



A new compound wind speed forecasting structure combining multi-kernel LSSVM with two-stage decomposition technique

Sizhou Sun^{1,2,3} · Jingqi Fu² · Ang Li² · Pinggai Zhang²

Published online: 4 August 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The aims of this study are to develop a novel compound structure that consists of two-stage decomposition (TSD), hybrid particle swarm optimization gravitational search algorithm (HPSOGSA) and multi-kernel least square support vector machine (MLSSVM) for improving forecasting accuracy. In the most previous wind speed forecasting studies, only one wind speed signal decomposition method is considered, which is insufficient. To better deal with the wind speed time series, TSD method combining complementary ensemble empirical mode decomposition with adaptive noise with wavelet transform is firstly employed in the proposed model to preprocess the wind speed samples; then, binary-valued particle swarm optimization gravitational search algorithm is exploited as feature selection to identify and eliminate the abnormal noise signal within the input candidate matrix that is determined by partial autocorrelation function. The kernel function and the kernel parameters have great influence on the regression performance of LSSVM. To solve these problems, integrations of radial basis function, polynomial (poly) and linear kernel functions by optimal weighted coefficients are constructed as multi-kernel function for LSSVM, namely MKLSSVM, and the parameter combination is tuned by conventional PSOGSA. The feature selection and parameter optimization are realized by hybrid PSOGSA (HPSOGSA) simultaneously. Finally, comprehensive comparison and analysis are carried out using the historical wind speed data from one wind farm of China to illustrate the excellent forecasting performance of TSD–HPSOGSA–MKLSSVM.

Keywords Wind speed forecasting · Two-stage decomposition · Complementary ensemble empirical mode decomposition with adaptive noise · Wavelet transform · Multi-kernel LSSVM · Hybrid particle swarm optimization gravitational search algorithm

1 Introduction

To reduce the greenhouse gas emission, many renewable resources have been exploited and utilized. Among the renewable resources, the utilization of wind energy has experienced a rapid growth in the last few years (Liu et al. 2017).

Reported by World Wind Energy Association (WWEA), the global cumulative capacity of wind power has increased from 283 GW in 2012 to 597 GW in 2018. As wind speed time series always present nonlinear and stochastic, the integration of the large scale of wind power is a serious unstable factor for power system, which brings about large power grid voltage fluctuations, flicker and other power quality, and also reduces the stability and reliability of power system. It is known that short-term wind power forecasting (ranging from 1 to 48 h) are critical to reduce the power reserves, lower the operation cost of wind farm, decrease the probability of wind power curtailment and increase the safety of the power grid (Sun et al. 2018c; Dong et al. 2017). However, influenced by temperature, humidity, height and other environmental and meteorological factors, it is hard to capture the internal tendency within wind speed times and there exist lots of difficulties in wind power prediction with high accuracy.

Communicated by V. Loia.

✉ Sizhou Sun
sszhou12345@i.shu.edu.cn

✉ Jingqi Fu

¹ Anhui Key Laboratory of Electric Drive and Control, Anhui Polytechnic University, Wuhu 241000, China

² School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China

³ School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China

In the last decades, various wind speed forecasting models have been developed and these models are mainly classified into four categories, namely physical approaches, statistical models, artificial intelligence and machine learning engines, and hybrid forecasting model. Physical approaches predict the wind power output using local terrain, humidity, temperature and other environmental and meteorological information through complex computation. Statistical models, mainly including Persistence, ARMA, ARIMA and Kalman filter, make wind speed forecasting by capturing the internal relationship among the historical wind speed time series. Artificial intelligence and machine learning engines, mainly including multilayer perceptron (MLP) (Aghajani et al. 2016; Bouzgou and Benoudjit 2011), least squares support vector machine (LSSVM) (Yan et al. 2018; Hu et al. 2015; Kumar et al. 2018), extreme learning machine (ELM) (Zhang et al. 2014, 2017; Sun et al. 2018a), back-propagation neural network (BPNN) (Wang et al. 2016), wavelet neural networks (WNN) (Sun et al. 2019), Elman neural network (ELMNN) (Du et al. 2017) and support vector regression (SVR) (Santamaria-Bonfil et al. 2016), have good capacity in processing nonlinear components within wind speed time series and are widely employed in wind speed forecasting; however, these approaches make forecasting in implicit ways.

In order to enhance the forecasting results, technicians and researchers always apply hybrid model which integrates and takes advantages of various individual approaches. For example, studies (Dong et al. 2017; Wang et al. 2016, 2017; Santamaria-Bonfil et al. 2016; Osorio et al. 2015) combine artificial intelligent models with signal decomposition and parameter optimization for wind speed forecasting. The case studies in the above literature illustrate that the hybrid models are successfully applied in the wind speed forecasting with satisfactory accuracy.

Among the aforementioned artificial intelligent models and machine learning engines, LSSVM is a powerful tool in dealing with small nonlinearity samples, and it has been widely used in system modeling (Stephen et al. 2014), time series prediction (Yan et al. 2018; Hu et al. 2015; Kumar et al. 2018; Sun et al. 2018b) and fault diagnosis (Zheng et al. 2011). The characteristics of LSSVM are to solve a quadratic programming problem by translating into linear equation, which can improve computational convergence. Thus, LSSVM is also adopted as the core forecasting engine in this study for wind speed forecasting.

LSSVM has better regression performance over the other ANN-based models and SVM (Hu et al. 2015); however, the regression performance of LSSVM is influenced by its configuration which includes the input training samples, the kind of kernel function and the kernel parameter. Therefore, researchers try to enhance the performance of LSSVM in short-term wind speed forecasting from these three aspects.

Yuan et al. (2015) selected a suitable kernel function through comparisons and applied GSA to optimize the kernel parameters. In Zhou et al. (2011), LSSVM with linear, polynomial (Poly) or Gaussian kernel function was separately constructed to make wind speed forecasting for finding which kernel function performs best. In Wang et al. (2015a), C–C method is firstly used to process the wind speed time series for phase space reconstruction, which automatically determine the input form for subsequent forecasting; then, the parameters in LSSVM are tuned by PSO-GSA and the outputs of LSSVM are corrected by Markov method.

However, there are still some shortages in LSSVM-based wind speed forecasting models.

- i: The existing studies on the application of LSSVM in wind speed forecasting are very limited because only one single particular kernel function (Yuan et al. 2015; Zhou et al. 2011; Wang et al. 2015a) or double kernel functions (Sun et al. 2018b) are considered. As for each kernel function has its characteristics in data process, these methods are not comprehensive.
- ii: Wavelet transform (WT) (Yan et al. 2018; Liu et al. 2014), wavelet packet decomposition (WPD) (Sun et al. 2018a), ensemble empirical mode decomposition (EEMD) (Wang et al. 2016; Sun et al. 2018b), empirical mode decomposition (EMD) (Zhang et al. 2016a), complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) (Peng et al. 2017) and variational mode decomposition (VMD) (Zhang et al. 2017; Naik et al. 2018) are commonly wind speed data preprocessing methods. In the previous most wind speed forecasting models, only single individual data preprocessing method is applied to decompose wind speed data, which cannot thoroughly deal with wind speed in that wind speed time series always present intermittence and instability (Sun et al. 2019; Yin et al. 2017).
- iii: Abnormal data or noise within wind speed time series is caused by malfunctioning sensors, measurement error or other factors, which is considered as an obstacle to obtain high forecasting accuracy (Liu et al. 2018).

To address the above problems, the concept of combined model is utilized to develop a new hybrid short-term wind speed forecasting method that is inspired by the forecasting mechanisms in the aforementioned studies.

- i: LSSVM model with a combination of linear, Poly and radial basis function (RBF) kernel functions by optimal weighted coefficients is developed as core forecasting engine, namely multiple kernel functions LSSVM (MKLSSVM), in the wind speed forecasting model. RBF has good local exploitation and Poly has good global exploration capacities in dealing with nonlinear data, and

linear kernel function can do well in linear signal processing, which can enhance forecasting performance.

- ii: A novel two-stage wind speed decomposition combining CEEMDAN with WT is utilized for data preprocessing. CEEMDAN is firstly utilized to decompose the original wind speed data into a few IMFs and one residual component with different frequencies; then, WT method is applied to break the highest frequency IMF1 into different subseries, which further lower the regression difficulties of LSSVM.
- iii: To eliminate the abnormal data, coral reefs optimization (CRO) was developed in Salcedo-Sanz et al. (2014) as feature selection to identify the effective input candidates for the forecasting engine ELM. A deep feature selection framework was developed to identify the most suitable candidates from the testing sample for machine learning models in the first layer (Feng et al. 2017). Apart from applying feature selection and parameter optimization on input candidates and artificial intelligent models, respectively, hybrid optimization algorithms have been exploited to realize these functions synchronously. In Luo et al. (2016), hybrid gravitational search algorithm (HGSA) integrating conventional GSA and binary-valued GSA (BGSA) were developed to realize the fault diagnosis of rolling element bearing and optimize the weights and bias parameters in ELM. In Zhang et al. (2017), ELM combined with hybrid backtracking search algorithm (HBSA) was developed for short-term wind speed forecasting, and the HBSA algorithm composes of real-valued BSA (RBSA) and binary-valued BSA (BBSA), which was also exploited to realize the function of feature selection and optimization. Hybrid PSO GSA (HPSOGSA) integrating conventional PSO GSA and binary-valued PSO GSA (BPSOGSA) is introduced to enhance the forecasting performance of MKLSSVM. HPSOGSA extracts the effective candidates from the testing samples and optimizes the parameter combination in MKLSSVM, simultaneously.

Apart from the above introduction, the other parts of this paper are arranged as follows. Section 2 provides the methodology that used in the proposed model. The working mechanism of the proposed forecasting strategy is presented in Sect. 3. Case studies that verify the effectiveness of the proposed forecasting model are carried out in Sect. 4. Conclusions are drawn in Sect. 5.

2 Methodology

2.1 Wind power preprocessing method

2.1.1 Complementary ensemble empirical mode decomposition with adaptive noise

The working mechanism of EMD is to decompose the nonlinear and nonstationary complicated signal into a few relatively stable intrinsic mode functions (IMFs) and one residual (Res) with different frequencies using Hilbert–Huang transform (HHT) approach (Zhang et al. 2016b). EMD method may not correctly extract effective characteristic information of signal in that it suffers from the drawback of the mode mixing.

To solve this problem, Wu and Huang proposed a noise-assisted EMD method (Wu and Huang 2009), namely EEMD. In EEMD, white Gaussian noise is added in the original signal at each shifting procedure for alleviation of the mode mixing in EMD and these white noise signals are eliminated by averaging in the end. However, there exists some white noise that may not be discarded after a finite shifting process, which affects signal reconstruction errors. To overcome this shortage, Torres et al. (2011) proposed a new complete EEMD with adaptive noise (CEEMDAN), which has been successfully applied in complicated time series signal processing (Peng et al. 2017) and fault feature extraction (Han et al. 2019). The decomposition process of any signal $x(t)$ by CEEMDAN can be presented as follows.

- Step 1: Add a number of white Gaussian noise $z(t)$ with $N(0, 1)$ into the original signal, which is expressed as Eq. (1).

$$y(t) = x(t) + \eta_0 z(t) \quad (1)$$

where η_0 denotes a noise coefficient;

- Step 2: Decompose the $y(t)$ by EMD method to obtain the corresponding IMF_1^i .
- Step 3: Calculate the first subseries of CEEMDAN by meaning the IMFs obtained by EMD approach, expressed as Eq. (2).

$$\text{IMF}_1^i(t) = \frac{1}{L} \sum_{i=1}^L \text{IMF}_1^i \quad (2)$$

- Step 4: Calculate the residual value $r_1(t) = x(t) - \text{IMF}_1^i(t)$
- Step 5: Calculate the first subseries of CEEMDAN by meaning the IMFs obtained by EMD approach again,

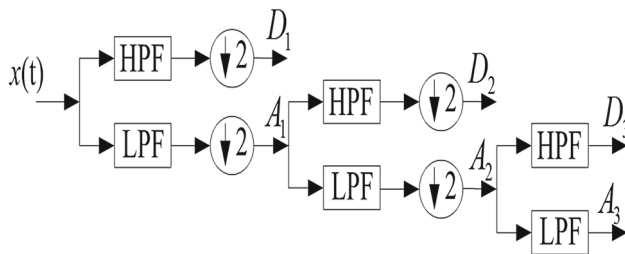


Fig. 1 Three-level decomposition of signal by WT

expressed as Eq. (3).

$$\text{IMF}_2^i(t) = \frac{1}{L} \sum_{i=1}^L E_1(r_1(t) + \eta_1 E_1(z^i(t))) \quad (3)$$

where $E_1(\cdot)$ represents the empirical mode decomposition EMD method.

