) method remedies the drawbacks of these decomposition methodologies to some extent. It can adaptively represent the processed signal and then decomposes the signal into a finite number of modes.

As for weighting-based approaches, the independent predictions of the forecasting models are directly connected by weight. Few studies have examined the combination of several independent forecasting techniques in a nonlinear way, which impedes proper usage of these models and causes them to suffer from limited expressiveness. A simple idea to overcome this problem is to first project the inputs into some high dimensional space using a set of basis functions and then apply the linear model in this space instead of directly on the inputs themselves.

Based on the shortcomings of current approaches, this paper outlines a combination forecasting approach for more precisely appraising wind speed. In this proposed hybrid approach, the EWT is employed to preprocess wind speed series by designing an appropriate wavelet filter bank. Next, the GPR (Gaussian Process Regression) model is utilized to combine the ARIMA, ELM (Extreme Learning Machine), SVM and LSSVM (Least Square SVM) models in a nonlinear way. The main novelties of this paper with respect to previous literature can be summarized as follows:

1) The model is proposed to fill a gap in existing weighting-based approaches and take advantage of hybrid algorithms that can enhance prediction precision. In this proposed model, forecasts made by the ELM, LSSVM and SVM models are mapped into feature space, and then deterministic forecasting is transformed into probabilistic forecasting by the GPR model. The proposed approach provides a method for tackling model selection, which has been challenging in wind speed forecasting.
2) The EWT, which is a novel signal processing tool, is adopted to eliminate various characteristics from the original time series. The EWT extracts only the meaningful components from the wind speed series.
3) The GPR is employed to combine the individual forecasting models. The Gaussian Process is a principled, practical, probabilistic approach, which is advantageous in the interpretation of model predictions. It possesses very good adaptability and is suitable for dealing with high dimensional spaces and small samples, as well as nonlinear complex problems. GPR is easy to implement, self-adaptive in terms of parameters, and flexible in making nonparametric inferences.
4) The GPR produces a predictive sample of wind speeds for the final probabilistic wind speed forecast. The predictive sample provides more information than a classical point forecast, including credible intervals and quantiles.
5) Wind speed has time-varying characteristics that should be exhibited by a wind speed forecasting model. This study employs the moving window method in prediction processes, which makes the proposed hybrid model adaptively respond to wind speed changes and thus better reflect the actual forecasting environment.

3. Methodology

This section first concisely introduces the individual models, including the EWT algorithm, ARIMA, ELM, LSSVM SVM and GPR models. Next, the operating process of the proposed hybrid approach is described.

3.1. EWT

The EWT, proposed by Jérôme Gilles [43], is used to identify and extract different intrinsic modes of a time series. This wavelet transform introduces a novel approach to build a family of empirical wavelets that can adaptively represent the processed signal, and then decomposes the signal into a finite number of modes. The central idea of this approach is to detect the Fourier supports from the information contained in the processed signal spectrum.

The empirical wavelets can be defined as bandpass filters on each V_n, where $V_n = [\omega_{n-1}, \omega_n]$ denotes each segment of the Fourier spectrum of the signal, and $\cup_{n=1}^{N} V_n = [0, \pi]$. $\forall n > 0$, the empirical scaling function and the empirical wavelets are defined by Eqs. (1) and (2), respectively.

$$\widehat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \le \omega_n - \tau_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}|\omega| - \omega_n + \tau_n\right)\right] & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and

$$\widehat{\varphi}_n(\omega)=\begin{cases}1 & if\ \omega_n+\tau_n\le|\omega|\le\omega_{n+1}-\tau_{n+1}\\ cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}|\omega|-\omega_{n+1}+\tau_{n+1}\right)\right] & if\ \omega_{n+1}-\tau_{n+1}\le|\omega|\le\omega_{n+1}+\tau_{n+1}\\ sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}|\omega|-\omega_n+\tau_n\right)\right] & if\ \omega_n-\tau_n\le|\omega|\le\omega_n+\tau_n\\ 0 & otherwise\end{cases}\tag{2}$$

The function $\beta(x)$ is an arbitrary $C^k([0,1])$ function and satisfies the following properties: $\beta(x)=\begin{cases}0\ if\ x\le 0\\ 1\ if\ x\ge 0\end{cases}$ and $\beta(x)+\beta(1-x)=1, \forall x\in[0,1]$.

The widely used function $\beta(x)$ is:

$$\beta(x)=x^4\left(35-84x+70x^2-20x^3\right)\tag{3}$$

Concerning the choice of τ_n, there are several possible options, and the simplest of which is to set τ_n proportional to ω_n : $\tau_n=\gamma\omega_n$ where $0<\gamma<1$. Consequently, when $\forall_n>0$, Eqs. (1) and (2) simplify to:

$$\widehat{\phi}_n(\omega)=\begin{cases}1 & if|\omega|\le(1-\gamma)\omega_n\\ cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}|\omega|-\omega_n+\tau_n\right)\right] & if(1-\gamma)\omega_n\le|\omega|\le(1-\gamma)\omega_n\\ 0 & otherwise\end{cases}\tag{4}$$

and

$$\widehat{\varphi}_n(\omega)=\begin{cases}1 & if(1+\gamma)\omega_n\le|\omega|\le(1+\gamma)\omega_{n+1}\\ cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}|\omega|-\omega_{n+1}+\tau_{n+1}\right)\right] & if(1-\gamma)\omega_{n+1}\le|\omega|\le(1-\gamma)\omega_{n+1}\\ sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}|\omega|-\omega_n+\tau_n\right)\right] & if(1-\gamma)\omega_n\le|\omega|\le(1-\gamma)\omega_n\\ 0 & otherwise\end{cases}\tag{5}$$

when

$$\gamma<min_n\left(\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n}\right),\tag{6}$$

the set $\{\phi_1(t),\{\varphi_n(t)\}_{n=1}^N\}$ is a tight frame of $L^2(R)$.

Because the appropriate frame set of empirical wavelets is built, the EWT can be implemented in the same way as the classic wavelet transform. The detail coefficients are obtained from the following inner products between the signal and the empirical wavelets:

$$\begin{aligned}w_f^\varepsilon(n,t)&=\langle f,\varphi_n\rangle=\int f(\tau)\overline{\varphi_n(\tau-t)}d\tau\\&=f(\tau)\left(\overline{\widehat{\varphi}_n(\tau-t)}\right)^\vee\end{aligned}\tag{7}$$

and the approximation coefficients are calculated by the following inner products with the scaling function:

$$\begin{aligned}w_f^\varepsilon(0,t)&=\langle f,\phi_1\rangle=\int f(\tau)\overline{\phi_1(\tau-t)}d\tau\\&=f(\tau)\left(\overline{\widehat{\phi}_1(\tau-t)}\right)^\vee\end{aligned}\tag{8}$$

The reconstruction is obtained by:

$$\begin{aligned}f(t)&=w_f^\varepsilon(0,t)*\phi_1(t)+\sum_{n=1}^N w_f^\varepsilon(n,t)*\varphi_n(t)\\&=\left(\widehat{w_f^\varepsilon}(0,\omega)*\phi_1(\omega)+\sum_{n=1}^N\widehat{w_f^\varepsilon}(n,\omega)*\varphi_n(\omega)\right)^\vee\end{aligned}\tag{9}$$

Following this formalism, the empirical mode is given by:

$$f_0(t)=w_f^\varepsilon(0,t)*\varphi_1(t)\tag{10}$$

$$f_k(t)=w_f^\varepsilon(k,t)*\varphi_k(t)\tag{11}$$

In decomposing the signal, it is important to designate the appropriate number of modes when no a priori information is available. In theory, the most important maxima in the magnitude of the Fourier transform of the input signal should be significantly larger than the other maxima. Let $\{M_i\}_{i=1}^{M}$ denotes the set of M detected maxima in the magnitude of the Fourier spectrum. This set is sorted by decreasing values ($M_1 \geq M_2 \cdots M_M$) and normalized in $[0, 1]$. The above idea results in keeping all maxima that exceed the threshold $M_M + \alpha(M_1 - M_M)$, where α corresponds to the relative amplitude ratio. Using α, the appropriate number of modes can be confirmed.

3.2. ARIMA

The ARIMA model is briefly introduced as follows. Given a time series $x_t, t = 1, \cdots, n$, an ARIMA model can be defined as:

$$\varphi(B)(1-B)^d x_t = \theta(B)a_t, \tag{12}$$

$$\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p, \tag{13}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q, \tag{14}$$

$\varphi(B)$ is the AR (autoregressive) model of order p and $\theta(B)$ is the moving average function of order q. The parameters p, d, q are integers. The parameters d, p and q denote the lag order of the non-seasonal differences, autoregressive terms and moving average terms, respectively. Details of Eqs. (12)–(14) are outlined in previous literature [44].

