Renewable and Sustainable Energy Reviews 44 (2015) 271–288

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Combined forecasting models for wind energy forecasting: A case study in China

Ling Xiao [a], Jianzhou Wang [b,*], Yao Dong [a], Jie Wu [a]

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

ARTICLE INFO

Article history:
Received 9 August 2014
Received in revised form
5 December 2014
Accepted 12 December 2014
Available online 6 January 2015

Keywords:
Wind speed forecasting
Review
Combined approach
No negative constraint theory
Artificial intelligence algorithm

ABSTRACT

As the energy crisis becomes a greater concern, wind energy, as one of the most promising renewable energy resources, becomes more widely used. Thus, wind energy forecasting plays an important role in wind energy utilization, especially wind speed forecasting, which is a vital component of wind energy management. In view of its importance, numerous wind speed forecasts have been proposed, each with advantages and disadvantages. Searching for more effective wind speed forecasts in wind energy management is a challenging task. As proposed, combined models have desirable forecasting abilities for wind speed. This paper reviewed the combined models for wind speed predictions and classified the combined wind speed forecasting approaches. To further study the combined models, two combination models, the no negative constraint theory (NNCT) combination model and the artificial intelligence algorithm combination model, are proposed. The hourly average wind speed data of three wind turbines in the Chengde region of China are used to illustrate the effectiveness of the proposed combination models, and the results show that the proposed combination models can always provide desirable forecasting results compared to the existing traditional combination models.

Contents

* Corresponding author. Tel.: +86 15339864602; fax: +86 41184710484.
E-mail address: wjzdufe@gmail.com (J. Wang).

http://dx.doi.org/10.1016/j.rser.2014.12.012
1364-0321/

1. Introduction

Energy is critically important to the social and economic development of any nation. With the increasing industrial and economic activity over the last several decades, energy demands have grown rapidly [1]. According to the *IEA World Energy Outlook 2010*, the energy demands of China and India will account for half of the growth in global energy demands by 2050. At that point, China's energy consumption will be nearly 70% greater than that of the United States, the second greatest energy-consuming country, thus placing China as the leading energy-consuming country in the world. However, China's per capita energy consumption will still be less than half that of the United States [2]. With increasing demands for energy, traditional resources such as oil and coal remain the dominant sources of energy worldwide. However, the use of traditional resources results in the release of significant amounts of carbon dioxide, which causes serious environmental problems, especially with respect to climate change. Climate change has been recognized as an international security threat, and as such, it no longer relates only to quality of life and the environment, but it also directly affects human and global security [3]. Energy crises and climate change make renewable energy necessary. Hydroelectric, wind, biomass, solar, terrestrial heat, clean coal and nuclear energies are rapidly being developed. As a clean energy source and because of its low cost of production, wind energy is often viewed as an attractive energy option. Accordingly, wind energy has achieved maturity in the energy market and, compared to the aforementioned alternatives, has experienced the greatest growth worldwide in the past several decades. Being both convenient and environmentally friendly, wind energy meets the growing demand for electricity. Furthermore, with the cost of electricity from non-renewable sources continuing to increase, wind energy is becoming increasingly competitive [4].

Introduced 20 years ago, China became one of the world's first countries to use wind energy to generate electricity, and its development since then has been rapid. In June 2012, China's grid-connected wind power reached 52.58 million kilowatts (KW), surpassing that of the United States, the country that previously ranked first in the world in wind energy capacity with a cumulative installed capacity of 44,733.29 MW (MW). In China's national power grid, dispatch reached 50.26 million kilowatts (kW), making it the largest global wind power grid and the fastest growing power grid [5]. Reaching 60.83 million kilowatts in 2012, China's total wind power grid has been ranked first in the world for two consecutive years. Wind power accounts for 2% of the country's total generating capacity with an annual generating capacity of over 100 billion kilowatts (kW). Thus, wind power has become the third largest source of power and plays a prominent role in optimizing the energy structure and promoting energy conservation in China [6]. Given that wind power generation depends on wind speed, the main problem with wind power is that the wind speed is subject to significant fluctuations that are harmful to the wind turbines. Because the fluctuating wind speeds must then be processed and used to generate power, obtaining accurate wind speeds is important. However, because wind is air motion whose driving force is due to the uneven heating and cooling of the earth's surface and the occurrence of wind is highly uncertain in time and space because it depends on many weather factors such as pressure and temperature, obtaining accurate forecasting results for wind speed can be challenging [7].

Accordingly, many significant studies have been devoted to developing efficient forecasts for wind speed that can be divided into two categories. The first is the physical method that uses numerous physical considerations for the best forecasting accuracy. The second is the statistical method that aims at finding the relationship between the online measured power data, including traditional statistical models (such as ARIMA models, ARCH models, Kalman Fitters (KF) etc.) and machine learning (ML) models. Artificial neural networks (ANNs), as a class of methods in ML models, have been widely used in a broad range of applications [8]. Moreover, ANNs are improved and combined with other methods for improved prediction accuracy. The ANN with a statistical weighted pre-processing method (SWP–ANN) can be used to predict ground source heat pump (GCHP) systems with the minimum data set, and the simulation results show that SWP–ANN performs better than ANN [9]. Fuzzy models can be combined with ANNs to create ANFIS, which demonstrate reliable forecasting applicability [10]. However, ANNs do not always perform well, and at times the proposed wavelet neural network (WNN) does better than ANN [11]. ANNs are also widely used in wind speed prediction where their performance depends on the training data sets [12]. Support vector machine (SVM), another ML method, has had rapid development in recent years, and its performance proved to be desirable. For example, SVM has been shown to have excellent performance compared to an ANN model and an ANFIS model [13]. The least-squares support vector machine (LS-SVM) method can be used to make a modeling study of the solar air heater (SAH) system and for estimating the efficiency of SAHs with reasonable accuracy [14]. Moreover, some hybrid wind speed forecasts have also been put forward [7,9]. Though the statistical method is more seasonal to forecast wind speed compared to the physical method, it does not always capture the relationship between different data and obtain accurate forecasting results. Therefore, some hybrid wind speed forecasts are proposed; when compared with individual forecasts, the hybrid forecasts have demonstrated outstanding forecasting results. However, hybrid forecasts are based on only two or three individual forecasts. When more individual forecasts are included, using hybrid forecasts becomes difficult. Therefore, to make full use of the advantages of individual forecasts while not increasing the simulation difficulties, combination forecasts have been proposed.

A combination forecast was proposed in 1969 by Bates and Granger with promising results [16]. After the 1970s, the study of combination forecasts received significant attention [17]. Many researchers devoted themselves to the study of combination forecasts, resulting in numerous articles about combining different forecasts and the application of combination forecasts [12–15]. This paper studies the combination of different individual forecasts to forecast wind speed. At present, the combination forecasts are grouped into the constant weight combination forecast method or the variable weight combination forecast method [18]. This paper focuses on the constant weight combination, which is based on the minimum sum of the forecast error square. Initially, the traditional combination method is adopted to combine the five individual forecasts. The simulation results show that if the combination methods are largely different or similar in forecasting accuracies, the best forecast would be assigned 1 and the rest of them would be assigned 0. To overcome the disadvantages of the traditional combination problem previously mentioned, the no negative constraint combination method and the PSO weight-determined combination method are proposed for combining the five individual forecasts. The simulation results proved to be better than the traditional combination method in forecasting accuracy.