Step 6: Repeat the above steps to yield the other IMFs until the residual component has no more than two extreme values.

2.1.2 Wavelet transform (WT)

WT, a mathematical method, has been widely used in the time series signal processing (Aghajani et al. 2016; Osorio et al. 2015). WT can extract the effective information in the sample signal without losing information, and it has the ability to capture both frequency and location information in a simultaneous manner. Low-resolution wavelets can approximately capture frequency components of the sample signal, while high-resolution wavelets can catch the high-frequency components. WT can be mainly divided into continuous WT (CWT) and discrete WT (DWT). For overlapping feature and duplicity of neighbor information, CWT is very slow; thus, DWT is adopted in this study to deal with the wind speed data. As shown in Eq. (4), DWT utilizes scale and position parameters by powers of two, which are named as dyadic dilation and translations.

$$\text{DWT}(m, n) = 2^{-\frac{m}{2}} \sum_{t=0}^{L-1} x(t) \psi \left(\frac{t - n \cdot 2^m}{2^m} \right) \quad (4)$$

where L denotes the total length of the sample signal $x(t)$, m and n are two integrate number, 2^m and $n \cdot 2^m$ are factor and shifted values, respectively. Three-level decomposition of sample signal by WT is illustrated in Fig. 1.

As shown in Fig. 1, the sample signal is broken into two components, namely approximation component A_i and detail component D_i , using low-pass filter (LPF) and high-pass filter (HPF), respectively. In the decomposition procedure, the sample signal is decomposed into high-frequency

component D_i and low-frequency component A_i ; then, the low-frequency component A_i is continuously decomposed, while the high-frequency component D_i is maintained.

The kind of the mother wavelet function plays great influence on the performance of WT (Haque et al. 2013). Among the common mother wavelet functions, including Morlet, Haar, Mexican Hat and Meyer wavelet functions, Daubechies of order 4 (DB4) can usually provide better results; thus, DB4 is adopted as the wavelet function in this study.

2.2 Particle swarm optimization gravitational search algorithm (PSOGSA)

2.2.1 Particle swarm optimization (PSO)

PSO is an optimization algorithm that determines the optimal solution by simulating the social behavior of birds swarm. A number of particles in PSO, standing for birds in the real world, look for the optimal solution in the region by adjusting their speed and location. PSO can be expressed as Eq. (5).

$$\begin{cases} v_i^{t+1} = \omega v_i^t + c_1 \times r_1 \times (p_{\text{best}i} - x_i^t) + c_2 \times r_2 \times (g_{\text{best}} - x_i^t) \\ x_i^{t+1} = x_i^t + v_i^{t+1} \end{cases} \quad (5)$$

where v_i^t and v_i^{t+1} are the velocity of the i th particle at t th and $(t+1)$ th iteration, respectively. ω denotes an inertial weight. c_1 and c_2 are learning factors; r_1 and r_2 stand for random number within $(0,1)$; and x_i^t and x_i^{t+1} are the position of the i th particle at t th and $(t+1)$ th iteration, respectively.

$$p_{\text{best}i}^t = \begin{cases} p_{\text{best}i}^{t-1} & \text{if } (f(x_i^t) \geq f(p_{\text{best}i}^{t-1})) \\ x_i^t(t) & \text{if } (f(x_i^t) < f(p_{\text{best}i}^{t-1})) \end{cases} \quad (6)$$

$$g_{\text{best}} = \arg \min_{p_{\text{best}i}^t} f(p_{\text{best}i}^t) \quad (7)$$

The ωv_i^t in Eq. (5) can offer good exploration capacity for PSO, while $c_1 * r_1 \times (p_{\text{best}i} - x_i^t) + c_2 \times r_2 \times (g_{\text{best}} - x_i^t)$ provides historical memory.

2.2.2 Gravitational search algorithm (GSA)

Inspired by Newton theory, a new heuristic optimization algorithm GSA was firstly proposed by Rashedi et al. (2009). Each agent in GSA stands for candidate solution to the objective function and is considered as an object with masses proportional to the values of the fitness function. In the iteration procedure, all agents obey laws of gravity and motion to attract each other and move.

It is assumed that a system with N agents and n dimensions is shown in Eq. (8) which scatter randomly in the search

space.

$$X_i = (x_i^1, x_i^2, \dots, x_i^d, \dots, x_i^N) \quad i = 1, 2, \dots, N \tag{8}$$

At t th iteration, the gravitational force $F_{ij}^d(t)$ of the i th agent acting on the j th agent is expressed as Eq. (9).

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij} + \varepsilon} (x_j^d(t) - x_i^d(t)) \tag{9}$$

where $M_{aj}(t)$ and $M_{pi}(t)$ are active mass and passive mass, respectively. ε denotes a small constant. $G(t)$ and R_{ij} are expressed as Eqs. (10) and (11), respectively.

$$G(t) = G_0 \times \exp(-\alpha \times iter/iter_{max}) \tag{10}$$

$$R_{ij} = \|X_i(t), X_j(t)\|_2 \tag{11}$$

where G_0 and α are the initial gravitational constant and the descending coefficient, respectively. $iter$ and $iter_{max}$ are the current iteration and the maximum iteration, respectively.

At t th iteration, the accelerated speed of i th agent with d dimension can be expressed Eq. (12).

$$a_i^d(t) = \frac{\sum_{j=1, j \neq i}^N r_j \times F_{ij}^d(t)}{M_i(t)} \tag{12}$$

where r_j is a random value within (0,1), and $M_i(t)$ is expressed as Eq. (13).

$$\begin{cases} m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \\ M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)} \end{cases} \tag{13}$$

where $fit(t)$, $best(t)$ and $worst(t)$ stand for the value, minimum value and maximum value of the t th agent fitness function, respectively.

In the end, the speed and location of the i th agent are mathematically expressed as Eq. (14).

$$\begin{cases} v_i^d(t+1) = rand \times v_i^d(t) + a_i^d(t) \\ x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \end{cases} \tag{14}$$

where $rand$ stands for a random number within (0,1).

2.2.3 Work principle of PSO GSA

To combine the exploitation ability of GSA with the exploration capacity of PSO, the $p_{besti} - x_i(t)$ in Eq. (5) is replaced by the acceleration $a_i(t)$ in Eq. (12) to construct a new PSO GSA hybrid algorithm; thus, PSO GSA is expressed as Eq. (15) (Zheng et al. 2017; Wang et al. 2015b). Like PSO

algorithm, PSO GSA updates its velocity and position by considering not only current state but also previous values.

$$v_i(t+1) = \omega \times v_i(t) + c_1 \times r_1 \times a_i(t) + c_2 \times r_2 \times (g_{best} - x_i(t)) \tag{15}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{16}$$

2.2.4 Binary PSO GSA

The traditional PSO GSA is a powerful organization algorithm in solving the real continuous problems; however, there exist binary problems that cannot be solved by the traditional PSO GSA. To solve these binary problems, Mirjalili and Hashimx (2012) proposed a binary-valued PSO GSA, namely BPSO GSA. In BPSO GSA algorithm, the velocity of each agent in PSO GSA is transformed into a probability function within (0,1) using the Hyperbolic tangent function, which is expressed as Eq. (17).

$$S(v_i^d(t)) = \left| \tanh(v_i^d(t)) \right| \tag{17}$$

where $\tanh(\cdot)$ represents hyperbolic tangent function. Then, the position of each agent is updated as Eq. (18).

$$\begin{cases} x_{i,j}^d(t+1) = complement(x_{i,j}^d(t)) \text{ rand} < S(x_{i,j}^d(t+1)) \\ x_{i,j}^d(t+1) = x_{i,j}^d(t) \text{ else} \end{cases} \tag{18}$$

where $complement(\cdot)$ stands for the logical negation function.

2.3 Multi-kernel function least squares support vector machine (MKLSSVM)

2.3.1 Working principle of LSSVM

LSSVM is developed on the basis of SVM by using square errors in the cost function and equality constraints instead of nonnegative errors and inequality constraints. As a result, LSSVM regresses by dealing with a linear system rather than quadratic programming problems (Kumar et al. 2018). In recent years, LSSVM has been successfully applied in time series prediction (Yan et al. 2018), fault diagnosis (Kumar et al. 2018) and system modeling (Hemmati-Sarapardeha et al. 2018).

It is assumed training samples $\{(x_i, y_i) | x_i \in R^d, y_i \in R\}$, where $i = 1, 2, \dots, N, N$ and d are the number of the total samples and the dimension of input variables, respectively. x_i and y_i are the input and output vectors, respectively. Using a nonlinear function $\varphi(x)$, the input vectors are mapped into a high-dimensional feature space Z where the regression

proression is carried out. The regression processions can be mathematically expressed as Eq. (19).

$$y = \omega^T \varphi(x) + b \tag{19}$$

where ω and b stand for the weight vector and bias term, respectively, which are estimated in the subsequent calculations in Eqs. (20) and (21).

$$C = 0.5\omega^T \omega + 0.5\gamma \sum_{i=1}^N e_i^2 \tag{20}$$

s.t.

$$y_i = \omega^T \varphi(x_i) + b + e_i \tag{21}$$

where $\gamma > 0$ is the penalty factor, e_i denotes an error.

To simply the calculation, Lagrange function is constructed as Eq. (22).

$$L(\omega, b, e, \alpha) = \frac{1}{2} \left(\|\omega\|^2 + \sum_{i=1}^N e_i^2 \right) - \sum_{i=1}^N \alpha_i (\omega^T \varphi(x_i) + b + e_i - y_i) \tag{22}$$

where α_i is the Lagrange multipliers.

Through partially differentiating ω, e, b and α , and eliminating ω and e , the regression results are expressed as Eq. (23).

$$y(x) = \omega^T \varphi(x) + b = \sum_{i=1}^N \alpha_i k(x, x_i) + b \tag{23}$$

where $k(x, x_i)$ stands for a positive definite kernel function.

2.3.2 Selection of kernel function

It has been proved that the regression performance and generalization capacity of LSSVM highly depend on the kind of kernel function and its parameters (Sun et al. 2018b). Among the common kernel functions, radial basis function (RBF) has good local exploitation capacity, while polynomial function (Poly) possesses excellent global exploitation ability, which has good capacity in processing nonlinear signal. To enhance the forecasting performance, a weighted multi-kernel function based on RBF, Poly and linear kernel functions is constructed for LSSVM, namely MKLSSVM, which takes advantages of individual approaches. The form of multi-kernel function can be expressed as Eq. (24).

$$k(x_i, x_j) = \mu_1 k_1(x_i, x_j) + \mu_2 k_2(x_i, x_j) + \mu_3 k_3(x_i, x_j) \tag{24}$$

where k_1, k_2 and k_3 stand for RBF, Poly and linear kernel function, respectively. μ_i denotes weight coefficient within (0, 1) and $\sum \mu_i = 1$. RBF, Poly and linear kernel functions are mathematically expressed as Eqs. (25), (26) and (27), respectively.

$$k_{\text{RBF}}(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\delta^2} \right) \tag{25}$$

$$k_{\text{Poly}}(x_i, x_j) = [(x_i \cdot x_j) + 1]^d \tag{26}$$

$$k_{\text{line}}(x_i, x_j) = x_i \cdot x_j \tag{27}$$

Thus, the outputs of MKLSSVM can be modified as Eq. (28).

$$y = \sum_{i=1}^N \alpha_i \left[\sum_{j=1}^3 \mu_j k_j(x_i, x) \right] + b \tag{28}$$

By tuning μ_j value, multi-kernel function exhibits optimal characteristics for different input variables. Although wind speed time series exhibits high nonlinearity, sometimes there exist some certain linear components (Li et al. 2018). The weighted combination of RBF, Poly and linear kernel functions not only processes the nonlinear components, but also solves the linear problems within wind speed data.

3 The working mechanism of the proposed forecasting strategy

3.1 Evaluation index for WSF

To illustrate the forecasting model with the best prediction performance, the definitions of four statistical indices, including RMSE, MAE, MAPE and MASE, are shown in Table 1 and utilized to evaluate and compare different forecasting approaches. Smaller MAE, RMSE, MAPE or MASE values mean that the prediction values are closer to the actual measured wind speed.