3.3. SVM

The SVM, proposed by Vapnik and coauthors, is based on statistical learning theory and the structural risk minimization principle. The basic principle of SVM for regression is to map the data into a high dimensional feature space via nonlinear mapping, after which a linear regression is performed in this feature space. The regression formula can be expressed as:

$$f(x) = \sum_{i=1}^{n} w_i \phi_i(x) + b \tag{15}$$

The coefficients $\{w_i\}_{i=1}^{n}$ can be obtained from the data by optimizing the following quadratic programming problem:

$$\underset{w,b,\xi,\xi^*}{minimize}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$s.t.\begin{cases} y_i - \langle w, \phi(x_i)\rangle - b \leq \varepsilon + \xi_i \\ \langle w, \phi(x_i)\rangle - y_i + b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \cdots n \end{cases} \tag{16}$$

By solving the optimization problem based on the conditions (KKTKarush–Kuhn–Tucker conditions) (see the Appendix), the estimation function can be obtained as follows:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle \phi_i(x)\phi(x)\rangle + b \tag{17}$$

In Eq. (15), $\{\phi_i(x)\}_{i=1}^{n}$ are named features and b is the bias term. In formula (16), ξ_i^* is a slack variable and $C > 0$ is a constant that determines penalties. In Eq. (17), $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C$.

SVM is characterized by the use of a technique known as the "kernel trick", which applies linear classification techniques to nonlinear classification problems. The kernel function represents the inner product in the D-dimensional feature space:

$$K(x, y) = \sum_{i=1}^{D} \phi_i(x)\phi_i(y) = \langle \phi(x)\phi(y)\rangle \tag{18}$$

It is not necessary to compute the feature ϕ_i; rather, what is needed is a simple kernel function, which has a known analytical form. The only necessary condition is that the kernel function has to meet Mercer's theorem (see the Appendix). Substituting the result of Eq. (18) into Eq. (17) leads to the following nonlinear regression function:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{D}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{19}$$

Eqs. (15)–(19) and more details on SVM can be found in the literature [45].

3.4. LSSVM

A modified version of SVM called LSSVM only considers equality constraints instead of inequalities. A brief review of the LSSVM algorithm for regression problems follows. Given a training set $\{x, y\}, i = 1, 2 \cdots N$, the regression formula can be constructed as follows:

$$y = W^T \Phi(x) + b \tag{20}$$

The weight vector W of the regression can be calculated by optimizing the following cost function containing a penalized regression error:

$$C = \frac{1}{2}W^T W + \frac{1}{2}\gamma\sum_{i=1}^{N} e_i^2 \tag{21}$$

subject to:

$$W^T \Phi(x) + b = 1 - e_i \tag{22}$$

Based on these equations, the Lagrange function is constructed as follows:

$$L(W, b, e, \alpha) = \frac{1}{2}\|W\|^2 + \gamma\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N}\alpha_i\left(W^T\Phi(x) + b - y_i + e_i\right) \tag{23}$$

The solution to Eq. (23) can be obtained by partially differentiating with respect to W, b, α_i and e_i. This yields the following result:

$$y = \sum_{i=1}^{N}\alpha_i\Phi(x_i)\Phi(x) + b = \sum_{i=1}^{N}\alpha_i\langle\Phi(x_i)\Phi(x)\rangle + b \tag{24}$$

In Eq. (20), W denotes the weight vector and b is the bias term. $\Phi(x)$ is the nonlinear mapping function that transfers the input to a higher-dimensional feature space. In Eq. (21) e_i is the error variable at time i and γ is a regulation constant. In Eq. (23), α_i are Lagrange multipliers.

Substituting the solution of Eq. (18) into Eq. (24) leads to the following nonlinear regression function:

$$y = \sum_{i=1}^{N} a_i K(x_i, x) + b \tag{25}$$

More details of LSSVM, including Eqs. (20)–(25) can be found in the literature [45].

3.5. ELM

The ELM, developed by Huang et al. [46], is a type of SLFN (single hidden-layer, feed-forward neural network) in which the hidden layer parameters do not need to be tuned.

When the m-output nodes are considered, the output formula for the ELM can be expressed as:

$$f(x) = h(x)\beta \tag{26}$$

The coefficient β can be obtained by optimizing the following quadratic programming problem:

$$\begin{aligned} &minimize \frac{1}{2}\|\beta\|^2 + \frac{1}{2}C\sum_{i=1}^{N}\|\xi_i\|^2 \\ &s.t.\ h(x_i)\beta = t_i^T - \xi_i^T, i = 1, \cdots, N \end{aligned} \tag{27}$$

According to the KKT theorem, the quadratic programming problem is equivalent to optimizing the following dual problem:

$$L_{ELM} = \frac{1}{2}\|\beta\|^2 + \frac{1}{2}C\sum_{i=1}^{N}\|\xi\|^2 - \sum_{i=1}^{n}\sum_{i=1}^{m}a_{i,j}(h(x_i)\beta_j - t_{i,j} + \xi_{i,j}) \tag{29}$$

After differentiating with respect to β, ξ and α, the output formula of the ELM is the following expression:

$$f(x) = h(x)H^T\left(\frac{I}{C} + HH^T\right)^{-1}T \text{ (when the amount of training samples is not large)} \tag{30}$$

Alternatively, the output expression of the ELM is:

$$f(x) = h(x)\left(\frac{I}{C} + HH^T\right)^{-1}H^TT \text{ (for a large amount of training samples)} \tag{31}$$

In Eq. (26), $f(x) = (f_1(x)\cdots f_m(x))$ and $h(x) = (h_1(x)\cdots h_m(x))$. The function $h(x)$ maps the n-dimensional input set into the L-dimensional hidden layer feature space. The coefficient β denotes the matrix $\beta = [\beta_1, \cdots \beta_m]$, where $\beta_j, j = 1\cdots m$ is the weight vector connecting the hidden layer with the jth output node. In Eq. (27), $C > 0$, $\xi_i = [\xi_{i,1}, \cdots \xi_{i,m}]^T$ is the training error vector of the mth output node, $t_i = [t_{i1}, t_{i2}, \cdots t_{im}]^T$ and $t_{ij} = \begin{cases} 1 & j = i \\ 0 & otherwise \end{cases}$. In Eqs. (30) and (31),

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \quad \text{and}$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \ddots & \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix}_{N\times L}$$

In this study, we take one output node case into account. Its output formula can be expressed as:

$$f(x) = h(x)\beta = \sum_{i=1}^{L}\beta_i h_i(x) \tag{32}$$

where $\beta = [\beta_1, \ldots, \beta_L]^T$ is the weight vector between the hidden layer and the output layer.

3.6. GPR model

GP (Gaussian processes) are powerful tools for probabilistic modeling purposes [47]. The models can be summarized as follows:

Observation model:

$$y|f, \phi \sim \prod_{i=1}^{n} p(y_i|f_i, \phi) \tag{33}$$

GP prior:

$$f(x)|\theta \sim gp(m(x), k(x, x'|\theta)) \tag{34}$$

Hyperprior:

$$\theta, \phi \sim p(\theta)p(\phi) \tag{35}$$

where $m(x)$ and $k(x, x'|\theta)$ denote the mean and covariance functions, and θ and ϕ are the parameters of the covariance function and the observation model, respectively. The covariance function is the crucial ingredient in GP predictors as it encodes the prior assumptions on the latent function, such as smoothness and scale of the variation. A function of input pairs is a valid covariance function as long as the covariance matrices that it produces are symmetric and positive semi-definite.

Assuming predicted values $\tilde{f}$ with new input $\tilde{x}$ to the latent function, the joint prior for latent variables f and $\tilde{f}$ is:

$$\begin{bmatrix} f \\ \tilde{f} \end{bmatrix} \Big| x, \tilde{x}, \theta \sim N\left(0, \begin{bmatrix} K_{f,f} & K_{f,\tilde{f}} \\ K_{\tilde{f},f} & K_{\tilde{f},\tilde{f}} \end{bmatrix}\right) \tag{36}$$

By defining the marginal distribution of $\tilde{f}$ as $p(\tilde{f}|\tilde{x}, \theta) = N(\tilde{f}|0, K_{\tilde{f},\tilde{f}})$, the conditional distribution of $\tilde{f}$ given f is:

$$\tilde{f}\Big|f, x, \tilde{x}, \theta \sim N\left(K_{\tilde{f},f}K_{f,f}^{-1}f, K_{\tilde{f},\tilde{f}} - k_{\tilde{f},f}K_{f,f}^{-1}K_{f,\tilde{f}}\right) \tag{37}$$

In Eqs. (36) and (37), $K_{f,\tilde{f}} = k(x, \tilde{x}|\theta)$ and $K_{\tilde{f},\tilde{f}} = k(\tilde{x}, \tilde{x}|\theta)$.

3.7. The proposed hybrid model

These forecasting engines (SVM, ELM and LSSVM) have good generalization performance because their algorithms can efficiently handle nonlinear problems. The ARIMA model has good prediction ability when the time series are linear. However, when these individual approaches are applied to short-term wind speed forecasting, they usually do not have good prediction performance owing to the complexity (randomness, intermittence, etc.) of the wind speed series. Hence, this study proposes a hybrid approach to tackle the forecasting task. Fig. 1 shows the general structure of the proposed hybrid wind speed forecasting method.

The hybrid approach includes three stages, and the general task in each stage is described as follows:

Stage *I*: Utilize the EWT algorithm to preprocess the original short-term wind speed series. In this stage, the wind speed series is first extended by mirroring, after which the extended wind speed series is computed by Fourier transform. Next, by segmenting of the Fourier spectrum based on the detected maxima, the empirical wavelet filter bank is established when Eq. (6) is satisfied. Finally, the signal is filtered to extract each subseries from the filter bank. The residual series are discarded because the residual is small and can be regarded as an uncorrelated white noise series, and the decomposed modes are aggregated into the new data series. This process de-noises the original data to improve prediction accuracy.

