Applied Energy 143 (2015) 472–488

Contents lists available at ScienceDirect

Applied Energy

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

The study and application of a novel hybrid forecasting model – A case study of wind speed forecasting in China

Jian-Zhou Wang [a], Yun Wang [b,*], Ping Jiang [a]

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

HIGHLIGHTS

- The Wavelet Packet Transform (WPT) methodology is exploited to process the original wind speed data.
- Phase Space Reconstruction is performed to select the input form.
- The LSSVM model whose parameters are tuned by an artificial intelligence (PSOSA) model is built to make forecast.
- Grey Relational Analysis and hypothesis test also indicate that the proposed model has the better forecast performance.

ARTICLE INFO

Article history:
Received 22 May 2014
Received in revised form 8 January 2015
Accepted 9 January 2015
Available online 9 February 2015

Keywords:
Wavelet Packet Transform
Least Square Support Vector Machine (LSSVM)
PSOSA algorithm
Grey Relational Analysis
Hypothesis test

ABSTRACT

Given the current increasingly serious energy crisis, the development and utilization of new energy resources are attracting increasing attention, and wind power is widely used among these renewable energy resources. However, the randomness of wind power can cause a series of problems in the power system. Furthermore, the integration of large-scale wind farms into the whole power grid can place a great burden on stability and security. Accurate wind speed forecasting would reduce the randomness of wind power, which could effectively alleviate the adverse effects on the power system. In this paper, a hybrid wind speed forecasting model is proposed with the hope of achieving better forecasting performance. Wavelet Packet Transform (WPT) was employed to decompose the wind speed series into several series with different frequencies. A Least Square Support Vector Machine (LSSVM), the parameters of which were tuned by a particle swarm optimization based on simulated annealing (PSOSA), was built to model those series. The optimal input form of the model was determined by Phase Space Reconstruction (PSR). To verify the effectiveness of the proposed model, the daily average wind speed series from four wind farms in Gansu Province, Northwest China, were used as a case study. The results of the simulation and Grey Relational Analysis indicate that the proposed model outperforms the comparison models, and the null hypothesis of the predicted series having the same mean of the real series was accepted.

1. Introduction

Along with the continuous increase in world energy consumption and vigorous development in traditional energy resources, the storage of fossil fuel is on the decrease, and the worldwide energy crisis is gradually becoming significant. Therefore, alleviating the energy crisis, developing renewable energy sources and achieving the sustainable development of energy has become a major initiative of the world's energy development strategy. Wind energy, an important category of renewable energy, is abundant, renewable, widely distributed and clean, resulting in wind power becoming an important renewable energy development direction. Presently, wind power is widely used not only in the developed countries but also in many developing countries, and even in some developed countries, wind power has partly replaced the traditional power generation modes and is providing the basic driving force of economic development.

China is the world's largest developing country and has abundant wind energy resources. With increasingly mature wind power technology and the government's strong support, wind power has become the fastest growing renewable energy in the country. According to the plan of the Chinese government, by 2020, the installed capacity of wind power will reach 30 GW [1]. In today's rapid development of wind power generation, the proportion of wind power in the whole power system is becoming larger.

* Corresponding author. Tel.: +86 18293191191; fax: +86 931 8912481.
E-mail address: wangyun12@lzu.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.apenergy.2015.01.038
0306-2619/

However, due to the randomness and intermittent of wind energy, random wind speed and wind direction lead to evident fluctuations of wind turbines' output power, such fluctuations adversely affect the grid frequency and voltage stability. When the proportion of wind power reaches a certain amount, it poses serious challenges to the security and stable operation of the power system and to the quality of the generated electric energy. In addition, to cope with the intermittent and random nature of wind power, sufficient backup power is necessary to protect the normal power supply to the users, which results in increases of the reserve capacity of the power system, which undoubtedly increases the operation cost of the power system. Thus, accurate wind speed forecasting can enhance the foreseeability of otherwise random wind power, reducing the reserve power requirements and accordingly increasing the reliability of the grid. A reduction of the operation cost and the spinning reserve would make it possible to increase the proportion of wind power in the grid.

To avoid the challenges of wind power integration, relatively accurate wind speed and wind power forecasting are extremely important. Currently, wind speed forecasting errors are approximately in the 25–40% range in Chinese wind farms [2,3]. These results are not satisfactory, and they are related not only to the forecasting methods but also to the forecasting period. Based on the requirements of wind power operation, forecasts can be grouped into four horizons [4,5]: ultra-short term, short term, mid and long term and long term. Ultra-short-term forecasts and short-term forecasts are mainly used for load tracking and preload sharing. The power system management and maintenance scheduling of the wind turbines are conducted using mid and long-term and long-term [6,7] forecasts, respectively.

Recently, many researchers have thoroughly studied wind speed and wind power forecasts, and many methods have been proposed and applied to wind farms. These methods can be classified into four categories [8]: physical models, statistical models, spatial correlation models, and artificial intelligence models.

Physical models not only make use of historical data but also consider the weather and geographical condition to aid wind speed forecasting with the expectation of achieving better forecasting accuracy [9,10]. On the contrary, statistical models, which are known as stochastic time series models, only employ historical wind speed. This model method is easy to apply and simple to implement. Thus, several types of time series models commonly appear in wind speed forecasting, including the autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA) [11], and autoregressive integrated moving average model (ARIMA) [12]. A survey of the literature leads to the conclusion that [4] the statistical models can perform well when applied to short-term, medium-term, and long-term wind speed forecasting in the vast majority of case studies, whereas the physical models present satisfactory results in the ultra-short-term and short-term horizons. Typically, the spatial correlation models [13,14] are mostly used when the available study information for wind farms is not sufficient, but the essential information in several adjacent wind farms is available. Unlike other models, to build the spatial correlation wind speed forecasting models, the wind speed and other necessary information, which contain delay times, have to be measured from multiple spatial correlated sites. Thus, the measurements and their delay times add complexity and cost to the implementation of spatial correlation forecasting. Recently, along with the rapid development and popularization of the artificial intelligence technology, it is common to apply different intelligence algorithms, including artificial neural networks (ANNs) [15–19], Support Vector Machine (SVM) [20–22] and fuzzy logic methods [23,24], to wind speed forecasting.

Hybrid wind speed forecasting models [25–28] are commonly and extensively adopted with satisfactory forecasting results. The reason for this that such models can mine the hidden information of different wind speed time series to a large extent. According to many studies, there are at least three methods that can enhance the wind speed forecasting accuracy. First, before being entered into the forecasting model, the original wind speed series is processed to achieve a relatively higher forecasting accuracy. The Wavelet Transform (WT) [29–31] and Empirical Mode Decomposition (EMD) [19,32,33] are the most common technology for processing wind speed series. The former can remove the irregular fluctuation of the original series, whereas the latter can decompose the original series into several intrinsic model functions (IMFs) for modeling. Second, as is well-known, the input form has a significant effect on the forecasting results. On occasion, the input form is decided by trial and error and the researchers' experience. In many papers, the Partial Auto Correlation function (PACF) [32,33] is applied to choose the best input form. In this paper, we use Phase Space Reconstruction [34] to determine the input form. Third, the parameters of models play a significant role in the forecasting accuracy of the modeling process. With the development of science and technology, common intelligent optimization algorithms, such as Genetic Algorithms (GAs) [29,35], simulated annealing (SA) [36,37], Practical Swarm Algorithms (PSOs) [5,35], and Ant Colony Optimization (ACO) [38] are widely used in parameter optimization. In addition, many papers have proved that the models that are optimized by intelligent optimization algorithms have a better forecasting performance.

In this paper, a hybrid model is proposed with the expectation of more precise forecasting of wind speed. In comparison with the WT methodology, the WPT methodology, which decomposes not only the low-frequency series but also the high-frequency series, can mine the traits of the original series more meticulously. First, the WPT methodology is adopted to process the original wind speed series to enhance the forecasting capacity, and several series with different frequencies that contain different characters of wind speed can be achieved. Then, Phase Space Reconstruction is performed for all the sub-series to select the input form except for the sub-series with the highest frequency, and the embedding dimension and delay time is calculated by the C–C method because of its easy operation and minimal calculation requirements. Finally, Least Square Support Vector Machine (LSSVM) models are built to forecast the selected sub-series. As all parameters in the LSSVM model haves a significant effect on the forecasting accuracy, the particle swarm optimization based on simulated annealing (PSOSA), which has the ability to avoid falling into local extreme points compared with the basic PSO algorithm, is employed to optimized the two parameters in the LSSVM model. The wind speed forecasting values can be obtained by adding up the forecasting values of all selected sub-series except for the one with the highest frequency. To evaluate the effectiveness of the hybrid method, case studies from four wind farms located in Gansu Province, Northwest China were conducted. In addition, Grey Relational Analysis and statistical hypothesis tests were used to evaluate the rationality of the forecasting series produced by the proposed model.

The advantages of the proposed model, which result in the better forecasting performance, are represented in the following several aspects. To begin with, many single methods make wind speed forecasting by using raw wind speed series directly, but the forecasting accuracy is not very satisfactory due to the influence of random noise in original series. In this paper, WPT is employed to preprocess the original wind speed series and reduce the effect of random noise. Then, the input form determination in the proposed model is smarter and more novel. By reviewing many papers, we found that determination of input form is based either on ones' experience or on the figure of PACF [29]. However, the fluctuation of wind speed affected by many factors such as temperature, pressure, and humidity is a

complex nonlinear dynamic system. So the innovation of this paper is trying to group the phase space theory into wind speed forecasting. Meanwhile, the input form can be automatically determined by C–C method, which can reduce the effects of personal selective input on the forecast results. Accordingly, instead of using the whole training set, we use a validation set to alleviate possible over fitting [39]. Additionally, it is suitable to select LSSVM as the main forecasting model by taking all aspects into consideration. In the literature, three very important and widely used wind speed forecasting models are ANNs, SVM and LSSVM. Though ANNs training algorithms have the ability to learn adaptively from the given data, they suffer from several disadvantages such as presence of local minima, over fitting and slow rate of convergence [40]. Moreover, LSSVM and SVM can also provide more stable forecasting results than some artificial neural networks because the weights of ANNs are initialized randomly. However, the major drawback of SVM is low speed in the training phase [41] while LSSVM applied in this paper can improve the training speed of solving the problem by transforming the quadratic programming problem of SVM into solving problem of linear equations. Unlike other LSSVM parameters determination methods, which only use single intelligent optimization algorithm such as single PSO or single SA, the combination of PSO and SA is applied in this paper to avoid falling into local extreme points. Finally, in addition to three common error criteria (MAE, MSE and MAPE) employed to evaluate the efficiency of the proposed hybrid model, Grey Relational Analysis and statistical hypothesis tests are also used to illustrate the superiority of the developed model from the perspective of geometric shape of forecasting series and statistics. On the whole, the novelty of developed model is reflected in the following aspects: (1) Data preprocessing method is exploited to process the original wind speed series; (2) Phase Space Reconstruction is performed to select the input form; (3) LSSVM model whose parameters are tuned by an artificial intelligence (PSOSA) model is built to make forecast. (4) Grey Relational Analysis and hypothesis test are selected as new tools to illustrate the forecast performance of the proposed model.

The rest of this paper is organized as follows. The methodology is described in detail in Sections 2, and 3 presents the modeling approaches of the proposed technique. Case studies are examined in Section 4 to verify the validity of the proposed model. In particular, the study area and data-sets are introduced in Section 4.1, whereas Section 4.2 presents the performance criteria of forecast accuracy. In Section 4.6, model comparisons are conducted; the results of Grey Relational Analysis are shown in Section 4.7, and the results of statistical hypothesis testing are presented in following section. Finally, the discussion and conclusions are provided in Section 5.