3.2 The proposed HPSOGSA-MKLSSVM

In the proposed HPSOGSA-MKLSSVM model, PSOGSA is employed to optimize the kernel parameters and weighted coefficients, which can make MKLSSVM avoid trapping in over-fitting or local optima. For measurement errors and malfunctioning sensors, there exists abnormal noise in the wind speed time series (Liu et al. 2018). To eliminate these ineffective components, BPSOGSA approach is exploited as feature selection to identify the abnormal data by encoding binary value “0” or “1” (“0” and “1” represent “discarded” and “selected”, respectively). The feature selection for selecting effective input candidates

Table 1 Forecasting indices and their definitions

Metrics	Formula	Definition
RMSE	$\sqrt{\frac{1}{N} \sum_{i=1}^N (v(i) - \hat{v}(i))^2}$	Root mean square error
MAE	$\frac{1}{N} \sum_{i=1}^N v(i) - \hat{v}(i) $	Mean absolute error
MAPE	$\frac{1}{N} \sum_{i=1}^N \frac{ v(i) - \hat{v}(i) }{v(i)} \times 100\%$	Mean absolute percent error
MASE	$\frac{1}{N} \sum_{i=1}^N \left(\frac{ v(i) - \hat{v}(i) }{\frac{1}{N-1} \sum_{i=2}^N v(i) - \hat{v}(i) } \right)$	Normalized mean absolute scaled error

$v(i)$ and $\hat{v}(i)$ stand for the actual measured wind speed data and the forecasting wind speed values, respectively. N means the quality of the testing samples

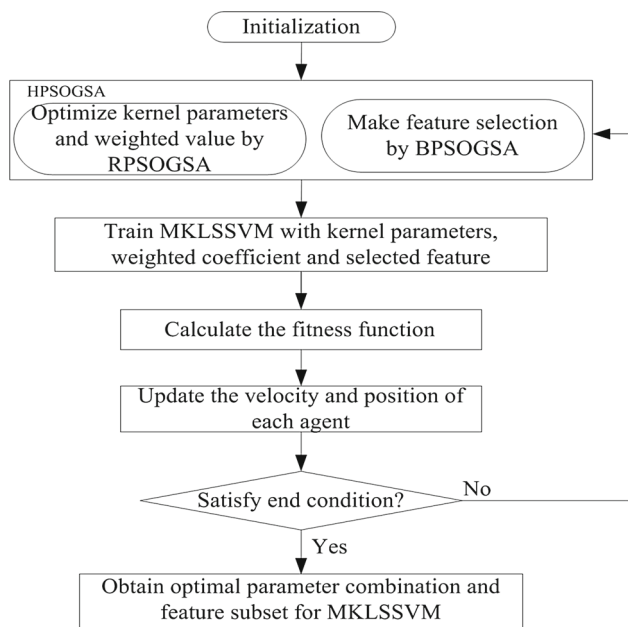


Fig. 2 Feature selection and parameter optimization for MKLSSVM by HPSOGSA

and parameter optimization for optimizing the parameter combination in MKLSSVM are carried out by HPSOGSA simultaneously. In this paper, RMSE is applied as the fitness function to evaluate the forecasting results. The specific working steps of HPSOGSA-MKLSSVM model for WSF are illustrated in Fig. 2, and descriptions are shown Algorithm 1.

$$\text{dim} = [B_1, B_2, \dots, B_l, \gamma, \delta, d, \mu_1, \mu_2, \mu_3]. \tag{29}$$

Algorithm 1 The pseudocode of HPSOGSA-MKLSSVM

- 1: Input: parameters including particle number N_1 dimension number D , maximum iteration number N_2 , gravitational constant G_0 , learning factors c_1 and c_2 , inertial weight ω and so on.
- 2: Output: optimal kernel parameter and weighted coefficients, and effective input candidates.
- 3: Construct the candidate feature set for some wind speed subseries.
- 4: Apply BPSOGSA algorithm to make feature selection.
- 5: Reconstruct the training and testing variables matrix after feature selection.
- 6: **for** $i = 1 : N_1$ **do**
- 7: Apply the proposed hybrid model to make wind speed forecasting using each population.
- 8: Compute the fitness function values according to RMSE.
- 9: **end for**
- 10: Determine the initial personal solution and global solution.
- 11: **for** $i = 1 : N_2$ **do**
- 12: Calculate the inertial mass $M_i(t)$ according to Eq. (13).
- 13: Calculate the force $F_i(t)$ according to Eq. (9).
- 14: Update acceleration and speed according to (12) and (15), respectively.
- 15: Update position according to Eq. (16).
- 16: Determine the personal solution and global solution.
- 17: Apply the proposed hybrid model to make wind speed forecasting and calculate the fitness function values.
- 18: Apply BPSOGSA algorithm to make feature selection and reconstruct the training candidates.
- 19: **end for**
- 20: Obtain the best global solution and effective candidates.

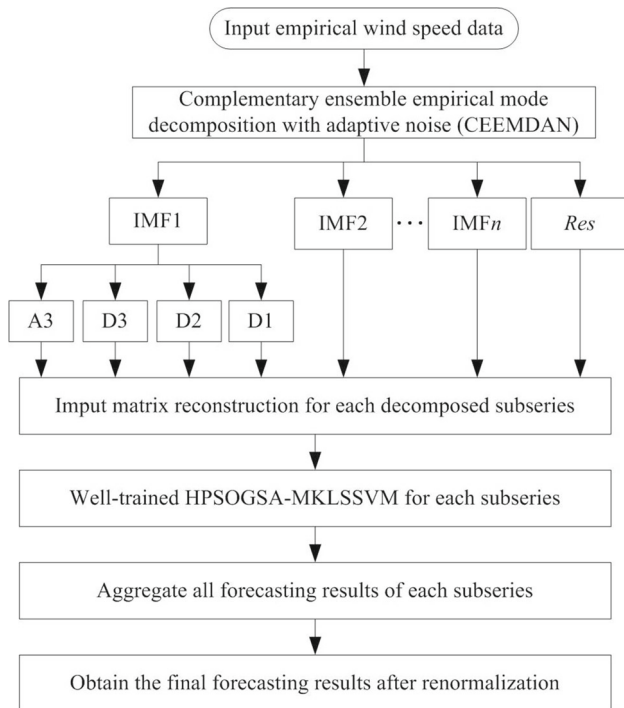
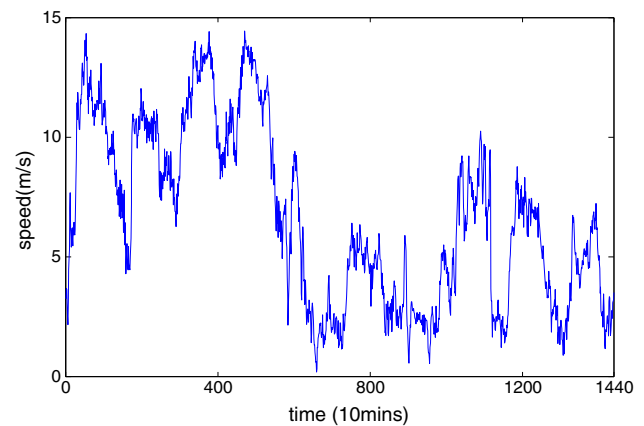


Fig. 3 Framework of the TSD-HPSOGSA-MKLSSVM model

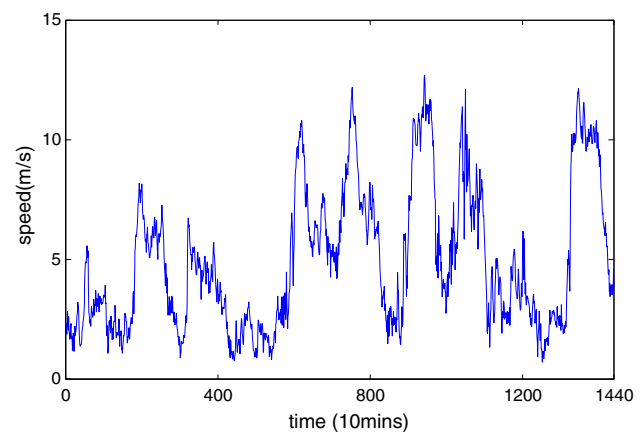
3.3 The working mechanism of the proposed hybrid forecasting model

The forecasting flowchart of the proposed TSD-HPSOGSA-MKLSSVM model is illustrated in Fig. 3, which can be divided into four stages.

- (i) *Stage I* Wind speed preprocessing. Employ CEEMDAN method to break the original empirical wind speed series into several IMFs and one residual (Res), which can relieve the degree of nonlinearity and fluctuation of wind speed time series. Employ WT approach to further decompose the highest frequency component IMF1 into a few subseries.
- (ii) *Stage II* Input matrix reconstruction. If the PACF of the i th subseries at lags bigger than parameter p are approximately independent $N(0, 1/n)$ random variables, the correlation coefficients among input values are obtained as p ; thus, the dimensions of the input variables are determined as p . For the i th subseries $X_i(t)$, the input variables of MKLSSVM can be represented as $X_i = \{x_i(t-1), x_i(t-2), \dots, x_i(t-p)\}$, which are utilized to predict the corresponding output of MKLSSVM. To lower the forecasting difficulties, all the subseries are linearly translated into the range $[0,1]$.
- (iii) *Stage III* HPSOGSA-MKLSSVM training. The confirmed input variables are divided into two parts, the first 1st to 1296th are utilized to train the HPSOGSA-



(a) Wind speed data set A



(b) Wind speed data set B

Fig. 4 Original wind speed time series

- MKLSSVM model and subsequent 1297th to 1440th are applied to test the model. To enhance the forecasting performance, artificial intelligent hybrid algorithm PSO-GSA is employed to tune the parameter combination in MKLSSVM, and BPSOGSA is used to identify the effective candidates and discard the abnormal data within the input candidate matrix determined by PACF.
- (iv) *Stage IV* Wind speed forecasting. The well-trained HPSOGSA-MKLSSVM model is employed to forecast short-term wind speed in different predicting steps. The outputs of MKLSSVM for each decomposed subseries are aggregated after denormalization, and the final forecasting results are obtained. In the end, comparisons between the proposed hybrid model with other forecasting models are carried out using statistical indices.

Table 2 Statistical description of the empirical wind speed data (m/s)

Wind power data	Max	Mean	Min	SD
Set A Entire samples	14.4418	6.6498	0.1868	3.6878
Training samples	14.4418	6.9515	0.1868	3.7268
Testing samples	7.2324	3.9343	0.8860	1.6797
Set B Entire samples	12.6914	4.9476	0.7104	2.9365
Training samples	12.6914	4.6529	0.7104	2.7476
Testing samples	12.1467	7.5996	1.9099	3.2463

4 Case studies

4.1 Description of the empirical wind speed data

This study develops a hybrid model using the historical wind speed data to forecast the subsequent future short-term wind speed. The wind turbines are installed on the top of the moun-

tain with around 120 m. Two sets of empirical historical wind speed data are collected from the Xuzhou wind farm in Jiangsu of China, illustrated in Fig. 4, to evaluate the proposed hybrid forecasting model. These wind speed data are measured every 10 min. The first 1st to 1296th continuous wind speed data are used to train the proposed model, and the subsequent 1297th to the 1440th wind speed data are utilized to test the proposed model. The statistical descriptions of the empirical wind speed data are listed in Table 2. As shown in Fig. 4 and Table 2, there exist highly nonlinearity and instability in wind speed time series.