Stage *II*: Perform wind speed forecasting in different forecasting horizons. Prior to building the forecasting models for wind speed prediction, the PACF (Partial Autocorrelation Function) [19],

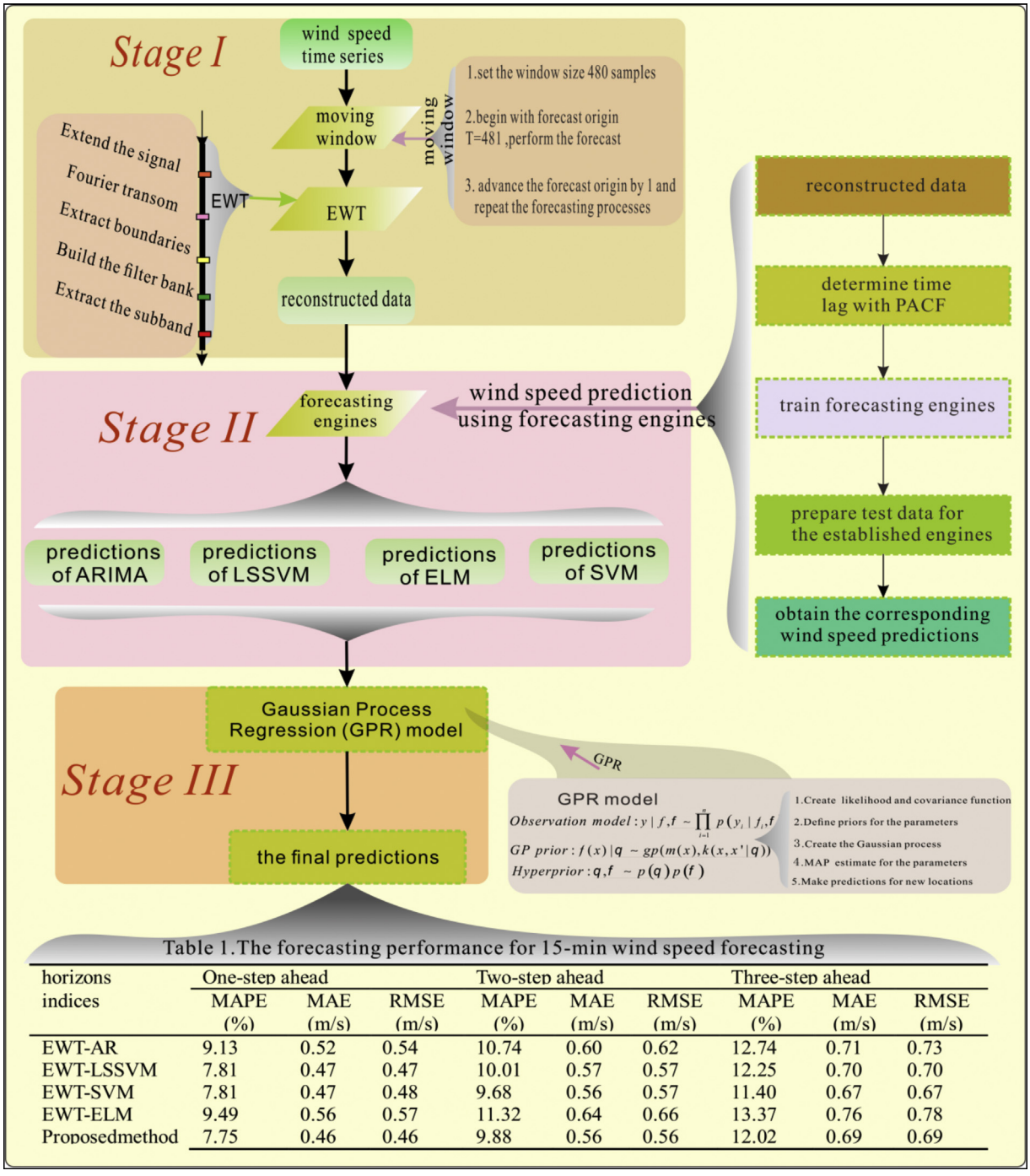

Table 1.The forecasting performance for 15-min wind speed forecasting

horizons	One-step ahead			Two-step ahead			Three-step ahead		
indices	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)
EWT-AR	9.13	0.52	0.54	10.74	0.60	0.62	12.74	0.71	0.73
EWT-LSSVM	7.81	0.47	0.47	10.01	0.57	0.57	12.25	0.70	0.70
EWT-SVM	7.81	0.47	0.48	9.68	0.56	0.57	11.40	0.67	0.67
EWT-ELM	9.49	0.56	0.57	11.32	0.64	0.66	13.37	0.76	0.78
Proposedmethod	7.75	0.46	0.46	9.88	0.56	0.56	12.02	0.69	0.69

Fig. 1. The overall framework of the hybrid EWT-GPR model.

a widely used lag identification method in AR (p) models, is employed to confirm and identify the inputs of the forecasting models. When the sample PACF values at lags greater than p approximate independent $N(0, 1/n)$ random variables, the lag order can be determined as p. After establishment of the forecasting engines, the well-trained forecasting engines are utilized to predict the wind speed in different forecasting horizons.

Stage *III*: Combine the independent forecasts generated by the aforementioned forecasting engines.

The fitting values derived from the forecasting engines are used to construct the Gaussian likelihood of the GPR model. With gradient-based optimization methods to the Gaussian likelihood, the maximum a posteriori estimate of the parameters can be obtained. Finally, the established GPR is utilized to predict the distribution of future wind speed with the independent forecasts generated by the forecasting engines as input in different forecasting horizons. Henceforth, the predicted mean wind speed can be obtained as well as the estimated interval of wind speed.

4. Case study

In this section, a case study is presented to demonstrate the effectiveness of the proposed hybrid approach through comparisons with other models. This case study includes four subsections:

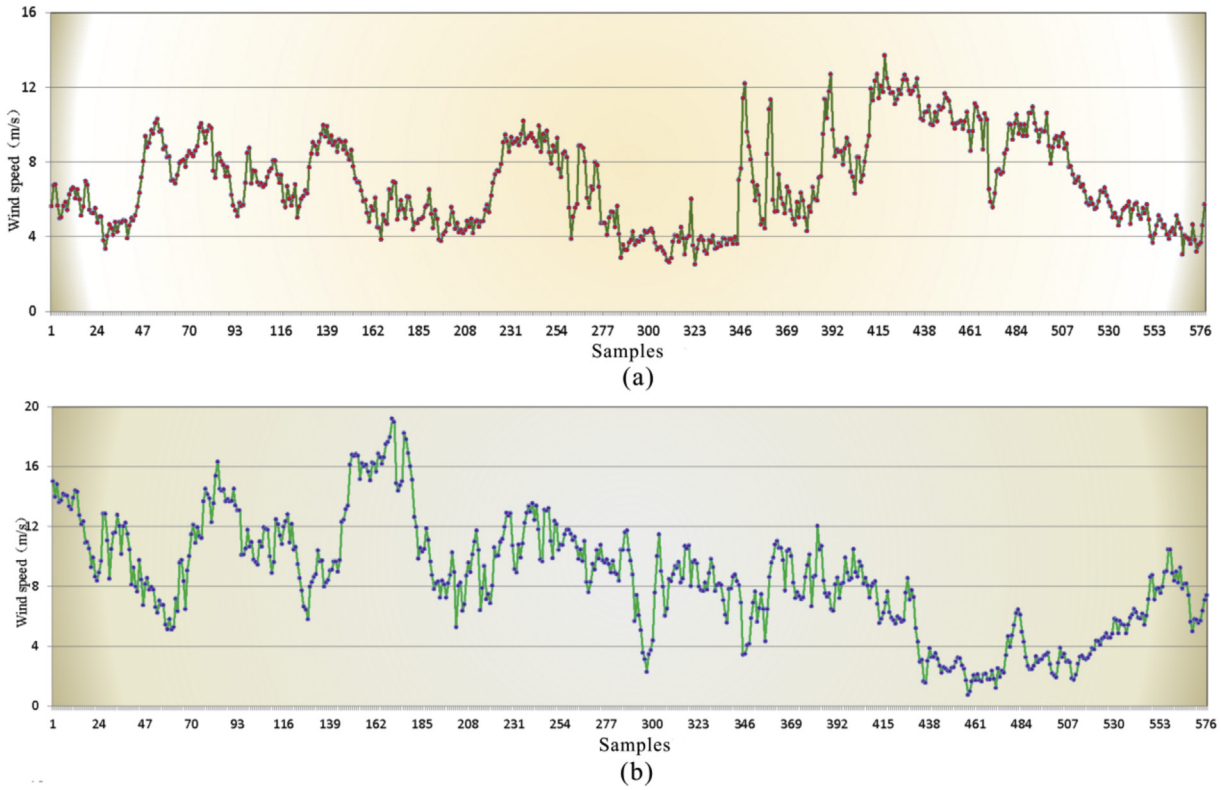

Fig. 2. Two wind speed datasets. Subplot (a) shows the 576 mean 15-min wind speed observations from site 1, and Subplot (b) shows the 576 mean 30-min wind speed observations from site 2.

collection of data, evaluation criteria for forecasting performance, simulation and comparison and discussion.