2. Methodology

2.1. Wavelet Packet Transform (WPT)

The Wavelet Packet Transform (WPT) is an effective method for wind data processing that can divide a wind speed series into a set of constitutive series. When using those constitutive series for time series forecasting, better forecasting accuracy can be achieved because of the filtering effect the WPT technique.

The concept of the wavelet packet was first proposed by Wickerhauser and others and is based on the theorem of wavelet transform. The structure of Wavelet Packet Transform is similar to discrete wavelet transform (DWT). Both have a framework involving multi-resolution analysis (MRA). However, the wavelet transform just decomposes the low frequency series, whereas the high frequency series remain unchanged. Part (*b*) of the Fig. 1 shows the schematic diagram of the wavelet transform. The result of the decomposition is $X_0 = H_1 + H_2 + \cdots + H_j + L_j$. That is, the original sequence X_0 is converted into an approximation component L_j and many detail components $H_i (i = 1, \ldots, j)$. The approximation component contains the low-frequency information. The most important part to provide the signal its identity and the detail components reveal the flavor of the signal. Because its decomposition scale is based on the binary mode changes, this type of decomposition produces poor frequency resolution at high frequencies and poor time resolution at low frequency. However, for Wavelet Packet Transform, as shown in part (*a*) of Fig. 1, it produces both further low frequency and high frequency decomposition. This effectively compensates for the lack of a wavelet transform. In Fig. 1, *L* and *H* represent the low frequency and high frequency, respectively, and the subscript indicates the decomposition layers of the wavelet packet. We can obtain $X_0 = LLL_3 + LLH_3 + LHL_3 + LHH_3 + HLL_3 + HLH_3 + HHL_3 + HHH_3$. The principle of WPT can be described as follows.

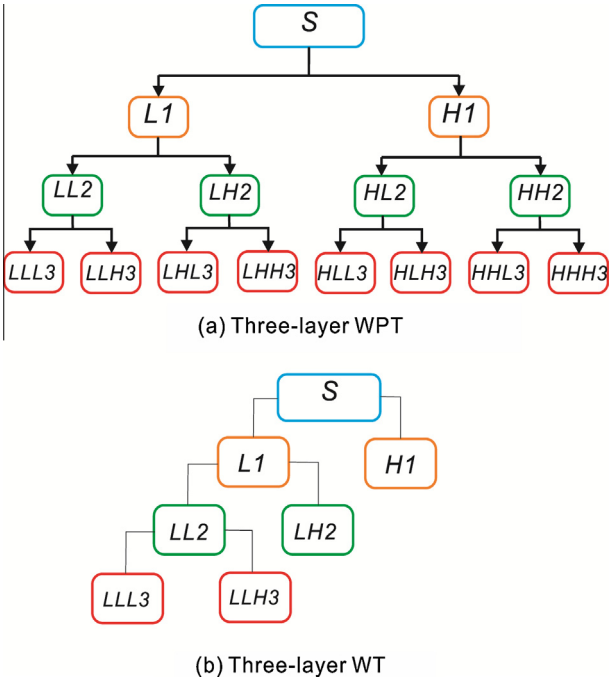


Fig. 1. Comparison between three-layer WPT and three-layer WT.

Let $\begin{cases} U_j^0 = V_j, & j \in Z \\ U_j^0 = W_j, & j \in Z \end{cases}$, with the subspace U_j^n defined as the closure space of the function $W_n(t)$; W_j^{2n} is the closure space of the function $W_{2n}(t)$, and let W_n meet the two-scale equation:

$$W_{2n}(t) = \sqrt{2}\sum_{k\in Z} h(k) W_n(2t-k) \tag{1}$$

$$W_{2n+1}(t) = \sqrt{2}\sum_{k\in Z} g(k) W_n(2t-k) \tag{2}$$

where $g(k) = (-1)^k h(1-k)$, which shows that the two coefficients $h(k)$ and $g(k)$ have orthogonal relationships, and its equivalent representation is

$$U_j^n = U_{j+1}^{2n} \oplus U_{j+1}^{2n+1}, \quad j \in Z, \quad n \in Z_+ \tag{3}$$

When $n = 0$, then we obtain $\begin{cases} W_0(t) = \varphi(t) \\ W_1(t) = \psi(t) \end{cases}$, where $\varphi(t)$ and $\psi(t)$ represent the scaling function and mother wavelet function, respectively. Therefore, we call $\{W_n(t)\}$ the wavelet packet of the orthogonal scaling function $W_0 = \varphi(t)$. The wavelet packet coefficients are given by the following:

$$S_p^{j,n} = \int_{-\infty}^{+\infty} x(t)2^{j/2}W_n(2^j t - p)dt \quad (4)$$

where $x(t)$, j, p and n are the signal, scale, position and surge parameter, respectively.

2.2. Phase Space Reconstruction – C–C method

The Phase Space Reconstruction theory was first proposed by Packard and Takens. Based on this theory, for a sequence of measured actual time $\{x_t|t = 1,2,\ldots,n\}$, if given the embedding dimension m and delay time τ, the number of point-phase space is N, which is calculated in Eq. (5), and the reconstructed phase space vector $X_i(i = 1,2,\ldots,N)$ is defined as Eq. (6).

$$N = n - (m-1)\tau \quad (5)$$

$$X_i = [x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)] \quad (6)$$

In the process of reconstruction, the selection of embedding dimension m and delay time τ is very important, and the proper choice of two parameters directly affects the quality of the Phase Space Reconstruction, thus affecting the prediction accuracy. In this paper, the C–C method is employed to determine the parameters of Phase Space Reconstruction. The C–C method can simultaneously estimate the delay time and embedding dimension by applying the correlation integral. Although the C–C method is based on statistical results and has no solid theoretical basis, it is easy to use, requires a small amount of calculation, and also has strong anti-noise capability. More information can be obtained in Ref. [42].

First, define the correlation integral

$$C(m,N,r,t) = \frac{2}{M(M+2)} \sum_{1 \leqslant i \leqslant j \leqslant M} \theta(r - d_{ij}) \quad r > 0 \quad (7)$$

where $d_{ij} = \|X_i - X_j\|$.

For a general time series with the t disjoint sub-sequence, the correlation integral $S(m,N,r,t)$ of each sub-sequence can be defined as follows:

$$S(m,N,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s\left(m,\frac{N}{t},r,t\right) - C_s^m\left(1,\frac{N}{t},r,t\right)\right] \quad (8)$$

It is used to describe the relevance of the nonlinear time series, and the best embedding dimension m and the best delay time τ is determined by its mean $\overline{S}(\tau)$. The C–C method can practically applied using the following steps:

Step 1: Select the appropriate data length and calculate the standard deviation of the given time series.
Step 2: Compute the statistics $\overline{S}(t)$ and $\Delta\overline{S}(t)$.

$$\overline{S}(\tau) = \frac{1}{16}\sum_{m=2}^{5}\sum_{1}^{4} S(m,r_j,\tau) \quad (9)$$

$$\Delta\overline{S}(t) = \frac{1}{4}\sum_{m=2}^{5} S(m,t) \quad (10)$$

where $r_j = \frac{j\sigma}{2}$, $j = 1,2,3,4$.
Step 3: $\overline{S}(t)$ and $\Delta\overline{S}(t)$ reflect the autocorrelation characteristics of the original time series. The best delay time τ is selected when the value of $\overline{S}(t)$ first reaches zero or $\Delta\overline{S}(t)$ reaches the first minimum point.
Step 4: Considering the $\overline{S}(t)$ and $\Delta\overline{S}(t)$ comprehensively, a new index is defined using the equation below:

$$S_{cor}(t) = \Delta\overline{S}(t) + |\overline{S}(t)| \quad (11)$$

The optimal time window ϖ can be obtained when $S_{cor}(t)$ reaches the global minimum. Thus, the embedding dimension m can be calculated by the formula $\varpi = (m-1)\tau$.

2.3. Least Square Support Vector Machine (LSSVM)

The Least Square Support Vector Machine was put forward by Suykens et al. in 1999 and is a variation of the standard Support Vector Machine (SVM). Solving the quadratic programming problem in the SVM model is transformed into solving system of linear equations, avoiding the insensitive loss function and greatly reducing the computational complexity. The estimation function of the LSSVM function used to solve the following problems as described. Consider the input and output training data set $T = \{(x_i,y_i)|x_i \in R^m,\ y_i \in R\}\ i = 1^l$. Then, use the high dimensional feature space structure optimal decision function $y(x) = \omega \cdot \varphi(x) + b$ to fit the sample set, where $\varphi(x)$ is the nonlinear mapping from input space to the high dimensional feature space; ω is the weight vector; b is the bias. When using the structural risk minimization principle, the regression problem is converted into a constrained quadratic optimization problem. An optimization problem for forecasting functions when applying the least squares support vector machine method can be formulated as follows:

$$J(\omega,\xi) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{i=1}^{l}\xi_i^2 \quad (12)$$

$$s.t. \quad y_i = \omega^T\varphi(x_i) + b + \xi_i; \quad i = 1,2,\ldots,l \quad (13)$$

where x_i and y_i represent the input and output variable, respectively. $\xi_i \in R$ is the error variable, and $\varphi(\cdot): R^m \to R^{m_h}$ represents the nonlinear mapping from input spaces to the feature space with a high dimension; ω is weight vector of the hyperplane; γ is the adjustable parameter; b is the bias. Introduce the Lagrange function

$$L(\omega,b,\xi,\alpha) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{l}\xi_i^2 - \sum_{i=1}^{l}\alpha_i(\omega^T\varphi(x_i) + b + \xi_i - y_i) \quad (14)$$

where $\alpha_i(i = 1,\ldots,l)$ is the Lagrange multiplication. According to the necessary conditions for the extremal existence, the following equations are obtained:

$$\frac{\partial L}{\partial\omega} = 0 \to \omega = \sum_{i=1}^{l}\alpha_i\varphi(x_i) \quad (15)$$

$$\frac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{l}\alpha_i = 0, \quad i = 1,\ldots,l; \quad (16)$$

$$\frac{\partial L}{\partial\xi_i} = 0 \to \alpha_i = \gamma\xi_i, \quad i = 1,\ldots,l; \quad (17)$$

$$\frac{\partial L}{\partial\alpha} = 0 \to y_i = \omega^T\varphi(x_i) + b + \xi_i, \quad i = 1,\ldots,l \quad (18)$$

Eliminating ω and ξ_i, we can obtain the following equation:

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_i,x_i) + \frac{1}{\gamma} & \cdots & K(x_1,x_l) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & K(x_l,x_1) & \cdots & K(x_l,x_l) + \frac{1}{\gamma} \end{bmatrix} \times \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_l \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_l \end{bmatrix} \quad (19)$$

where $K(\cdot,\cdot)$ represents the kernel function and satisfies Mercer's theorem. Then, the parameters b and a_i $(i = 1,2,\ldots,l)$ can be solved

from the equation, and the regression prediction function can be expressed by the following equation:

$$y = \sum_{i=1}^{l} \alpha_i \cdot K(x, x_i) + b \tag{20}$$

As observed from solving the implementation process, the least squares support vector machine set the inequality constraints of the SVM model to equality constraints, and, in addition, meanwhile the quadratic programming problem is converted into solving a system of linear equations. Then, the least squares method is used to determine the estimated function parameter, which greatly simplifies the computation complexity.

There are a number of kernel functions that can be chosen, and the following three kernel functions are common:

(a) Sigmoid kernel function

$$K(x_i, x_j) = \tanh[b(x_i \cdot x_j) + c] \tag{21}$$

(b) polynomial kernel function

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \tag{22}$$

(c) RBF kernel function

$$K(x_i, x_j) = \exp[-(x_i - x_j)^2/(\sigma^2)] \tag{23}$$

In this paper, the RBF kernel function is employed. Thus, for the LSSVM algorithm, the selection of the kernel function parameter σ^2 and the regularization parameter γ has an important influence on the establishment of the model, and to obtain better forecasting results, it is necessary to optimize these parameters.