4.2 Wind speed data preprocessing

In this study, TSD method is employed to preprocess the original wind speed to eliminate the uncertainty. Firstly, the wind speed is disassembled by CEEMDAN technique into different IMFs and one Res, which are displayed in Fig. 5a, c. Suggested by Torres, etc. Torres et al. (2011), noise stan-

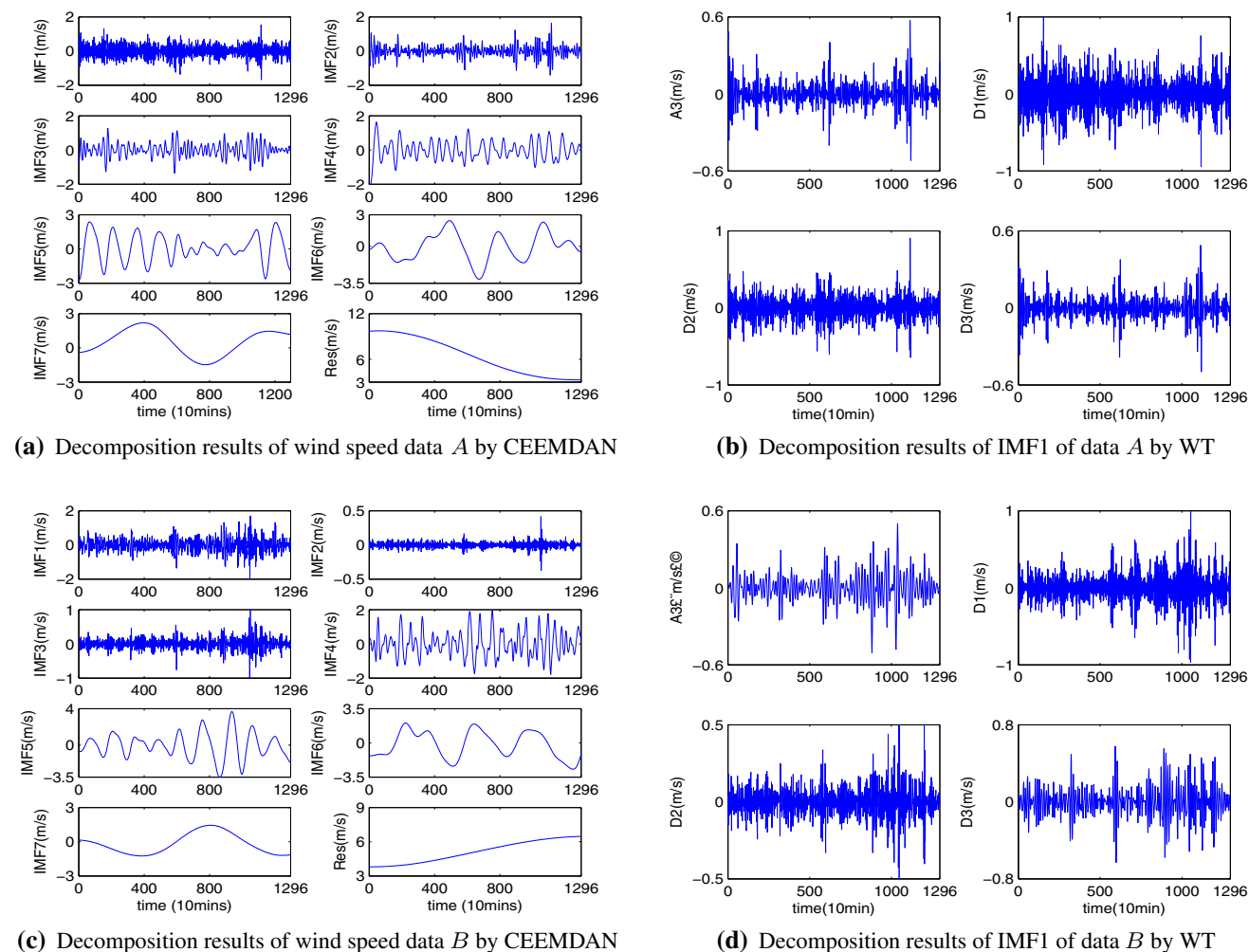
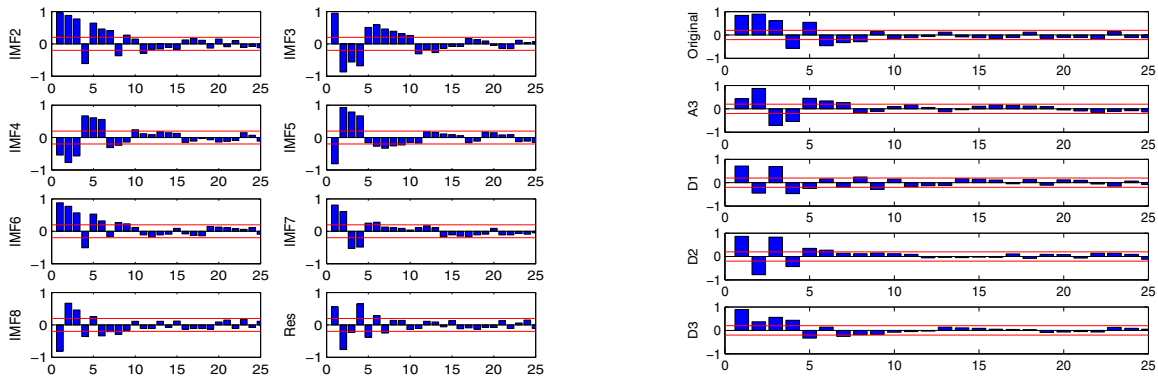
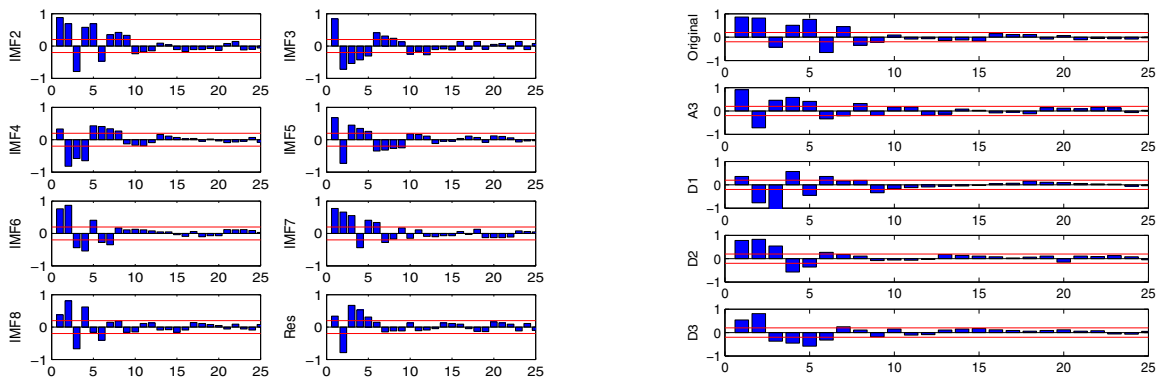


Fig. 5 Decomposition results of the empirical wind speed time series by TSD



(a) PACF values of the decomposed results by CEEMDAN for data set A (b) PACF values of the original wind speed and decomposed results of IMF1 for data set A



(c) PACF values of the decomposed results by CEEMDAN for data set B (d) PACF values of the original wind speed and decomposed results of IMF1 for data set B

Fig. 6 PACF values with 95% confidence

standard deviation ε and ensemble size I are set as 0.02 and 500, respectively. As seen from the figures, the IMF1–IMF3 with higher frequency reveal the nonlinear information of the empirical wind speed. IMF4–IMF7 reflect the periodic information of the original wind speed, while the Res components illustrate the general tendency. WT is applied to further decompose the highest frequency component IMF1 into four sets of relatively subseries at three levels, which is named as A3, D1, D2 and D3. The decomposed subseries are shown in Fig. 5b, d. After decomposition by TSD technique, the decomposed subseries are converted to $[0,1]$, which can reduce the regression difficulties of MKLSSVM (Hu et al. 2015).

4.3 Input matrix reconstruction and parameter selection for HPSOGSA-MKLSSVM

4.3.1 Determination of the input dimension

Prior to submit the decomposed subseries to MKLSSVM for wind speed forecasting, the appropriate dimension of

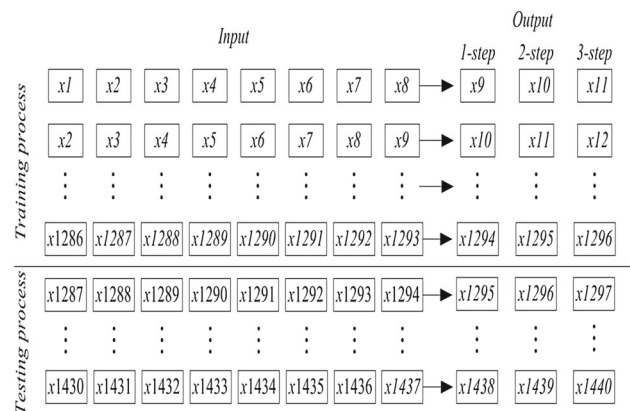


Fig. 7 Inputting and outputting data format for original samples A

the input variables should be determined in advance. In this study, partial autocorrelation function (PACF) is applied to calculate the correlation coefficients among each decomposed subseries, and the correlation coefficients can be considered as the dimension of the input variables for MKLSSVM (Hu et al. 2015; Zhang et al. 2017). The PACF

Table 3 Feature selection results by BPSOGSA

Samples	Time series	Lag order												
		1	2	3	4	5	6	7	8	9	10	11	12	13
A	IMF2	●	●	●	●	○	●	●	○	●	●	●		
	IMF3	●	○	●	●	●	○	●	●	●	●	●	●	●
	IMF4	●	○	○	●	●	●	●	○	●	●			
	IMF5	●	●	●	○	●	●	●	●	●				
	IMF6	●	●	●	○	●	●	●	●	●				
	IMF7	●	●	●	○	○	●							
	IMF8	○	●	○	●	●	●	●	●					
	Res	●	●	○	●	●	●	●						
	A3	○	●	●	○	○	●	●						
	D1	●	●	○	●	●	●	●	●	●				
	D2	●	○	●	●	○	●							
	D3	●	●	○	●	●	○	●						
	Original	●	●	●	○	●	●	○	●					
	B	IMF2	●	○	●	●	○	●	●	●	●	●		
IMF3		●	●	●	○	○	●	●	●	○	●	●	●	
IMF4		○	●	●	○	●	●	●	●					
IMF5		●	○	●	○	●	●	●	○	●				
IMF6		●	○	●	●	●	●	●						
IMF7		●	○	●	●	●	●	○						
IMF8		●	●	○	●	●	●							
Res		●	○	●	○	●								
A3		●	○	○	●	●	●	○	●					
D1		●	●	○	●	●	○	●	●	●				
D2		●	●	●	●	●	●							
D3		●	●	○	●	●	●	●						
Original		●	●	○	●	○	●	●	●	●				

● represents “select,” ○ denotes “discard”

values of the empirical time series for two sets are computed from lags 0 to 25 and the calculated results with 95% confidence interval line are illustrated in Fig. 6. As shown in Fig. 6, the PACF values of original wind speed data A at lags bigger than 8 are within the red horizontal lines, which means that 8 antecedent wind speed time series severely influence the corresponding subsequent multi-step ahead forecasting (Hu et al. 2015; Peng et al. 2017). Thus, the input dimension of the original wind speed for MKLSSVM can be determined as 8, and the input and output data formats are shown in Fig. 7. AS shown in Fig. 7, a rolling forecasting process is adopted in this study. In the 1-step horizontal forecasting, the previous consecutive 8 time series are utilized to predict the 9th point data; then, the forecasting value is used as the historical

time series to forecast the next value. The same forecasting strategies are executed for the multi-step forecasting, and the dimensions of the other time series are determined in the similar way.

4.3.2 Construction of the HPSOGSA-MKLSSVM

As described in Sect. 2.3, the regression performance of MKLSSVM is influenced greatly by the kernel parameters and penalty value. These real parameters are optimized by conventional PSOGSA, while BPSOGSA technique is exploited to identify the effective candidates and discard the abnormal components; these parameter optimization and feature selection are executed simultaneously by HPSOGSA.

Table 4 Parameter selection

Parameters	Values
Iteration maximum number	100
Population size	30
Learning factor c_1	1.3
Learning factor c_2	1.8
Gravitational constant G_0	2
Weighted coefficients μ_1, μ_2, μ_3	(0,1)
Penalty factor γ	(0,200)
Kernel parameters δ, d	(0,20)

The effective input candidates identified by BPSOGSA method are shown in Table 3.

The parameters in the HPSOGSA-MKLSSVM model are set in Table 4, which are tuned by calculating the RMSE values. The inertial weight ω in PSO decreases linearly within (0.1, 0.9), which is expressed as Eq. (30).

$$\omega(t) = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{it_{\max}} \times t \quad (30)$$

where ω_{\max} and ω_{\min} denote constant values 0.9 and 0.1, respectively, and it_{\max} and t stand for maximum iteration number and iteration time, respectively.

4.4 Comparisons and analysis

All the algorithms are carried out in MATLAB 2014a under Windows 8 Operating System environment. The statistical indices including RMSE, MAE, MAPE and MASE are utilized to measure the forecasting error. All the tests are carried out 20 times and the averages are considered as the final forecasting results to eliminate the statistical errors. A systematic investigation is carried out to illustrate how the LSSVM configuration influences its prediction performance in short-term wind speed forecasting. The regression plots of actual wind speed and forecasting values are shown in Fig. 8.