4.1. Data collection

Wind speed data from two wind farms in China are used to demonstrate the effectiveness and reliability of the proposed hybrid forecasting approach. These data include 576 mean 15-min wind speed observations from site 1 and 576 mean 30-min wind speed observations from site 2 (shown in Fig. 2). In our simulation, the moving window approach is adopted to respond to wind speed changes and limit the size of the training dataset. Here, the size of the moving window is set as 480 samples. The samples in the moving window are used to build or train the models. As the window moves forward, a corresponding model is established for wind speed forecasting. The process of modeling repeats 96 times with the movement of the window.

4.2. Evaluation indices for forecasting performance

To evaluate the proposed hybrid approach, three statistical indices are utilized to measure the forecasting accuracy. These indices are the MAE (mean absolute error), RMSE (root mean square error) and MAPE (mean absolute percent error), for which small values indicate high forecast performance. These indices are defined as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|p_t^{true} - p_t^{forecast}\right| \tag{38}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{p_t^{true} - p_t^{forecast}}{p_t^{true}}\right| \tag{39}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(p_t^{true} - p_t^{forecast}\right)^2} \tag{40}$$

where p_t^{true} is the observed value for the time period t and $p_t^{forecast}$ is the predicted value for the corresponding period. The MAE reveals how similar the predicted values are to the observed values, whereas the RMSE measures the overall deviation between the predicted and observed values. The MAPE is a unit-free measure of accuracy for the predicted wind series and is sensitive to small changes in the data.

Table 1
The number of detected modes for each signal according to different values of α.

α	Number of modes	
	Mean 15-min wind speed series	Mean 30-min wind speed series
0.1	31	30
0.2	16	14
0.3	6	7
0.4	3	4
0.5	3	3

4.3. Simulation

In this simulation, the wind series are first submitted to the individual forecasting models to obtain individual predicted outcomes, and then the corresponding predictions are submitted to the GPR model to obtain the final forecast.

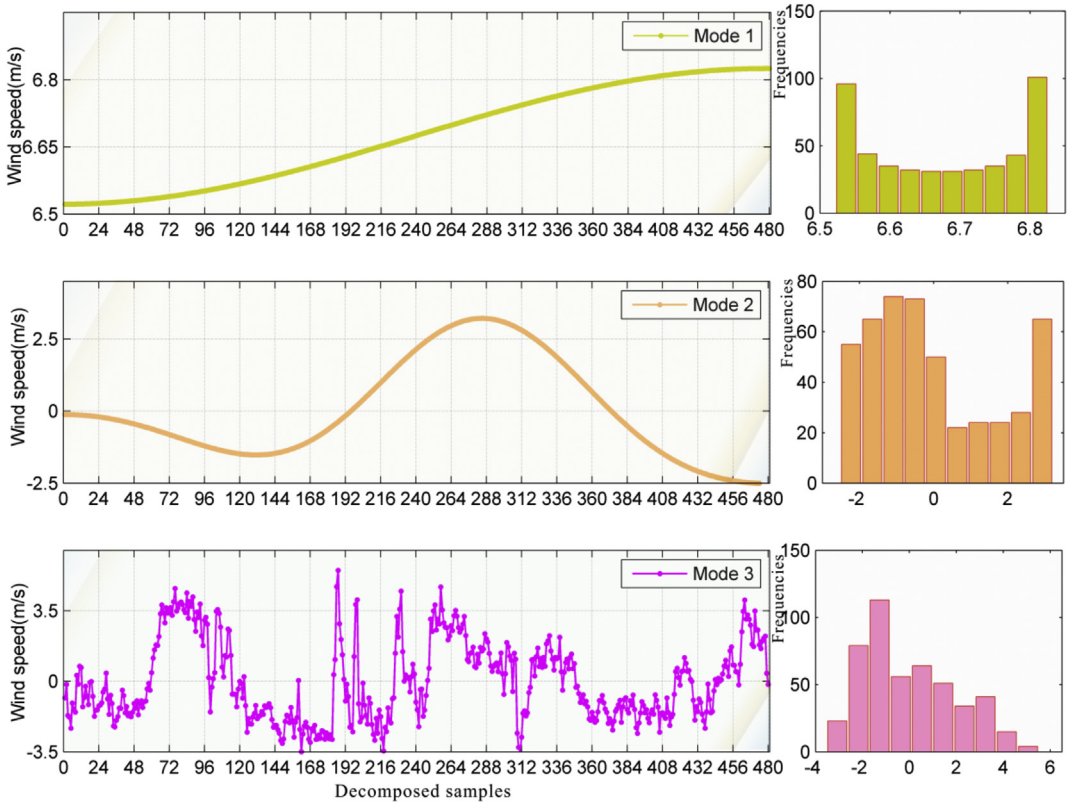

Fig. 3. The subseries of the mean 15-min dataset decomposed by EWT. The left column shows the decomposed subseries, and the right histograms denote the distribution of the corresponding decomposed subseries.

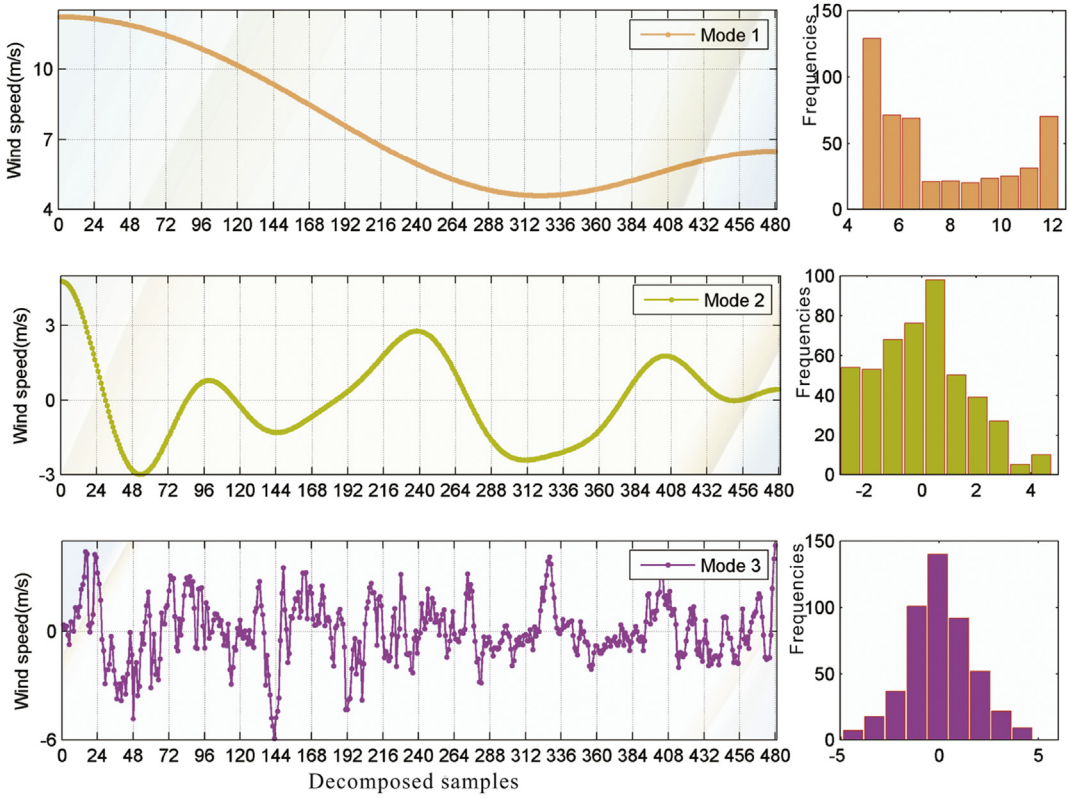

Fig. 4. The decomposed subseries of the mean 30-min dataset by EWT. The left column denotes the decomposed subseries and the right histograms denote the distribution of the corresponding decomposed subseries.