Section 2 of this paper gives a review of wind speed prediction approaches. In Section 3, the combination theory is introduced. Section 4 presents the five individual forecasting algorithms applied in the wind speed forecasting, and Section 5 proposes a CPSO–BP model for wind speed forecasting. Section 6 focuses on a case study and analyzes the simulation results. Finally, Section 7 presents the conclusions of this paper.

2. A review of the wind speed prediction approaches

In the last few decades, many wind speed forecasting approaches have been studied and proposed; each model uses different techniques and performs well based on different data sets and prediction horizons. However, according to many studies in the wind speed prediction field, the fact that these models cannot obtain satisfactory results for all situations is widely accepted. This section presents an overview of the wind speed forecasting models.

2.1. A review of the existing wind speed prediction approaches

The proposed models can be categorized as physical models, traditional statistical models, artificial intelligence and machine learning models or other hybrid models [19]. Physical models use the meteorological information to predict the long-term wind speed and are only the first step in forecasting the wind speed. Physical models include the numerical weather prediction (NWP) model that, because of the limitation of the currently computer facilities and the shortcomings in horizontal resolution, physical parameterizations, initial and boundary conditions [20], is undesirable to be used for forecasting wind speed alone and is always used for auxiliary input of statistical models [21]. As an efficient forecasting method, the weather research forecasting (WRF) model is also used to forecast wind speed by taking the complexity of the terrain and the resolution of the selected domain into account, which has proved to have good performance in short-term wind speed forecasting [22]. The most commonly used traditional statistical model is the autoregressive integrated moving average (ARIMA) model that uses historical data to establish a forecasting model, which can capture the linear relationship of the training data set well. With the development of statistical models and artificial techniques, artificial intelligence models have been proposed [14] to develop forecasting models and are used to optimize the parameters of the established models to form the hybrid models, which have desirable performance [23]. The newly developed models based on artificial intelligence and machine learning such as the artificial neural network (ANN), the fuzzy logic (FL) methods and the support vector machine (SVM) are used widely in the wind speed forecasting field [14]. However, no method is always suitable for all data sets [23]. In further studies, hybrid models based on conventional statistical models and artificial intelligent models have been proposed to improve wind speed forecasting accuracies [18,19]. Large amounts of research have been conducted on hybrid models that integrate the single prediction models in order to provide advanced forecasting models. However, reviewing many studies about wind speed forecasting revealed that the universally accepted definitions for these approaches have not been agreed upon [19].

2.2. The combined wind speed prediction approaches

In many simulation results, a final combined approach that takes advantage of component forecasts should be better than the individuals, or at least equivalent to the best one, making it desirable to combine individuals to forecast wind speed [Ismael Sánchez]. Combined forecasting methodologies aggregate individual forecasting methods and take advantage of component models in order to improve the final forecasting performance. The literature contains controversy and confusion about the definition and structure of the combined models. The most widely accepted procedure is to assign a weighting coefficient to each method according to past forecasting performance. Meanwhile,

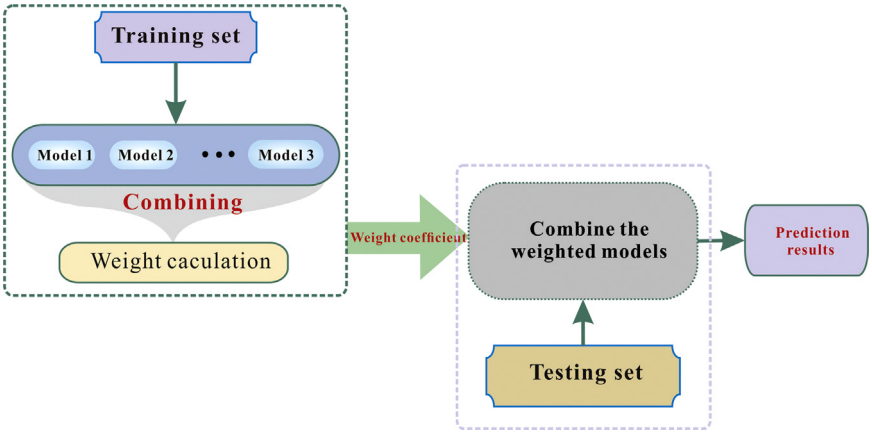


Fig. 1. Flowchart for the weighting-based combined approaches.

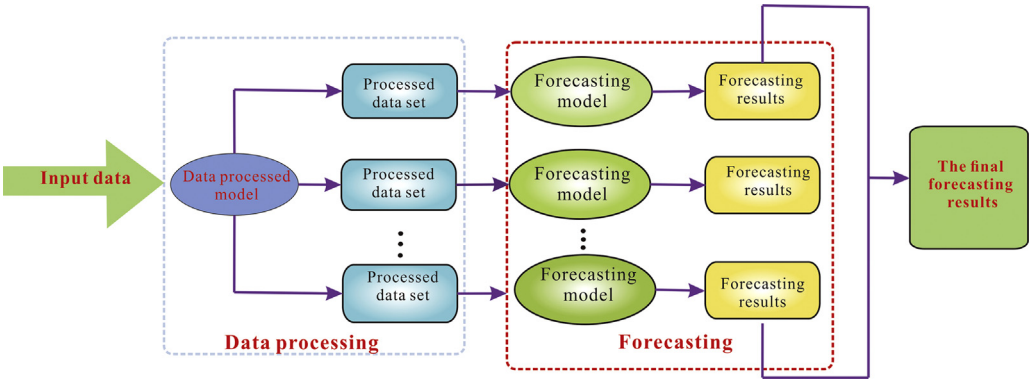


Fig. 2. Flowchart for the data pre-processing technique combination approaches.

other forecasting models based on aggregating different individual forecasting methodologies are also regarded as combined forecasting models, such as some hybrid approaches based on parameter optimization, data set pre-processing technique and error-processing technique [23]. In much of the literature, these approaches consist of two or three different models, with one for the main forecasting task and the rest for auxiliary processes. In this way, the contributions of the component models are anthropogenic rather than determined by their past performance.

2.2.1. Weighting-based combined approaches

The most widely accepted approach for combining individual models is to assign a weighting coefficient to each individual, namely, weighting-based combined approaches. As shown in Fig. 1, these approaches assign a weighting coefficient to each component model according to their performance in the fitting procedure (by examining the measured and predicted values) and the final combination forecasting results can be calculated as follows:

$$\hat{F}_{t+h} = \sum_{i=1}^{M} w_i \hat{f}_{i,t+h|t} \tag{1}$$

where $\hat{F}_{t+h}$ is the combination forecasting result at h steps ahead based on the performance at time t; w_i is the weight of the ith component model, which should satisfy $\sum_{i=1}^{M} w_i = 1$; $\hat{f}_{i,t+h|t}$ is the forecasting value of the ith component model at h steps ahead based on the performance at time t and M is the number of combined models. The weighting coefficient can be calculated in various ways, and then the combination results can be calculated by the determined weighting coefficients, such as weight average (WA) [18] (namely, each w_i is $1/M$), time-varying weighted combination model [25] etc.