4.4.1 Verification of the effectiveness of TSD

This section mainly investigates the forecasting results of HPSOGSA-MKLSSVM model combining with the further decomposition methods of IMF1-, IMF1-2- and IMF1-3-based TSD method to determine the appropriate quantity of IMFs that should be disassembled, which are given in Table 5. Like IMF1, the subseries IMF2 and IMF3 are also decomposed by WT at three levels. As given in Table 5, the forecasting results of the TSD(IMF1)-

HPSOGSA-MKLSSVM model are better than those of TSD(IMF1-2)-HPSOGSA-MKLSSVM and TSD(IMF1-3)-HPSOGSA-MKLSSVM regardless of 1-step or multi-step forecasting.

Compared with HPSOGSA-MKLSSVM with TSD (IMF1-2), the RMSE errors of the HPSOGSA-MKLSSVM with TSD (IMF1) in 1-step, 2-step and 3-step are reduced by 0.0546 m/s, 0.0409 m/s and 0.0387 m/s for wind speed data A, respectively, 0.0560 m/s, 0.0769 m/s and 0.0811 m/s for data B, respectively. Compared with HPSOGSA-MKLSSVM with TSD (IMF1-3), the RMSE errors of the HPSOGSA-MKLSSVM with TSD (IMF1) in 1-step, 2-step and 3-step are reduced by 0.0619 m/s, 0.0519 m/s and 0.0492 m/s for wind speed data A, respectively; and 0.0708 m/s, 0.0856 m/s and 0.0930 m/s for data B, respectively.

Remark This forecasting results illustrate that the highest frequency components within IMF1 are the main distributing factors for the forecasting accuracy, and further decomposition of IMF2 and IMF3 also influences the forecasting accuracy.

Compared with HPSOGSA-MKLSSVM with CEEM-DAN, the RMSE errors of the proposed model in 1-step, 2-step and 3-step are reduced by 0.0488 m/s, 0.0443 m/s and 0.0468 m/s for wind speed data A, respectively; and 0.05 m/s, 0.0496 m/s and 0.0472 m/s for wind speed data B, respectively.

Remark These forecasting results can be explained that the irregularity of IMF1 with highest frequency is the main factor that influences the forecasting accuracy and the further decomposition by WT can effectively address the irregularity issue of IMF1. Thus, the further decomposition of IMF1 in the proposed TSD-HPSOGSA-MKLSSVM model is appropriate.

4.4.2 Comparisons with other different decomposition methods

To further illustrate the effectiveness of the proposed TSD, the popular wind speed preprocessing methods, including EEMD, EMD, WD and VMD, are employed to combine the HPSOGSA-MKLSSVM model to make multi-step wind speed forecasting. The parameters in EEMD and VMD are set according to Refs. Sun et al. (2018b) and Sun et al. (2019), respectively. The statistical forecasting errors of the decomposition-based model are given in Table 6. As given in Tables 5 and 6, CEEMDAN-based forecasting model outperforms EEMD- and EMD-based forecasting models. For example, compared with the EEMD-based forecasting model, the RMSE values of the proposed CEEMDAN-HPSOGSA-MKLSSVM model in 1-step, 2-step and 3-step are cut by 0.0380 m/s, 0.0235 m/s and 0.0143 m/s for data set

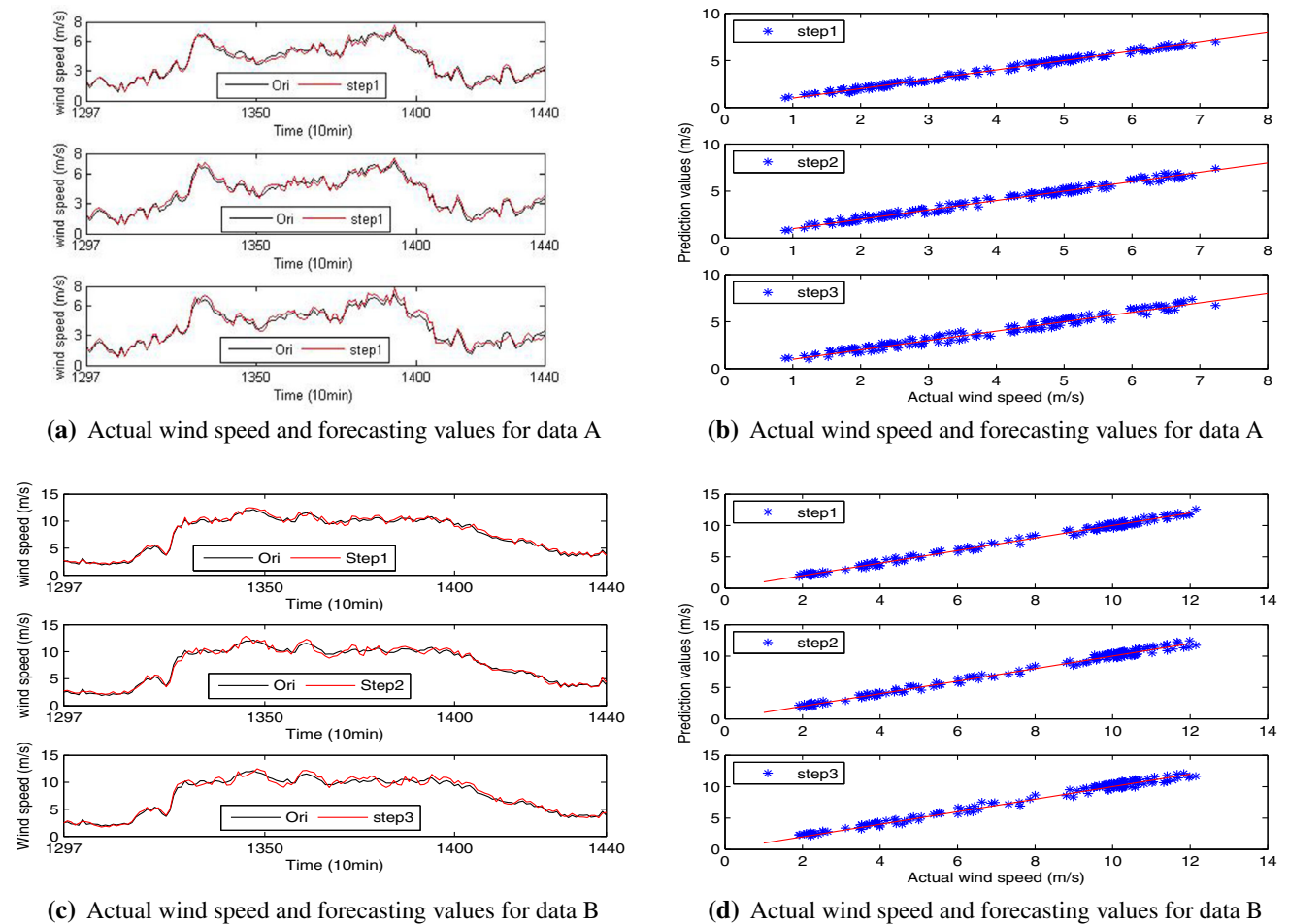


Fig. 8 Forecasting results by the proposed model

Table 5 Forecasting errors of the HPSOGSA-MKLSSVM with CEEMDAN and further decomposition

Data set	Index	CEEMDAN			TSD(IMF1)			TSD(IMF1-2)			TSD(IMF1-3)		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.2098	0.2585	0.3126	0.1824	0.2276	0.2848	0.2161	0.2572	0.3085	0.2246	0.2682	0.3242
	RMSE (m/s)	0.2389	0.2839	0.3463	0.1901	0.2396	0.2995	0.2447	0.2806	0.3383	0.2520	0.2915	0.3488
	MAPE (%)	5.3865	7.7113	8.3878	5.7325	7.0393	8.4897	5.8037	7.6198	8.9515	6.1744	7.8642	9.0707
	MASE	0.5853	0.7211	0.8721	0.5091	0.6350	0.7944	0.6028	0.7176	0.8606	0.6266	0.7481	0.9044
B	MAE (m/s)	0.3461	0.3815	0.4306	0.3006	0.3541	0.3912	0.3467	0.3894	0.4399	0.3622	0.4032	0.4575
	RMSE (m/s)	0.3726	0.4245	0.4682	0.3226	0.3749	0.4210	0.3786	0.4518	0.5021	0.3934	0.4605	0.5140
	MAPE (%)	5.2221	5.6895	6.0462	4.8812	5.5381	6.2350	5.4664	5.7123	5.9435	5.6451	5.9593	6.2433
	MASE	0.9654	1.0641	1.2012	0.8386	0.9877	1.0912	0.9672	1.0864	1.2272	1.0105	1.1249	1.2765

Table 6 Forecasting errors of the HPSOGSA-MKLSSVM with other different decomposition methods

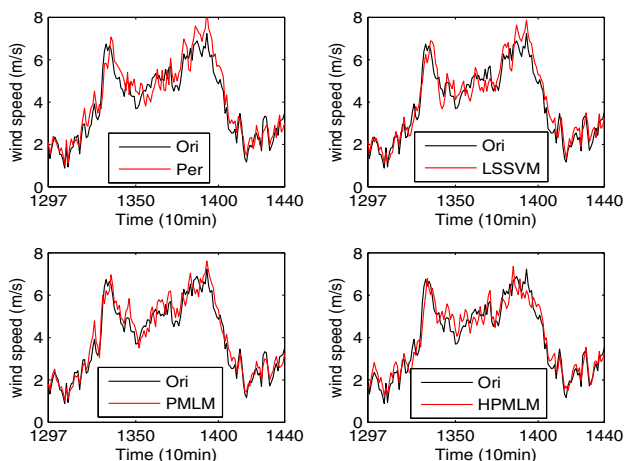
Data set	Index	EEMD			EMD			VMD			WT		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.2430	0.2807	0.3251	0.3237	0.3803	0.4122	0.2584	0.2932	0.3472	0.2928	0.3197	0.3662
	RMSE (m/s)	0.2770	0.3074	0.3606	0.3673	0.4121	0.4461	0.2929	0.3212	0.3852	0.3098	0.3492	0.4070
	MAPE (%)	6.1939	8.3521	8.5985	8.7339	10.6845	12.3206	6.7238	8.6848	9.4817	8.5378	8.9699	9.4549
	MASE	0.6780	0.7833	0.9070	0.9032	1.0611	1.1502	0.7208	0.8181	0.9686	0.8168	0.8918	1.0218
B	MAE (m/s)	0.3734	0.4181	0.4546	0.4107	0.4597	0.5132	0.3919	0.4425	0.5023	0.3935	0.4630	0.5084
	RMSE (m/s)	0.4086	0.4520	0.4907	0.4467	0.5115	0.5687	0.4113	0.4631	0.5190	0.4497	0.5029	0.5491
	MAPE (%)	5.5035	6.1776	6.6562	6.3921	6.8521	7.2277	6.8114	7.2767	8.1191	5.1645	6.7869	7.5144
	MASE	1.0416	1.1661	1.2681	1.1459	1.2826	1.4318	1.0931	1.2343	1.4012	1.0977	1.2916	1.4184

Table 7 Forecasting errors of the TSD-HPSOGSA-LSSVM based on different kernel functions

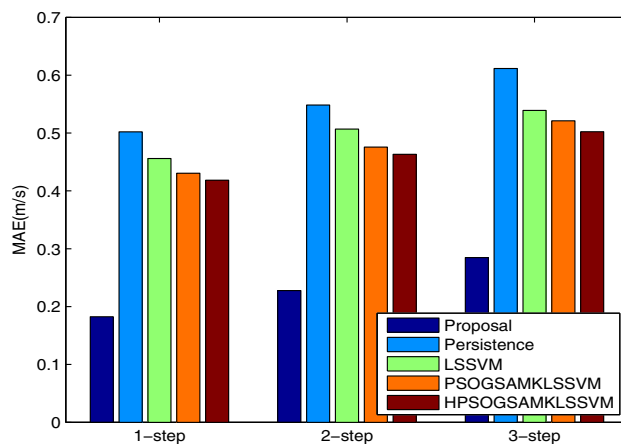
Samples	Index	RBF			Poly			Line			RBF&Poly		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.2287	0.2573	0.3108	0.2395	0.2734	0.3295	0.2775	0.3240	0.3811	0.1976	0.2391	0.2906
	RMSE (m/s)	0.2419	0.2748	0.3464	0.2562	0.2947	0.3724	0.3035	0.3453	0.4029	0.2145	0.2517	0.3213
	MAPE (%)	6.4390	7.2126	8.6492	7.4222	7.9364	8.7343	8.7598	9.6058	11.6353	5.3840	6.9289	8.0023
	MASE	0.6380	0.7180	0.8671	0.6682	0.7626	0.9193	0.7742	0.9040	1.0633	0.5513	0.6669	0.8107
B	MAE (m/s)	0.3512	0.3881	0.4535	0.3515	0.4125	0.4534	0.3999	0.4431	0.4995	0.3315	0.3927	0.4156
	RMSE (m/s)	0.4008	0.4326	0.5052	0.3908	0.4432	0.4916	0.4389	0.4713	0.5415	0.3603	0.4194	0.4661
	MAPE (%)	5.3483	5.6023	6.9271	5.2700	6.0512	6.9994	5.4473	6.5406	7.7502	5.0339	5.7644	6.4775
	MASE	0.9799	1.0828	1.2651	0.9808	1.1508	1.2648	1.1156	1.2363	1.3936	0.9246	1.0954	1.1594