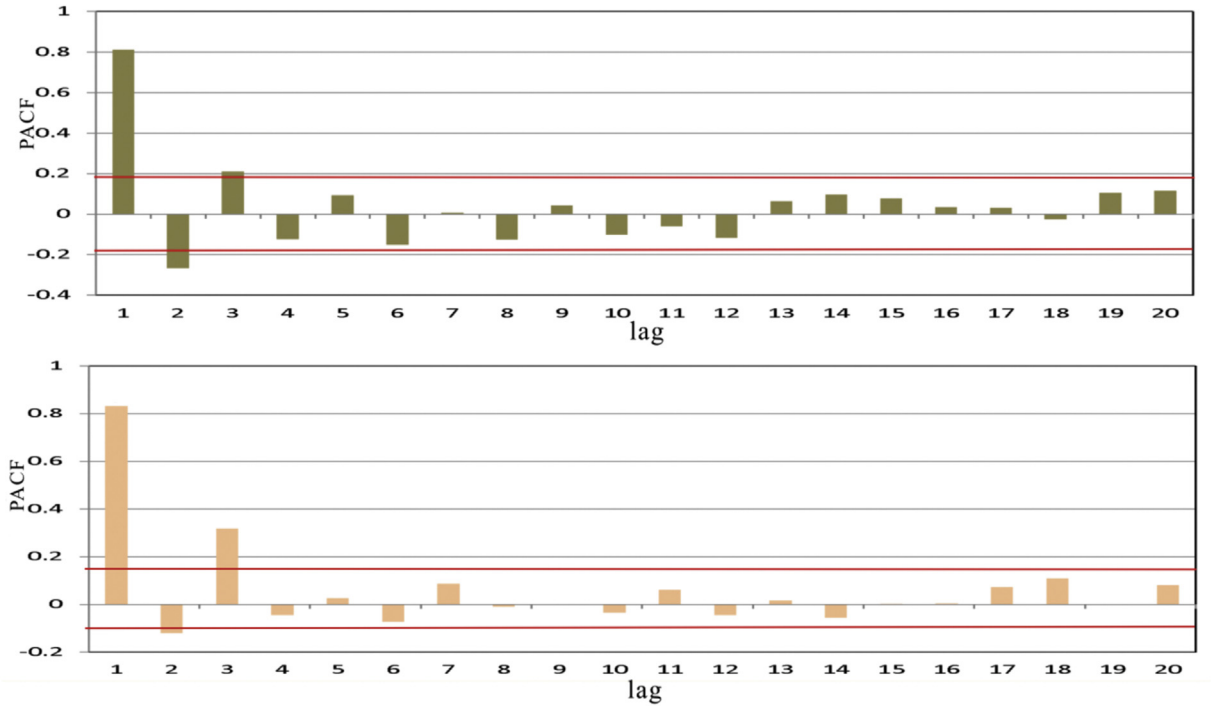

Fig. 5. The PACF values of the samples. (The top subplot shows the PACF values of the 15-min wind speed series from site 1, and the bottom subplot shows the PACF values of the 30-min wind speed series from site 2.)

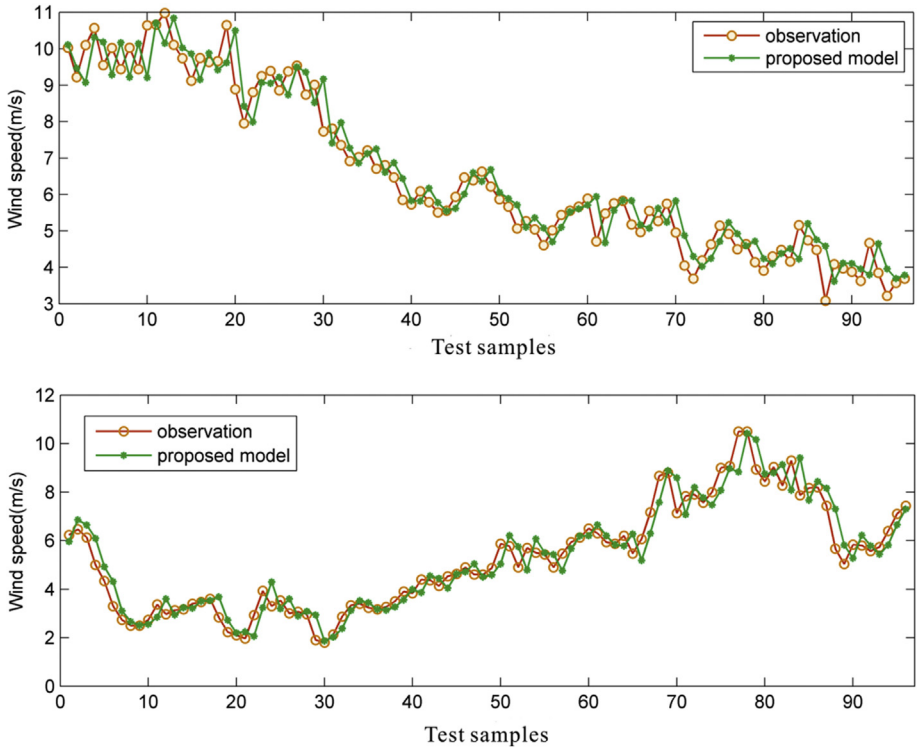

Fig. 6. Comparison between the actual wind speed series and the one-step ahead forecasting results from the proposed model. (The top subplot shows the forecast for the 15-min wind speed series from site 1, and the bottom subplot shows the forecast for the 30-min wind speed series from site 2.)

4.3.1. Data preprocessing

Fig. 2 demonstrates that the original wind speed series fluctuates greatly. The method mentioned in Subsection 3.1 is used to obtain the number of modes automatically detected for different α for wind speed series (results are shown in Table 1). It is apparent that values of α of approximately 0.4 and 0.5 give consistent results, corresponding to a tradeoff between too much detection and a good separation of the information in the Fourier spectrum. We add

Table 2
Performance evaluation of different models for forecasts using the 15-min wind speed data from site 1.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)
EWT-AR	9.13	0.52	0.54	10.74	0.60	0.62	12.74	0.71	0.73
EWT-LSSVM	7.81	0.47	0.47	10.01	0.57	0.57	12.25	0.70	0.70
EWT-SVM	7.80	0.47	0.48	9.68	0.56	0.57	11.40	0.67	0.67
EWT-ELM	9.49	0.56	0.57	11.32	0.64	0.66	13.37	0.76	0.78
Proposed method	7.75	0.46	0.46	9.88	0.56	0.56	12.02	0.69	0.69

Table 3
Performance evaluation of different models for forecasts using the 30-min wind speed data from site 2.

Horizons	One-step ahead			Two-step ahead			Three-step ahead		
Indices	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)
EWT-AR	14.92	0.61	0.64	18.86	0.79	0.81	21.51	0.87	0.90
EWT-LSSVM	9.92	0.48	0.48	15.71	0.72	0.72	20.12	0.87	0.87
EWT-SVM	10.96	0.52	0.53	16.02	0.75	0.76	20.58	0.92	0.93
EWT-ELM	13.42	0.60	0.64	21.03	0.92	0.97	25.08	1.05	1.10
Proposed method	9.84	0.48	0.48	15.44	0.72	0.72	20.05	0.86	0.88

the prior information available to help determine the appropriate mode. Previous work [26] suggests that three-level decomposition describes the wind speed series in a meaningful way. Therefore, three final modes are determined for both the mean 15-min wind speed series and the mean 30-min wind speed series. Using the EWT algorithm, the three uncorrelated modes are extracted from the wind speed series (shown in Figs. 3 and 4) and one residual is obtained after extraction. Prior to forecasting the wind speed, the residual is ignored, which helps to remove noisy data from the wind speed series. All three modes are incorporated into the new series by the EWT.

4.3.2. Individual forecasting

Prior to the forecasting task, the wind speed data series are used as input for the forecasting engines. Fig. 5 shows that the PACF is significant from lag 1 to lag 3 after the elimination of internal correlation. Therefore, the values from lag-1 to lag-3 are selected as the input to train the ELM, LSSVM and SVM models for wind speed forecasting. The available inputs are linearly normalized to the range [0, 1] to overcome the saturation phenomenon. Then, by employing the forecasting engines, the forecasting results of the corresponding model series are achieved. The previous two stages are repeated until the last sample of the dataset is incorporated. The forecasting process for the ARIMA model is detailed in previous work [44].

4.3.3. Combining the forecasts

Individual forecasts from each forecasting engine are used as input for the connecting model GP. Implementations of the Bayesian inference are conducted using priors defined in advance for the hyperparameters of the GPR model. Finally, the established GPR model is used to obtain the forecasting results for the combined series. The final forecasting results are shown in Fig. 6. Comparison between the data predicted by the proposed model and the observational data demonstrate the time lag between the predicted data and the observational data. This can be illustrated with the sample PACF (Fig. 5). We use the mean 30-min wind speed series as an example to explain these comparisons. Subsection 4.2 mentioned that the lag order of the autoregressive process is three. The PACF value at lag 1 ranges as high as 0.81, while the PACF values at lags 2–3 are 0.13 and 0.31, respectively. These values indicate that the data at time t strongly affect the prediction that occurs one time step ahead, while the effect of the actual data at time $t - l$ to the one-step ahead prediction is not so obvious, where l is lag order. Therefore, the correlation leads to the time lag of the predicted data against the actual data in visual way.

4.4. Results and comparisons

Tables 2 and 3 show the evaluation results for the individual forecasting and combination forecasting for site 1 and site 2. It is apparent that the forecast accuracy of the proposed combination approach is similar to or exceeds the best component model in all three metrics for both sites. This result reflects the reliability of the proposed combined method in view of the stochastic nature of wind and its spatial and temporal variations.

More detailed analyses are described as follows. For site 1, the EWT-SVM approach outperforms any other component model of

Table 4
Performance evaluations of different models with or without EWT for the forecast for the 15-min wind speed series from site 1.