As an example of weighting-based combined approaches, a recent research paper presented the multiple architecture system (MAS), which was developed by combining different wind speed forecasting models. This literature has taken a data fusion approach, and three fusion strategies including average strategy (AS), weighted strategy (WS) and non-linear strategy are proposed. Assuming that training sample $x_i \in R^d (i = 1, 2, \ldots, N)$, a target $y_i \in R (i = 1, 2, \ldots, N)$ is the real value of wind speed (used to calculate the fitting errors and forecasting errors). Let us consider a set of M predictors $f_i(x) (i = 1, 2, \ldots, M)$. ϕ is a function defined on the real data set. The final problem is to define a combination strategy as follows:

$$F(x) = \varnothing\{f_1(x), f_1(x), \ldots, f_1(x)\} \tag{2}$$

AS is a very simple strategy, and the results can be estimated by Eq. (3). The final prediction formula is given by Eq. (4).

$$F_{AS}(x) = \frac{1}{M}\sum_{i=1}^{M} f_i(x) \tag{3}$$

$$F_{WS}(x) = \sum_{i=1}^{M} \beta_i f_i(x) \tag{4}$$

where β_i is the reliability weight assigned to the predictor. The problem in Eq. (4) is the determination of the weight, which can be written as

$$\begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_M(x_1) \\ f_1(x_2) & f_2 d(x_2) & \cdots & f_M(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ f_1(x_N) & f_2 d(x_N) & \cdots & f_M(x_N) \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_M \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_M \end{bmatrix} \tag{5}$$

and can also be rewritten as

$$\overline{F} \cdot \overline{\beta} = \overline{Y} \tag{6}$$

An interesting property of WS, when the proper values of the reliability weight are adopted, is that the combination model is less sensitive to each predictor respective bias and variance than the AS strategy. The non-linear strategy is very useful when the output predictors have a complex relationship that cannot be simulated by the simple linear model. The motivation behind the MAS is to combine the complementary predictive power of multiple models [19].

2.2.2. Data pre-processing technique combined approaches

Unlike the other combination approaches, these combination approaches process data sets by decomposing the wind speed time series into more stationary and regular subseries that are easy to identify [25] and then give each decomposed part a suitable prediction (without considering weights), before combining the decomposed prediction results (Fig. 2). The measured wind speeds are preprocessed and produce standard deviation, average and slope; these three parts are then put into a fuzzy or neural predictor, and, finally, the estimated wind speeds are collected, which demonstrates the desirable forecasting ability [26]. Wavelet transformation (WT) is a new technique for time–frequency analysis. CAO and LI have used WT to decompose wind speed data into different frequency signals, before the decomposed time series are described as ARIMA models and the combined method is used to obtain the forecasting results of original wind speed [28]. As a form of adaptive time series decomposition technique, empirical mode decomposition (EMD) is used to decompose the wind speed time series data, artificial neural networks (ANNs) are used to forecast each subseries and the prediction results of the

subseries are summed to formulate an ensemble forecast for the original wind speed series [29]. Zhou et al. used statistical clustering approaches to classify wind data before modeling. With the help of characteristic parameters on anticipated days provided by the metrological department, a group of data is filtered according to the criterion of maximal similarity, which means that the wind speed on these days has strong similarity to that of the anticipated day. Then, filtered data are used for modeling [30]. The ANN–MC model for forecasting wind speed in a very short-term time scale has also been developed [31]. In this model, two ANNs are utilized; the first ANN (ANN-1) is used for primary prediction and capturing short-term trends in wind speed signal, and the second ANN (ANN-2) for the final prediction. Two FNIs, two BNIs and TPM, generated by the MC approach together with the primary predictions, are fed as input variables to a second ANN.

2.2.3. Parameter selection and optimization technique combined approaches

A great deal of research on wind speed forecasting show that certain parameter selection and optimization approaches can also make a considerable contribution to the prediction performance during the training process [19]. As mentioned in many studies, artificial intelligence optimization algorithms are very popular in parameter selection models. Much literature has focused on genetic algorithms (GAs) for their fast convergence to global optima. Considering that back-propagation neural network (BPNN) could fall into local minimum easily and could not obtain the optimal value when the structure of the network is complex, slowing convergence, GA is used to optimize the weights and bias of BPNN [32]. In many other papers, GA is also used to optimize or select parameters of statistical models, such as support vector machines (SVMs) and autoregressive and moving average (ARMA) models. The particle swarm optimization (PSO) algorithm has been applied to optimize the parameters of the ARIMA model for its advantage of few assumptions needed and no a priori postulation of the models required [24]. Another popular artificial intelligence optimization algorithm is the simulated annealing (SA) algorithm, which has been used to select the appropriate parameters of support vector regression model [33]. Fig. 3 shows the flowchart of the parameter selection and optimization technique combined approaches.

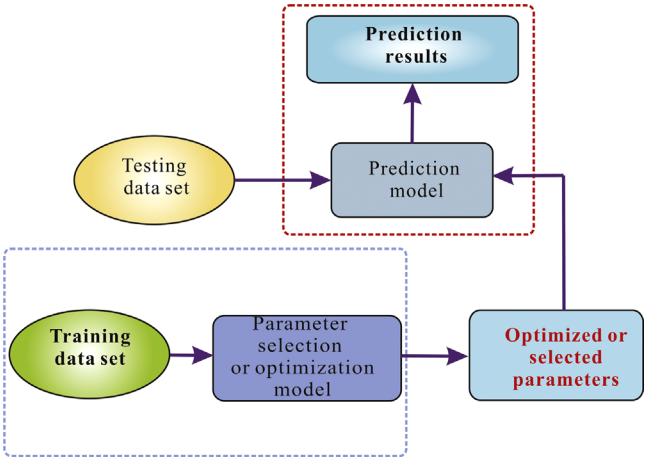


Fig. 3. Flowchart for the parameter optimization technique combined approaches.

2.2.4. Data post-processing technique combined approaches

The above approaches combine individual models before giving a final prediction. However, these combined models do not take into account the residual error values obtained from a forecasting model. Other literature has focused on how to diminish the negative effects of the systematic errors [19]. In these models, the final prediction models are revised by the residual error prediction models. The basic process of these models has three steps: (1) construct the primary prediction model and calculate the residual errors, (2) analyze the residual errors and construct the residual errors prediction models and (3) use the residual errors prediction models to revise the primary models and obtain the final prediction models (Fig. 4). Louka et al. used Kalman filtering as a post-processing method in numerical predictions of wind speed. The application of the Kalman filter to these data leads to the elimination of any possible systematic errors, and the results obtained showed a remarkable improvement in the model forecasting skill [34]. Similarly, the Grey–Markov chain model and the least squares support vector machine model are used to predict wind speeds; the Grey model is used to forecast wind speed values and the Markov chain model, which is used as the post-processing method, forecasts the wind speed residual errors [35].

2.2.5. A brief comparison of the combined approaches for wind speed

As shown in Sections 2.2.1–2.2.4, the combined approaches for wind speed are listed, and each combined approach has good performance in certain situations. However, the literature has proven that no single model is always better than others for different data sets. Table 1 shows a brief evaluation of combined approaches for wind speed prediction.

3. Combination theory

As the widely accepted method to improve forecasting accuracies, the combination model is very popular. This study focuses on the combination models based on different weight-determined ways. Let $\hat{x}_t$ denote the output of the individual forecasts at time t; then the output of the combination models can be denoted as [27]

$$\hat{x}_t^c = \sum_{i=1}^{m} l_{it}\hat{x}_t \tag{7}$$

where m is the number of the combined predictions and l_{it} is the weight of the model i. In this light, determining the weight of the combined models is the most important. To obtain the optimal weight, the combination method tries to minimize the sum of

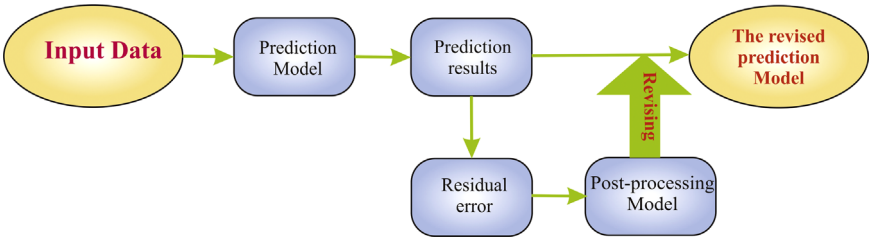


Fig. 4. Flowchart for the post-processing technique combined approaches.