2.4. Optimization algorithms

In this section, a brief introduction of the principles of the basic Particle Swarm Optimization (PSO) and simulated annealing (SA) are introduced, followed by the strategy of the combined optimization algorithm (PSOSA).

2.4.1. Particle Swarm Optimization (PSO)

Inspired by the flocking behavior of animals, Particle Swarm Optimization (PSO) was first introduced by Kennedy and Eberhart and widely used for parameters optimization. In the PSO algorithm, each particle represents a potential optimization problem solution. First, after initializing a swarm of random particles, each particle updates their positions by tracking the individual extremum (the optimal solution found by the single particle) and the global extremum (the optimal solution found by all particles) until the optimal solution was found.

Suppose that a group is composed of m particles. The Information of each particle is represented by a d-dimensional vector, so the position vector and vector of the ith particle are $X_i = (x_{i1}, x_{i2}, \ldots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, \ldots, v_{id})$, respectively. The historical optimal position vector of the ith particle is denoted as $P_i = (p_{i1}, p_{i2}, \ldots, p_{id})$, and the vector of the group is $P_g = (p_{g1}, p_{g2}, \ldots, p_{gd})$. In the process of iteration, the particles update their velocity and position by tracking the individual extremum P_i and global extremum P_g. The specific update formula is as follows:

$$v_{tm}(t+1) = \eta(t) \cdot v_{tm}(t) + c_1 r_1 (p_{tm} - x_{tm}(t)) + c_2 r_2 (p_{gm} - x_{tm}(t)) \tag{24}$$

$$x_{tm}(t+1) = x_{tm}(t) + \varpi \cdot v_{tm}(t+1) \tag{25}$$

where $m = 1, 2, \ldots, d$; t is the number of iterations; c_1 and c_2 represent the learning factor, and r_1 and r_2 are two random values, typically ranging between 0 and 1. ϖ stands for the elasticity. $\eta(t)$ is the inertia weight at iteration t, which can control the impact of the velocity vector as each iteration and is defined by Eq. (26).

$$\eta(t) = \eta_{\max} - (\eta_{\max} - \eta_{\min}) \times \frac{t}{t\max} \tag{26}$$

where $\eta_{\min}$ and $\eta_{\max}$ are the maximum and minimum of inertia weight, respectively; t represents the iteration number, and t max is the maximum iteration number.

2.4.2. Simulated annealing (SA)

The Simulated annealing (SA) algorithm, a stochastic optimization algorithm based on the Monte Carlo iterative solution strategy, was first used in the areas of combinatorial optimization by Kirkpatrick et al. The algorithm follows the similarity between the annealing processes of the solid matter and the process of general combinatorial optimization. It then draws on the physical principle of annealing of solid matter and applies this thermodynamic theory to statistics. Starting from a higher initial temperature and employing the Metropolis sampling strategy for random search in the solution space, the sampling process is repeated along with the gradual decreasing temperature, and eventually, the global optimal solution of the problem can be obtained. The basic steps of the SA algorithm are as follows:

Step 1: Parameter initialization. Set the range of parameters for the model; x_0 is randomly generated within the range as the initial solution, and the target value $E(x_0)$ is calculated. Set the initial temperature T_0, final temperature T_t and the cooling function $T(t+1) = \gamma T(t)$, where $\gamma \in (0, 1)$ denotes the annealing coefficient; t is the number of iteration.
Step 2: Generating the new solution. Add an increment Δx to the current solution x to generate a new solution $x' = x + \Delta x$, and then compute the increment of the target value $\Delta E(x) = E(x') - E(x)$.
Step 3: When $\Delta E(x) < 0$, accept the x' as the new solution; otherwise accept the x' as the new solution with the probability $p = e^{-\Delta E(x)/T}$.
Step 4: Continue to generate new solutions in the neighborhood of current solution x and repeat step 3.
Step 5: Comply with the cooling rule to reduce the temperature T.
Step 6: Repeat step 2 to step 5, until the temperature T reaches 0 or a given temperature.

The SA algorithm complies with certain annealing solutions in terms of lowering the temperature gradually by repeating the above steps. When the temperature is low enough, the global optimal state is reached. The SA algorithm achieves good results in solving combinatorial optimization problems and can solve some of the problems that are difficult for traditional optimization methods.

2.4.3. The combined optimization algorithm-PSOSA

The basic design idea of the hybrid optimization algorithm is detailed as follows. First, improve the basic PSO algorithm by using a particle velocity updating formula with a shrinkage factor updating method. Then, combine the simulated annealing algorithm and the particle swarm algorithm and establish a relation between the initial temperature and initial particle swarm. In addition, the genetic algorithm-related Roulette Wheel Selection Strategy is used in this paper.

The PSO algorithm is simple, easy to implement, requires fewer parameters and has a fast convergence speed. However, the PSO algorithm can easily fall into local extreme points, resulting in the problem deviating from the optimal position. To improve the

ability to avoid falling into local extreme points, we attempt to choose a position from the individual best position p_i to replace the global optimal position p_g in the velocity update formula, which is recorded as p'_g.

First, improve the basic PSO algorithm by using the shrinkage factor χ, which was proposed by Clerc and a function of the learning factors c_1 and c_2. The velocity and position update formula of the PSO algorithm with the shrinkage factor is defined as follows:

$$v_{tm}(t+1) = \chi \cdot (v_{tm}(t) + c_1 r_1 (p_{tm} - x_{tm}(t)) + c_2 r_2 (p_{gm} - x_{tm}(t))) \tag{27}$$

$$x_{tm}(t+1) = x_{tm}(t) + \varpi \cdot v_{tm}(t+1) \tag{28}$$

where χ is calculated by the following equation:

$$\chi = \frac{2}{|2 - C - \sqrt{C^2 - 4C}|}, \quad C = c_1 + c_2 \tag{29}$$

The question arises of how to select an effective optimal position p_i. Obviously, a p_i with good performance should be given a higher selection probability. According to the mechanism of the SA algorithm, the best position of each particle p_i is believed to be a special position worse than the global optimal position p_g, so the leap probability can be calculated by Eq. (30) when the temperature is T.

$$P_{leap}(p_i) = e^{-(f(p_i) - f(p_g))/T} \tag{30}$$

where $f(\cdot)$ is the particle position fitness value.

Therefore, the probability that the best position of each particle p_i is the new global optimal position p'_g under the current temperature is calculated according to the equation below:

$$P(p_i) = \frac{e^{-(f(p_i) - f(p_g))/T}}{\sum_{j=1}^{N} e^{-(f(p_j) - f(p_g))/T}} \tag{31}$$

N presents the size of the population.

According to the alternative probabilities above, the Roulette Wheel Selection Strategy is used to randomly determine which p_i should be selected as the new global optimal position p'_g.

The principles of the Roulette Wheel Selection Strategy are shown in Fig. 2.

We also know that the initial temperature has a greater impact on the performance of simulated annealing algorithm. If the initial temperature is too low, it will lead to insufficient leap strength of the algorithm; inversely, it will produce an excessive circuitous search. In this paper, a certain relationship between the initial temperature T_0 and the performance of the initial population is established. Specifically, after randomly generating the initial population, the fitness value of the global optimal position $f(p_g)$ can be obtained, and generally, the initial temperature can be defined as

$$T_0 = -f(p_g)/\ln(0.2) \tag{32}$$

Clearly, the whole design is helpful for overcoming some weaknesses of the basic PSO algorithm. Here, the PSO algorithm with the simulated annealing strategy is referred to as the PSOSA algorithm. The specifics of the PSOSA algorithm are described as in the following steps, and its flowchart is depicted in Fig. 3.

Step 1: Determine the initial temperature, and initialize the particle swarm.

(1.1) Initialize the positions and velocities of each particle in the population;

(1.2) Evaluate the fitness value of all particles;

(1.3) Let the best position p_i and the fitness $f(p_i)$ value of each particle be its current position and fitness value, and let the global best position p_g and fitness value $f(p_g)$ be the best position and fitness value in all the p_i;

(1.4) Determine the initial temperature, namely $T_0 = -f(p_g)/\ln(0.2)$.

Step 2: Update the particle's position and velocity.

(2.1) Calculate the probability $P(p_i)$;

(2.2) Judge the relationship of the probability $P(p_i)$ and *rand()*; if $P(p_i) >$ *rand()*, the new global optimal position p'_g is accepted as generated by the Roulette Wheel Selection Strategy; let $p_g = p'_g$;

(2.3) Update the velocity and position of each particle according to the location and velocity update formula;

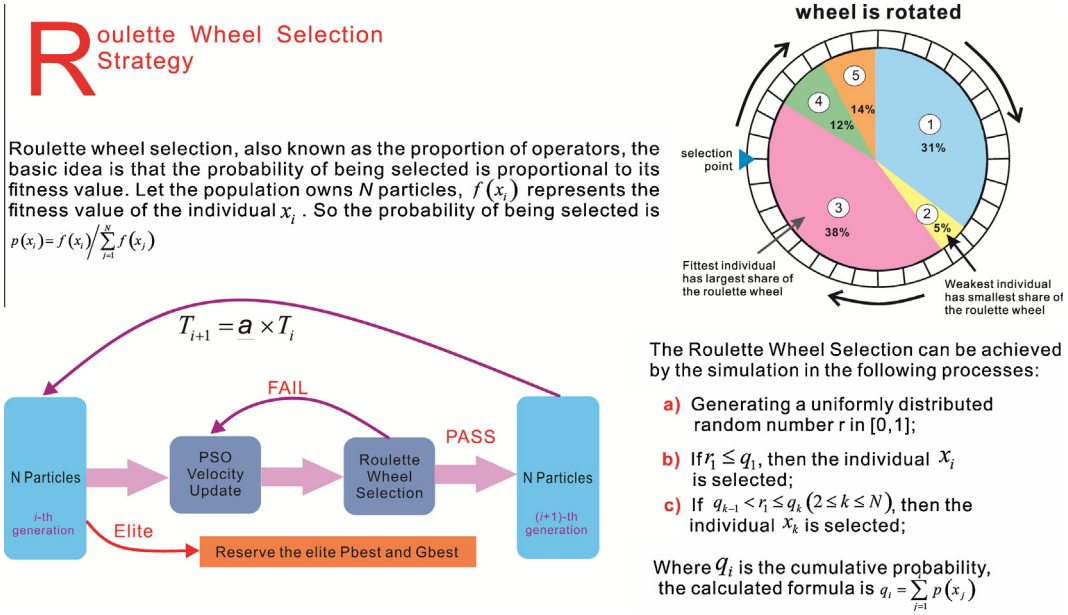


Fig. 2. The theorem of the Roulette Wheel Selection Strategy.