Table 8 Forecasting results of the other single forecasting models (I)

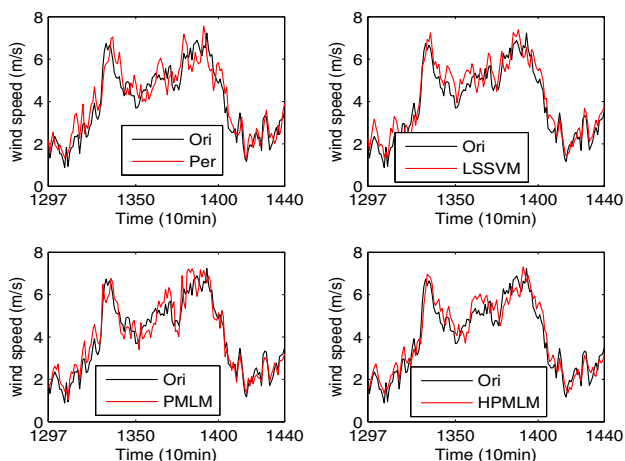
Data set	Index	Persistence			LSSVM(RBF)			CSA-LSSVM		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.5021	0.5484	0.6115	0.4559	0.5069	0.5390	0.4442	0.5014	0.5288
	RMSE (m/s)	0.5587	0.5990	0.6509	0.4982	0.5426	0.5912	0.4813	0.5345	0.5747
	MAPE (%)	13.3756	15.6816	17.7538	12.9821	15.6294	15.2724	12.7226	15.4633	15.0439
	MASE	1.4010	1.5301	1.7062	1.2720	1.4142	1.5037	1.2392	1.3987	1.4754
B	MAE (m/s)	0.6382	0.6568	0.7654	0.5882	0.6417	0.6965	0.5641	0.6282	0.6729
	RMSE (m/s)	0.7045	0.7503	0.8416	0.6342	0.6840	0.7581	0.6019	0.6656	0.7273
	MAPE (%)	9.5897	9.7667	10.8788	9.0856	10.2179	11.0682	8.7597	10.0788	10.8339
	MASE	1.7806	1.8320	2.1351	1.6408	1.7901	1.9429	1.5734	1.7523	1.8771



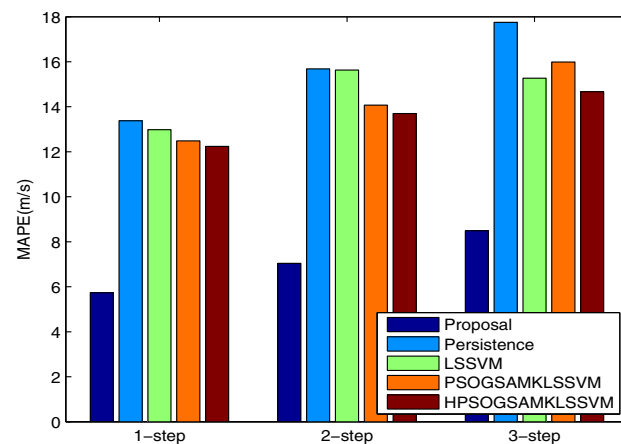
(a) 1-step forecasting results



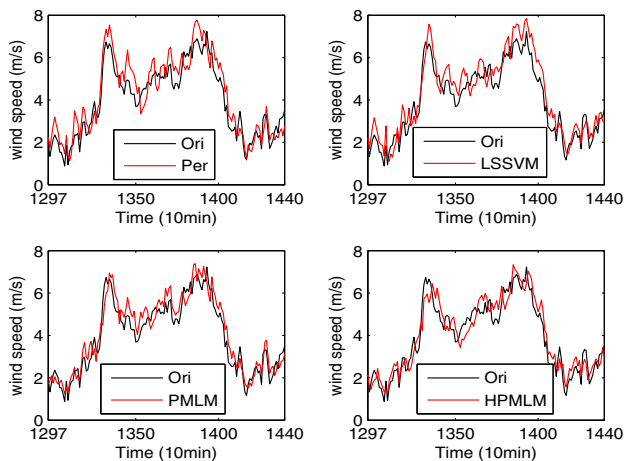
(b) MAE values



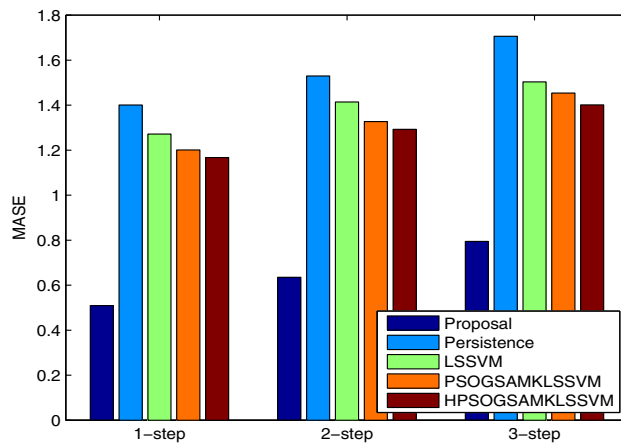
(c) 2-step forecasting results



(d) MAPE values

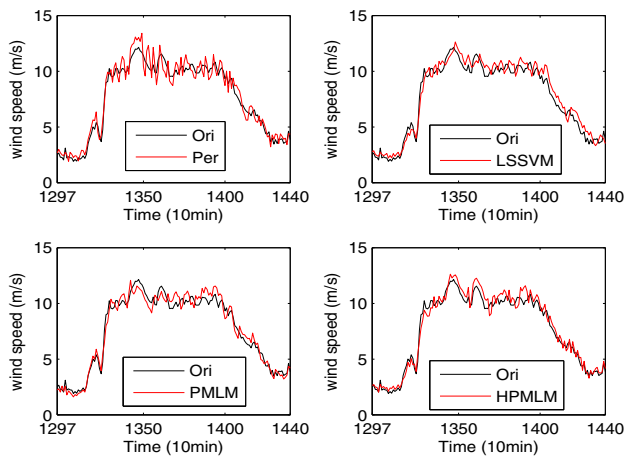


(e) 3-step forecasting results

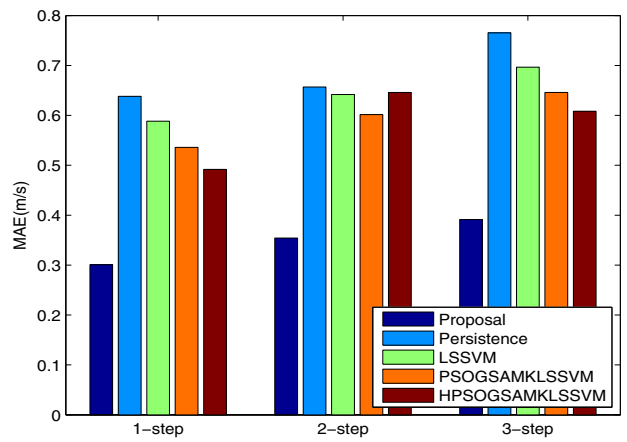


(f) MASE values

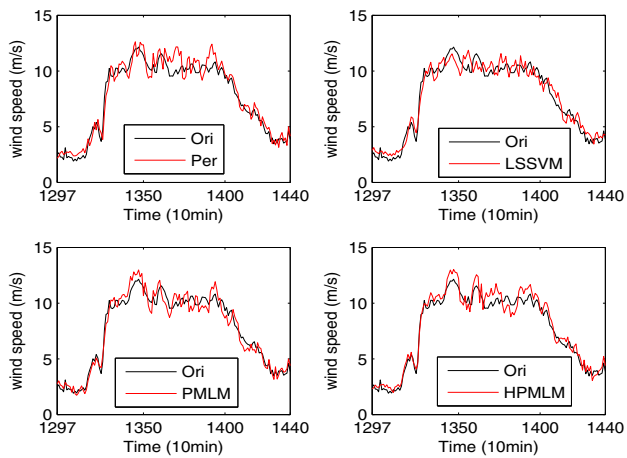
Fig. 9 Forecasting results of the other forecasting models for data set A (ori, Per, PMLM and HPMLM stand for original, Persistence, PSOGSA-MKLSSVM and HPSOGSA-MKLSSVM, respectively. The same definitions are in Fig. 10)



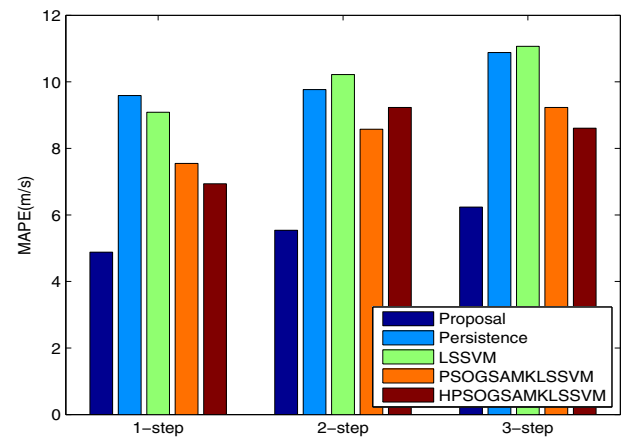
(a) 1-step forecasting results



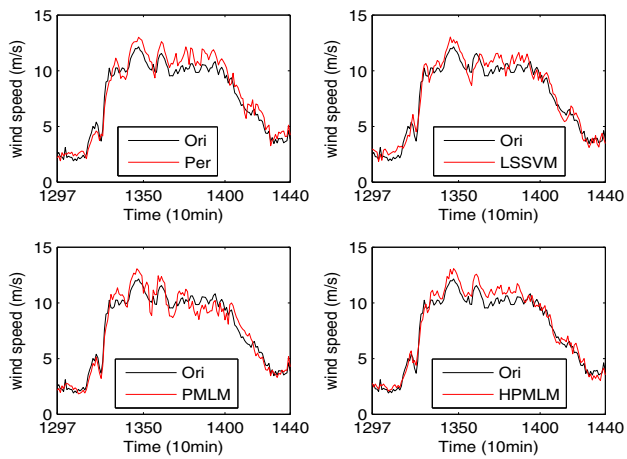
(b) MAE values



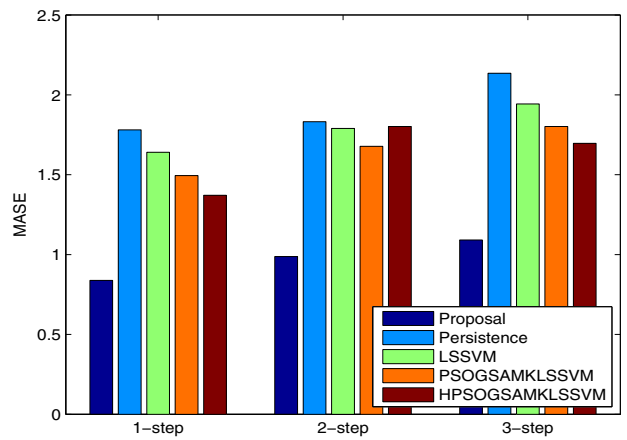
(c) 2-step forecasting results



(d) MAPE values



(e) 3-step forecasting results



(f) MASE values

Fig. 10 Forecasting results of the other forecasting models for data set B

Table 9 Forecasting results of the other single forecasting models (II)

Data set	Index	PSOGSA-MKLSSVM			HPSOGSA-MKLSSVM			Proposal		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.4305	0.4757	0.5212	0.4184	0.4633	0.5024	0.1824	0.2276	0.2848
	RMSE (m/s)	0.4728	0.5283	0.5522	0.4613	0.5044	0.5450	0.1901	0.2396	0.2995
	MAPE (%)	12.4831	14.0740	15.9885	12.2406	13.7004	14.6695	5.7325	7.0393	8.4897
	MASE	1.2009	1.3271	1.4541	1.1673	1.2927	1.4015	0.5091	0.6350	0.7944
B	MAE (m/s)	0.5359	0.6015	0.6459	0.4917	0.5354	0.6083	0.3006	0.3541	0.3912
	RMSE (m/s)	0.5936	0.6598	0.7101	0.5614	0.6334	0.6803	0.3226	0.3749	0.4210
	MAPE (%)	7.5477	8.5755	9.2306	6.9381	6.7162	8.6064	4.8812	5.5381	6.2350
	MASE	1.4948	1.6777	1.8016	1.3714	1.4935	1.6967	0.8386	0.9877	1.0912

A, respectively; and 0.0359 m/s, 0.0274 m/s and 0.0226 m/s for data set B, respectively. Compared with the HPSOGSA-MKLSSVM with EEMD, the RMSE errors of the proposed model for wind speed data A in 1-step, 2-step and 3-step are reduced by 0.0869 m/s, 0.0678 m/s and 0.0610 m/s, respectively; and 0.0869 m/s, 0.0678 m/s and 0.061 m/s for wind speed data B, respectively.

