Energy Conversion and Management 114 (2016) 75–88

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm

Anbo Meng *, Jiafei Ge, Hao Yin, Sizhe Chen

School of Automation, Guangdong University of Technology, Guangzhou 510006, China

ARTICLE INFO

Article history:
Received 1 December 2015
Accepted 6 February 2016
Available online 16 February 2016

Keywords:
Short-term wind speed forecasting
Crisscross optimization algorithm
Wavelet packet decomposition
Artificial neural network

ABSTRACT

Wind speed forecasting is of great significance for wind farm management and safe integration into electric power grid. As wind speed is characterized by high autocorrelation and inherent volatility, it is difficult to predict with a single model. The aim of this study is to develop a new hybrid model for predicting the short wind speed at 1 h intervals up to 5 h based on wavelet packet decomposition, crisscross optimization algorithm and artificial neural networks. In the data pre-processing phase, the wavelet packet technique is used to decompose the original wind speed series into subseries. For each transformed components with different frequency sub-bands, the back-propagation neural network optimized by crisscross optimization algorithm is employed to predict the multi-step ahead wind speed. The eventual predicted results are obtained through aggregate calculation. To validate the effectiveness of the proposed approach, two wind speed series collected from a wind observation station located in the Netherlands are used to do the multi-step wind speed forecasting. To reduce the statistical errors, all forecasting methods are executed 50 times independently. The results of this study show that: (1) the proposed crisscross optimization algorithm has significant advantage over the back-propagation algorithm and particle swarm optimization in addressing the prematurity problems when applied to train the neural network. (2) Compared with the previous hybrid models used in this study, the proposed hybrid model consistently has the minimum mean absolute percentage error regardless of one-step, three-step or five-step prediction.

1. Introduction

As a promising renewable energy source, wind energy is of vital importance among the low-carbon energy technologies. The main motivation for developing wind energy is to reduce dependence on fossil fuels with their adverse effect on the environment. According to Global Wind Energy Council [1], Wind energy's contribution to the global electricity supply is expected to reach 12% by 2020 and 22% by 2030. With the advances in new wind energy harvesting techniques such as mean flow acoustic engine [2], it will also contribute much to the reduction of greenhouse gas emissions. Duic and Rosen [3] further state that the figure is expected to fall 80–95% by 2050. However, due to the inherent volatility and intermittence of the wind source, the integration efficiency of wind power into electricity grids is limited. As stated by Zhang et al. [4], this problem can be significantly mitigated if the operation of the wind farms can be controlled based on accurate information of the dynamic wind speed prediction. As a result, accurate forecasting of the short-term wind speed is a critical issue and it has drawn much attention of system operators, utilities and researchers toward the state-of-the-art wind speed forecasting methods.

In the past few decades, various wind speed forecasting approaches have been proposed in the literature. They can be divided into physical, statistical and intelligent methods. Every kind of methods has their own advantages and disadvantages. The physical model predicts wind speed mainly through the terrain features, ambient temperature, atmospheric pressure and other meteorological information. As stated by Ren et al. [5], the physical model is usually only used as an auxiliary input for other statistical models, which is known as stochastic time series model, Tascikaraoglu and Uzunoglu [6] state that statistical method aims at describing the relation between historical time series of wind speed at the location of interest by generally recursive techniques. To name a few here, Huang and Chalabi [7] present a well-known Time-varying autoregressive (AR) model, which can be given as a typical example in the statistical methods. Erdem and Shi [8] explore the effectiveness of auto regressive moving average (ARMA) based approaches to prediction of the tuple of wind speed and direction. Su et al. [9] present a hybrid wind speed forecasting

* Corresponding author. Tel.: +86 13928799469; fax: +86 020 39322563.
E-mail address: menganbo@vip.sina.com (A. Meng).

http://dx.doi.org/10.1016/j.enconman.2016.02.013
0196-8904/

Nomenclature

Abbreviations

WPD	wavelet packet decomposition
CSO	crisscross optimization
PSO	particle swarm optimization
GA	genetic algorithm
ANN	artificial neural network
BP	back propagation
SVM	support vector machine
MAPE	mean absolute percentage error
MAE	mean absolute error
RMSE	root mean squared error
HC	horizontal crossover
VC	vertical crossover
MS	moderation solution

Symbols

X	initialized population
c_1, c_2	expansion coefficients
r_1, r_2	uniformly distributed random values between 0 and 1
MS_{hc}	moderation solutions of horizontal crossover
MS_{vc}	moderation solutions of vertical crossover
D	number of dimension
P_v	vertical crossover probability
$p_t, \hat{p}_t$	target and output of the tth output neuron
n, h, m	neuron number in input, hidden and output layers
IW_{hn}	weight of connection from input n to hidden neuron h
OW_{mh}	weight of connection from hidden neuron h to output neuron m
b_{Hh}, b_{Mm}	bias values in the hidden and output layers

model based on autoregressive integrated moving average (ARIMA) and Kalman filter. Lydia et al. [10] have developed four linear and non-linear ARMA models with eXogenous inputs (ARMAX) for short-term wind speed forecasting. In their study, the parameters of the linear models are obtained using the Gauss–Newton algorithm and the non-linear autoregressive models are developed using three different data mining algorithms. However, these conventional statistical methods assume that the wind speed data is normally distributed, but it is well known that wind speed series is not a normally distributed. In addition, Kani and Ardehali [11] further state that the intermittent and stochastic characteristic of wind speed series need more complex functions for capturing the nonlinear relations, but these models are based on the assumption that a linear correlation structure exists among time series values.

To overcome this limitation of statistical approaches, the AI approaches, mainly including artificial neural networks (ANNs) and support vector machines (SVMs), have attracted more attention for time series-based wind speed forecasting. For example, Cadenas and Rivera [12] apply ANN to the hourly wind speed time series and aimed to enhance prediction accuracy for each month of the year. Men et al. [13] present an ensemble of mixture density neural networks (MDN) used for short-term wind speed and power forecasting. In this study, the ensemble MDN model provides uncertainty quantification for the wind speed and wind turbine power forecasts in the form of a confidence interval at a prescribed confidence level. Kong et al. [14] present a wind speed prediction model using reduced support vector machines (RSVM) with feature selection. They find that the parameters of SVM optimized by particle swarm optimization algorithm (PSO) can avoid the excessive manual intervention for parameter tuning. More introduction to AI methods can be found in [6] presented by Tascikaraoglu and Uzunoglu. According to the results reported in the literature, the AI methods appear to be more accurate as compared to traditional statistical models. As a popular forecasting technique, ANNs have been widely applied into different forecasting fields. Compared with other prediction methods, ANNs exhibit some advantage in high data error tolerance, easy adaptability to online measurements, no need for excess information other than wind speed history. As a typical ANN proposed by Rumelhart et al. [15], the back propagation neural network (BP-NN) can implement any complex nonlinear mapping function proved by mathematical theories and approximate an arbitrary nonlinear function with satisfactory accuracy, according to Zhang et al. [16].

However, BP is used in conjunction with an optimization method such as gradient descent, leading to a poor convergence rate and tends to get stuck in local minima easily in large-scale ANN training. To overcome the disadvantages of conventional BP algorithm, some heuristic algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO) and ant colony optimization (ACO), have been introduced to optimize the ANNs, which have been proved feasible and effective in different forecasting applications. For example, Yu and Xu [17] present an improved BP neural network combined with real-coded GA for short-term gas load forecasting. In the combinational model, GA is used to determine the ANN's initial weights and thresholds. The authors indicate that such improvements facilitate forecasting efficiency and exert maximum performance of the model. Pousinho et al. [18] present a new hybrid approach by combining PSO and adaptive-network-based fuzzy inference system for short-term wind power prediction in Portugal. They state that PSO-based hybrid model can significantly improve the prediction accuracy by comprehensive parameter selection. Rahmani et al. [19] present a hybrid approach consisting of ANN, ACO and PSO to forecast the energy output of a real wind farm located in Binaloud, Iran. The authors indicate that the hybridization of ACO and PSO to optimize the forecasting model leads to a higher quality result with a faster convergence profile. Chitsaz et al. [20] have developed a Wavelet neural network (WNN) with the activation functions of the hidden neurons constructed based on multi-dimensional Morlet wavelets. In this study, a new improved Clonal selection algorithm is used to optimize the WNN's free parameters. The proposed forecaster is effectively validated on real-world hourly data of system level wind power generation in Alberta, Canada.

The predicted results reported in the literature demonstrate that the performance of ANNs optimized by heuristic algorithms can be improved. In this paper, a new heuristic algorithm called crisscross optimization algorithm (CSO) is firstly attempted to be applied into the ANNs' training for short-term wind speed forecasting. The motivation stems from the CSO's advantage over other heuristic algorithms in addressing the high-dimensional non-convex problems. For example, Meng et al. [21] apply the CSO algorithm to solve a large scale combined heat and power economic dispatch (CHPED) problem and in [22], they have also successfully addressed a more challenging dynamic economic dispatch problem with 1000 generators by CSO. The results show CSO's significant advantage over other heuristic algorithms like GA and PSO. Therefore, CSO is expected to have potential to overcome the convergence premature problem when applied to train ANNs with many local minima. The empirical tests in Section 4 also confirm the superiority of CSO-based neural network (CSO-NN) in terms of prediction accuracy.