Table 1
The brief evaluation of the combined approaches applied in wind speed prediction [17].

Combined approaches	Strategy	Advantages	Disadvantages
Weighting-based combined	Assigning weight factors to models according to their performance	Easy to implement and code, suitable for a wide range of prediction time, adaptive to new data	Does not guarantee the best predictions along the prediction horizon, requires an extra model for determining the weights
Data pre-processing techniques-based combined approaches	Forecasting of the subseries obtained by decomposition models	Higher performance, easy to find literature examples, robustness to rapid changes in wind speed	Requires a detailed mathematical knowledge on decomposition models, provides slow response to new data
Parameter selection and optimization techniques-based combined approaches	Optimization of the parameters of forecasting model	Easy to find literature examples, a relatively basic structure	Harder to code, dependent on designer's knowledge about the optimization problems, computationally intensive
Data post-processing techniques-based combined approaches	Forecasting of residual error caused by forecasting model	High accuracy, effective in reducing systematic error	Computational time inefficiency

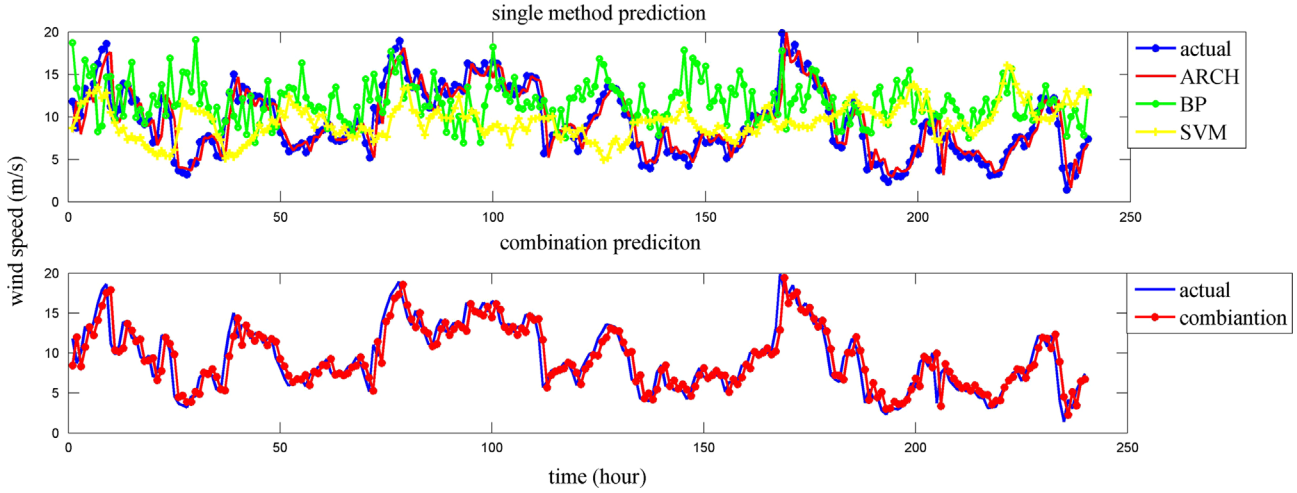


Fig. 5. Comparison of the individual forecasts and combination forecasting.

squared errors (SSE) of the combined output, which can be denoted as the following [20]:

$$SSE = \sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} \tag{8}$$

$$e_t = x_t - \hat{x}_t = \sum_{i=1}^{m} l_i e_i \tag{9}$$

where e_{it} is the error of the ith method at time t and $\hat{x}_t$ is the forecasting value of the ith method at time t.

3.1. Traditional combination theory

The traditional forecasting combination method attempts to find the best weight of the combined models based on minimizing SSE:

$$\min\ \mathbf{J} = \mathbf{L}^T\mathbf{E}\mathbf{L} = \sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} \begin{cases} \mathbf{R}^T\mathbf{L} = 1 \\ \mathbf{L} \geq 0 \end{cases} \tag{10}$$

where $\mathbf{L} = (l_1,\ l_2, \ldots,\ l_m)^T$ is the weight vector; $\mathbf{R} = (1,\ 1, \ldots,\ 1)^T$ is a column vector where all elements are 1 and $\mathbf{E}_{ij} = \mathbf{e}_i^T\mathbf{e}_j$ and $\mathbf{e}_i = (e_{i1}, e_{i2}, \ldots,\ e_{iN})$. $\mathbf{E}$ is called the combination forecasting $\mathbf{E} = (\mathbf{E}_{ij})_{m\times m}$ error information matrix [18].

3.2. The proposed no negative constraint combination theory (NNCT)

While many articles [30–36] have discussed how to determine the weights of combination forecasts, some of those articles are based on positive weight determination and have a negative attitude toward negative weights. Whether the combination forecasting weights can be negative is a controversial issue because many researchers believe that weights are perceived as having forecasting credibility or as representing investment proportion and if weights are negative, the meaning becomes confusing and convoluted. Determining positive weights has been the subject of numerous studies [33,35,36], and positive weights have been applied in many areas [38]. While there is no more convincing proof or reason for the weights of the forecasting combination to be positive, the articles referenced and discussed thus far are based on positive weights.

However, in traditional forecasting combination, the forecasting results indicate that if the combination forecasts are largely different or similar in forecasting accuracy, the weight of the best forecast would be assigned 1 and the worst would be assigned 0. The results show that these phenomena do not occur when the weights are negative. Furthermore, the forecasting accuracies improve when negative weights are assigned. From the overall forecasting trend, when the trend of the forecasting series approaches the original value series, the forecasting result is more accurate. Thus, if the forecasting series trend is opposite that of the original series trend, a negative weight is assigned to the forecast; otherwise, a positive weight is assigned. This theory explains part of the phenomena of forecasting combinations when three or more forecasts are combined. An example is presented in Fig. 1 that compares the forecasting results of three individual models (ARCH, BP and SVM) and their combination model for wind turbine 1 in the Chengde region of China. Fig. 5 shows a positive relationship between the ARCH model and the actual time series. In contrast, the SVM model forecasting trend reveals a negative

relationship to the actual series, while the BP model shows an uncertain relationship. The performance of the combination model of ARCH, SVM and BP is better than the best individual models.

3.3. Traditional combination method based on the proposed no negative constraint theory (TCM–NNCT)

Basing on the proposed no negative constraint theory (NNCT), the improvement of the traditional combination, TCM–NNCT, is given as the following:

$$\min \ \mathbf{J} = \mathbf{L}^T \mathbf{E} \mathbf{L} = \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} l_i l_j e_{it}$$
$$\text{st} \ \ \mathbf{R}^T \mathbf{L} = 1 \tag{11}$$

In Eq. (11), the weight vector has no limitation of in the range [0, 1]. The experiment results show that the combination model can obtain desirable results when the weight vector has a value in the range $[-2, 2]$. This section provides a weight-determined method that was assessed by experiment simulation rather than a theoretical proof.

3.4. Artificial intelligence algorithms that determine weights based on no negative constraint theory (AI–CM–NNCT)

In this section, two artificial intelligence algorithms, the chaos particle optimization algorithm and the genetic algorithm, are proposed to determine weights of combined models.