Remark The underlying reasons are that CEEMDAN method effectively resolves the mode mixing problems existing in decomposition method EMD and completely neutralizes the added white Gaussians noise composing of paired positive and negative signals that added to the original samples. HPSOGSA-MKLSSVM with CEEMDAN also has better forecasting results than HPSOGSA-MKLSSVM with WD and VMD in terms of the statistical indices. In terms of the statistical indices listed in the tables, HPSOGSA-MKLSSVM with TSD also outperforms that with VMD, EMD or WT. Thus, TSD is more effective wind speed pre-processing method.

4.4.3 Verification of the effectiveness of multi-kernel function

The main aims of this section are to illustrate the forecasting performance of TSD-HPSOGSA-LSSVM based on multi-kernel function over that based on RBF, Poly, linear or RBF&Poly kernel function. The forecasting results of TSD-HPSOGSA-LSSVM based on different kernel functions are shown in Table 7. As given in Table 7, the LSSVM with weighted RBF&Poly kernel functions performs best in terms of the fitness values RMSE among that with linear, RBF and Poly kernel functions, because RBF&Poly kernel function takes advantages of the local exploitation capacity of RBF function and the global exploration ability of Poly function by optimal weighted coefficients. Overall, the LSSVM with the linear kernel function performs the worst in that there

are a lot of nonlinear information hiding in the wind speed data that the linear kernel function cannot catch. As given by the statistical error indices in Tables 5 and 7, MKLSSVM integrating Poly, RBF and line kernel functions by optimal weighted coefficients has better forecasting accuracy than LSSVM based on the combination of RBF and Poly kernel functions by optimal weighted coefficients, the reasons of which are that there exist some linear information within the wind speed data.

4.4.4 Comparisons of TSD-HPSOGSA-MKLSSVM with Persistence and the other LSSVM-based forecasting models

In this section, Persistence, LSSVM, PSOGSA-MKLSSVM, HPSOGSA-MKLSSVM are constructed to assess the forecasting performance of TSD-HPSOGSA-MKLSSVM using the four statistical indexes and the forecasting results are listed in Tables 8 and 9. The RBF function is adopted as kernel function in LSSVM and its parameters σ and γ are set according to Ref. Yuan et al. (2015) and the parameters in couple simulated annealing (CSA)-LSSVM are set according to Ref. Rostami et al. (2019).

The well-known Persistence, a simple approach for time series prediction, is generally utilized as a benchmark to compare with a new proposed forecasting models (Hu et al. 2015; Wang et al. 2017; Zheng et al. 2017). Persistence employs the current values at time t as the forecasting values at the future time $t+k$, namely $\hat{u}_{t+k} = u_t$ where k denotes the forecasting interval. In this study, the forecasting quality of the proposed hybrid model is also compared with the Persistence method. Based on the forecasting error indices listed in Tables 5 and 8, it can be obtained that the RMSE errors differences between the proposed hybrid model and Persistence are 0.3686 m/s, 0.3593 m/s and 0.3514 m/s in 1-step, 2-step and 3-step for data set A, respectively; and 0.3819 m/s, 0.3753 m/s and

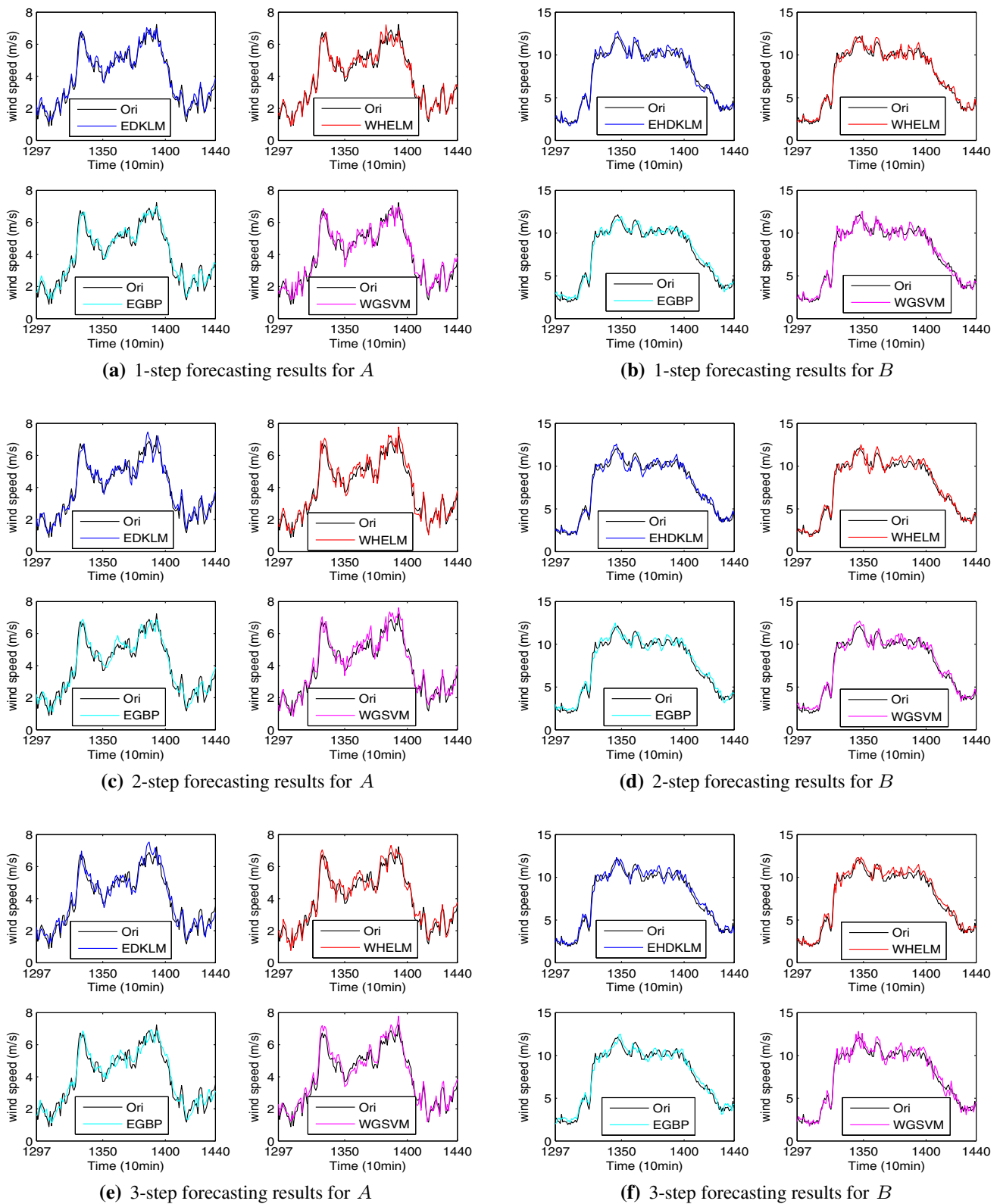


Fig. 11 Forecasting results of the recently developed forecasting models (EDKLM, WHELM, EGBP and WGSVM stand for EEMD-HGSA-DKLSVM, WPD-HPSOGSA-ELM, EEMD-GA-BP and WD-GA-SVM, respectively)

Table 10 Forecasting results of the other hybrid forecasting models

Data set	Index	EEMD-HGSA-LSSVM			WPD-HPSOGSA-ELM			EEMD-GA-BP			WT-GA-SVM		
		1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step	1-step	2-step	3-step
A	MAE (m/s)	0.2517	0.3010	0.3645	0.2423	0.3144	0.3707	0.2707	0.3374	0.3974	0.2995	0.3548	0.3872
	RMSE (m/s)	0.2745	0.3364	0.3906	0.2776	0.3401	0.4021	0.2931	0.3572	0.4181	0.3201	0.3860	0.4332
	MAPE (%)	8.3976	9.1049	10.8903	6.3288	8.9250	10.3515	8.8057	10.6921	12.4065	9.2944	10.3405	9.8701
	MASE	0.7023	0.8397	1.0170	0.6761	0.8771	1.0342	0.7552	0.9413	1.1087	0.8356	0.9899	1.0802
B	MAE (m/s)	0.3960	0.3683	0.3886	0.4171	0.4266	0.4124	0.4341	0.4755	0.4770	0.4616	0.4889	0.5166
	RMSE (m/s)	0.4197	0.4011	0.4386	0.4672	0.4719	0.4537	0.4806	0.5156	0.5183	0.4956	0.5332	0.5635
	MAPE (%)	6.2716	5.4779	5.8256	5.9267	6.4246	6.0938	6.4926	7.3297	6.6910	6.7825	7.0398	7.5548
	MASE	1.1048	1.0274	1.0842	1.1635	1.1901	1.1504	1.2112	1.3264	1.3306	1.2875	1.3640	1.4410

0.4206 m/s for data set *B*, respectively. From the above error indices in the tables, the proposed TSD-HPSOGSA-MKLSSVM model obtains the largest improvement over Persistence.

As given in Tables 5 and 8, the single individual model LSSVM without input data preprocessing and parameter optimization performs much worse than that with parameter optimization and feature selection. Compared with LSSVM, the RMSE errors of PSO-GSA-MKLSSVM for data set *A* in 1-step, 2-step and 3-step predictions are cut by 0.0254 m/s, 0.0143 m/s and 0.039 m/s, respectively. Compared with PSO-GSA-MKLSSVM model, the RMSE errors of HPSOGSA-MKLSSVM for data set *A* in 1-step, 2-step and 3-step predictions are cut by 0.0115 m/s, 0.0239 m/s and 0.0072 m/s, respectively; and 0.0322 m/s, 0.0263 m/s and 0.0298 m/s for wind speed data *B*, respectively (Fig. 9).

Remark The reasons of these results are (1) TSD decomposes the nonstationary and nonlinear wind speed time series into a few relatively stable components that can lower the regression difficulty of LSSVMs. (2) BPSOGSA algorithm identifies and discards the noise components that can relieve the negative effect of input candidates on the forecasting results. (3) LSSVM has good capacity in addressing small sample and nonlinear problems (Hu et al. 2015); in addition, the kernel function in LSSVM takes advantages of RBF, Poly and linear functions and the parameter combination is optimized by PSO-GSA algorithm. These combination of different kernel functions by optimal parameters can enhance the forecasting results. Tables 5, 6 and 8 show that HPSOGSA-MKLSSVM without signal decomposition method performs much worse than that with signal decomposition method regardless of WT, EMD, EEMD, VMD or TSD when applied in wind speed forecasting. These signal decomposition methods break the wind speed data into the relatively stable subseries that contribute to higher forecasting accuracy; thus, the proposed compound structure taking

advantages of various approaches is an effective wind speed forecasting model.

4.4.5 Comparing TSD-HPSOGSA-MKLSSVM with the other recently developed forecasting models

In order to further assess the forecasting performance of the proposed TSD-HPSOGSA-MKLSSVM, the recently developed forecasting models, including EEMD-HGSA-DKLSSVM (Sun et al. 2018b), WPD-HPSOGSA-ELM (Sun et al. 2018a), EEMD-GA-BP (Wang et al. 2016) and WD-GA-SVM (Liu et al. 2014), are constructed as comparative reference and these referenced forecasting models are simply described as follows.