As pointed out by Zhang et al. [4], forecasting the wind speed data with the noisy directly is usually subject to large errors. To further enhance the wind speed forecasting accuracy, the multi-scale decomposition of original wind time series is indispensable. The previous studies show that the prediction accuracy of wind speed can be greatly improved by applying different decomposition techniques such as wavelet decomposition (WD), wavelet packet decomposition (WPT), empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD). For example, Wang et al. [23] present a hybrid EMD–ENN approach for wind speed forecasting based on EMD and Elman neural network (ENN). In their study, the original wind speed datasets are decomposed into a collection of intrinsic mode functions (IMFs) and a residue by EMD. both IMF components and residue are applied to establish the corresponding ENN models. The results show that the EMD–ENN model outperforms the persistent model, BP-NN and ENN. Liu et al. [24] present a short-term wind speed forecasting approach by using WD and SVMs optimized by genetic algorithm. In their study, WD is exploited to decompose the wind speed signal into two components, an approximation signal to maintain the major fluctuations and a detail signal to eliminate the stochastic volatility. The SVMs are built to model the approximation signal and their parameters were fine-tuned by GA to ensure the generalization of SVMs. Hu et al. [25] investigate the possibility of improving the quality of wind speed forecasting by using a hybrid model named EEMD–SVM based on SVM and the Ensemble Empirical Mode Decomposition (EEMD). Their developed tool shows a good promise for the forecasting of intricate time series which are intrinsically highly volatile and irregular. Liu et al. [26] present a hybrid model named EMD–ANN for wind speed prediction based on ANNs and EMD. The results show that EMD–ANN method is robust in dealing with jumping samplings in non-stationary wind series. Liu et al. [27] present three hybrid models based on WD, WPD, time series and ANNs to predict the wind speed. They state that WPD-based hybrid models have better performance. Recently in [28], Liu et al. have developed another four different hybrid models by combining four signal decomposing algorithms (e.g., WD/WPD/EMD/EEMD) and Extreme Learning Machines (ELMs). The authors investigate the promoted percentages of the ELMs by those mainstream signal decomposing algorithms in the multiple step wind speed forecasting. The obtained results show that all the proposed hybrid algorithms have better performance than the single Extreme Learning Machines by utilizing the decomposing algorithms.

Unlike WD, which only decomposes the previous approximation coefficients, the WPD decomposes both of the previous approximation and detail coefficients. It is obvious that the latter offers richer signal analysis. Therefore, WPD is selected as the decomposition methods in this paper in view of the complex nonlinear characteristics of original wind speed series. By combining the advantages of WPD and CSO-NN, a novel hybrid model named WPD–CSO-NN is proposed for short-term wind speed forecasting. In the new prediction model, WPD is used to realize the non-stationary wind speed decomposition and the CSO-NN is employed to forecast the decomposed wind speed subseries. The proposed WPD–CSO-NN model is validated by forecasting the multi-step ahead wind speed of a wind farm located in the Netherlands. The prediction results show that the proposed approach has advantage in varying degrees over other forecasting models used in this study.

This paper is organized as follows: Section 2 presents the proposed hybrid WPD–CSO-NN approach. Section 3 describes the empirical studies on the parameters selection. Section 4 gives the forecasting results and comparative analysis. The conclusions are made in Section 5.

2. The hybrid WPD–CSO-NN approach

2.1. Data set

The hourly wind speed series samples are collected from a wind observation station of Rotterdam located in the Netherlands (Latitude: 51.955°N, Longitude: 4.444°E). Those wind speed data provided by the Royal Netherlands Meteorological Institute are available in [29]. In this work, we randomly selected two months of sampling points in 2014 as the data set. Figs. 1 and 2 show the wind speed time series measured in January, 2014 and in October, 2014. According to our experimental tests and the suggestion made by Liu et al. [30], the 1st–600th sampling points were used to train the proposed forecasting model and the subsequent 601st–700th data were applied to test the foresting performance of the well-trained WDP–CSO-NN model.

2.2. Crisscross optimization algorithm

Crisscross optimization (CSO) algorithm [31] is our newly developed heuristic algorithm for numeric optimization. Compared with most of other heuristic algorithms like PSO and GA, It has been proved that CSO has obvious advantage in terms of solution accuracy and convergence speed especially when applied to solve high-dimensional non-convex optimization problems.

The global search ability of CSO owes much to two distinct search operators, namely horizontal crossover (HC) and vertical crossover (VC). The HC operator searches for a new solution in a half population of separate hyper-cubes with a large probability while on their respective peripheries with a decreasing probability. Such cross-border search approach was proved to be an effective

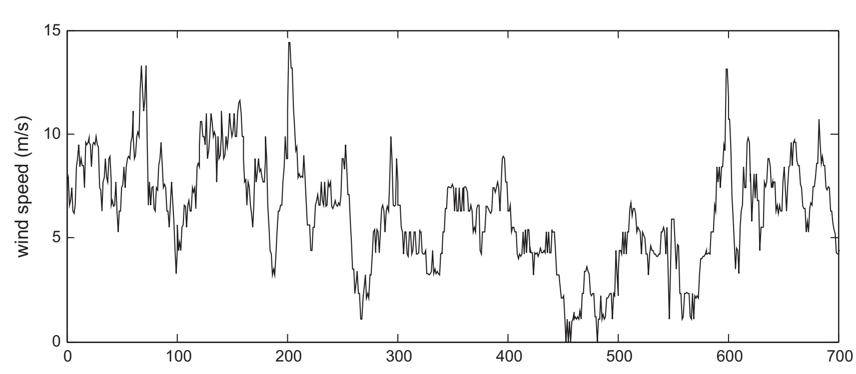

Fig. 1. Original wind speed time series in January, 2014.

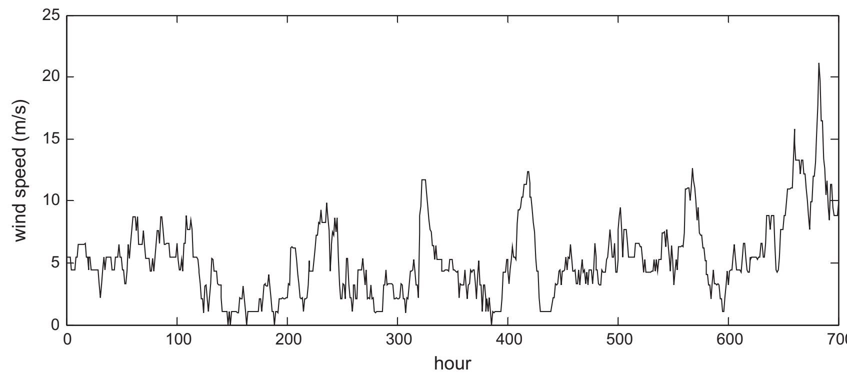

Fig. 2. Original wind speed time series in October, 2014.

way to reduce the search blind spots and enhance the global search ability. Unlike HC that executes an arithmetic crossover in the horizontal direction, Instead, The VC operator generates a new solution by performing a crossover in the opposite direction. This unique crossover way plays a key role in facilitating the stagnant dimensions to escape out of the local minima.

In CSO, the new solutions generated by HC or VC are called moderation solution (MS), which cannot enter the next generation directly. Only those that outperform their counterparts in the parent population, survive in the next generation. Otherwise, they are eliminated by their parent competitors. It is clear that CSO always maintains a population of personal best solutions in the evolving iterative process. This elite selection strategy makes CSO approximate toward the global optimum with a fast-converging speed. The basic procedure of CSO can be described as below.

Step 1. Initialize the population

The initialized population X consists of M randomly generated individuals in the D-dimensional space.

Step 2. Perform horizontal crossover

All the individuals in X are randomly divided into $M/2$ pairs without repetition. Suppose the ith parent individual $X(i)$ and the jth parent individual $X(j)$ are used to carry out the horizontal crossover operation at the dth dimension. The moderation solutions are reproduced by (1).

$$\begin{cases} MS_{hc}(i,d) = r_1 \cdot X(i,d) + (1-r_1) \cdot X(j,d) + c_1 \cdot (X(i,d) - X(j,d)) \\ MS_{hc}(j,d) = r_2 \cdot X(j,d) + (1-r_2) \cdot X(i,d) + c_2 \cdot (X(j,d) - X(i,d)) \end{cases} \tag{1}$$

where r_1 and r_2 are uniformly distributed random values between 0 and 1, c_1 and c_2 are expansion coefficients, which are uniformly distributed random values between -1 and 1. $MS_{hc}(i)$ and $MS_{hc}(j)$ are the moderation solutions, which are the offspring of $X(i)$ and $X(j)$, respectively.

According to Meng et al. [31], the HC operator divides the multidimensional problem-solving space into half-population of hypercubes that take the paired parent individuals (e.g. $X(i)$ and $X(j)$) as their diagonal vertices. Each pair of the parent individuals reproduces the offspring in the space of their own hypercube to a greater extent. To reduce the blind spot that cannot be reached, the HC operator also searches the periphery of each hypercube with a decreasing probability. Apparently, the expansion coefficients determine the size of peripheral space. Fig. 3 illustrates the distribution of probability density of the moderation solutions in 2-D space.