Horizons		One-step ahead			Two-step ahead			Three-step ahead		
Indices		MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)
Without EWT	AR	9.61	0.55	0.57	11.21	0.62	0.65	13.17	0.73	0.76
	LSSVM	7.88	0.46	0.46	10.16	0.57	0.57	12.32	0.69	0.69
	SVM	8.19	0.48	0.50	9.89	0.56	0.58	11.13	0.64	0.66
	ELM	8.36	0.47	0.48	10.03	0.56	0.57	11.45	0.64	0.65
	Combination	7.84	0.46	0.46	9.96	0.56	0.56	11.68	0.66	0.66
With EWT	EWT-AR	9.13	0.52	0.54	10.74	0.60	0.62	12.74	0.71	0.73
	EWT-LSSVM	7.81	0.47	0.47	10.01	0.57	0.57	12.25	0.70	0.70
	EWT-SVM	7.80	0.47	0.48	9.68	0.56	0.57	11.40	0.67	0.67
	EWT-ELM	9.49	0.56	0.57	11.32	0.64	0.66	13.37	0.76	0.78
	Pro. method	7.75	0.46	0.46	9.88	0.56	0.56	12.02	0.69	0.69

Table 5
Performance evaluations of different models with or without EWT for the forecast for the 30-min wind speed series from site 2.

Horizons		One-step ahead			Two-step ahead			Three-step ahead		
Indices		MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)
Without EWT	AR	16.62	0.66	0.70	20.47	0.84	0.87	23.06	0.92	0.96
	LSSVM	10.09	0.48	0.48	15.97	0.73	0.73	20.88	0.90	0.90
	SVM	11.92	0.56	0.58	16.62	0.76	0.77	20.97	0.91	0.93
	ELM	15.34	0.67	0.72	19.48	0.84	0.87	22.85	0.95	0.98
	Combination	9.90	0.48	0.48	15.77	0.72	0.72	20.66	0.89	0.89
With EWT	EWT-AR	14.92	0.61	0.64	18.86	0.79	0.81	21.51	0.87	0.90
	EWT-LSSVM	9.92	0.48	0.48	15.71	0.72	0.72	20.12	0.87	0.87
	EWT-SVM	10.96	0.52	0.53	16.02	0.75	0.76	20.58	0.92	0.93
	EWT-ELM	13.42	0.60	0.64	21.03	0.92	0.97	25.08	1.05	1.10
	Pro. method	9.84	0.48	0.48	15.44	0.72	0.72	20.05	0.86	0.88

the proposed approach. For example, the EWT-SVM approach outperforms the other component models in three-step ahead forecasting and has a lower RMSE value of 0.67 m/s in contrast to RMSEs of 0.73 m/s, 0.70 m/s and 0.78 m/s for the EWT-AR, EWT-LSSVM and EWT-ELM models, respectively. The proposed method performs slightly better than the best component model, EWT-SVM, in one-step ahead and two-step ahead forecasting and performs slightly worse in three-step ahead forecasting. For site 2, the EWT-LSSVM model obtains the best forecasting performance among the component models. The proposed method performs better in one-step ahead and multiple steps-ahead forecasting than the best component model, EWT-LSSVM. Overall, the proposed approach has higher reliability at both sites in terms of the forecasting performance.

Tables 4 and 5 show the evaluation results of predictions obtained from models that used EWT and models that did not use EWT for both site 1 and site 2. The model comparisons demonstrate that the preprocessing EWT method is effective in boosting the forecasting accuracy of short-term wind speed prediction. For example, compared with the LSSVM model, the EWT − LSSVM model leads to reductions in all of the evaluation indices (MAE, RMSE and MAPE). Likewise, the proposed model results in lower evaluation indices with EWT than without EWT.

The above analysis focuses on the forecasting accuracy of single-value predictions. The GPR model, as a powerful, nonparametric, probabilistic approach, offers probabilistic predictive distribution beyond single-value predictions. Figs. 7 and 8 show the Gaussian probability density functions and the 95% confidence interval of the wind speed forecast, which illustrate the uncertainties in the predicted wind speed. Fig. 9 shows the forecast values obtained from the suggested model and the corresponding 95% confidence intervals. It can be observed from Fig. 9 that most of the actual observations fall within the confidence intervals. Such interval estimation can assist in the decision-making process for production

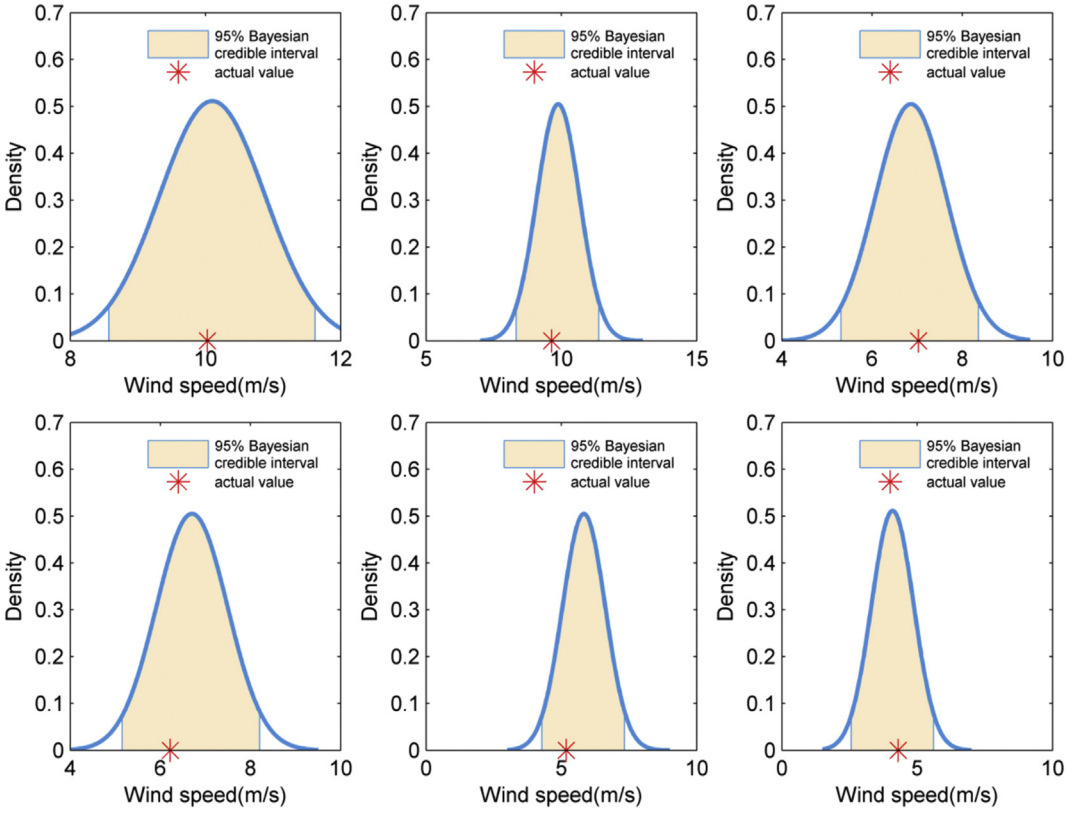

Fig. 7. The predictive pdf of the mean 15-min wind speed for the 1st, 17th, 33rd, 49th, 65th and 81st tests of data from site 1.

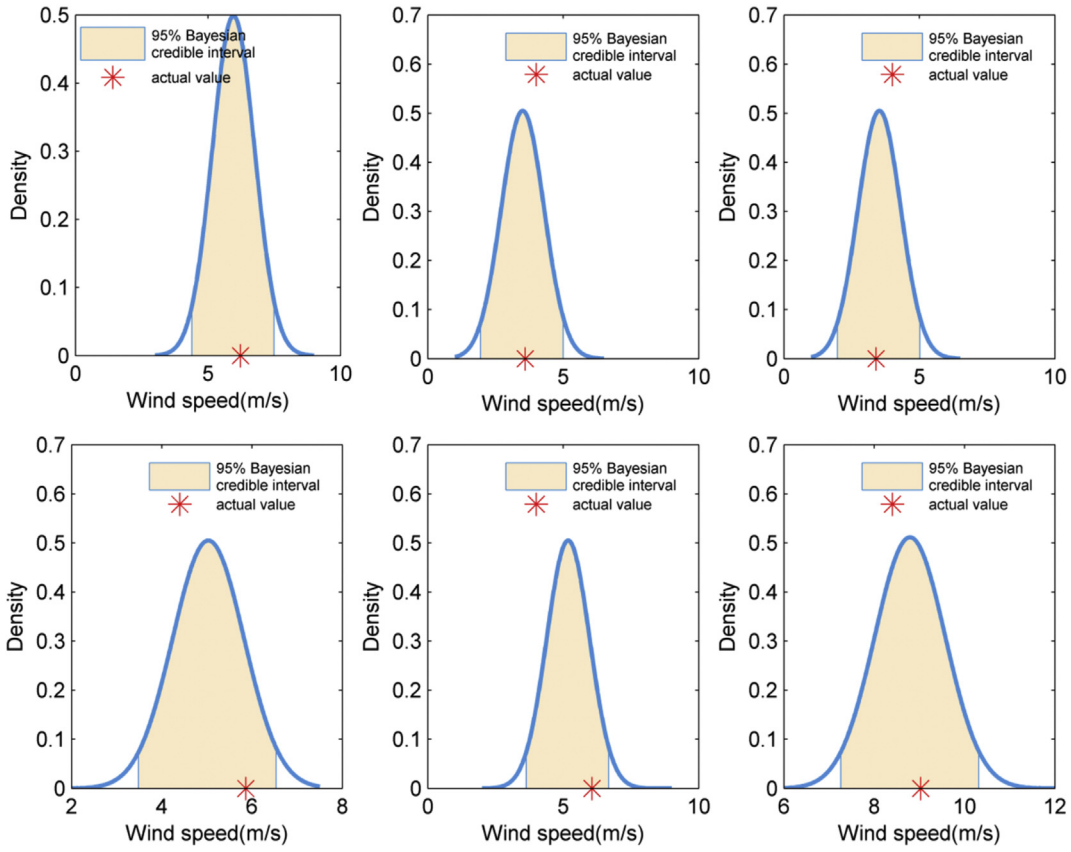

Fig. 8. The predictive pdf of the mean 15-min wind speed for the 1st, 17th, 33rd, 49th, 65th and 81st tests of data from site 2.