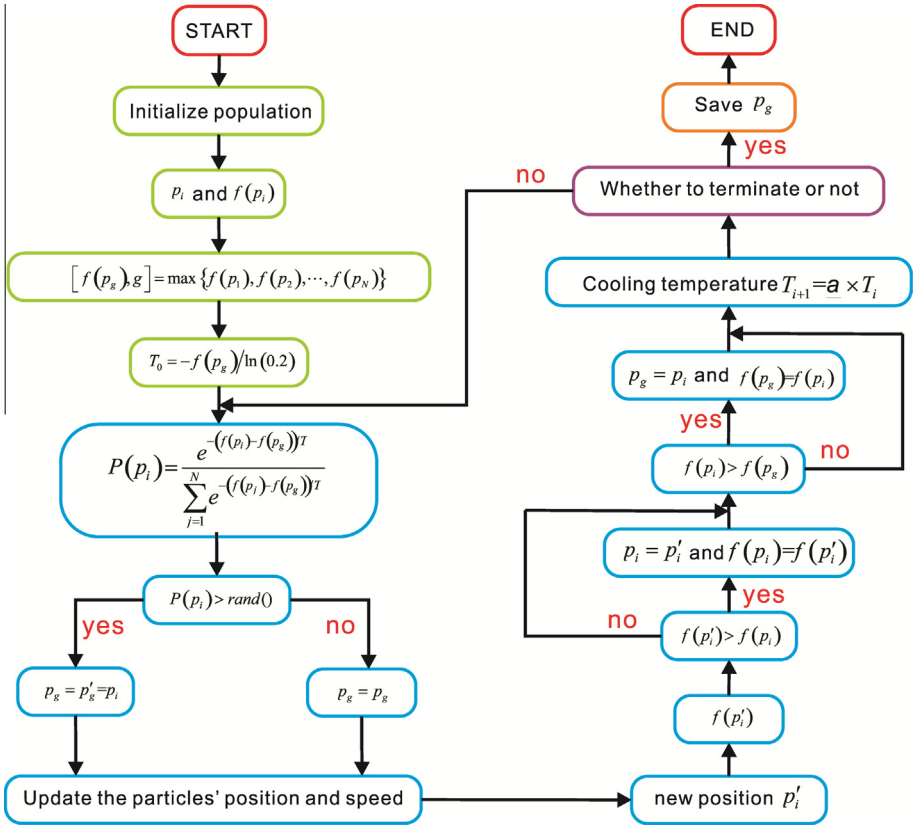


Fig. 3. The flowchart of the PSOSA algorithm.

(2.4) Compute the fitness value $f(p'_i)$ of each particle with new position;
(2.5) Compare the fitness value $f(p'_i)$ with the historical optimal fitness value $f(p_i)$; if $f(p'_i) > f(p_i)$, let p'_i be the current optimal position p_i, and let $f(p'_i)$ be the current optimal fitness value $f(p_i)$, namely $p_i = p'_i$ and $f(p_i) = f(p'_i)$;
(2.6) Compare the fitness value $f(p_i)$ of the current optimal position p_i of each particle with the fitness value $f(p_g)$ of the current global optimal position p_g. If $f(p_i) > f(p_g)$, let p_i be the current global optimal position p_g, and let $f(p_i)$ be the current global optimal fitness value $f(p_g)$, namely $p_g = p_i$ and $f(p_g) = f(p_i)$;
(2.7) Comply with the cooling rule to reduce the temperature.

Step 3: Judge whether the terminated conditions are met (usually the default calculation accuracy or iterations), and go to step 2 and continue to search if reached; otherwise, continue to step 4;
Step 4: Output the relevant information of p_g, and then end the algorithm.

2.5. Grey Relational Analysis

The theory of Grey Relational Analysis was first introduced by Deng in 1986. The basic idea of the Grey Relational Analysis is to check whether the different series have close contact according to the similarity degree of the series geometric shape. The higher the similarity degree of geometric shape is, the greater the correlation is, and the inverse is also true. The main steps in the Grey Relational Analysis are as follows:

Step 1: Establish the reference and comparison sequences

Define a reference time series:

$$X_0 = (X_0(1), X_0(2), \ldots, X_0(n)) \tag{33}$$

Then, define$m$ time series:

$$X_i = (X_i(1), X_i(2), \ldots, X_i(n)), \quad i = 1, 2, \ldots, m \tag{34}$$

Step 2: Achieve dimensionless variables

Change the reference and comparison time series using the equations below:

$$x_i(t) = \frac{X_i(t) - \overline{X}_i}{S_i} \tag{35}$$

$$\overline{X}_i = \frac{1}{n}\sum_{t=1}^{n} X_i(t) \tag{36}$$

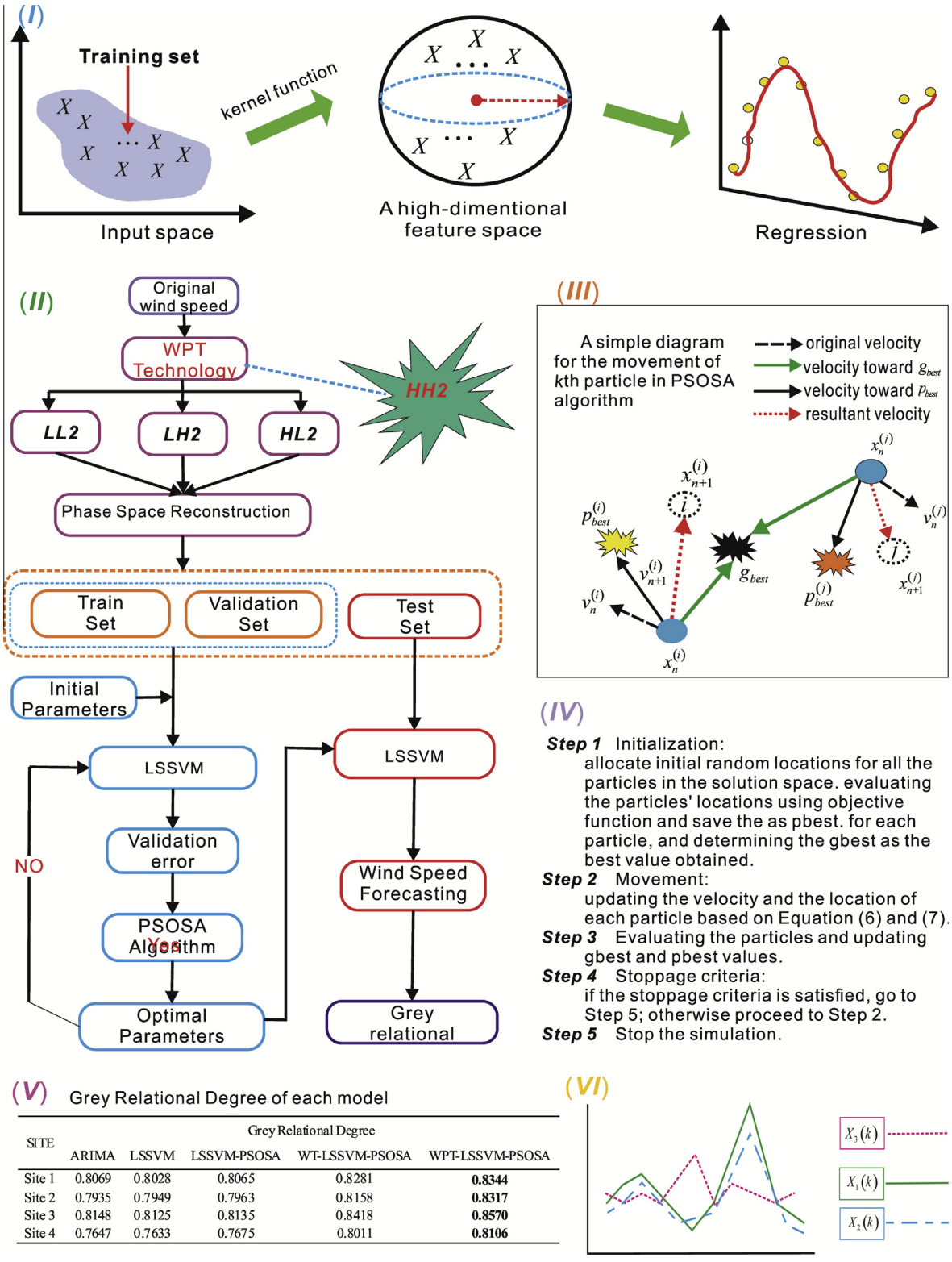
$$S_i = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(X_i(t) - \overline{X}_i)} \tag{37}$$

Reference time series $x_0 = (x_0(1), x_0(2), \ldots, x_0(n))$;
Comparison time series $x_i = (x_i(1), x_i(2), \cdots, x_i(n))$, $i = 1, 2, \ldots, m$;

Step 3: Calculate the correlation coefficient

The correlation coefficient between x_0 and x_i is

$$\xi_i(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}$$
$$\rho \in (0, \infty) \tag{38}$$



SITE	Grey Relational Degree				
	ARIMA	LSSVM	LSSVM-PSOSA	WT-LSSVM-PSOSA	WPT-LSSVM-PSOSA
Site 1	0.8069	0.8028	0.8065	0.8281	**0.8344**
Site 2	0.7935	0.7949	0.7963	0.8158	**0.8317**
Site 3	0.8148	0.8125	0.8135	0.8418	**0.8570**
Site 4	0.7647	0.7633	0.7675	0.8011	**0.8106**

Fig. 4. The flowchart of the proposed model. *Notes*: In this figure, part (*I*) depicts the basic principles of the LSSVM model. Part (*II*) presents the flowchart of the proposed model. A simple diagram for the movement of *k*th particle for the PSOSA algorithm is shown in part (*III*). The steps of the standard PSO algorithm are described in part (*IV*). Part (*VI*) is an ordinary description of the Grey Relational Analysis results, and the grey relational degrees of each model at four sites are presented in the table in part (*V*).

where ρ is the distinguishing coefficient. The smaller the ρ value is, the higher the resolution is. Generally, the value of ρ ranges from 0 to 1 and depends on the specific circumstances. Here, let $\rho = 0.5$.

Step 4: Compute the relational degree

$$r_i = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k) \quad k = 1, 2, \ldots, n \tag{39}$$

where r_i represents the relational degree between x_0 and x_i.

Fig. 5. A simple map of the study area.

Table 1
The statistical properties for each wind speed series.

Sites	Mean (m/s)	Std. dev. (m/s)	Maximum velocity (m/s)	Minimum velocity (m/s)
Wuwei	1.9944	0.9027	7	0
Jiuquan	4.4055	2.9252	13.5	0.5
Zhangye	1.9817	0.5976	3.3	0.3
Mazong Mountain	1.9461	0.5112	6	0.3

Step 5: Rank the relational degree
The relational degree is sorted by size. If $r_i < r_j$, the curve of j th series has greater similarity to the curve of the reference series.

3. The hybrid WPT–LSSVM–PSOSA model

In this section, the proposed hybrid model (WPT–LSSVM–PSOSA) will be introduced in detail, and the flowchart of the model is shown in Fig. 4. The hybrid methodology consists of four parts.

Part one: Data preprocessing. The WPT is employed to decompose the wind data series into a low-frequency component and several high-frequency components. Compared with the low-frequency component, the high-frequency components have large randomness and have little effect on future wind speed. Therefore, the highest frequency series is removed, and the other series are reconstructed and then fed into wind speed forecasting models as the independent variable as they maintain the major fluctuations of the wind speed series.

Part two: Input selection. Using the Phase Space Reconstruction method to rewrite the original wind speed series, the embedding dimension and delay time are determined by the C–C method. Then, each series is divided into three sets: the training set, the validation set and the test set for model training, validation and testing, respectively.

Part three: Model training and model validation. Here, the LSSVM–PSOSA model is applied for wind speed forecasting, with the specific steps are as follows:

Step 1: Determine the parameters of PSOSA.
Step 2: Determine the fitness function, and let the Mean Square Error (MSE) of validation set be the fitness of the particles.

$$fitness = MSE = \frac{1}{N}\sum_{i=1}^{N}(y^{p}_{validation} - y_{validation})^2 \quad (40)$$

where N is the number of validation set; $y^{p}_{validation}$ and $y_{validation}$ represent the predictive value and original wind speed data of the validation set, respectively.

Step 3: Initialize the structure of the LSSVM model.
Step 4: Update the individual extremum of each particle and global extremum, and then, generate the next generation of particles.
Step 5: Determine the stop condition of this algorithm. First, evaluate the fitness of the new particle population. If reaching the maximum number of iterations or expected error, then, go to step 6; otherwise, go to step 4 for continue iteration.
Step 6: Stop iteration, the global extremum must be reached and thus the corresponding optimal parameters of the LSSVM model.

Part four: Wind speed forecasting. In this part, the LSSVM model with optimal parameters tuned by PSOSA are used to forecast each series decomposed by WPT, and then, the wind speed forecasting series can be obtained by superimposing the forecasting values of each sub-series.