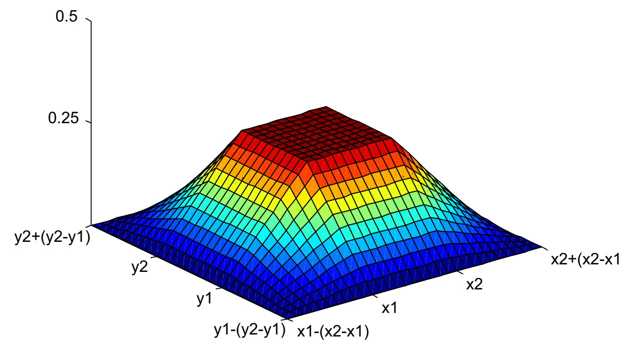

Fig. 3. Probability density of the moderation solutions in 2-D space.

Step 3. Perform vertical crossover

All the dimensions in X are randomly divided into $D/2$ pairs without repetition. Suppose the $d1$th and $d2$th dimensions of the individual i are used to carry out the vertical crossover operation, their offspring $MS_{vc}(i)$ can be reproduced by (2).

$$MS_{vc}(i,d1) = r \cdot X(i,d1) + (1-r) \cdot X(i,d2) \quad i \in N(1,M), \\ d1, d2 \in N(1,D) \tag{2}$$

where r is a uniformly distributed random value between 0 and 1.

One of CSO's greatest contributions lies in the finding that only a few dimensions of the swarms for most heuristic algorithms are possibly struck into stagnancy when applied to solve the multimodal problems. In view of this fact, the VC operator uses the vertical crossover probability P_v to control how many paired dimensions could participate in the arithmetic crossover. According to Meng et al. [31], P_v is usually set in the range [0.2, 0.8]. Unlike HC, in VC, each crossover reproduces only one offspring. The motive is to provide an opportunity for the stagnant dimension to jump out of the local optimum and not destroy another dimension that is possibly normal. Besides, to perform VC, the normalization and reverse normalization operation are necessary since different dimensions may have different units or upper–lower limits.

Step 4. Terminal condition

If the number of iterations is larger than the predefined maximum number, the process terminates. Otherwise, go to Step 2 for a new round of iteration.

2.3. Framework of modeling

The framework of the proposed WPD–CSO-NN approach is illustrated in Fig. 4. The main procedure for wind speed forecasting can be described as below.

(1) Determine the number of decomposition levels for WPD and the vertical crossover probability P_v for CSO-NN by a number of empirical studies.
(2) Use the wavelet packet to decompose an original wind speed series into several sub-series with different frequency subbands.
(3) Train the ANNs using the proposed CSO methods for every transformed subseries. The 1st–600th sampling points in Figs. 1 and 2 are selected as the training dataset.
(4) Apply the optimized CSO-NN models to do the multi-step ahead forecasting in each subseries. The 601th–700th sampling points are selected as the test dataset.
(5) Conduct aggregate calculation to obtain the final forecasting result.
(6) To compare the performance between WPD–CSO-NN and other methods, the same wind speed subseries are also predicted by other four hybrid models including WPD–BP-NN, WPD–PSO-NN, EMD-NN and WD–GA–SVM.
(7) To make a full comparison, one-hourly wind speed series from one wind farm in the Netherlands in January and October are provided in case studies.

2.4. Wavelet packet decomposition

Wavelet decomposition (WD) proposed by Mallat [32] is a mathematical technology used to analyze signals by decomposition into various frequencies. Wavelet packet decomposition (WPD) can be seen as a special kind of wavelet decomposition. In the orthogonal wavelet decomposition procedure, the generic step splits the approximation coefficients into two parts. After splitting we obtain a vector of approximation coefficients and a vector of detail coefficients, both at a coarser scale. The information lost between two successive approximations is captured in the detail coefficients. Then the next step consists of splitting the new approximation coefficient vector; successive details are never reanalyzed. However, in the corresponding wavelet packet situation, each detail coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. The complete three-layer binary trees of WD and WPD are illustrated in Fig. 5(a) and (b). Considering that WPD can offer the richer analysis, we use WPD to build the hybrid model combined with CSO-NN. The results of wavelet packet decomposition of the 1st–600th original wind speed time series data (see Fig. 1) at level 3 are demonstrated in Fig. 6.

2.5. ANNs optimized by CSO

The standard BP neural network uses gradient descent algorithm to minimize the mean square error between the target value and network output. Many studies show that the BP algorithm has slow training speed and it is vulnerable to the problem of local minimum which make it difficult to converge to the global minimum point. To address the above problems, we use the proposed CSO algorithm to replace the conventional BP algorithm for optimizing the network parameters.

Fig. 7 shows a three-layer BP-NN structure. Suppose that the neuron number in input, hidden and output layers are n, h, m, respectively. It is clear that the total number of decision variables (i.e., weights and bias) is $D = n \times h + m \times h + h + m$. The fitness of every individual solution can be measured as follows.

$$E = \frac{1}{N}\sum_{k=1}^{N}\sum_{t=1}^{m}(p_t - \hat{p}_t)^2 \tag{3}$$

where p_t, $\hat{p}_t$ are the target and output of the tth output neuron, respectively, N is the number of training samples.

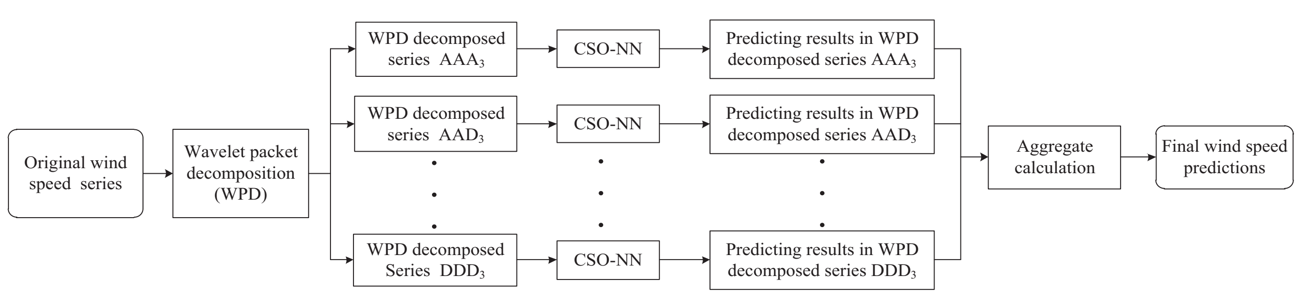

Fig. 4. The framework of the WPD–CSO-BP model.

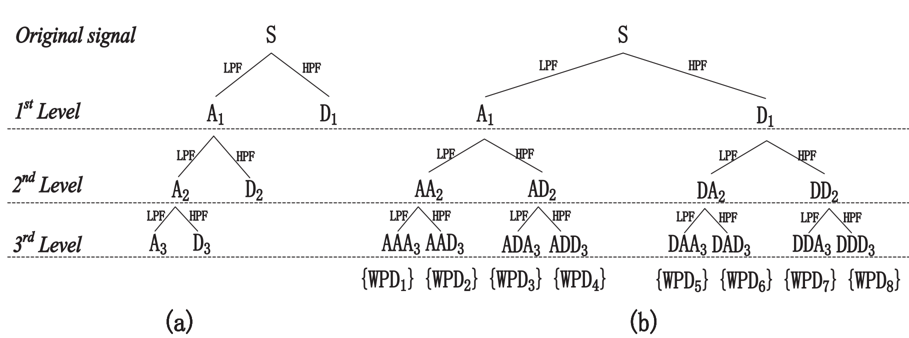

Fig. 5. Decomposition tree at level 3: (a) wavelet decomposition and (b) wavelet pocket decomposition.