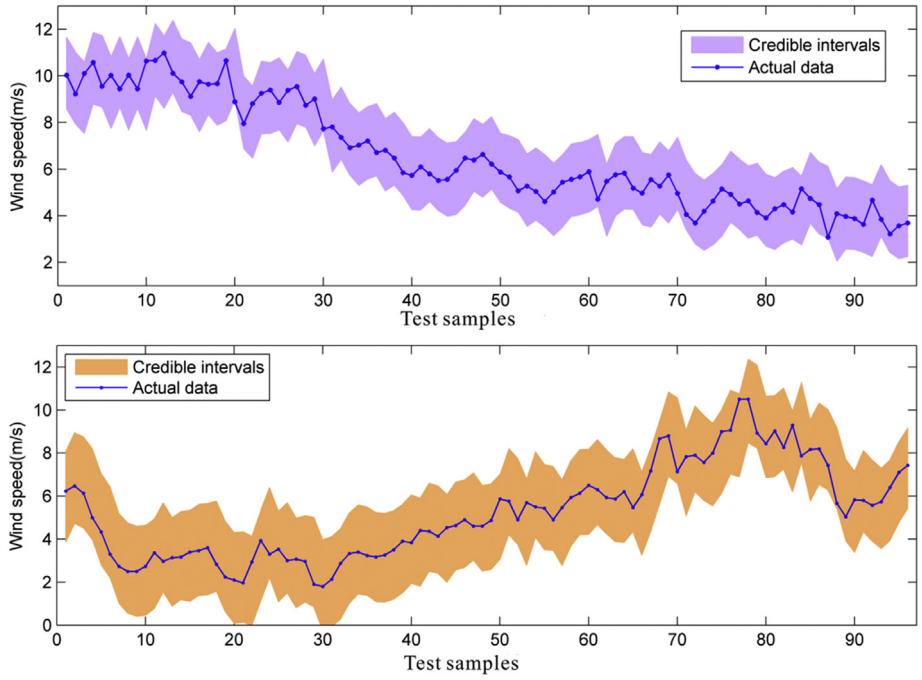

Fig. 9. The 95% confidence interval of the forecast against the observational data. (The top subplot shows the forecast for the 15-min wind speed series, and the bottom subplot shows the forecast for the 30-min wind speed series.)

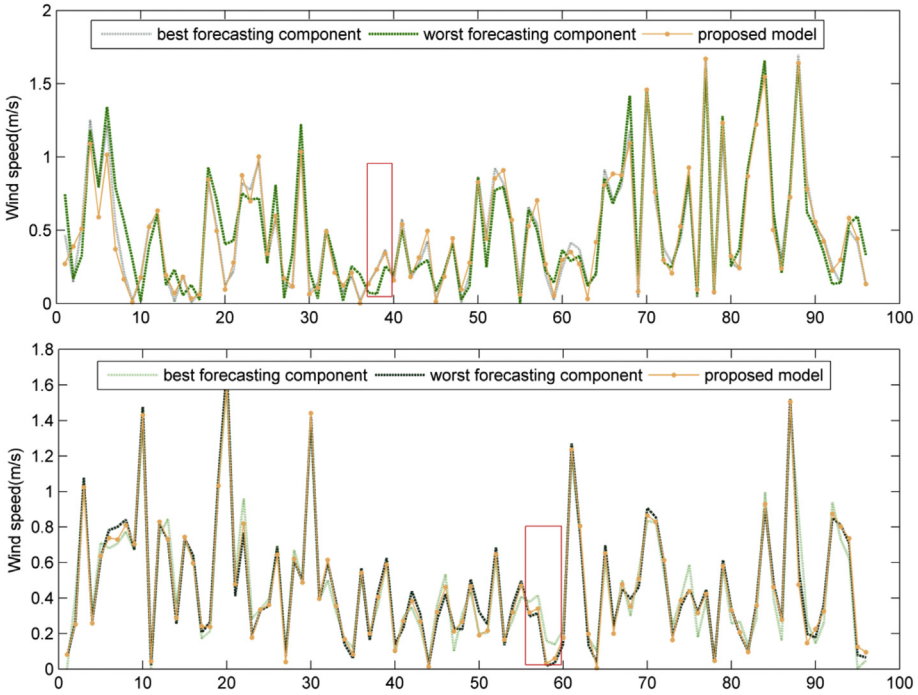

Fig. 10. Forecast errors of the BMA model and the two models generating the smallest/largest RMSE values.

scheduling and control. It could also reduce the opportunity costs of overly conservative bidding in the forward market that can result from uncertain availability.

4.5. Discussion

Fig. 10 shows the forecasts generated by the best and worst forecasting component models, as well as the combination forecasts for 96 consecutive data points. It can be observed that, although the best forecasting component performs better and with more accuracy than the worst forecasting component in terms of the RMSE, the worst component model (ELM) still has smaller forecast errors at individual times, for example, time points 38 and 39. This indicates that the performance of each forecast model varies with time. The forecasting engines have similar results for the mean 30-min wind speed series. Moreover, the performance of each forecasting engine varies from site to site [48]. These variations further support the necessity of developing a robust and accurate combination method for short-term wind speed forecasting. Due to the inconsistency in performance with time among the component forecast models, a Bayesian analysis-based forecast can always achieve an overall performance close to or better than the best component model. This confirms the reliable nature of Bayesian combination forecasts.

The performance of individual forecasting engines might change over time for different sites, which poses a challenge to the application of forecast models in industry. One solution is to generate a final single forecast that could take advantage of a set of plausible forecasts. For example, forecasts from alternative forecast agencies should be considered before making a decision. In such cases, an agency also has to combine all of the available information to provide a single forecast [34]. Therefore, an efficient forecast combination procedure is crucial for short-term wind speed forecasting. Ideally, this single time series forecast should have close-to-optimal performance under each evaluation metric.

The proposed methodology evaluates Bayesian analysis-based GPR forecasting in addition to various forecasting models. The simulation results verify that by combining the forecasts from each specific model with their corresponding posterior probabilities, the GPR model can generate reliable and comparatively high-accuracy forecasts. Moreover, the hybrid GPR model can provide probability information such as credible intervals and quantiles corresponding to a forecast based on predictive probability distribution.

5. Conclusions

Greater incorporation of wind energy into power systems has necessitated development of accurate and reliable techniques for short-term wind speed forecasting. Owing to the influence of various meteorological factors, wind speed series have intermittent and stochastic features, making it difficult to forecast wind speed using a single model, which mostly cannot achieve accurate predictions across various sites. Recent research has focused on developing new approaches and combining methods. This study estimates the wind profile using a combined method that integrates the ELM, SVM, LSSVM and GPR models. The ELM, SVM and LSSVM models are used to forecast the short-term wind speed, and the GPR is used to connect the forecasting engines and improve wind speed forecasting. Applying the GP method to the candidate models yields the most likely value and probability information corresponding to the forecast based on the predictive probability distribution. The most likely value forecast achieves close to (or better than) optimal performance as defined by three evaluation metrics (MAE, RMSE, and MAPE). 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.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant 71171102/G0107.

Appendix

Mercer's theorem

Let (x, μ) be a finite measure space and $K \in L_\infty(x^2, \mu^2)$ be a kernel such that $T_K : L_2(x, \mu) \to L_2(x, \mu)$ is positive definite let $\phi_i \in L_2(x, \mu)$ be the normalized eigenfunctions of T_K associated with the eigenvalues $\lambda_i > 0$, then:

1) the eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ are absolutely summable
2) $K(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i^*(x')$, hold μ^2 almost everywhere, where the series converges absolutely and uniformly μ^2 almost everywhere

Karush–Kuhn–Tucker conditions (KKT conditions)

In mathematical optimization, the Karush–Kuhn–Tucker conditions are first order necessary conditions for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. Allowing for inequality constraints, the KKT approach to nonlinear programming generalizes the method of Lagrange multipliers, which allows only equality constraints. The system of equations corresponding to the KKT conditions usually is not solved directly, except in a few special cases where a closed-form solution can be derived analytically. In general, many optimization algorithms can be interpreted as methods for numerically solving the KKT system of equations.