After the forecasting values were obtained by the models presented in this paper, Grey Relational Analysis was conducted. Then, hypothesis testing was performed to evaluate whether the mean of the forecasting series was the same as the real series at four sites.

4. Experimentation design and results

4.1. Study area and data sets

In this paper, the study area chosen contains four wind farms: Zhangye, Wuwei, Jiuquan, and Mazong Mountain, which are in northwest China–Gansu Province. A simple map of the study area is shown in Fig. 5. Because of their geographical characteristics, those areas contain the world's leading wind energy resources and have become home for China's biggest wind power projects [28]. Therefore, it is highly essential to achieve accurate wind speed forecasting for evaluating wind energy resources.

Here, the mean daily wind speed for approximately six years sampled from the four sites is employed to illustrate the effective performance of the proposed model. The total number of daily wind speeds is 2191. Before being fed into the model, the series

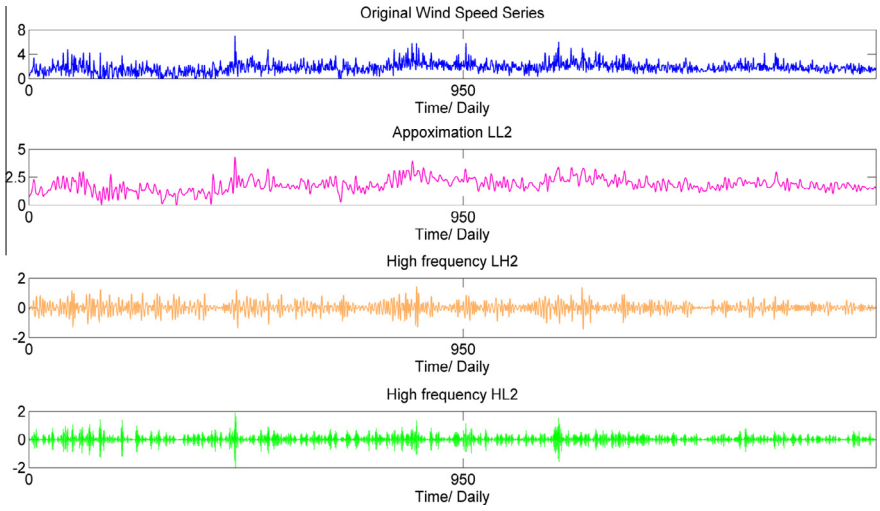


Fig. 6. Original wind speed series and the decomposed series with different frequencies.

Table 2
Results of the C–C method for the Wuwei site and Jiuquan site.

Sites	Series	Index	Delay time											
			1	2	3	4	5	6	7	8	9	10	11	12
Site 1	***LL2***	S_{cor}	0.342871	0.223956	0.152869	0.128085	0.106658	**0.101335**	0.105152	0.112244	0.116788	0.109056	0.105176	0.107673
		$\overline{S}(t)$	0.184174	0.109371	0.069453	0.056914	0.046424	0.043848	0.045885	0.049518	0.05115	0.047677	0.046022	0.047797
		$\Delta\overline{S}(t)$	0.158697	0.114585	0.083416	0.071172	0.060234	**0.057488**	0.059266	0.062726	0.065638	0.061379	0.059154	0.059876
	LH2	S_{cor}	0.18562	0.135423	0.119215	0.032889	0.029451	0.027692	0.019814	0.022573	0.017945	**0.013965**	0.019556	0.0227
		$\overline{S}(t)$	0.090605	0.064808	0.054846	0.01493	0.012831	0.011556	0.008277	0.009694	0.007652	0.00554	0.008106	0.009752
		$\Delta\overline{S}(t)$	0.095015	0.070615	0.064369	0.017959	0.01662	0.016136	**0.011536**	0.012879	0.010293	0.008425	0.011451	0.012948
	HL2	S_{cor}	0.284599	0.141384	0.052559	0.02649	0.019886	0.014817	0.01281	0.016617	0.014201	0.014804	0.014137	**0.006152**
		$\overline{S}(t)$	0.149592	0.067396	0.023414	0.011375	0.008083	0.005877	0.004935	0.006597	0.005525	0.005816	0.005977	0.002644
		$\Delta\overline{S}(t)$	0.135006	0.073988	0.029145	0.015115	0.011803	0.008939	**0.007876**	0.01002	0.008676	0.008989	0.008161	0.003509
Site 2	***LL2***	S_{cor}	0.34246	0.230722	0.164696	0.135833	0.115523	0.108067	0.110893	0.11258	0.114835	0.108917	**0.103101**	0.103453
		$\overline{S}(t)$	0.187758	0.115149	0.075856	0.061106	0.049631	0.046342	0.048381	0.050036	0.050874	0.047753	0.04486	0.045639
		$\Delta\overline{S}(t)$	0.154702	0.115573	0.08884	0.074727	0.065891	**0.061726**	0.062512	0.062544	0.063961	0.061163	0.058241	0.057813
	LH2	S_{cor}	0.190855	0.143426	0.129121	0.043182	0.04257	0.037208	0.030009	0.024089	0.025766	0.023665	**0.021298**	0.026418
		$\overline{S}(t)$	0.091453	0.067943	0.057846	0.018194	0.01809	0.014912	0.011905	0.010485	0.010814	0.009699	0.008819	0.010986
		$\Delta\overline{S}(t)$	0.099402	0.075483	0.071275	0.024988	0.02448	0.022296	0.018104	**0.013604**	0.014952	0.013966	0.012479	0.015432
	HL2	S_{cor}	0.271048	0.128055	0.04343	0.026237	0.028327	0.022524	0.014391	0.011216	**0.006848**	0.014276	0.01179	0.015027
		$\overline{S}(t)$	0.140581	0.060452	0.019066	0.010858	0.011468	0.009094	0.005984	0.004749	0.002671	0.005398	0.004356	0.006223
		$\Delta\overline{S}(t)$	0.130467	0.067603	0.024364	**0.015379**	0.016859	0.01343	0.008407	0.006467	0.004177	0.008878	0.007433	0.008804

Notes: The bolded entries indicate the global minimum of S_{cor} and the first minimum point of $\Delta\overline{S}(t)$.

were divided into three parts: training set, validation set and testing set, with their proportions set to approximately 75%, 25% and 25%, respectively. Table 1 below shows the mean, the standard deviation and the maximum and minimum velocities for the four wind speed series. It can be clearly observed that the statistical properties of the series are different from each other, and thus, we can conclude from the forecasting results if the proposed model can be widely applied for real-world use.

In the rest of this paper, we use the designations site 1, site 2, site 3 and site 4 to replace Wuwei site, Jiuquan site, Zhangye site, and Mazong Mountain site, respectively.

4.2. Performance criteria of forecast accuracy

To evaluate the efficiency of the proposed hybrid model, other models were built for comparison purpose. Recently, in the literature, many different error criteria have been widely used, but no one criterion has been proven as a universal standard method that can be applied to all cases. Therefore, multiple criteria are used to comprehensively evaluate the forecasting performance in the present study, and it is interesting to evaluate if the different criteria reflect the same performance level of different models. In this paper, three criteria are employed: mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{41}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{42}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{43}$$

where N is the number of test samples; $\hat{y}_i$ is the forecast value for a time period i, whereas y_i represents the actual wind speed at the same time.

Table 3
Results of the C–C method for the Zhangye site and Mazong Mountain site.

Sites	Series	Index	Delay time											
			1	2	3	4	5	6	7	8	9	10	11	12
Site 3	***LL2***	S_{cor}	0.328865	0.193395	0.109192	0.076231	0.052521	0.046703	0.051771	0.056289	0.049961	0.050809	0.048926	**0.045582**
		$\overline{S}(t)$	0.174241	0.093212	0.049135	0.033368	0.022372	0.019792	0.022358	0.024081	0.021479	0.021892	0.020494	0.019756
		$\Delta\overline{S}(t)$	0.154624	0.100183	0.060057	0.042863	0.030148	**0.026911**	0.029412	0.032208	0.028481	0.028917	0.028432	0.025826
	LH2	S_{cor}	0.184691	0.131259	0.119192	0.035007	0.027682	0.025314	0.019498	0.015178	0.014938	**0.015379**	0.010198	0.015679
		$\overline{S}(t)$	0.089088	0.061844	0.054236	0.015916	0.012032	0.01047	0.008125	0.007132	0.005936	0.0036	0.004201	0.006753
		$\Delta\overline{S}(t)$	0.095602	0.069416	0.064956	0.019091	0.01565	0.014843	0.011373	**0.008046**	0.009002	0.005673	0.005997	0.008926
	HL2	S_{cor}	0.291962	0.150954	0.061899	0.027889	0.017815	0.019295	0.01428	0.013799	**0.00916**	0.00955	0.011984	0.01683
		$\overline{S}(t)$	0.154896	0.072717	0.027938	0.011977	0.007078	0.007952	0.006041	0.005767	0.004191	0.004006	0.005364	0.006767
		$\Delta\overline{S}(t)$	0.137066	0.078237	0.033962	0.015912	**0.010737**	0.011343	0.008239	0.008032	0.004969	0.005543	0.006619	0.010063
Site 4	***LL2***	S_{cor}	0.30693	0.164182	0.076192	0.040828	0.024194	0.020688	0.020079	**0.019659**	0.021068	0.023656	0.021041	0.0215
		$\overline{S}(t)$	0.159363	0.07748	0.033139	0.017705	0.009483	0.008033	0.00751	0.008086	0.008383	0.00983	0.008107	0.008778
		$\Delta\overline{S}(t)$	0.147567	0.086703	0.043053	0.023123	0.014711	0.012655	0.012569	**0.011573**	0.012685	0.013826	0.012934	0.012722
	LH2	S_{cor}	0.179742	0.125359	0.113808	0.019496	0.017565	0.02223	0.008214	0.007683	0.004042	0.005453	**0.003073**	0.011447
		$\overline{S}(t)$	0.087278	0.059371	0.052182	0.008999	0.00765	0.00929	0.003455	0.002917	0.001397	0.001682	0.001	0.004461
		$\Delta\overline{S}(t)$	0.092464	0.065989	0.061626	0.010497	**0.009915**	0.01294	0.00476	0.004766	0.002645	0.003771	0.002073	0.006986
	HL2	S_{cor}	0.276085	0.119418	0.024013	**0.002686**	0.004409	0.002735	0.003816	0.006906	0.009566	0.00334	0.006796	0.008045
		$\overline{S}(t)$	0.139798	0.055214	0.010617	-0.00047	-0.00159	-0.00101	-0.00128	-0.00268	-0.00394	-0.00066	0.001475	0.001888
		$\Delta\overline{S}(t)$	0.136287	0.064204	0.013396	**0.002215**	0.002818	0.001725	0.002534	0.004231	0.005624	0.002679	0.005321	0.006158

Notes: The bolded entries indicate the global minimum of S_{cor} and the first minimum point of $\Delta\overline{S}(t)$.

Table 4
The optimal values of ϖ, τ and m using the C–C method.

Site	Max Lag	Parameters of ***LL2***			Parameters of ***LH2***			Parameters of ***HL2***		
		ϖ	τ	m	ϖ	τ	m	ϖ	τ	m
Site 1	12	6	6	2	10	7	2	12	7	3
Site 2	12	11	6	3	11	8	2	9	4	3
Site 3	12	12	6	3	10	8	2	9	5	3
Site 4	12	8	8	2	11	5	3	4	4	2

Currently, wind speed forecasting errors approximately range from 25% to 40% [2,3], which is related to not only the forecasting methods but also the forecasting horizon and the characteristics of the wind speed in a location. Generally, the shorter the forecasting horizon is or the more stable the wind speed variation is, the smaller the forecasting errors will be. Otherwise, the forecasting errors will be larger [43].