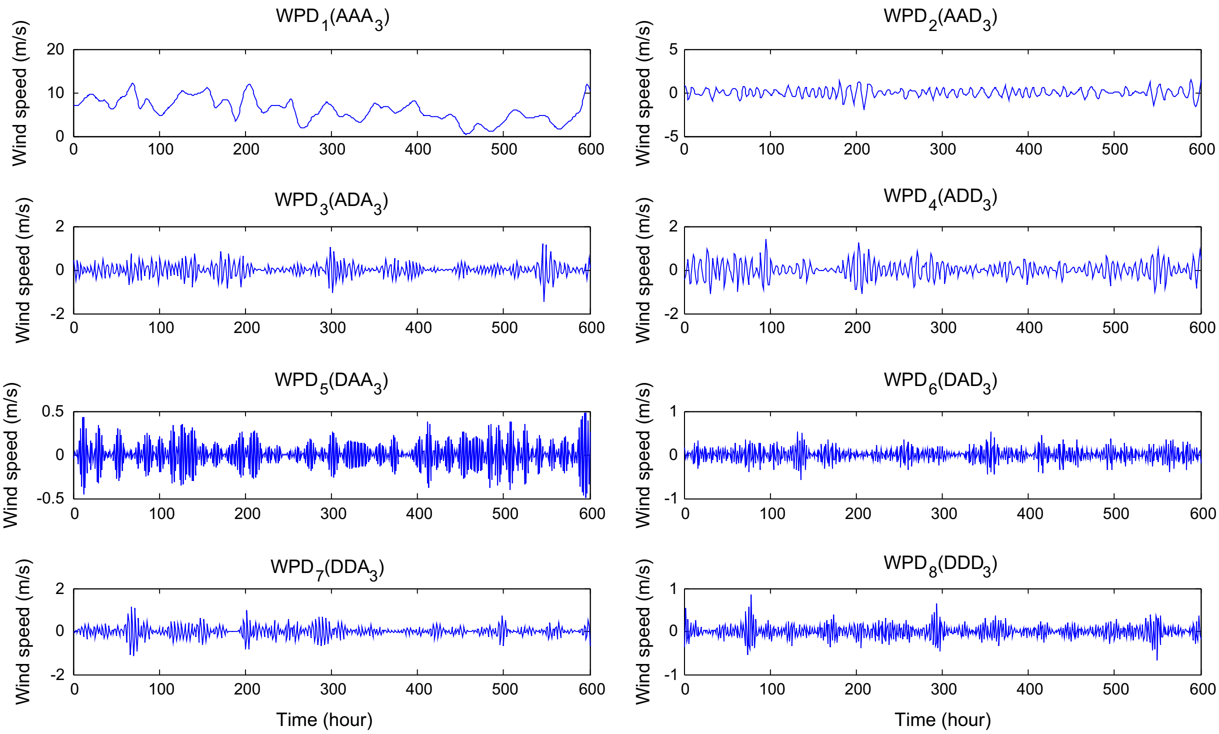

Fig. 6. Results of WPD for the 1st–600th original weed speed data.

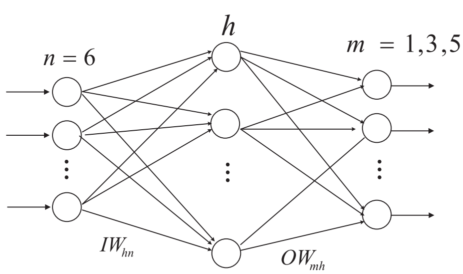

Fig. 7. A three-layer BP-NN structure.

The step-by-step of BP-NN training procedure using the CSO algorithm can be summarized as follows.

(1) *Prepare the sample data*

To train CSO-NN, the 1st–600th sampling points depicted in Figs. 1 and 2 are used as the training sample dataset and the rest of 100 sampling points are selected as the testing dataset. For every WPD transformed subseries, we do the one-step, three-step and five-step prediction. Take one-step prediction for example, the format of input–output sample pairs is illustrated in Fig. 8.

(2) *Determine network topology*

In this study, for every ANN used in all subseries, the input number is set to 6. Since the one-step, three-step and five-step prediction are done in case studies, obviously the number of output neurons should be 1, 3, and 5, respectively. The neuron number of hidden layer is set to 10 according to the empirical expression suggested by Liu et al. [33]. The sigmoid functions are selected as the activation function in hidden and output layers.

(3) *Initialize the population and parameter setting*

The population X is initialized with M individual solutions consisting of D randomly generated weights and bias. X can be represented as $X = [X_1, X_2, \ldots, X_i, \ldots, X_M\,]^T$

$$X_i = [IW_{11}\ \cdots\ IW_{hn}\ OW_{11}\ \cdots\ OW_{mh}\ b_{H1}\ \cdots\ b_{Hh}\ b_{M1}\ \cdots\ b_{Mm}]$$
$$i = 1, 2, \ldots, M \qquad (4)$$

where IW_{hn} is the weight of connection from input n to hidden neuron h; OW_{mh} is the weight of connection from hidden neuron h to output neuron m. b_{Hh} and b_{Mm} are the bias values in the hidden and output layers, respectively.

For CSO, there is only one adjustable parameter, namely the vertical crossover probability P_v. In this study, we determine this parameter by conducting many empirical tests in Section 3.2.

(4) *Compute fitness value*

Every individual's fitness value is calculated according to (3).

(5) *Horizontal crossover*

Perform the horizontal crossover operation according to (1). After all the moderation solutions in matrix MS_{hc} are generated, the population X needs to be updated by comparing the fitness valve between individuals in X and MS_{hc}.

(6) *Vertical crossover*

Perform the vertical crossover operation according to (2). It is noted that only a few paired dimensions have the chance to perform the arithmetic crossover operation according to the vertical crossover probability P_v. After all the moderation solutions in matrix MS_{vc} are generated, the population X needs to be updated by comparing the fitness valve between individuals in X and MS_{vc}.

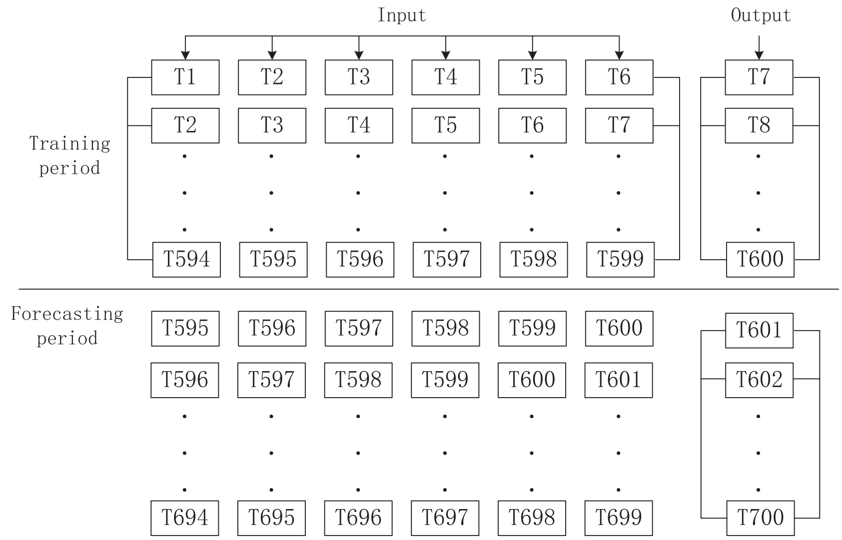

Fig. 8. Data format.

(7) *Terminal condition*

If the error E is less than 0.01, the algorithm terminates. Otherwise, go to (3) for a new round of iteration.

3. Parameters selection

3.1. Evaluation criteria

To validate the effectiveness of the proposed approach, WPD–CSO-NN is applied to one-step, three-step and five-step ahead wind speed forecasting. All simulations are developed in MATLAB 2010a and carried out on executed on a PC with a Core (TM) CPU running at 2 GHz with memory capacity of 8 GB under Windows 8 Operating System. In order to reduce the statistic errors, all the experiments will be carried out 50 times independently. To evaluate the prediction accuracy, three metrics are used in this study: Mean absolute percentage error (MAPE), mean absolute error (MAE), root mean-squared error (RMSE).

$$\mathrm{MAPE} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{p_t - \widehat{p}_t}{p_t}\right| \times 100\% \tag{5}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}\left|p_t - \widehat{p}_t\right| \times 100\% \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(p_t - \widehat{p}_t\right)^2} \tag{7}$$

where p_t and $\widehat{p}_t$ represent observed value and predictive value of the wind speed, N is the total number of data used for performance evaluation and comparison.

3.2. Parameters selection in WPD–CSO-NN

For WPD–CSO-NN, there are two key parameters, namely the vertical crossover probability P_v and the layer number of WPD, which may affect the performance of WPD–CSO-NN. Therefore, the following section will focus on the parameters selection by empirical tests. In this study, the input number the neurons number of hidden layer is decided by using a empirical expression: $h = \sqrt{n+m} + a$, $a \in [1,10]$. For the sake of convenience, the neurons number is uniformly set to be 10 in all the following experimental tests, which is proved to meet the prediction accuracy requirement. It is noted that the parameters selection is based on the 1st–600th original data illustrated in Fig. 1.

3.2.1. Selection of wavelet packet decomposition levels

Experimental tests show that the WPD levels have a great impact on the prediction accuracy. If the decomposition levels are too small WPD does not function well. Conversely, when too many levels are provided, some original data may be forcibly destroyed. So far, there are few investigations into the effect of WPD levels on the forecasting results in the literature.

To determine the appropriated levels of WPD, we do 50 times independent computer experiments for one-step, three-step and five-step wind forecasting using WPD–CSO-NN based on the original wind speed series shown in Fig. 1. The parameters of WPD–CSO-NN are listed as follows.

(1) Number of input: 6.
(2) Number of hidden neurons: 10.
(3) Number of output neurons: 1 (one-step); 3 (three-step); 5 (five-step).
(4) Vertical crossover probability P_v: 0.5.

Fig. 9 shows the MAPE errors at different decomposition levels (1–7). The results reveal that the MAPE errors decrease with the increase of decomposition levels. But there is a turning point regardless of one-step, three-step or five-step prediction. Considering both the prediction accuracy and computation time, WPD is used to decompose the original wind speed series for one-step, three-step or five-step prediction at levels 4, 5 and 6, respectively.