- (i) EEMD-HGSA-DKLSSVM: The original empirical wind speed data are decomposed by EEMD, HGSA make feature selection and parameter optimization, and all the decomposed subseries are used to forecast by well-trained DKLSSVM.
- (ii) WPD-HPSOGSA-ELM: The original empirical wind speed data are decomposed by WPD, HPSOGSA make feature selection and parameter optimization, and all the decomposed subseries are used to forecast by well-trained ELM.
- (iii) EEMD-GA-BP: The original empirical wind speed data are decomposed by EEMD, and all the decomposed subseries are used to forecast by BPNN optimized by GA.
- (iv) WT-GA-SVM: The original empirical wind speed data are decomposed by WT, and all the decomposed subseries are used to forecast by SVM optimized by GA.

All the parameters of each referenced forecasting model are set according to the corresponding paper. These four referenced forecasting models make multi-step wind speed forecasting using the same wind speed data displayed in

Fig. 4 and the forecasting results are shown in Table 10.

As given in Tables 5 and 10, it is obviously obtained that: Compared with EEMD-HGSA-DKLSSVM, WPD-HPSOGSA-ELM, EEMD-GA-BP and WD-GA-SVM for data set *A* in 1-step, the RMSE values of the proposed model are cut by 0.0844 m/s, 0.0875 m/s, 0.103 m/s and 0.13 m/s, respectively; in the 2-step forecasting, the RMSE values of the proposed model are cut by 0.0967 m/s, 0.1005 m/s, 0.1175 m/s and 0.1463 m/s, respectively; in the 3-step forecasting, the RMSE values of the proposed model are cut by 0.0911 m/s, 0.1026 m/s, 0.1185 m/s and 0.1337 m/s, respectively. For data set *B*, it can be also found that the proposed model outperforms EEMD-HGSA-DKLSSVM, WPD-HPSOGSA-ELM, EEMD-GA-BP and WD-GA-SVM. These forecasting results further illustrate the effectiveness of TSD-HPSOGSA-MKLSSVM in short-term wind speed forecasting (Fig. 11).

5 Conclusion

In this study, a new compound HPSOGSA-MKLSSVM wind speed forecasting model combined with decomposition method TSD is proposed. To illustrate the effectiveness of TSD-HPSOGSA-MKLSSVM, two sets of 10-min wind speed time series selected randomly from one wind farm in China are employed to test the proposed model and other forecasting approaches. Based on the aforementioned comparisons and analysis with other single and hybrid forecasting models, some conclusions can be obtained as follows.

- i: Considering that TSD signal processing method can decompose thoroughly wind speed into more stable components, we adopt TSD method combining CEEMDAN with WT to break the wind speed sample into different components. The testing results illustrate that the decomposed performance of TSD outperforms that of both CEEMDAN and WT. HPSOGSA-MKLSSVM with TSD method also has better forecasting results than that with EEMD, EMD, VMD or WT. Thus, TSD is very suitable to be used in the hybrid proposed model.
- ii: Because wind speed time series always present large irregularity, not only single individual LSSVM without signal processing technique but also MKLSSVM with parameter optimization and feature selection cannot make wind speed forecasting with high accuracy. Feature selection and parameter optimization contribute to enhance the forecasting performance of the proposal.
- iii: MKLSSVM takes advantages of RBF kernel function with good capacity in local exploitation, Poly kernel function with excellent ability in global exploration and linear

kernel functions by optimal coefficient. Thus, TSD-HPSOGSA-LSSVM based on multiple kernel function has obtained higher forecasting accuracy than that based on RBF, Poly, linear or double kernel function.

- iv: Compared with the recently developed EEMD-HGSA-DKLSSVM, WPD-HPSOGSA-ELM, EEMD-GA-BP and WD-GA-SVM forecasting models, the proposed hybrid model has better forecasting performance.

What is more, the proposed model makes great improvement compared with the base model Persistence. Thus, the proposed TSD-HPSOGSA-MKLSSVM is an effective method for wind speed forecasting. For further studies, it can be applied in energy demand prediction and other similar domain.

Acknowledgements This work was supported by the Key Laboratory of Electric Drive and Control of Anhui Higher Education Institutes, Anhui Polytechnic University (DQKJ202003); the Projects of Science and Technology Commission of Shanghai Municipality (Grant No. 15JC1401900 and No.17511107002); the National Natural Youth Fund Project (Grant No. 61803001); the Open Research Fund of Wanjiang Collaborative Innovation Center for High-end Manufacturing Equipment, Anhui Polytechnic University (Grant No. GCKJ2018010); and the Natural youth fund project of Anhui Province (Grant No. 1808085QF194).

Compliance with ethical standards

Conflict of interest There is no conflict of interest existing in the submission of this manuscript which has been approved by all authors for publication, and there is no information that might be relevant for the editors and reviewers of soft computing. This study develops a new wind speed forecasting model using two-stage signal decomposition, hybrid PSOGSA algorithm and multi-kernel LSSVM. This article does not contain any studies with human participants performed by any of the authors. This article also does not contain any studies with animals performed by any of the authors. I would like to declare on behalf of my co-authors that the work described is original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

Appendix

All abbreviations are collected, and a brief explanation is given in Table 11.

Table 11 Abbreviations

BBSA	Binary backtracking search algorithm	MAE	Mean absolute error
BPNN	Back-propagation neural network	MAPE	Mean absolute percent error
CEEMDAN	Complementary ensemble empirical mode decomposition with adaptive noise	MASE	Normalized mean absolute scaled error
CWT	Continuous wavelet transforms	MLP	Multilayer perceptron
DB4	Daubechies of order 4	MKLSSVM	Multi-kernel least square support vector machine
DWT	Discrete wavelet transforms	PACF	Partial autocorrelation function
EMD	Empirical mode decomposition	Poly	Polynomial
EEMD	Ensemble empirical mode decomposition	PSO	Particle swarm optimization
ELM	Extreme learning machine	PSOGSA	Particle swarm optimization gravitational search algorithm
ELMNN	Elman neural network	RBF	Radial basis function
GSA	Gravitational search algorithm	RMSE	Root mean square error
HGSA	Hybrid gravitational search algorithm	SVR	Support vector regression
HPF	High-pass filter	TSD	Two-stage decomposition
HPSOGSA	Hybrid particle swarm optimization gravitational search algorithm	WNN	Wavelet neural network
HHT	Hilbert–Huang transform	WPD	Wavelet package decomposition
IMF	Intrinsic mode functions	WT	Wavelet transform
LPF	Low-pass filter	VMD	Variational mode decomposition
LSSVM	Least square support vector machine		

References

- Aghajani A, Kazemzadeh R, Ebrahimi A (2016) A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. *Energy Convers Manag* 121:232–240
- Bouzgou H, Benoudjit N (2011) Multiple architecture system for wind speed prediction. *Appl Energy* 88:2463–2471
- Dong QL, Sun YH, Li PZ (2017) A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: A case study of wind farms in China. *Renew Energy* 102:241–257
- Du P, Wang JZ, Guo ZH et al (2017) Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. *Energy Convers Manag* 150:90–107
- Feng C, Cui MJ, Hodge B et al (2017) A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl Energy* 190:1245–1257
- Han T, Liu QN, Zhang L et al (2019) Fault feature extraction of low speed roller bearing based on Teager energy operator and CEEMD. *Measurement* 138:400–408
- Haque AU, Mandal P, Meng J et al (2013) A novel hybrid approach based on wavelet transform and fuzzy ARTMAP networks for predicting wind farm power production. *IEEE Trans Ind Appl* 49:2253–2261
- Hemmati-Sarapardeha A, Varameshb A, Huseinc M et al (2018) On the evaluation of the viscosity of nanofluid systems: modeling and data assessment. *Renew Sustain Energy Rev* 81:313–329
- Hu JM, Wang JZ, Ma KL (2015) A hybrid technique for short-term wind speed prediction. *Energy* 81:563–574
- Kumar L, Krishna Sripada S, Surekac A et al (2018) Effective fault prediction model developed using least square support vector machine (LSSVM)". *J Syst Softw* 137:686–712
- Li HM, Wang JZ, Lu HY et al (2018) Research and application of a combined model based on variable weight for short term wind speed forecasting. *Renew Energy* 116:669–684
- Liu D, Niu D, Wang H et al (2014) Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew Energy* 62:592–597
- Liu JQ, Wang XR, Lu Y (2017) A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system. *Renew Energy* 103:620–629
- Liu TH, Wei HK, Zhang KJ (2018) Wind power prediction with missing data using Gaussian process regression and multiple imputation. *Appl Soft Comput* 71:905–916
- Luo M, Li CS, Zhang XY et al (2016) Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. *ISA Trans* 65:556–566

- Mirjalili S, Hashim S (2012) A new hybrid PSO-GSA algorithm for function optimization. In: International conference on computer and information application, pp 374–377
- Naik J, Dash S, Dash PK et al (2018) Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network. *Renew Energy* 118:180–212
- Orosio GJ, Matias JCO, Catalao JPS (2015) Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. *Renew Energy* 75:301–307
- Peng T, Zhou JZ, Zhang C et al (2017) Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. *Energy Convers Manag* 153:589–602
- Rashedi E, Nezamabadi S, Saryazdi S (2009) GSA: a gravitational search algorithm. *Inform Sci* 179:2232–2248
- Rostami S, Rashidi F, Safari H (2019) Prediction of oil-water relative permeability in sandstone and carbonate reservoir rocks using the CSA-LSSVM algorithm. *J Petrol Sci Eng* 173:170–186
- Salcedo-Sanz S, Pastor-Sanchez A, Prieto L et al (2014) Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization-extreme learning machine approach. *Energy Convers Manag* 87:10–18
- Santamaria-Bonfil G, Reyes-Ballesteros A, Gershenson C (2016) Wind speed forecasting for wind farms: A method based on support vector regression. *Renew Energy* 85:790–809
- Stephen JE, Kumar SS, Jayakumar J (2014) Nonlinear modeling of a switched reluctance motor using LSSVM-ABC. *Acta Polytech Hung* 11:143–158
- Sun SZ, Fu JQ, Zhu F (2018a) A hybrid structure of an extreme learning machine combined with feature selection, signal decomposition and parameter optimization for short-term wind speed forecasting. *Trans Inst Meas Control* 40:1–18
- Sun SZ, Fu JQ, Zhu F, Xiong N (2018b) A compound structure for wind speed forecasting using MKLSSVM with feature selection and parameter optimization. *Math Probl Eng* 9287097(11):1–21
- Sun GP, Jiang CW, Cheng P et al (2018c) Short-term wind power forecasts by a synthetical similar time series data mining method. *Renew Energy* 115:575–584
- Sun SZ, Wei LS, Xu J et al (2019) A new wind power forecasting modeling strategy using two-stage decomposition, feature selection and DAWNN. *Energies* 12334:1–24
- Torres ME, Colominas MA, Schlotthauer G et al (2011) A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 4144–4147
- Wang Y, Wang JZ, Wei X (2015a) A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: A case study of wind farms in northwest China. *Energy* 91:556–572
- Wang Y, Wang JZ, Wei X (2015b) A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: A case study of wind farms in northwest China. *Energy* 91:556–572
- Wang SX, Zhang N, Wu L et al (2016) Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew Energy* 94:629–36
- Wang DY, Luo HY, Grunder O et al (2017) Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. *Renew Energy* 113:1345–1358
- Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adaptive Data Anal* 1:1–41
- Yan J, Huang GQ, Peng XY et al (2018) A novel wind speed prediction method: hybrid of correlation-aided DWT, LSSVM and GARCH. *J Wind Eng Ind Aerod* 174:28–38
- Yin H, Dong Z, Chen YL et al (2017) An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. *Energy Convers Manag* 150:108–121
- Yuan XH, Chen C, Yuan YB et al (2015) Short-term wind power prediction based on LSSVM-GSA model. *Energy Convers Manag* 101:393–401
- Zhang GY, Wu YG, Liu YQ (2014) An advanced wind speed multi-step ahead forecasting approach with characteristic component analysis. *J Renew Sustain Energy* 053139(6):1–14
- Zhang C, Wei HK, Zhao JS et al (2016a) Short-term wind speed forecasting using empirical mode decomposition and feature selection. *Renew Energy* 96:727–737
- Zhang C, Wei HK, Zhao JS et al (2016b) Short-term wind speed forecasting using empirical mode decomposition and feature selection. *Renew Energy* 96:727–737
- Zhang C, Zhou JZ, Li CS et al (2017) A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Convers Manag* 143:360–376
- Zheng H, Liao R, Grzybowski S et al (2011) Fault diagnosis of power transformers using multi-class least square support vector machines classifiers with particle swarm optimization. *IET Electr Power Appl* 5:691–696
- Zheng WQ, Peng XG, Lu D et al (2017) Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: A new approach. *Energy Convers Manag* 151:737–752
- Zhou JY, Shi J, Li G (2011) Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers Manag* 52:1990–1998

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.