4.3. Results of WPT

To improve the forecasting performance of the original wind speed series, many methods, such as WT, WPT and EMD, are available. WPT is employed to address the original wind speed series. There are many known wavelet functions available to decompose raw data, and according to the characteristics of the data of case study, the 'db4' wavelet function is employed to decompose the original wind speed into two floors that contain four subsequences with different frequencies. Fig. 6 shows the original wind speed series and its subsequences with their frequency from low to high by WPT for Site 1 (Wuwei site). From Fig. 6, it can be clearly observed that the sub-series (***LL2***) with the lowest frequency that shows the major fluctuation of the original wind speed time series has a high similarity with the original wind speed time series. However, the other three subsequences with relative higher frequency reveal some irregular changes of wind speed. In this paper, the three sub-series (***LL2***, ***LH2*** and ***HL2***) on the second floor are taken into consideration with the hope of mining the hidden characters of the original wind speed as much as possible, and the sub-series ***HH2*** with relative higher frequency is removed.

4.4. Phase Space Reconstruction (PSR)

To forecast the wind speed, the LSSVM model was built as follows for the selected subsequences. For a LSSVM model, there are approximately four parts that will affect the forecasting performance: input forms, forecasting horizon, kernel function and the model parameters. In this section, the Phase Space Reconstruction is used to determine the input form. In this method, two important parameters of PSR, embedding dimension m and delay time τ, are selected by the C–C method. However, the time window ϖ and the delay time τ can be obtained. The time window ϖ can be obtained when S_{cor} reaches the global minimum, and the delay time τ is selected when $\Delta\overline{S}(t)$ reaches the first minimum point. Then, the embedding dimension m is obtained indirectly by the formula

Table 5
The PSOSA algorithm parameters employed in this paper.

Parameters	Value	Parameters	Value
N	30	M	200
Vcmax	0.5	popcmin	0.1
Vcmin	−0.5	popcmax	1000
Vgmax	0.9	popgmin	0.01
Vgmin	0.3	popgmax	100
c_1	2	ϖ	1
c_2	1.8	λ	0.7

Notes: Particles are initialized between (popcmin, popcmax) and (popgmin, popgmax) randomly. Velocities are initialized between (Vcmin,Vcmax) and (Vgmin,Vgmax) randomly. c_1 and c_2 indicate the learning factors. N and M present the population size and the evolution number. ϖ is the elasticity of population regeneration formula, whereas λ is the annealing constant.

Table 6
The optimal parameters in the LSSVM model.

Site	*LL2*		*LH2*		*HL2*	
	γ	σ^2	γ	σ^2	γ	σ^2
Site 1	0.1000	0.1189	0.1000	1.656	1.1192	53.2439
Site 2	2.9509	100	500.1000	2.4604	1000	17.5007
Site 3	1.3398	1.8403	357.8090	100	1.9658	1.3493
Site 4	53.0688	8.5519	1000	7.4528	4.0054	1.8745

$\varpi = (m-1)\tau$. Here, the max delay time in the C–C method is 12, which is selected arbitrarily. Tables 2 and 3 present the results of the C–C method as applied for the four sites. According to the results shown in Tables 2 and 3, the value of ϖ and delay time τ can be obtained. In addition, the calculated embedding dimension is shown in Table 4.

4.5. Selection of LSSVM model

As discussed above, we know that the parameters and the kernel function are important aspects of the LSSVM model when used to for wind speed forecasting. Here, the RBF kernel function is applied, and the PSOSA optimization algorithm is employed to optimize the parameters (γ, σ^2) by minimizing errors generated in the validation set. Some parameters of the PSOSA algorithm are listed in Table 5. The fitness function of the PSOSA is the MSE value generated in the validation set for selecting the optimal parameters of the LSSVM model for time series forecasting. Table 6 shows the optimal parameters (γ, σ^2) of the LSSVM models achieved by using the PSOSA algorithm.

After the best input form is decided and the optimal LSSVM model with best parameters is obtained, the testing set is applied to evaluate the forecasting performance of the proposed model. Table 7 shows the errors of the proposed model in the validation set and the testing set for the four sites.

Table 8 provides some statistical indices of the forecasting values and the real values. We can clearly see that the mean values have high similarity to the mean values of the true wind speed series. At the same time, the variance of the forecasting values is lower than the true variance, indicating that the forecasting series are relatively stable.

4.6. Models comparison

In this section, we compare the ARIMA model, LSSVM model, LSSVM–PSOSA model and the WT–LSSVM–PSOSA models, which

Table 7
Validation set and testing set results for the proposed WPT–LSSVM–PSOSA.

Site	Validation set			Testing set		
	MAE (m/s)	RMSE (m/s)	MAPE (%)	MAE (m/s)	RMSE (m/s)	MAPE (%)
Site 1	0.3056	0.3939	18.4622	0.4155	0.5530	25.4678
Site 2	0.3692	0.4808	18.4273	0.5055	0.7199	20.9877
Site 3	0.3868	0.5189	20.9673	0.3809	0.5237	20.3748
Site 4	0.8189	1.0583	21.0095	1.0180	1.3325	24.3096

Table 8
Some statistical indices in the testing step of the proposed model.

Test set	Index	Site			
		Site 1	Site 2	Site 3	Site 4
Forecasting series	Average	1.8209	2.3609	1.9821	4.6952
	Maximum	3.3955	5.5295	4.4585	9.7126
	Minimum	0.9099	0.8266	0.9490	1.4699
	Variance	0.1855	0.5564	0.3048	1.9274
Real series	Average	1.8287	2.3879	1.9766	4.6959
	Maximum	5.4000	6.8000	6.3000	13.5000
	Minimum	0.4000	1.0000	0.7000	1.1000
	Variance	0.5147	1.0586	0.5757	3.3815

Table 9
Error comparison among several forecasting models.

Area	Model errors	Forecasting models				
		ARIMA	LSSVM	LSSVM–PSOSA	WT–LSSVM–PSOSA	WPT–LSSVM–PSOSA
Site 1	MAPE (%)	32.5982	33.5425	33.1691	27.6086	**25.4678**
	RMSE (m/s)	0.6993	0.6885	0.6883	0.6062	**0.5530**
	MAE (m/s)	0.5146	0.5215	0.5130	0.4426	**0.4155**
Site 2	MAPE (%)	27.5445	27.5474	27.0291	23.5859	**20.9877**
	RMSE (m/s)	0.9442	0.9380	0.9584	0.7801	**0.7199**
	MAE (m/s)	0.6655	0.6605	0.6607	0.5608	**0.5055**
Site 3	MAPE (%)	29.3884	30.1150	29.6996	23.3224	**20.3748**
	RMSE (m/s)	0.7287	0.7353	0.7324	0.5988	**0.5237**
	MAE (m/s)	0.5314	0.5391	0.5357	0.4348	**0.3809**
Site 4	MAPE (%)	32.3436	32.3551	31.7645	26.1755	**24.3096**
	RMSE (m/s)	1.7192	1.7528	1.7297	1.4266	**1.3325**
	MAE (m/s)	1.3499	1.3691	1.3389	1.0909	**1.0180**

Table 10
The decreases in MAPE, RMSE and MAE reductions in comparison with the ARIMA, LSSVM and LSSVM–PSOSA models.

Area	Model errors	The proportion of reduction			
		ARIMA	LSSVM	LSSVM–PSOSA	WT–LSSVM–PSOSA
Site 1	MAPE (%)	21.87	24.07	23.21	7.75
	RMSE (m/s)	20.92	19.68	19.65	8.77
	MAE (m/s)	19.25	20.32	19.01	6.12
Site 2	MAPE (%)	23.80	23.81	22.35	11.01
	RMSE (m/s)	23.75	23.25	24.88	7.71
	MAE (m/s)	24.04	23.46	23.49	9.86
Site 3	MAPE (%)	30.6	32.34	31.39	12.63
	RMSE (m/s)	28.13	28.77	28.49	12.54
	MAE (m/s)	28.32	29.34	28.89	12.39
Site 4	MAPE (%)	24.83	24.86	23.46	7.12
	RMSE (m/s)	22.49	23.97	22.96	6.59
	MAE (m/s)	24.58	25.64	23.96	6.68

Table 11
The forecasting results and the computational time (in seconds) of the WPT–SVM–PSOSA and WPT–LSSVM–PSOSA model at different evolution times in four sites.

Evolution times	Study Areas	WPT–SVM–PSOSA						WPT–LSSVM–PSOSA					
		Forecasting errors			CPU time (s)			Forecasting errors			CPU time (s)		
		MAPE (%)	RMSE (m/s)	MAE (m/s)	***LL2***	***LH2***	***HL2***	MAPE (%)	RMSE (m/s)	MAE (m/s)	***LL2***	***LH2***	***HL2***
200	Site1	25.0875	0.5476	0.4085	4652.4852	4670.1089	4725.9376	25.4678	0.5530	0.4155	3837.0630	3836.3174	3956.3452
	Site2	21.3862	0.7334	0.5244	4689.4346	4723.9637	4750.5451	20.9877	0.7199	0.5055	3995.1337	4123.1330	4260.2030
	Site3	20.0685	0.4882	0.3611	4810.5656	4925.0692	4911.8528	20.3748	0.5237	0.3809	4020.2273	4220.1640	4290.0715
	Site4	23.8434	1.2997	1.0002	4692.2583	4705.9976	4715.6020	24.3096	1.3325	1.0180	3889.2565	3950.2712	3944.8735
50	Site1	25.4097	0.5395	0.4190	1408.9898	1540.7896	1642.9774	25.5189	0.5559	0.4211	958.5264	1082.3457	1149.8364
	Site2	20.8899	0.7199	0.5100	1519.2104	1559.4613	1616.1918	21.0005	0.7301	0.5210	991.7194	1069.4828	1175.1277
	Site3	20.4388	0.5488	0.3987	1602.3353	1592.248	1757.2747	20.4189	0.5412	0.3951	1070.6051	1116.2611	1289.0903
	Site4	24.1432	1.3281	1.0022	1493.6517	1561.7168	1571.4023	24.3402	1.3425	1.0359	956.7822	1031.7099	1095.0222
20	Site1	25.4025	0.5401	0.4115	471.6006	540.9638	594.1407	25.5256	0.5589	0.4223	337.4597	352.1082	359.565
	Site2	21.0040	0.7301	0.5312	477.9638	530.8268	634.7108	21.0259	0.7285	0.5252	330.4240	371.5627	392.9024
	Site3	20.0013	0.4899	0.3589	578.1407	595.9120	707.9145	20.4248	0.5323	0.3909	426.6767	398.9709	426.6455
	Site4	24.5544	1.3528	1.1035	465.8547	516.6370	588.4411	24.3452	1.3444	1.0431	339.5969	349.4874	367.3183

Table 12
Grey relational degree of each model.

Site	Grey relational degree				
	ARIMA	LSSVM	LSSVM–PSOSA	WT–LSSVM–PSOSA	WPT–LSSVM–PSOSA
Site 1	0.8069	0.8028	0.8065	0.8281	**0.8344**
Site 2	0.7935	0.7949	0.7963	0.8158	**0.8317**
Site 3	0.8148	0.8125	0.8135	0.8418	**0.8570**
Site 4	0.7647	0.7633	0.7675	0.8011	**0.8106**

were created for comparisons with the proposed model. The results are provided in Table 9 based on three criteria, MAE, RMSE and MAPE. In Table 9 below, the values in bold indicate the values of MAE, RMSE and MAPE that are the smallest among the various forecasting models. In this model comparison, it can be clearly observed that the proposed hybrid model WPT–LSSVM–PSOSA performs the best among all models, illustrating that the hybrid model can capture features of single models, capturing the characteristics of the wind data series, and achieving good wind speed forecasting performance.