3.2.2. Selection of vertical crossover probability P_v

Compared with other heuristic algorithm, CSO is easy to implement and has only one parameter P_v to be adjusted. To investigate the effects of different vertical crossover probabilities on the performance of CSO, we do 50 times independent computer experiments for different vertical crossover probability using WPD–CSO-NN. The parameters of WPD–CSO-NN are listed as follows:

(1) Number of input: 6.
(2) Number of hidden neurons: 10.

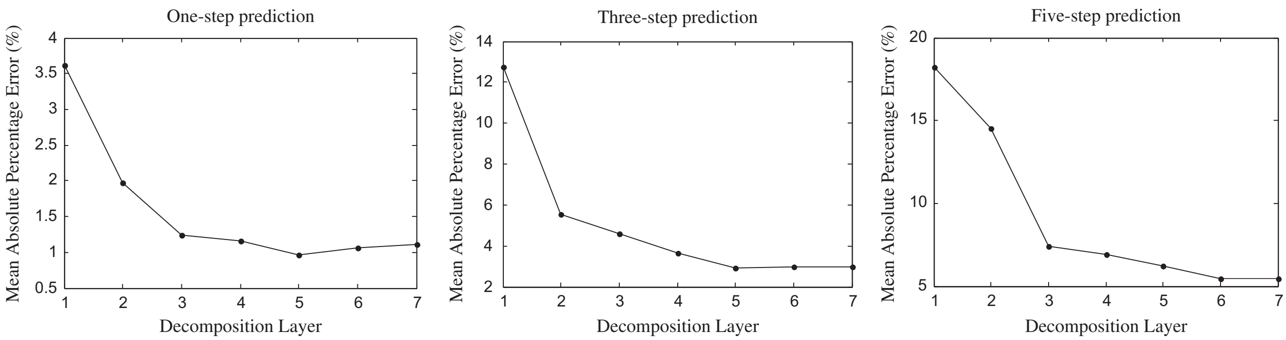

Fig. 9. The MAPE errors with different decomposition levels in multi-step prediction.

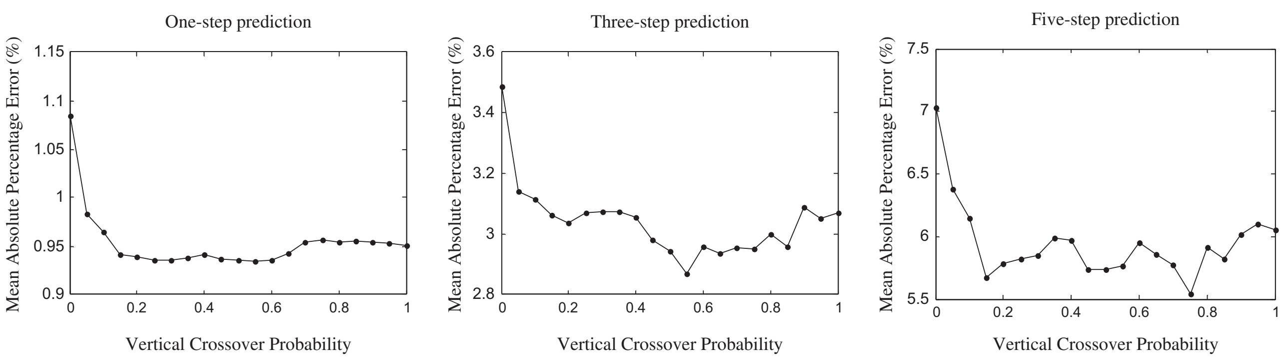

Fig. 10. The MAPE errors with different vertical crossover probabilities in multi-step prediction.

(3) Number of output neurons: 1 (one-step); 3 (three-step); 5 (five-step).
(4) The number of decomposition layers: 4 (one-step); 5 (three-step); 6 (five-step).

Fig. 10 shows the errors with different vertical crossover probabilities. It is seen that the prediction error is sharply increased when the vertical crossover is removed from CSO (i.e., $P_v = 0$). This means that CSO suffers the premature convergence problem in various degrees. However, once the P_v value is set in the rang [0.2, 1], the prediction error changes flatly and choice of the P_v has little effect on the prediction accuracy. As a result, the P_v value is set to be 0.5 for all the tests in this work.

4. Forecast results and comparative analysis

To validate the effectiveness of the proposed WPD–CSO-NN approach, two wind speed series shown in Figs. 1 and 2 are used to do the multi-step wind speed forecasting in case 1and 2. It is worthy to note that the 601st–700th data are used as the test data in both case 1 and case 2. To further investigate the performance of the proposed hybrid method, two groups of forecasting comparisons are provided in this section. To reduce the statistical errors, all forecasting methods are executed 50 times independently.

4.1. Case 1

4.1.1. WPD–CSO-NN vs. BP-NN vs. WPD–BP-NN

In case 1, the original data shown in Fig. 1 are collected in January, 2014. The first group of hybrid models by using BP-NN, WPD–BP-NN and WPD–CSO-NN are provided to do the multi-step wind speed prediction. To make a fair comparison, the parameters of all the prediction models are set as follows.

4.1.1.1. For all the prediction models.
(1) Number of inputs: 6.
(2) Number of hidden neurons: 10.
(3) Number of output neurons: 1 (one-step); 3 (three-step); 5 (five-step).

4.1.1.2. For the WPD-based prediction models. Number of decomposition layers: 4 (one-step); 5 (three-step); 6 (five-step).

4.1.1.3. For the CSO algorithms.
(1) Population size: 20.
(2) Maximum iterative times: 1000.
(3) Vertical crossover probability of CSO: 0.5.

Figs. 11–13 show the forecasting results at 601th–700th sampling points for one-step, three-step or five-step prediction. Table 1 shows the mean errors of 50 independent computer experimental tests.

As shown in Table 1 and Figs. 11–13, the BP-NN model performs much worse than the WPD-based models. The MAPE errors of BP-NN in one-step, three-step and five-step predictions are 10.532%, 12.934% and 22.666%, respectively. The trend of errors demonstrates that it is difficult for BP-NN to obtain good multi-step forecasting results by direct use of the original data without multi-scale decomposition. The reason of the results is because WPD decompose the non-stationary wind speed into a series of relatively stable subseries which lower the forecasting difficulty of the ANNs. The results shown in Table 1 confirm the conclusion that the prediction accuracy can be greatly improved by WPD when applied to the ANNs trained by regardless of the BP or CSO algorithms.

In Table 1, it is also observed that WPD–CSO-NN obtains better forecasting results than WPD–BP-NN in varying degrees. The MAPE

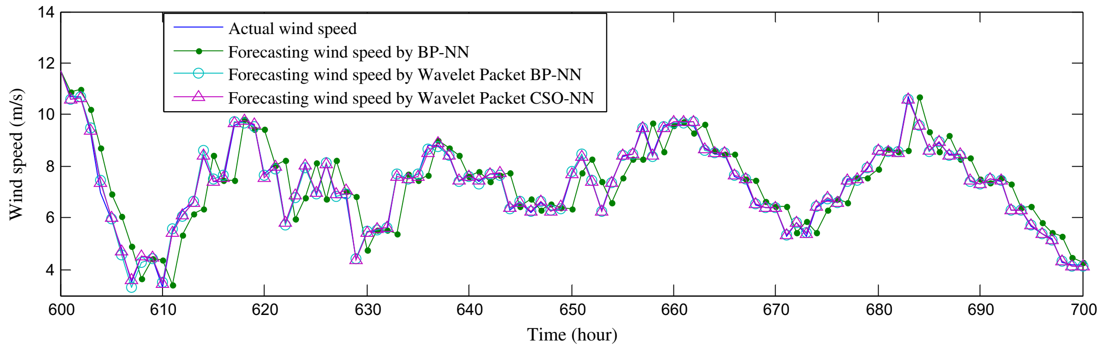

Fig. 11. One-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 1.

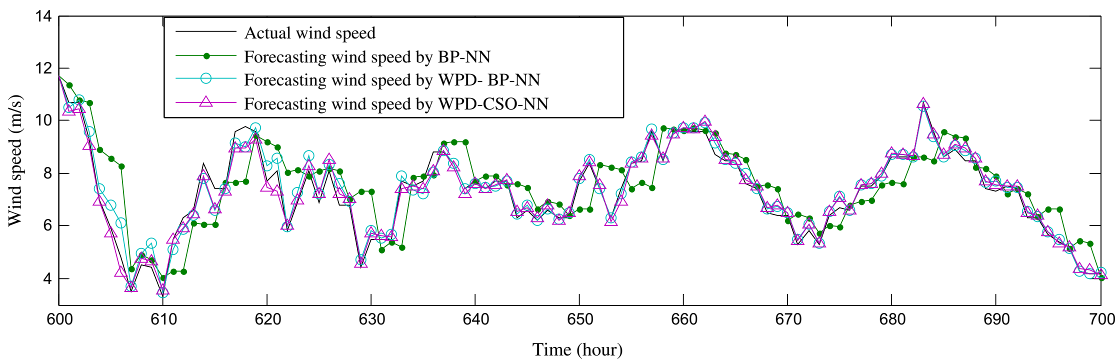

Fig. 12. Three-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 1.