3.4.1. Chaos particle swarm optimization algorithm for weights based on no negative constraint theory (CPSO–CM–NNCT)

The particle swarm optimization algorithm was introduced by James Kennedy and Russell Eberhart in 1995 as a computational intelligence technique [39] and has been applied to non-linear and non-continuous optimization problems with continuous variables. Each individual is assigned a randomized velocity according to its own and its companions' experiences. While the particle swarm optimization algorithm is simple, it easily falls into local extreme points such as slow convergence of the latter's evolution and a low degree of accuracy. To remove particles from the location of the local optimum value, the chaos particle swarm optimization (CPSO) is used to determine the weight of the combination forecasts. Chaos is a common non-linear phenomenon in nature, whose variable changes seemingly messily but in fact contains an inherent regularity and randomness whereby the ergodicity and regularity of chaotic variables is used to optimize the search. In this study, chaos optimization theory is applied in the particle swarm optimization algorithm to generate a random chaotic sequence based on the optimal location of the current particle swarm using the particle in the best location of this chaotic sequence to replace the position of a particle of the current particle swarm [40]. Such a logistic equation is a chaotic system, and the chaotic sequence is expressed as

$$z_{n+1} = \mu z_n (1 - z_n), \quad n = 0, \ 1, \ 2, \ \ldots \tag{12}$$

where μ is the control parameter, $z_0 \in [0, 1]$. The individuals, called particles, are then flown through hyperspace at their specific certain speed and move to the best positions according to the fitness values. In this paper, the weight of the forecasting combination is treated as a randomized individual value and SSE is regarded as the fitness value. Accordingly, the particles adjust their speed based on their experience. The ith particles are expressed as $X = (x_{i1}, x_{i2}, \ldots, x_{iN})$, and the best position can be written as $P_i = (P_{i1}, P_{i2}, \ldots, P_{iN})$, also called P_{best}. The speed of the particles is expressed as $V_i = (v_{i1}, v_{i2}, \ldots, v_{iN})$, where the new velocity of each particle is calculated as Eq. (13) and the new position of each particle is calculated as Eq. (14):

$$v_{id}(t+1) = w v_{id}(t) + c_1 rand1()(p_{id} - x_{id}(t)) + c_2 rand2()(p_{gd} - x_{id}(t)), \quad i = 1, 2, \ldots, N \tag{13}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1), \quad i = 1, 2, \ldots, N \tag{14}$$

where c_1 and c_2 are acceleration coefficients, w is the inertia factor, $rand1()$ and $rand2()$ are two independent random numbers uniformly distributed in the range of [0, 1] [28] and N represents the number of particles. In this paper, N is assigned 20, $c_1 = 1.5$, $c_2 = 1.5$, and w changes from 0.4 to 0.9.

The CPSO algorithm flowchart is shown in Fig. 6, and the steps are as follows:

Step 1: Determine the learning parameters c_1 and c_2, the population size N, the number of evolution times and the chaos optimization times;
Step 2: Randomly generate N particle populations;
Step 3: Make particles fly under the control of Eqs. (13) and (14);
Step 4: Find the best location $P_g = (p_{g,1}, p_{g,2}, \ldots, p_{g,N})$ and generate a chaotic sequence;
Step 5: Calculate the fit value of the chaotic sequence and keep the best particle P^*;

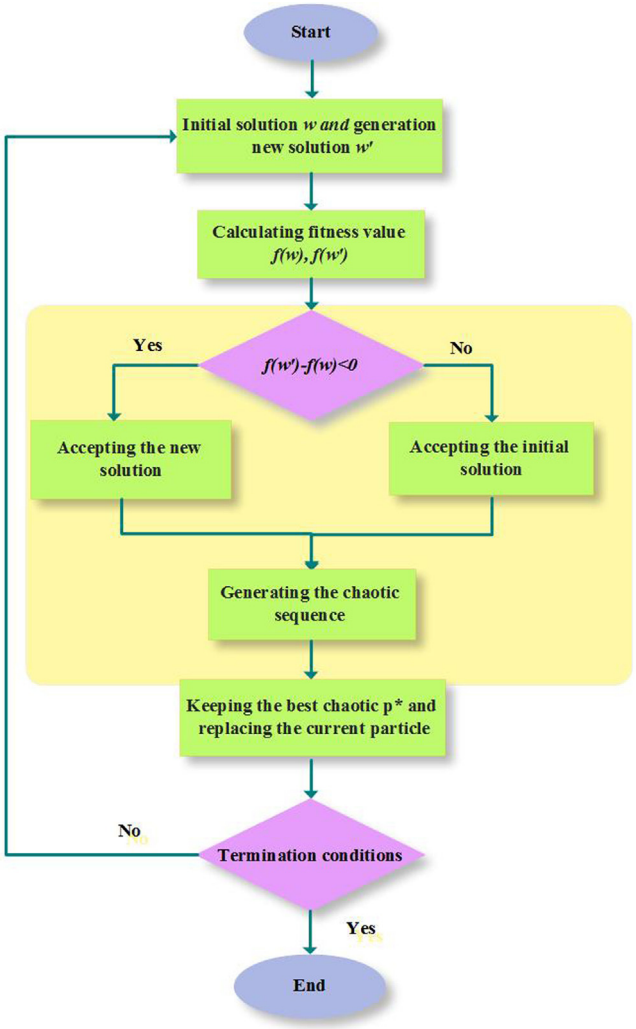


Fig. 6. CPSO algorithm flowchart.

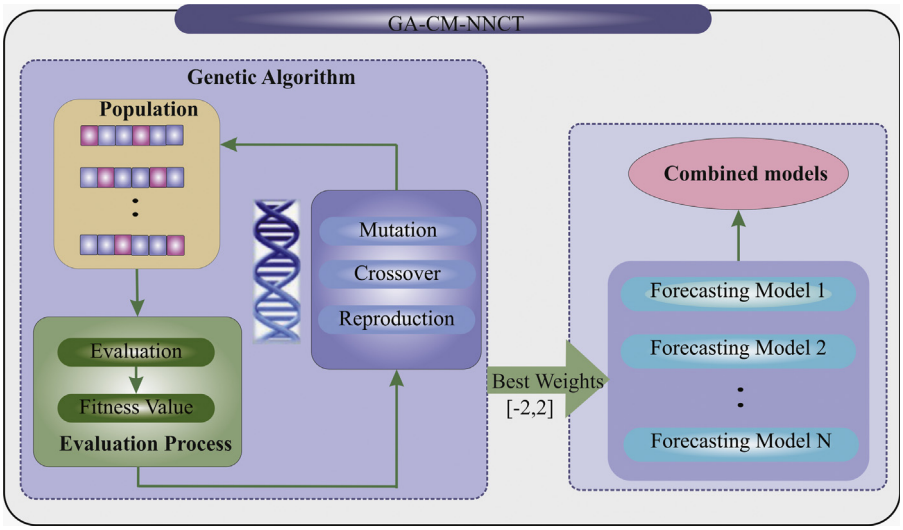


Fig. 7. Description of GA–CM–NNCT.

Step 6: Elect a particle from the current particles and use the replaced P^*;
Step 7: Map $p_{g,i}(i=1, 2, \ldots, N)$ to the domain [0,1] using the formula $z_i=(p_{g,i}-a_i)/(b_i-a_i), \quad (i=1, 2, \ldots, n)$ and then produce the chaotic variable sequence $z_t^{(m)}$. This reverts to the original solution space according to $P_{g,t}^{(m)}=a_t+(b_t-a_t)z_t^{(m)}$, $m=1,2,\ldots$, obtain $P_g{}^m=(p_{g,1}^{(m)}, p_{g,2}^{(m)}, \ldots, p_{g,N}^{(m)})$, calculate the fit value of each particle and retain the best particle, P^*;
Step 8: Randomly choose a particle from the current population and replace it with P^*;
Step 9: Stop the iteration or go back to Step 3 if the maximum number of iterations has been reached or the best solution has been identified.