As can be clearly seen from Table 9, the hybrid model WPT–LSSVM–PSOSA has the highest accuracy compared with the ARIMA, LSSVM, LSSVM–PSOSA and WT–LSSVM–PSOSA models. More precisely, at site 1, which is the Wuwei site, the proposed model achieves reductions of 21.87%, 24.07%, 23.21% and 7.75% in the total MAPE in comparison with the ARIMA, LSSVM, LSSVM–PSOSA and WT–LSSVM–PSOSA models. At the same time, the decreases in total RMSE are 20.92%, 19.68%, 19.65% and 8.77%, whereas for MAE, they are 19.25%, 20.32%, 19.01% and 6.12%, respectively. For the other three sites, the decreases of MAPE, RMSE and MAE are clearly listed in Table 10.

More through comparison research is described below. First of all, the ARIMA and LSSVM single models provide sometimes unsatisfactory performance in wind speed forecasting. This phenomenon reveals that the single models cannot contain the comprehensive features of the original wind speed series. When hybrid models are taken into consideration, the LSSVM∗∗∗–PSOSA model outperforms LSSVM model all the time, reflecting that the PSOSA algorithm is an efficient method for optimizing the parameters in the LSSVM model. Moreover, among all the forecasting models, the proposed hybrid WPT–LSSVM–PSOSA model achieves the best performance. Especially when compared with the LSSVM–PSOSA

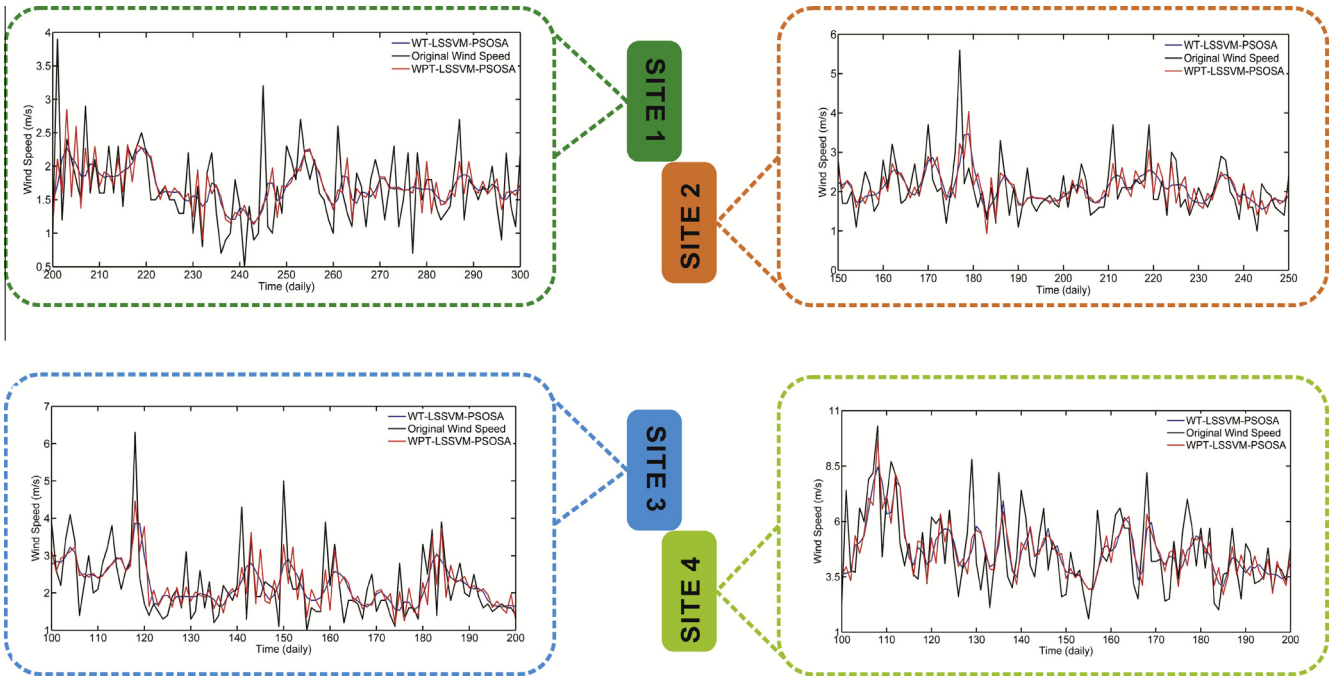


Fig. 7. Forecasting results of the WT–LSSVM–PSOSA model and WPT–LSSVM–PSOSA model at the four study sites.

Table 13
The hypothesis test results including the computed values of *T* statistic and *F* statistic in bold.

Area	Compare groups	Levene's test for equality of variances	*t*-test for equality of means					α
			F	$F_{\frac{\alpha}{2}}$	$F_{1-\frac{\alpha}{2}}$	t	$t_{\frac{\alpha}{2}}$	
Site 1	$\mu_1$and μ_2	Equal variance assumed	**4.0065**	0.8359	1.1962	**0.1880**	1.6471	0.1
		Equal variance not assumed				**0.1880**	1.6471	
	$\mu_1$and μ_3	Equal variance assumed	**2.7752**	0.8359	1.1962	**0.1717**	1.6471	
		Equal variance not assumed				**0.1717**	1.6471	
Site 2	$\mu_1$and μ_2	Equal variance assumed	**2.2668**	0.8359	1.1962	**0.3943**	1.6471	0.1
		Equal variance not assumed				**0.3943**	1.6471	
	$\mu_1$and μ_3	Equal variance assumed	**1.9025**	0.8359	1.1962	**0.3896**	1.6471	
		Equal variance not assumed				**0.3896**	1.6471	
Site 3	$\mu_1$and μ_2	Equal variance assumed	**2.6992**	0.8359	1.1962	**−0.1537**	1.6471	0.1
		Equal variance not assumed				**−0.1537**	1.6471	
	$\mu_1$and μ_3	Equal variance assumed	**1.8890**	0.8359	1.1962	**−0.1074**	1.6471	
		Equal variance not assumed				**−0.1074**	1.6471	
Site 4	$\mu_1$and μ_2	Equal variance assumed	**2.1215**	0.8359	1.1962	**−0.0006**	1.6471	0.1
		Equal variance not assumed				**−0.0006**	1.6471	
	$\mu_1$and μ_3	Equal variance assumed	**1.7544**	0.8359	1.1962	**0.0052**	1.6471	
		Equal variance not assumed				**0.0052**	1.6471	

model and WT–LSSVM–PSOSA model, we found that WPT methodology could effectively enhance the original wind speed series forecasting ability.

As is known, LSSVM model employs sum of error square as the empirical loss function, and changes the inequality constraints of the traditional SVM to equality constraint, thus the problem of solving quadratic programming is transformed into solving problem of linear equations, which results in faster convergence speed. The simulation results of WPT–SVM–PSOSA and WPT–LSSVM–PSOSA at different evolution times are presented in Table 11. It can be clearly seen from Table 11 that the forecasting accuracy of the WPT–SVM–PSOSA is generally a little better than that of the proposed model at different evolution times, but the gaps between them are very small. Also, with the increase of evolution times, the forecasting accuracy of WPT–LSSVM–PSOSA almost keeps the same, so does the WPT–SVM–PSOSA model. Therefore, in reality, PSOSA evolving nearly 20 times can find the optimal parameters of SVM and LSSVM, which not only can save time greatly but also can achieve the same good forecasting accuracy as the model with 200 evolution times. When comparing the CPU computational time of the two methods at different evolution times, we find that WPT–LSSVM–PSOSA is truly a much faster algorithm. And generally, the CPU computational time increases along with the increase of sub-series (***LL2***, ***LH2***, ***HL2***) frequency, and the lowest frequency sub-series ***LL2*** owns almost the least CPU computational time. It is not surprise to see that phenomenon for the more random fluctuations in high frequency series, which enhance the computational complexity of the proposed model. On the whole, WPT–LSSVM–PSOSA not only saves much time, but also achieves good forecasting accuracy compared with WPT–SVM–PSOSA at lower evolution times.

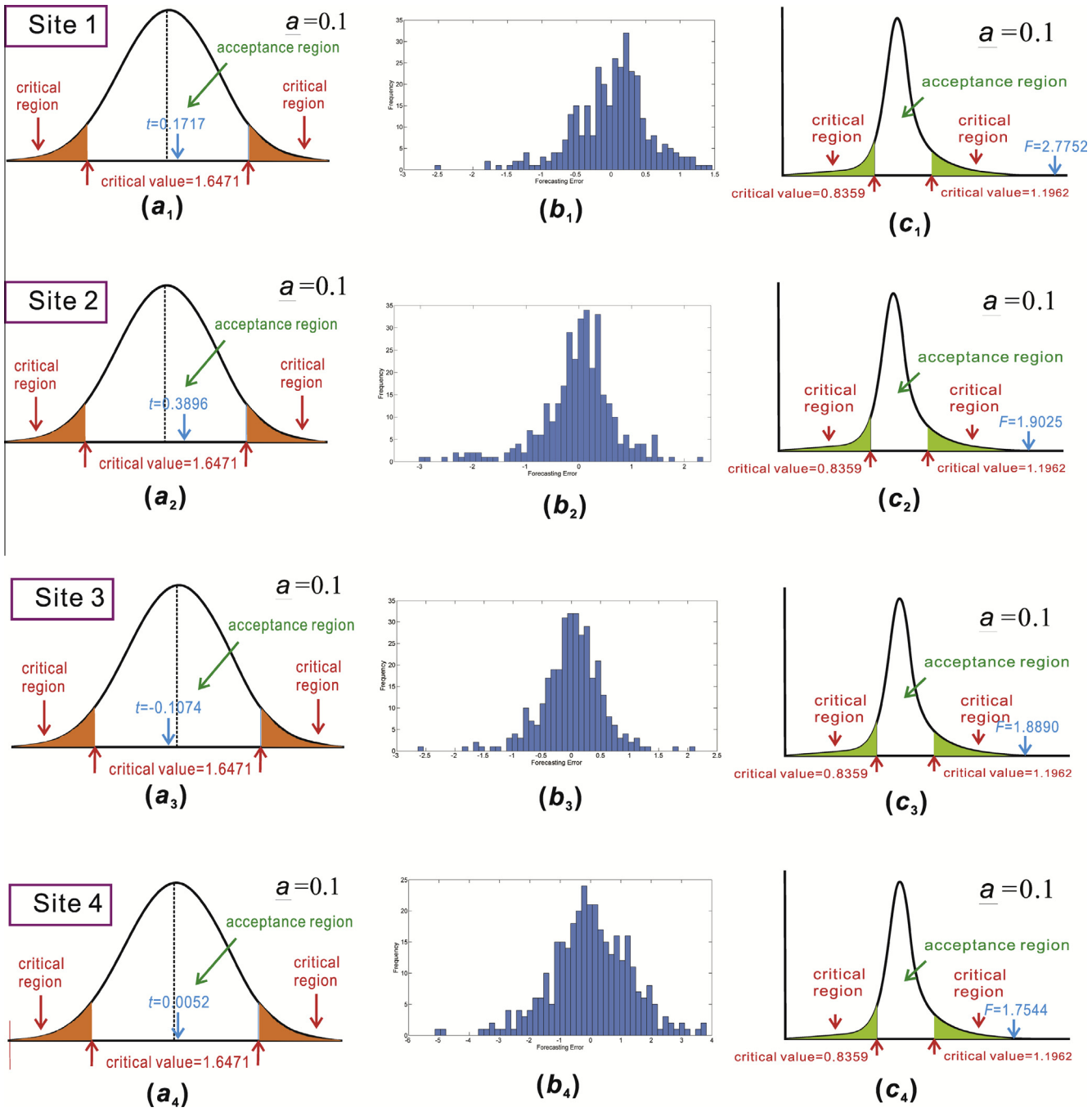


Fig. 8. The location of the statistics' values of the WPT–LSSVM–PSOSA model. Notes: (a_i) $(i = 1,2,3,4)$ shows the locations of the values of the T statistic, whereas (c_i) reveals the locations of the values of the F statistic; (b_i) is the frequency distribution of the forecasting errors produced by the proposed WPT–LSSVM–PSOSA model.