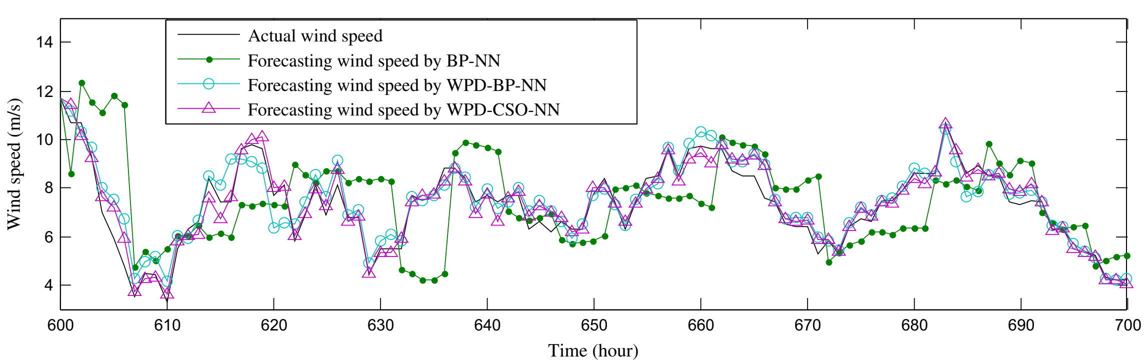

Fig. 13. Five-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 1.

Table 1
Results of errors for multi-step prediction shown in Figs. 11–13.

Index	BP-NN			WPD–BP-NN			WPD–CSO-NN		
	One-step	Three-step	Five-step	One-step	Three-step	Five-step	One-step	Three-step	Five-step
MAE (m/s)	0.704	0.865	1.550	0.064	0.247	0.398	0.063	0.198	0.363
MAPE (%)	10.532	12.934	22.666	1.015	3.683	6.409	0.939	2.928	5.408
RMSE (m/s)	0.938	1.135	1.933	0.088	0.377	0.639	0.084	0.280	0.531

errors listed in Table 1 show that compared to WPD–BP-NN, the MAPE errors of WPD–CSO-NN are reduced by 21.596%, 23.861%, 26.126%, respectively. It can be concluded that CSO is more effective than the conventional BP algorithm when applied to optimize the ANNs.

4.1.2. WPD–CSO-NN vs. other wind forecasting methods

To further estimate the performance of the proposed models, the WPD–CSO-NN is compared with WPD–PSO-NN, EMD-NN and WD–GA–SVM. For WPD–PSO-NN, it differentiates from the proposed WPD–CSO-NN in that the PSO-NN [5] instead of CSO-NN is used to forecast each WPD transformed subseries. WD–GA–SVM [24] and EMD-NN [26] are two recently developed hybrid models for wind speed prediction. The former integrates wavelet decomposition and support vector machines optimized by genetic algorithm and the latter combines empirical mode decomposition and artificial neural networks. In this test, the parameters of the compared models are set according to the corresponding references.

The results for one-step, three step and five step wind prediction are shown in Table 2 and Figs. 14–16. It is observed that the prediction accuracy is decreased as the forecast step increases. Obviously this is an inevitable phenomenon for any wind speed forecasting method. Compared with WPD–PSO-NN, the MAPE errors of WPD–CSO-NN are cut by 7.88%, 13.21% and 13.42% in one-step, three-step and five-step prediction, respectively. The results show that CSO has significant advantage over PSO in addressing the prematurity problems when applied to the training of large-scale ANNs with many local minima. As seen from Table 2, it is found that WPD–CSO-NN also outperforms the previous hybrid models like EMD-NN and WD-GA–SVM regardless of one-step, three-step or five-step prediction. The results further substantiate the effectiveness of the proposed hybrid approach for short-term wind speed prediction.

4.2. Case 2

To further verify the performance of WPD–CSO-NN, another case from the same wind farm at autumn season is used to establish models and do the multi-step ahead predictions. The data in case 2 are collected from October, 2014, as illustrated in Fig. 2. It is obvious that the wind speed series in this case is more non-stationary than that in case 1. The parameters are set the same as case 1. The forecasting results are shown in Figs. 17–22 and Tables 3 and 4. According to those results, the same conclusions to case 1 can be made.

Table 2
Results of errors for multi-step prediction shown in Figs. 14–16.

Index	One-step	Three-step	Five-step	One-step	Three-step	Five-step
	WPD–PSO-NN			*EMD-NN*		
MAE (m/s)	0.065	0.2163	0.3721	0.0655	0.2486	0.3821
MAPE (%)	1.013	3.3147	6.134	1.034	3.4121	6.2031
RMSE (m/s)	0.091	0.3113	0.5753	0.0927	0.3641	0.6573
	WD–GA–SVM			*WPD–CSO-NN*		
MAE (m/s)	0.065	0.2234	0.3682	0.063	0.198	0.363
MAPE (%)	1.011	3.2269	5.812	0.939	2.928	5.408
RMSE (m/s)	0.087	0.3081	0.5855	0.084	0.280	0.531

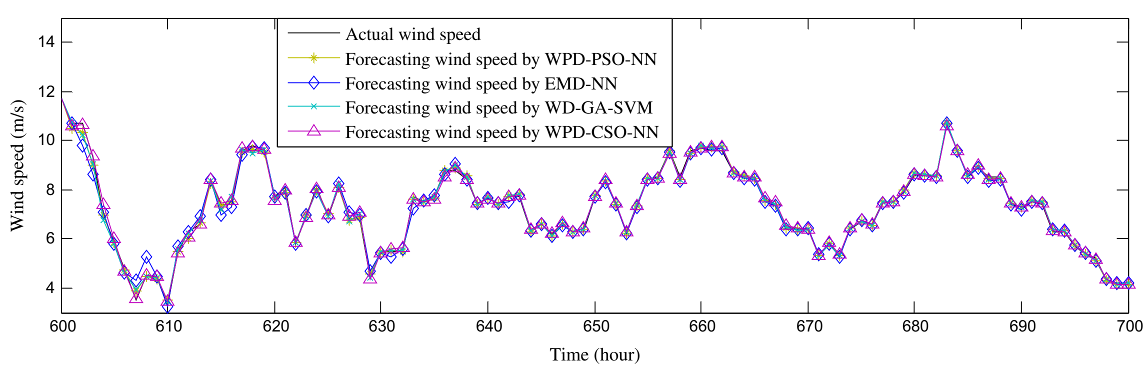

Fig. 14. One-step wind speed forecasting results by WPD–PSO-NN, EMD–NN, WD–GA–SVM and WPD–CSO-NN for case 1.

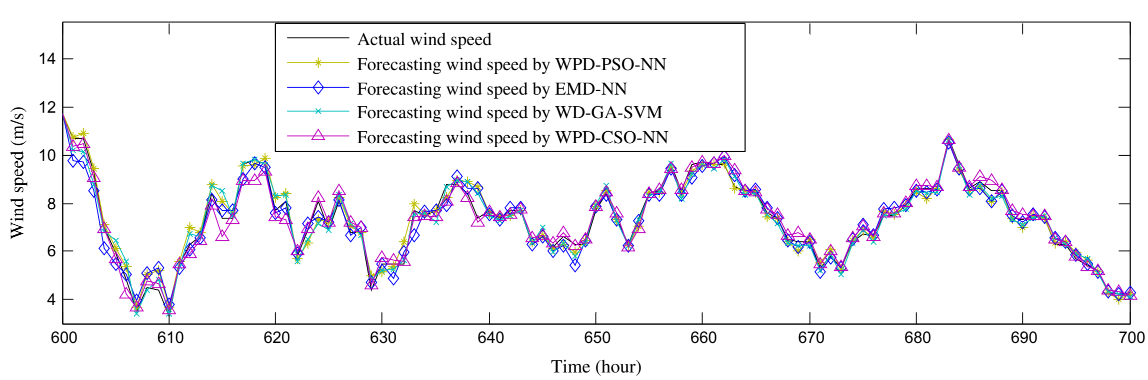

Fig. 15. Three-step wind speed forecasting results by WPD–PSO-NN, EMD–NN, WD–GA–SVM and WPD–CSO-NN for case 1.

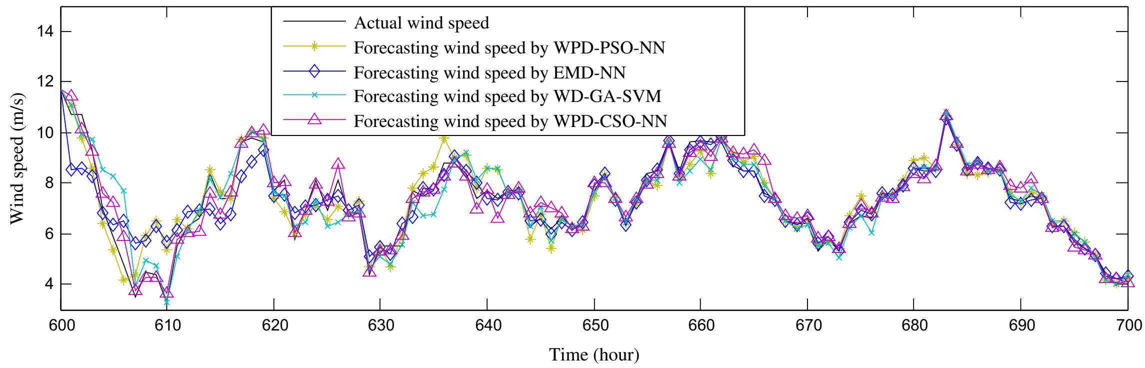

Fig. 16. Five-step wind speed forecasting results by WPD–PSO-NN, EMD-NN, WD–GA–SVM and WPD–CSO-NN for case 1.