The particles, or weights, in this paper take values in the range $[-2, 2]$ because numerous experiments conducted are desirable when weights are in this range. In addition, limiting the particles searching optimal solutions with restraint $\sum_{i=1}^{m} l_i=1$ is difficult, so CPSO–CM–NNCT is given as

$$\min \mathbf{J}=\mathbf{L}^T\mathbf{E}\mathbf{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} \begin{cases} \mathbf{R}^T\mathbf{L}\approx 1 \\ -2\le \mathbf{L}\le 2 \end{cases} \tag{15}$$

3.4.2. Genetic algorithm for weights based on no negative constraint theory (GA–CM–NNCT)

The genetic algorithm (GA) is a heuristic and stochastic optimization algorithm based on evolution theory and genetic principles. It is an aggressive search approach that quickly converges to find the optimal solution in a large solution domain by the genetic manipulations [41]. Usually, GAs receive work on an initial population involving the individual solutions represented by "chromosomes," which are strings that include all of the genes (i.e., variables) involved in a possible solution. The chromosomes are evaluated based on the objective function, which is the desired objective of the problem [42]. A conventional single-objective genetic algorithm was used in this study to select the optimal weights of the combined models where the square sum of the error (SSE) is defined as the fitness value so that CPSO–CM–NNCT is similar to that given in Eq. (15). The basic operation of CPSO–CM–NNCT includes selecting, crossing and mutation, and the following process gives rise to the implementation of GA in this study (Fig. 7):

Determine some parameters and initialize the population: In this study, the encoding length can be determined by the number of combined models when the maximum iteration is 20, the chromosome population is 10, the crossover probability (p_c) is 0.2 and the mutation rate (p_m) is 0.1. In addition, in this study the chromosomes should be encoded in $[-2, 2]$.
·*Selection* chooses the better individual (weight) from the initial population that can be applied to the subsequent iteration according to the lowest fitness value.
Crossover exchanges part of the chromosomes between a pair of parent individuals with the p_c and produces two new individuals (weights).
Mutation generates new individuals by changing one or some gene values of the chromosomes according the p_m.

4. Individual forecasts and algorithms

4.1. ARIMA model

Introduced by Box and Jenkins [43], the ARIMA model, also called the Box–Jenkins model, is one of the most popular forecasting models. An ARIMA model can be expressed as follows:

$$y_t=\phi_1 y_{t-1}+\phi_1 y_{t-2}+\cdots+\phi_p y_{t-p}+\varepsilon_t-\theta_q\varepsilon_{t-q} \tag{16}$$

where $y_i(i=1,2,\ldots,t)$ is the actual value, $\varepsilon_i(i=1,2,\ldots,t)$ is the random error at time t, θ_i and θ_i represent the coefficients and p and q are inter numbers that are often referred to as autoregressive and moving average polynomials, respectively [44]. The ARIMA model treats the target data series with the lapse of time as a random series and uses the specific mathematical model to approximate the series. The method consists of three phases: model identification, parameter estimation and diagnostic checking. The established model is then used to forecast the future series.

4.2. Autoregressive conditional heteroscedasticity model

The ARCH model (autoregressive conditional heteroscedasticity model), first proposed by Professor Engle in 1982 [45], developed rapidly in the field of econometrics. The ARCH model is a noise series in a certain moment following a normal distribution based

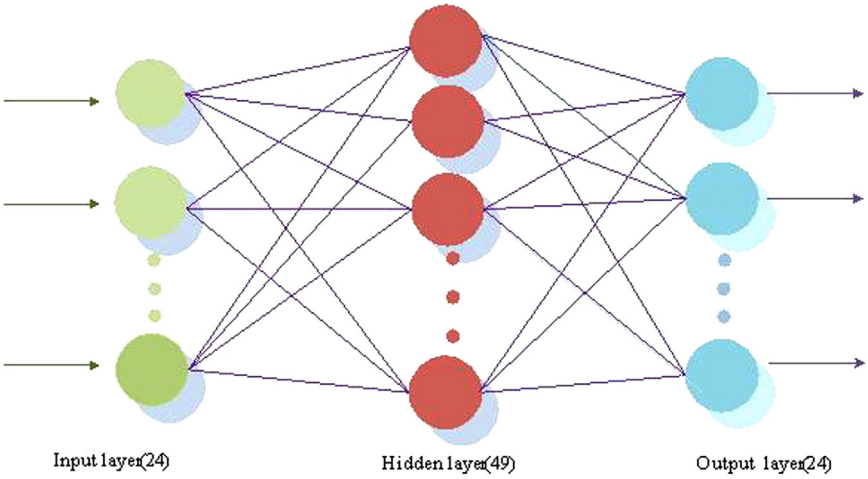


Fig. 8. The topology of the BP neural network.

on a previous set of information. The normal distribution has a zero mean and time-varying variance that is the linear combination of the finite term's past noise value. As a new theory, the ARCH model has been widely used to describe the process of verifying the financial theory and forecasting in the financial market. In this paper, the ARCH model is used to forecast wind speed. While the ARCH process can be defined in various ways in this paper, the main equation is defined as the ARIMA model in Section 3.1, and the ε_t in the model is defined by [46]

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \eta_t, \quad t = 1,\ 2, \ldots \tag{17}$$

where η_t is independent and satisfies $E(\eta_t) = 0$, $D(\eta_t) = \lambda^2$. Thus, Eq. (11) is the ARCH (q) model.

4.3. BP neural network

The BP neural network is a type of multilayer feed-forward neural network with a wide variety of applications. It is based on a gradient descent method that minimizes the sum of the squared errors between the actual and the desired output values. The transfer function of the neuron is S type. The output function is between 0 and 1 and can transform input to output for continuous non-linear mapping [47]. The BP neural network consists of three layers: an input layer, a hidden layer and an output layer. While neurons from the layers connect with each other in a one-way system, the inner neurons are independent of each other. The topology of the BP neural network, as used in this study, is shown in Fig. 8.

To ensure that the input values are compatible when there are significant differences in their magnitudes, we determine the input vector by normalizing each input value as follows:

$$P = \{P_i\} = \frac{v_i - v_{i\,min}}{v_{i\,max} - v_{i\,min}}, \quad i = 1, 2, \ldots, 720 \tag{18}$$

where $v_{i\,min}$ and $v_{i\,max}$ are the minimal and maximal values of each input factor, respectively [15].

4.4. Support vector machine (SVM) model

Proposed by Vapnik, the support vector model (SVM) centers on mapping input vectors into a high dimensional feature space where a maximal margin hyperplane is constructed [48]. Based on the structured risk minimization (SRM) principle, the SVM model seeks to minimize an upper bound of the generalization error rather than the empirical error as in other neural networks. Additionally, the SVM models generate the regress function by applying a set of high dimensional linear functions. The SVM regression function is formulated as following [44]:

$$y = w\varphi(x) + b \tag{19}$$

where $\varphi(x)$ is called the feature, which is non-linear and mapped from the input space x. The penalty parameter C is introduced, and the coefficients w and b are then estimated by minimizing.

$$\min Q = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) \tag{20}$$

$$\text{s.t.} \quad \begin{cases} y_i - (w\varphi(x_i)) - b \le \varepsilon + \xi_i^* \\ (w\varphi(x_i)) + b - y = \varepsilon_i + \xi_i \\ \xi_i^*, \xi_i \ge 0, \ \ i = 1, 2, \ldots, \ s \end{cases} \tag{21}$$

where ξ_i denotes the upper training error and ξ_i^* is the lower training error subject to ε-incentive tube $|y - (\langle w, x_i \rangle + b| \le \varepsilon$.