Nonlinear optimization problem

Consider the following nonlinear optimization problem:

Maximize $f(x)$
Subject to
$g_i(x) \le 0, h_i(x) = 0$

where x is the optimization variable, f is the objective or cost function, $g_i(i = 1, \cdots, m)$ are the inequality constraint functions, and $h_i(i = 1, \cdots, l)$ are the equality constraint functions. The numbers of inequality and equality constraints are denoted m and l, respectively.

Necessary conditions

Suppose that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable at a point x^*. If x^* is a local minimum that satisfies some regularity conditions (see below), then there exist constants $\mu_i(i = 1, \cdots, m)$ and $\lambda_i(i = 1, \cdots, l)$, called KKT multipliers, such that the inequality constraint diagram for optimization problems is as follows:

Stationarity

For maximizing $f(x)$: $\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*)$
For minimizing $f(x)$: $-\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*)$

Primal feasibility

$$g_i(x^*) \le 0, \text{for all } i = 1, \cdots, m$$
$$h_i(x^*) = 0, \text{for all } i = 1, \cdots, l$$

Dual feasibility

$$\mu_i \ge 0, \text{for all } i = 1, \cdots, m$$

Complementary slackness

$$\mu_i g_i(x^*) = 0, \text{for all } i = 1, \cdots, m$$

In the particular case $m = 0$, i.e., when there are no inequality constraints, the KKT conditions turn into the Lagrange conditions, and the KKT multipliers are called Lagrange multipliers.

If some of the functions are non-differentiable, subdifferential versions of the KKT conditions are available.

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] Nor Khalid Mohamed, Shaaban Mohamed, Rahman Hasimah Abdul. Feasibility assessment of wind energy resources in Malaysia based on NWP models. Renew Energy 2014;62:147–54.
[3] Cassola Federico, Burlando Massimiliano. Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. Appl Energy 2012;99:154–66.
[4] Erdem Ergin, Shi Jing. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[5] Liu Hui, Tian Hong-qi, Li Yan-fei. An EMD-recursive ARIMA method to predict wind speed. J Wind Eng Ind Aerodyn 2015;141:27–38.
[6] Kavasseri Rajesh G, Seetharaman Krithika. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[7] Yang Dazhi, Sharma Vishal, Ye Zhen, Lim Lihong Idris, Zhao Lu, Aryaputera Aloysius W. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy 2015;81:111–9.
[8] Li Yunhua, Ling Lina, Chen Jiantao. Combined grey prediction fuzzy control law with application to road tunnel ventilation system. J Appl Res Technol 2015;13:313–20.
[9] Wang Jujie, Zhang Wenyu, Li Yaning, Wang Jianzhou, Dang Zhangli. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl Soft Comput 2014;23:452–9.
[10] Ren Chao, An Ning, Wang Jianzhou, Li Lian, Hu Bin, Shang Duo. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowledge-Based Syst 2014;56:226–39.
[11] Shi Jing, Guo Jinmei, Zheng Songtao. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew Sustain Energy Rev 2012;16:3471–80.
[12] Cao Qing, Ewing Bradley T, Thompson Mark A. Forecasting wind speed with recurrent neural networks. Eur J Oper Res 2012;221:148–54.
[13] Sideratos George, Hatziargyriou Nikos D. Probabilistic wind power forecasting using radial basis function neural networks. IEEE Trans Power Syst 2012;27:1788–96.
[14] López Paz, Maseda Francisco. Wind speed estimation using multilayer perceptron Ramón Velo. Energy Convers Manage 2014;81:1–9.
[15] Zhu Bo, Chen Minyou, Wade Neal, Li Ran. A prediction model for wind farm power generation based on fuzzy modelling. Int Conf Environ Sci Eng 2012;12:122–9.
[16] Yin Xiu-xing, Lin Yong-gang, Li Wei, Gu Ya-jing, Liu Hong-wei, Lei Peng-fei. A novel fuzzy integral sliding mode current control strategy for maximizing wind power extraction and eliminating voltage harmonics. Energy 2015;85:677–86.
[17] 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.
[18] Kong Xiaobing, Liu Xiangjie, Shi Ruifeng, Lee Kwang Y. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2 December 2015;169:449–56.
[19] Hu Jianming, Wang Jianzhou, Zeng Guowei. A hybrid forecasting approach applied to wind speed time series. Renew Energy 2013;60:185–94.
[20] Osório GJ, Matias JCO, Catalão JPS. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle

swarm optimization, wavelet transform and mutual information. Renew Energy 2015;75:301–7.
[21] Kanna Bhaskar S, Singh N. AWNN-assisted wind power forecasting using feedforward neural network. IEEE Trans Sustain Energy 2012;3:15.
[22] Tascikaraoglu A, Uzunoglu M, Vural Tascikaraoglu B. The assessment of the contribution of short-term wind power predictions to the efficiency of stand-alone hybrid systems. Appl Energy 2012;94:156–65.
[23] Huang Norden E, Shen Zheng, Long Steven R, Wu Manli C, Shih Hsing H, Zheng Quanan, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc R Soc A Math Phys Eng Sci 1998;454:903–95.
[24] Guo Zhenhai, Zhao Weigang, Lu Haiyan, Wang Jianzhou. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37:241–9.
[25] Liu Hui, Chen Chao, Tian Hong-qi, Li Yan-fei. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. Renew Energy 2012;48:545–56.
[26] Liu Hui, Tian Hong-qi, Li Yan-fei. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. Energy Convers Manage 2015;89:1–11.
[27] Flores Juan J, Graff Mario, Rodriguez Hector. Evolutive design of ARMA and ANN models for time series forecasting. Renew Energy 2012;44:225–30.
[28] Li Ran, Ke Yongqin, Zhang Xiaoqian. Forecasting of wind speed with least squares support vector machine based on genetic algorithm. In: Proceedings of the International conference on consumer electronics, communications and networks (CECNet); 2011. p. 358–61.
[29] Amjady Nima, Keynia Farshid, Zareipour Hamidreza. Short-term wind power forecasting using ridgelet neural network. Electr Power Syst Res 2011;81:2099–107.
[30] Pandit Manjaree, Chaudhary Vishal, Dubey Hari Mohan, Panigrahi BK. Multi-period wind integrated optimal dispatch using series PSO-DE with time-varying Gaussian membership function based fuzzy selection. Electr Power Energy Syst 2015;73:259–72.
[31] Hervás Martínez C, Gutiérrez PA, Fernández1 JC, Salcedo-Sanz S, Portilla-Figueras A, Pérez-Bellido A, et al. Hyperbolic tangent basis function neural networks training by hybrid evolutionary programming for accurate short term wind speed prediction. In: Proceedings of the Ninth international conference on intelligent systems design and applications (ISDA'09); 2009. p. 193–8.
[32] Han Shuang, Li Jinshan, Liu Yongqian. Tabu search algorithm optimized ANN model for wind power prediction with NWP. Energy Proc 2011;12:733–40.
[33] Jabbari ghadi M, Hakimi gilani S, sharifiyan A, Afrakhteh H. A new method for short-term wind power forecasting. In: Proceedings of 17th conference on electrical power distribution networks (EPDC); 2012. p. 1–6.
[34] Sánchez Ismael. Adaptive combination of forecasts with application to wind energy. Int J Forecast 2008;24:679–93.
[35] Han Shuang, Liu Yongqian. The study of wind power combination prediction. In: Proceedings of the Asia-Pacific power and energy engineering conference (APPEEC); 2010. p. 1–4.
[36] Ali Ehab S. Speed control of induction motor supplied by wind turbine via imperialist competitive algorithm. Energy 2015:1–8.
[37] Jiang Yu, Song Zhe, Kusiak Andrew. Very short-term wind speed forecasting with Bayesian structural break model. Renew Energy 2013;50:637–47.
[38] Bessa Ricardo J, Miranda V, Botterud A, Zhou Z, Wang J. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renew Energy 2012;40:29–39.
[39] Pinson Pierre. Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions. J R Stat Soc Ser C Appl Stat 2012;61:555–76.
[40] Pinson Pierre, Kariniotakis George. Conditional prediction intervals of wind power generation. IEEE Trans Power Syst 2010;25:1845–56.
[41] Kou Peng, Gao Feng, Guan Xiaohong. Sparse online warped Gaussian process for wind power probabilistic forecasting. Appl Energy 2013;108:410–28.
[42] Baran Sándor. Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components. Comput Stat Data Anal 2014;75:227–38.
[43] Gilles Jérôme. Empirical wavelet transform. IEEE Trans signal Process 2013;61:3999–4010.
[44] Tsay Ruey S. Analysis of financial time series. 3rd ed. John Wiley & Sons, Inc.; 2010.
[45] Wang Haifeng, Hu Dejin. Comparison of SVM and LS-SVM for regression. In: Neural Networks and Brain, 2005. ICNN&B'05. International Conference on, IEEE, vol. 1; 2005. p. 279–83.
[46] Huang GuangBin, Zhou Hongming, Ding Xiaojian, Zhang Rui. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern 2012;42:513–28.
[47] Vanhatalo Jarno, Riihimäki Jaakko, Hartikainen Jouni, Jylänki Pasi, Tolvanen Ville, Vehtari Aki. GPstuff: Bayesian modeling with Gaussian processes. J Mach Learn Res 2013;14:1175–9.
[48] Li Gong, Shi Jing, Zhou Junyi. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. Renew Energy 2011;36:352–9.