The simulations of different algorithms on all data sets are carried out in MATLAB R2012a environment running on Windows 7 with 2.50 GHz Intel Core i5 3210M, 64 bit having 4 GB RAM. It can be concluded from Table 11 that the forecasting results evolving 20 times are almost the same as those of the proposed model with evolving 200 times, but the CPU computational time slumps greatly. So it is easy to implement the developed wind speed forecasting program with lower evolving times, which can save more computational time using parallel computing by high-performance computers in wind farms. In fact, the single models such as ARIMA and LSSVM take less time when compared with the hybrid models. However, their forecasting accuracy is lower than the hybrid models. Therefore, considering the security of the whole power system when wind power is connected to the power grid, it is wise to take a few more minutes to make more accurate wind speed forecasting and thus alleviate the possible risks of the grid collapse.

According to recent papers, the wind speed forecasting errors at Chinese wind farms are approximately 25–40%. Although the features of the original wind speed series are different, the comparison indicates that the proposed hybrid model can perform effectively when applied to mean daily wind speed forecasting in northwest China. Therefore, the hybrid approaches provide a new viable option for wind speed forecasting applications in that they perform better than the single forecasting models in many instances. Thus, they should be included for consideration in searches of the best forecasting model [25].

4.7. Grey Relational Analysis results

Grey Relational Analysis was used to verify whether the forecasting result curve from the proposed model had the highest similarity to the curve of the true wind speed series. The grey relational degrees among the forecasting results of different models are presented in Table 12, and the bold values reflect the largest grey relational degree values. In Table 12, it can be seen clearly that the forecasting curves generated by the WT–LSSVM–PSOSA and the proposed hybrid WPT–LSSVM–PSOSA model have the greatest similarity to the real wind speed curve at the four wind farms examined in this paper. However, the grey relational degree of the proposed model is the highest of all. Samples from the forecasting series and the original wind speed series are shown in Fig. 7.

4.8. Statistical hypothesis testing

Hypothesis testing is also known as confirmatory data analysis and is different from exploratory data analysis. With frequency probability methods, we almost always make our decisions depending on the null-hypothesis tests. One use of hypothesis testing is deciding whether the experimental results contain enough information to invalidate current beliefs and conclusions. The critical region of a hypothesis test is the set of all outcomes that, if they occur, will lead us to decide there is a difference, that is, if the null hypothesis is false and to be rejected with the acceptance of the alternative hypothesis [44].

In this section, we detail the results of applying a two-tailed hypothesis test to test the efficiency of the proposed model (WPT–LSSVM–PSOSA). The hypothesis test is defined as follows:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

where μ_1 represents the mean of the real data, whereas μ_2 and μ_3 are the means of the forecasting series of the hybrid WT–LSSVM–PSOSA model and WPT–LSSVM–PSOSA model, respectively.

Some statistical indices (i.e., Average, Maximum, Minimum and Variance) in the testing step of the proposed model and the real series are shown in Table 8. Here, the value of alpha, our significance level, was arbitrarily selected. It is typical to let alpha be 0.05 or 0.1. For this paper, a 5% level was selected, and the alpha was selected as 0.05. Thus, if the value of the T statistic is in the interval $\left[-t_{\frac{\alpha}{2}}, t_{\frac{\alpha}{2}}\right]$ with 90% confidence, which covers the acceptable region, this indicates we can accept the null hypothesis that the means are the same. The results of the hypothesis test are presented in Table 13, and the corresponding values are marked in Fig. 8. It can be concluded that the mean of the forecasting series by the proposed model and the real series can be considered to be equivalent with 90% confidence. From Fig. 8, it can be clearly seen that the values of the T statistic are all located in the accepted region for the four sites; however, the values of the F statistic are all located in the critical region. Therefore, the forecasting series generated by the hybrid WT–LSSVM–PSOSA and WPT–LSSVM–PSOSA models in this paper are acceptable. The means of the forecasting series generated by the WT–LSSVM–PSOSA and the WPT–LSSVM–PSOSA models are considered to be equivalent to the means of the real series with 90% confidence.

5. Conclusions

Along with the world's unprecedented economic growth, population, and pollution problems, wind energy as a green and renewable resource has been gradually attracting attention. To aid in the understanding of the inner characteristics of wind speed time series and conduct wind speed forecasting, a considerable number of forecasting models have been developed. In this paper, a hybrid model was built with the hope of enhancing the wind speed forecasting accuracy. A WPT methodology was employed to process the original wind speed data to improve the forecasting capacity. Phase Space Reconstruction was used to determine the input form. Then, the LSSVM model was applied to perform wind speed forecasting, with parameters optimized by the PSOSA algorithm. The proposed hybrid model performed better than ARIMA, LSSVM, LSSVM–PSOSA and WT–LSSVM–PSOSA models.

The reasons for why the proposed hybrid model outperforms the other models can be summarized in terms of three aspects. First and foremost, the forecasting capacity of the original wind speed series can be largely enhanced after having been processed by the WPT method. Thus, the hybrid models including the WPT method can often achieve a relatively higher accuracy regardless of the original sequence. Second, the optimal structure and parameters of the model also play a decisive role in wind speed forecasting. Thus, intelligent algorithms are widely and commonly employed to optimize wind speed forecasting models. Third, the hybrid models possess the advantages of several single models and can also comprehensively retrieve the features of the original wind speed, whereas the single models can only mine the limited inner characteristics of the original wind speed series. In this context, it is not surprising to see that the hybrid model performs better than the ARIMA, LSSVM, LSSVM–PSOSA and WT–LSSVM–PSOSA models in terms of three types of statistical errors (MAPE, RMSE and MAE). On the other hand, the larger grey relational degree values also demonstrate that the hybrid model performs better. When applying this proposed hybrid model to the Wuwei, Jiuquan, Zhangye and Mazong Mountain sites, desirable results are achieved. Thus, the developed model not only can display satisfactory forecasting accuracy, it also can easily be implemented at wind farms. In addition, the statistical hypothesis test for the forecasting series of the four sites indicate that the there is no difference between the means of the forecasting series and the real series, and thus, the wind speed forecasting values are acceptable and credible.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 71171102) and the National Social Science Foundation of China (Grant No. 12CTJ012).

References

[1] International Energy Agency. Technology Roadmap: China Wind Energy Development Roadmap 2050. Source OECD Energy; 2012.
[2] Yang X, Xiao Y, Chen S. Wind speed and generated power forecasting in wind farm. Proc Chinese Soc Elect Eng 2005;25:1–5.
[3] Guo ZH, Wu J, Lu HY, Wang JZ. A case study on a hybrid wind speed forecasting model using BP neural network. Knowledge-Based Syst 2011;24:1048–56.
[4] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. Renew Sustain Energy Rev 2014;31:762–77.
[5] Ren C, An N, Wang JZ, 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.
[6] Carro-Calvo L, Salcedo-Sanz S, Kirchner-Bossi N, Portilla-Figueras A, Prieto L, Garcia-Herrera R, et al. Extraction of synoptic pressure patterns for long-term wind speed estimation in wind farms using evolutionary computing. Energy 2011;36:1571–81.
[7] Barbounis TG, Theocharis JB. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 2006;69:466–96.
[8] Ma L, Luan SY, Jiang CW, Liu HL, Zhang Y. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13:915–20.
[9] Kariniotakis G, Pinson P, Siebert N, Giebel G, Barthelmie R. The state of the art in short-term prediction of wind power–from an offshore perspective. 2004 Sea Tech Week; 2004.
[10] Lange M, Focken U. Physical approach to short-term wind power prediction. New York, NY: Springer; 2006.
[11] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.

[12] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renewable Energy 2009;34:1388–93.
[13] Bechrakis DA, Sparis PD. Correlation of wind speed between neighboring measuring stations. IEEE Trans Energy Convers 2004;19:400–6.
[14] Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. IEEE Trans Energy Convers 2004;19:352–61.
[15] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87:2313–20.
[16] Blonbou R. Very short-term wind power forecasting with neural networks and adaptive Bayesian learning. Renew Energy 2011;36:1118–24.
[17] Barbounis T, Theocharis J. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 2006;69:466–96.
[18] Barbounis T, Theocharis J. Locally recurrent neural networks for wind speed prediction using spatial correlation. Inf Sci 2007;177:5775–97.
[19] Guo ZH, Zhao WG, Lu HY, Wang JZ. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy 2012;37:241–9.
[20] Zhou JY, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manage 2011;52:1990–8.
[21] Salcedo-Sanz Sancho, Ortíz-García Emilio G, Pérez-Bellido Ángel M, Portilla-Figueras Antonio, Prieto Luis. Short term wind speed prediction based on evolutionary support vector regression algorithms. Expert Syst Appl 2011;38:4052–7.
[22] Mohandes M, Halawani T, Rehman S, Hussain AA. Support vector machines for wind speed prediction. Renewable Energy 2004;29:939–47.
[23] Hong YY, Chang HL, Chiu CS. Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. Energy 2010;35:3870–6.
[24] Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Trans Power Syst 2007;22:258–65.
[25] Shi J, Guo J, Zheng S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew Sustain Energy Rev 2012;16:3471–80.
[26] Cadenas E, Rivera W. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. Renew Energy 2010;35:2732–8.
[27] Liu H, Tian HQ, Li YF. Comparison of two new ARIMA–ANN and ARIMA–Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[28] Guo ZH, Zhao J, Zhang WY, Wang JZ. A corrected hybrid approach for wind speed prediction in Hexi Corridor of China. Energy 2011;36:1668–79.
[29] Liu D, Niu DX, Wang H, Fan LL. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renewable Energy 2014;62:592–7.
[30] Catalao JPS, Pousinho HMI, Mendes VMF. Short-term wind power forecasting in portugal by neural networks and wavelet transform. Renewable Energy 2011;36:1245–51.
[31] Catalao JPS, Pousinho HMI, Mendes VMF. Hybrid intelligent approach for short-term wind power forecasting in Portugal. IET Renew Power Gener 2011;5:251–7.
[32] Liu H, Tian HQ, Chen C, Li YF. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. Renewable Energy 2012;48:545–56.
[33] Hu JM, Wang JZ, Zeng GW. A hybrid forecasting approach applied to wind speed time series. Renewable Energy 2013;60:185–94.
[34] Wang JP, Xie YL, Zhu CH, Xu XB. A solar radiation prediction based on phase space reconstruction of wavelet neural network. Proc Eng 2011;15:4603–7.
[35] Liu H, Tian HQ, Chen C, Li YF. An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization. Int J Electr Power Energy Syst 2013;52:161–73.
[36] Pai PF, Hong WC. Software reliability forecasting by support vector machines with simulated annealing algorithms. J Syst Softw 2006;79:747–55.
[37] Pai PF, Hong WC. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers Manage 2005;46:2669–88.
[38] Niu DX, Wang YL, Wu DD. Power load forecasting using support vector machine and ant colony optimization. Expert Syst Appl 2010;37:2531–9.
[39] Xue XW, Yao M, Wu ZH, Yang JH. Genetic ensemble of extreme learning machine. Neurocomputing 2014;129:175–84.
[40] Balasundaram S. On extreme learning machine for ε-insensitive regression in the primal by Newton method. Neural Comput Appl 2013;22:559–67.
[41] Kang S, Cho S. Approximating support vector machine with artificial neural network for fast prediction. Expert Syst Appl 2014;41:4989–90.
[42] Kim HS, Eykholt R, Salas JD. Nonlinear dynamics, delay times, and embedding windows. Physica D 1999;127:48–60.
[43] Zhang WY, Wang JJ, Wang JZ, Zhao ZB, Tian M. Short-term wind speed forecasting based on a hybrid model. Appl Soft Comput 2013;13:3225–33.
[44] Introduction of the Statistical Hypothesis Testing, <http://phdtutor.com/stat_course/Hypothesis_Testing.aspx>. [Accessed 02.08.14].