4.5. Hybrid Kalman filter (HKF) model

The KF model is derived from Dr. Kalman's Ph.D. paper written in the 1960s [49]. The filter, used in many areas, includes filtering, forecasting and smoothing. In this paper, the forecasting model can be applied to wind speed forecasting over the next 10 days (240 data). The KF model consists of a set of two equations, the observation Eq. (22) and the system equation, Eq. (23). The system can be described as a linear stochastic differential equation, also called the state equation [37]. This paper has adopted the ARIMA model to determine parameters, and then developed into the hybrid Kalman filter (HKF) model.

$$X(k) = A * X(k-1) + B * U(k) + w(k) \tag{22}$$

We then add the measured value of the system:

$$Z(k) = H * X(k) + v(k) \tag{23}$$

In the above equations, $X(k)$ and $Z(k)$ are the actual state and measurement vectors at time k, $U(k)$ is the system controlling value at time k, A and B are system parameter arrays, H is the array of measurement parameters of the system and $w(k)$ and $v(k)$ represent the process series and measurement noise series, respectively, and are assumed to be white Gaussian noise.

5. CPSO–BP model

The CPSO algorithm is used to optimize the weights and biases of BP models in the proposed CPSO–BP model. The BP model is

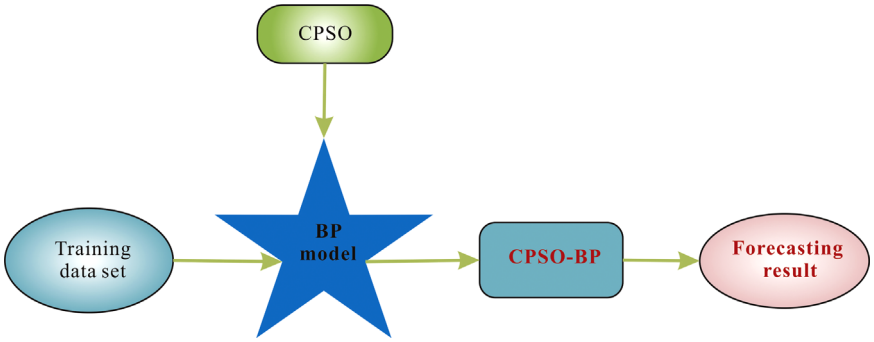


Fig. 9. A description of the CPSO–BP model.

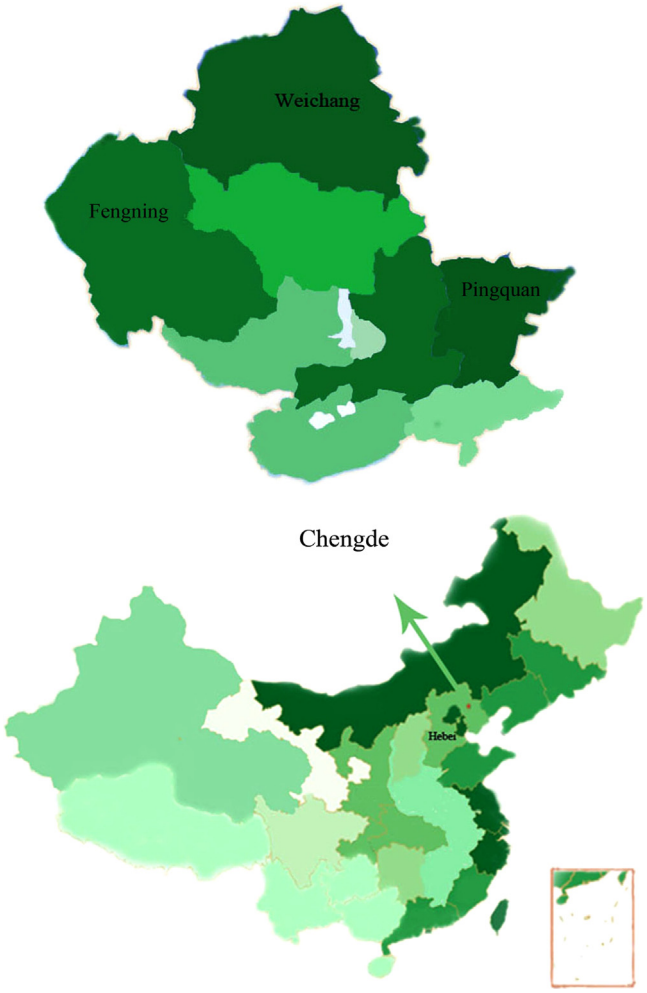


Fig. 10. Chengde area of China.

referred to in Section 4.3 and the CPSO algorithm is in Section 3.4. In Section 4.3, the weights and bias of the BP are randomly determined. However, in this section, the CPSO algorithm is adopted to select the optimal parameters of the BP model by minimizing the square sum of the error (SSE) of the training data sets. Fig. 9 gives a brief description of the CPSO–BP model.

6. A case study

The study data were collected from the Chengde area of China (shown in Fig. 8).

6.1. Study area and data collection

Chengde is located northeast of Hebei Province at longitude 115°54′–119°15′ and north latitude 40°11′–42°40′ (Fig. 10). Chengde is not only famous for its beautiful scenery, but it is also known for its abundant wind energy resources. Because of its prime location, Chengde possesses one of the five largest wind farms with an installed capacity of 200,000 kW in 2008. Furthermore, according to the forecasting, capacity could potentially reach 6,000,000 kW in the future. The wind energy resources of Chengde are mainly concentrated in northern and western Weichang County, northern and northwestern Fengning County and western Pingquan County. Additionally, the planned wind farm area is 3029 km^2. The available area for wind farm building in Weichang County is 2400 km^2. Some areas in Fengning County are also suitable for building wind farms as there are no obstacles in the direct path of the wind. Thus, the total area available for development could host approximately 350,000 kW. In addition, the area available for development in Pingquan County could produce approximately 80,000 kW. Thus, as evidenced, there are a number of large-scale investments in wind energy projects, and as a result, grid-connected wind power is increasing [50].

For this study, the hourly wind speed of three wind turbines from April 5, 2010 to April 24, 2011 were selected in the Chengde area for model construction and hourly wind speeds are then predicted for the following ten days.