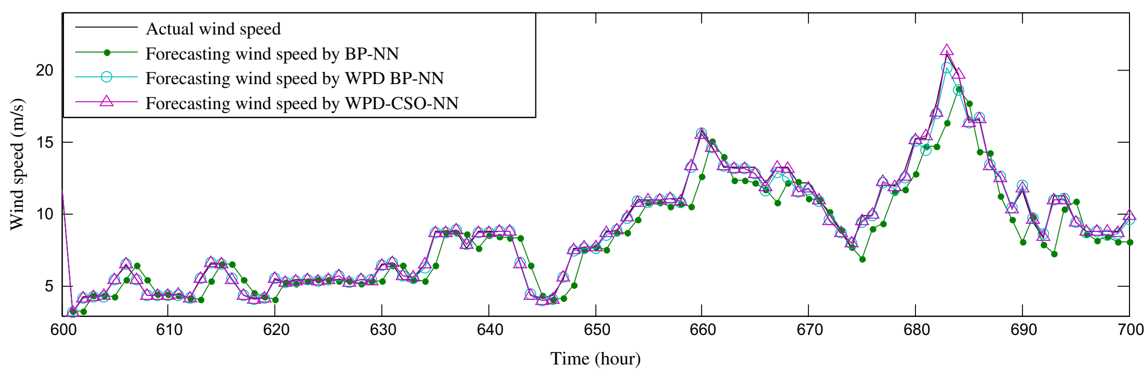

Fig. 17. One-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 2.

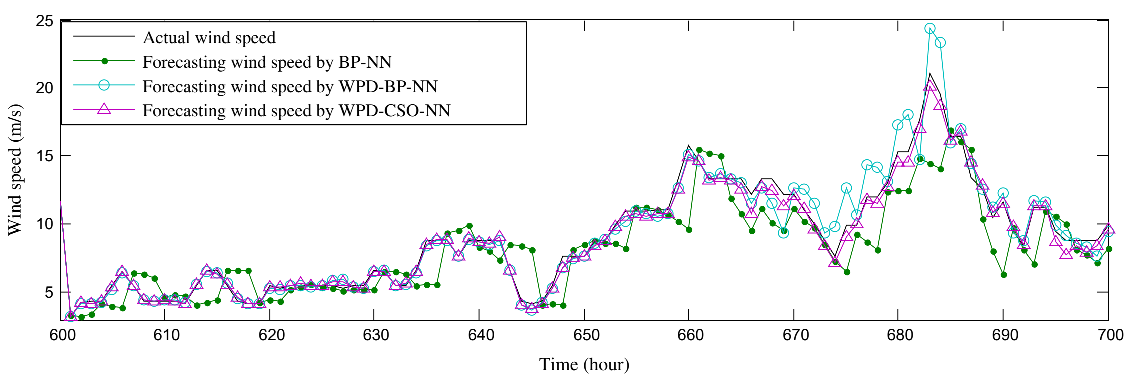

Fig. 18. Three-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 2.

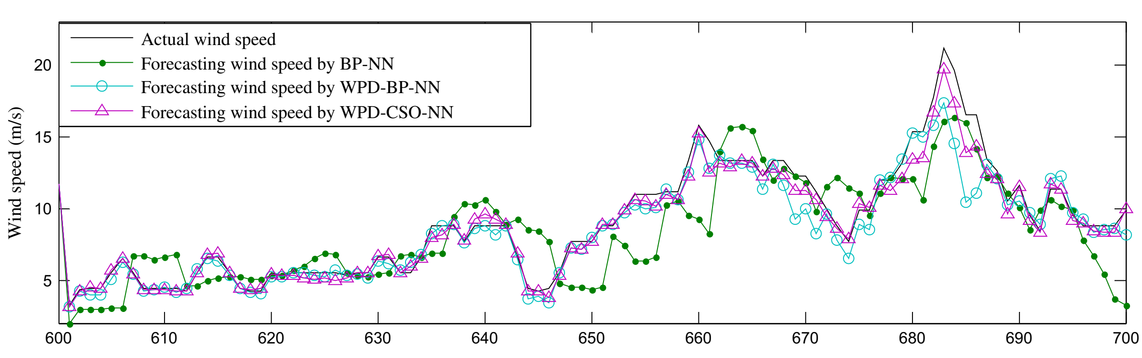

Fig. 19. Five-step wind speed forecasting results by BP-NN, WPD–BP-NN and WPD–CSO-NN for case 2.

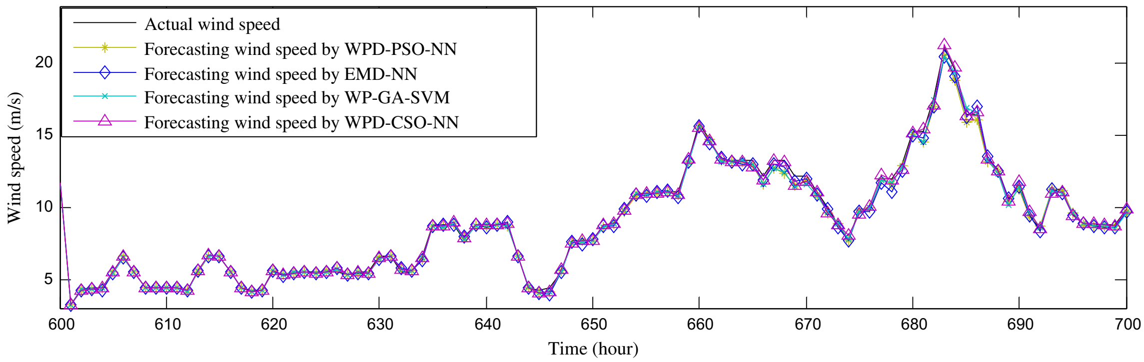

Fig. 20. One-step wind speed forecasting results by WPD–PSO-NN, EMD-NN, WD–GA–SVM and WPD–CSO-NN for case 2.

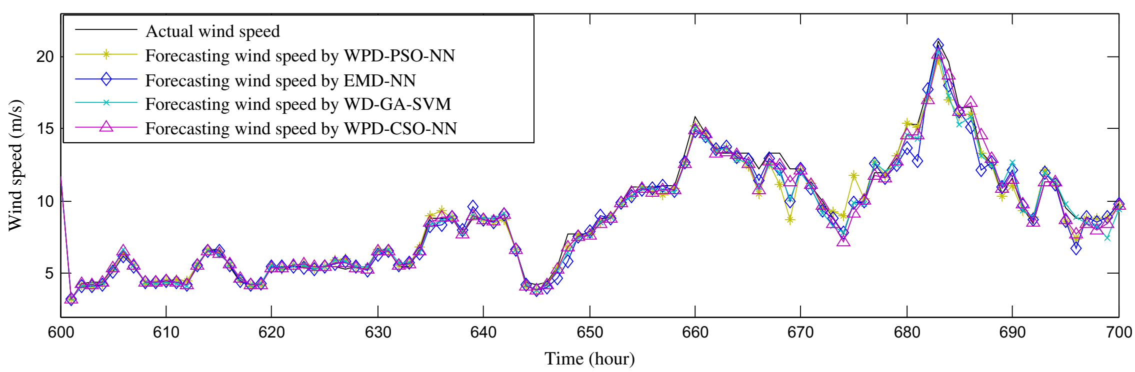

Fig. 21. Three-step wind speed forecasting results by WPD–PSO-NN, EMD-NN, WD–GA–SVM and WPD–CSO-NN for case 2.

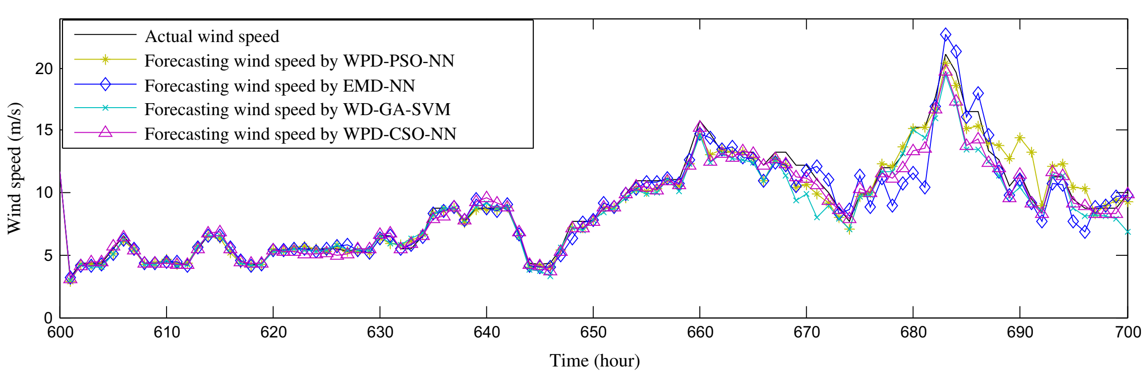

Fig. 22. Five-step wind speed forecasting results by WPD–PSO-NN, EMD-NN, WD–GA–SVM and WPD–CSO-NN for case 2.

Table 3
Results of errors for multi-step prediction shown in Figs. 17–19.

Index	BP-NN			WPD–BP-NN			WPD–CSO-NN		
	One-step	Three-step	Five-step	One-step	Three-step	Five-step	One-step	Three-step	Five-step
MAE (m/s)	0.9173	1.5471	1.8819	0.1705	0.4224	0.6840	0.1365	0.2791	0.4895
MAPE (%)	10.058	17.628	23.240	1.7363	4.4334	6.426	1.214	2.9181	4.7443
RMSE (m/s)	1.3332	2.0994	2.4387	0.2660	0.7306	1.0955	0.2216	0.4525	0.7248

Table 4
Results of errors for multi-step prediction shown in Figs. 20–22.

Index	One-step	Three-step	Five-step	One-step	Three-step	Five-step
	WPD–PSO-NN			*EMD-NN*		
MAE (m/s)	0.1578	0.3786	0.5176	0.1441	0.3819	0.5582
MAPE (%)	1.5696	3.8733	5.388	1.5291	3.9863	5.5034
RMSE (m/s)	0.2446	0.6747	0.7974	0.228	0.638	0.8322
	WD–GA–SVM			*WPD–CSO-NN*		
MAE (m/s)	0.1554	0.3665	0.5125	0.1365	0.2791	0.4895
MAPE (%)	1.549	3.7325	5.3243	1.214	2.9181	4.7443
RMSE (m/s)	0.2411	0.5574	0.7656	0.2216	0.4525	0.7248

5. Conclusions

In the literature, there exist many studies on the predictions of short-term wind speed. As pointed out by Tascikaraoglu and Uzunoglu [5], it is not possible to say exactly which method is the most appropriate candidate due to the fact that the prediction models developed are generally site-specific. To address this problem, combining several methods been deemed as a promising research area. In this paper, a new hybrid model named WPD–CSO-NN is proposed to predict the short-term wind speed at 1 h intervals up to 5 h. Some conclusions and contributions of this research are summarized as below.

(1) Many studies show that evolutionary algorithms such as GA and PSO tend to perform well for the selection and optimization of the prediction system parameters. In view of the superiority of CSO over other heuristic algorithms in addressing multimodal problems, we attempt to apply it to the optimization of ANN's weights and thresholds. The results shown in Tables 1–4 confirm CSO's powerful global search ability in addressing non-linear neural network optimization problems with many local minima.

(2) Considering that WPD can decompose both of the previous approximation and detail coefficients, we choose it as the data pre-processing method in the hybrid models. Correspondingly, three WPD-based prediction models including WPD–BP-NN, WPD–PSO-NN and WPD–CSO-NN are built and then tested in the multi-step wind speed prediction on a wind farm of the Netherlands. The results listed in Tables 1–4 show that the prediction performance can be improved by WPD when applied to the ANNs trained by regardless of the BP, PSO or CSO algorithms. It is also observed that WPD–CSO-NN outperforms WPD–BP-NN and WPD–PSO-NN in varying degrees, which further confirm the superiority of CSO over PSO and the conventional BP algorithm when applied to optimize the ANNs.

(3) As is known that the effectiveness of the developed models for wind speed prediction is relatively location dependent. Therefore, the direct comparison of results among different models makes no sense. To further compare WPD–CSO-NN with previous studies, we use two recently developed hybrid models (i.e., WD–GA–SVM [24] and EMD-NN [26]) to do the same wind speed forecasting in case 1–2. As seen from Tables 2 and 4, the proposed WPD–CSO-NN model consistently has the minimum value of MAE, MAPE and RMSE regardless of one-step, three-step or five-step prediction. It can be concluded that the proposed hybrid model can improve the forecasting performance in an effective way.

(4) As seen from Fig. 10, the only control parameter Pv of CSO seems to have little influence on the prediction accuracy once it is set in the rang [0.2, 1]. Such feature exhibits the strong robustness of the proposed WPD–CSO-NN model.

In view of CSO's good performance on the neural network training, we are preparing to use it to optimize other prediction tools like Elman neural network and support vector machine. Another direction is to explore the possibility of combining other decomposition techniques with CSO-based ANNs for different forecasting purposes in the new energy field.

Acknowledgements

This research was supported by the National Natural Science Fund Project of China (Grant No. 51307025). The authors would like to thank Royal Netherlands Meteorological Institute for providing the original wind speed data of this research. In addition, the authors also appreciate the editor and the anonymous reviewers whose useful editorial suggestions and valuable comments contribute much to improve the quality of this paper.

References

[1] Akda SA, Güler N. A novel energy pattern factor method for wind speed distribution parameter estimation. Energy Convers Manage 2015;106:1124–33.
[2] Sun D, Xu Y, Chen H, Wu K, Liu K, Yu Y. A mean flow acoustic engine capable of wind energy harvesting. Energy Convers Manage 2012;63:101–5.
[3] Duic N, Rosen MA. Sustainable development of energy systems. Energy Convers Manage 2014;87:1057–62.
[4] Zhang W, Wang J, Wang J, Zhao Z, Tian M. Short-term wind speed forecasting based on a hybrid model. Appl Soft Comput 2013;13:3225–33.
[5] Ren C, An N, Wang J, Li L, Hu B, Shang D. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl-Based Syst 2014;56:226–39.
[6] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34:243–54.
[7] Huang Z, Chalabi ZS. Use of time-series analysis to model and forecast wind speed. J Wind Eng Ind Aerodyn 1995;56:311–22.
[8] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[9] Su Z, Wang J, Lu H, Zhao G. A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. Energy Convers Manage 2014;85:443–52.
[10] Lydia M, Kumar S Suresh, Selvakumar A Immanuel, Kumar G Edwin Prem. Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers Manage 2016;112:115–24.
[11] Kani SA Pourmousavi, Ardehali MM. Very short-term wind speed prediction: a new artificial neural network–Markov chain model. Energy Convers Manage 2011;52:738–45.
[12] Cadenas E, Rivera W. Short term wind speed forecasting in La Venta, Oaxaca, Mexico, using artificial neural networks. Renew Energy 2009;34:274–8.
[13] Men Z, Yee E, Lien FS, Wen D, Chen Y. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. Renew Energy 2016. http://dx.doi.org/10.1016/j.renene.2015.10.014.
[14] Kong X, Liu X, Shi R, Lee K. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[15] Rumelhart DE, Hintont GE, Williams RJ. Learning representations by back propagating errors. Nature 1986;323:533–6.
[16] Zhang L, Wu K, Zhong Y, Li P. A new sub-pixel mapping algorithm based on a BP neural network with an observation model. Neurocomputing 2008;71:2046–54.
[17] Yu F, Xu X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. Appl Energy 2014;134:102–13.

[18] Pousinho HMI, Mendes VMF, Catal OJPS. A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal. Energy Convers Manage 2011;52:397–402.
[19] Rahmani R, Yusof R, Seyedmahmoudian M, Mekhilef S. Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting. J Wind Eng Ind Aerodyn 2013;123:163–70.
[20] Chitsaz H, Amjady N, Zareipour H. Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm. Energy Convers Manage 2015;89:588–98.
[21] Meng A, Mei P, Yin H, Peng X, Guo Z. Crisscross optimization algorithm for solving combined heat and power economic dispatch problem. Energy Convers Manage 2015;105:1303–17.
[22] Meng A, Hu H, Yin H, Peng X, Guo Z. Crisscross optimization algorithm for large-scale dynamic economic dispatch problem with valve-point effects. Energy 2015;93:2175–90.
[23] Wang J, Zhang W, Li Y, Wang J, Dang Z. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl Soft Comput 2014;23:452–9.
[24] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62:592–7.
[25] Hu J, Wang J, Zeng G. A hybrid forecasting approach applied to wind speed time series. Renew Energy 2013;60:185–94.
[26] Liu H, Chen C, Tian H, Li Y. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. Renew Energy 2012;48:545–56.
[27] Liu H, Tian H, Pan D, Li Y. Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. Appl Energy 2013;107:191–208.
[28] Liu H, Tian H, Li Y. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Convers Manage 2015;100:16–22.
[29] Wind climate assessment of the Netherlands. URL: http://www.knmi.nl/samenw/hydra.
[30] Liu H, Tian H, Li Y, Zhang L. Comparison of four Adaboost algorithm based artificial neural works in wind speed predictions. Energy Convers Manage 2015;92:67–81.
[31] Meng A, Chen Y, Yin H, Chen S. Crisscross optimization algorithm and its application. Knowl-Based Syst 2014;67:218–29.
[32] Mallat SG. A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 1989;11:674–93.
[33] Liu H, Tian H, Chen C, Li Y. 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.