6.2. Statistical measures of forecasting performance

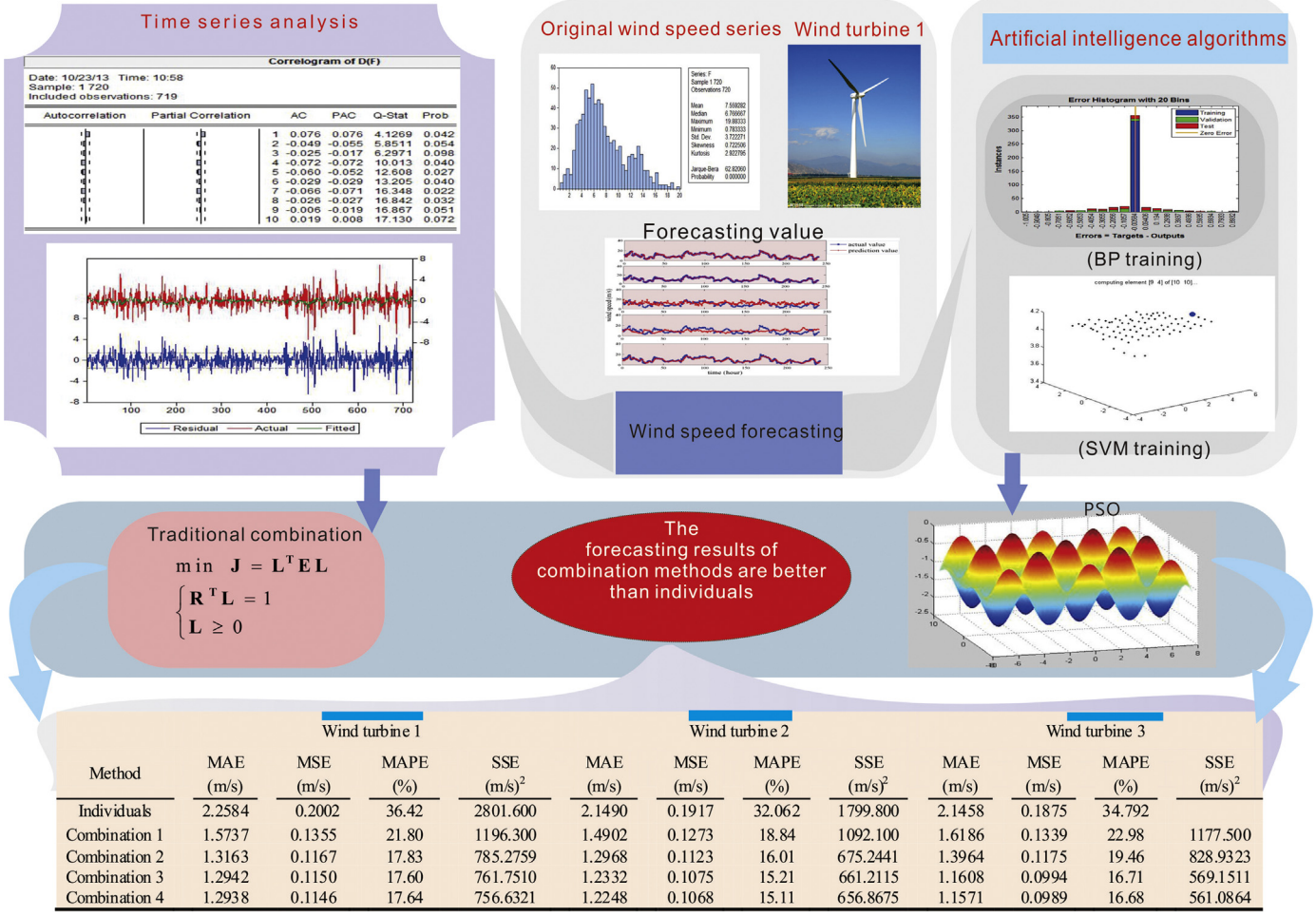
To evaluate and compare forecasts, four forecasting error measures were adopted: the mean square error (MSE), shown in Eq. (25), the mean absolute error (MAE) (Eq. (26)), the mean absolute percentage error (MAPE) (Eq. (27)) and the square sum of the error (SSE) (Eq. (28)). Y_t is the actual observation value for a time period t and y_t is the forecasting value for the same period. Hence, the error is defined as

$$e_t = Y_t - y_t \tag{24}$$

$$MSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n} e_i^2} \tag{25}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{26}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - y_i}{Y_i} \times 100\%\right| \tag{27}$$



Method	Wind turbine 1 MAE (m/s)	MSE (m/s)	MAPE (%)	SSE (m/s)^2	Wind turbine 2 MAE (m/s)	MSE (m/s)	MAPE (%)	SSE (m/s)^2	Wind turbine 3 MAE (m/s)	MSE (m/s)	MAPE (%)	SSE (m/s)^2
Individuals	2.2584	0.2002	36.42	2801.600	2.1490	0.1917	32.062	1799.800	2.1458	0.1875	34.792	
Combination 1	1.5737	0.1355	21.80	1196.300	1.4902	0.1273	18.84	1092.100	1.6186	0.1339	22.98	1177.500
Combination 2	1.3163	0.1167	17.83	785.2759	1.2968	0.1123	16.01	675.2441	1.3964	0.1175	19.46	828.9323
Combination 3	1.2942	0.1150	17.60	761.7510	1.2332	0.1075	15.21	661.2115	1.1608	0.0994	16.71	569.1511
Combination 4	1.2938	0.1146	17.64	756.6321	1.2248	0.1068	15.11	656.8675	1.1571	0.0989	16.68	561.0864

Fig. 11. The schematic forecasting process.

$$SSE = \sum_{t=1}^{N} (Y_t - y_t)^2 \tag{28}$$

The selection of the best forecasting methods follows the criteria of the four measures of accuracy, for which the lowest values represent better accuracy.

6.3. Simulations

The collected wind speed data in the study area were simulated by the proposed forecasting models. The simulation process consisted of the following steps:

Step 1: Conduct five individual method forecasts and collect forecast results (for all three wind turbines).

This paper treats the wind speed series of the Chengde region as a random series, and 20 days of data (480 data) from April 5 to 24 of 2011 are used to build the model. The models are then used to forecast the wind speed for the next 10 days. The embellishment of the ARIMA model and the ARCH model is based on the *Eviews* toolbox. ARIMA(2,1,1), ARIMA(2,1,2) and ARIMA(2,1,2) were established for wind speed forecasting for the three wind turbines, and ARCH (1), ARCH (1) and ARCH (2) were then adopted to forecast the wind speed of the three turbines.

For the BP model, 720 wind speed data for 30 days (24 data per day) from the Chengde region were chosen as the input matrix, with the corresponding wind speed for the next day as the target matrix. To forecast 240 wind speeds for the next 10 days, we used 480 wind speed data from the past 20 days to train the net. Accordingly, we choose the number of nodes in the hidden layer to be 49, which can be determined by $2i+1$ (i is the number of input nodes), while the transfer functions adopted in the hidden layer and the output layer are $\tan sig(x)$ and $\log sig(x)$, respectively.

In the SVM model, given the Gaussian Kernals' good performance under general smoothness assumptions, we adopt Gaussian Kernals to estimate the non-linear behavior of the forecasting data set. Due to the negative effect of the systematic errors, the predicted values of support vector machine (SVM) forecasting are underestimated most of the time, so we increased every forecasting value by 4.0895 m/s to revise the forecasting result. This correction factor is the mean value of the fitting systematic errors according to the simulations.

The HKF model is a basic one-step forecasting. To forecast more accurately, 10 previous hourly data were adopted to forecast next step values and to consistently update the latest value with the actual value and so on. The coefficients of the ARIMA model in Section 3.2 were used to calculate$A$, B and H.

Step 2: Combine every forecast for a total of 26 combination models using the traditional combination method.

Step 3: Combine every method for a total of 26 combination models using the proposed TCM–NNCT method.

Step 4: Combine every method for a total of 26 combination models using the proposed CPSO–CM–NNCT method.

Table 2
Comparison between traditional combination forecast and TCM–NNCT forecast for wind turbine 1.

Weight	Traditional combination forecast methods						TCM–NNCT forecast methods					
Methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two method combination	0.0000	1.0000	–	–	–	776.894	−0.0646	1.0646	–	–	–	776.7376
	0.9098	–	0.090	–	–	768.209	0.9098	–	0.0902	–	–	768.2096
	0.9192	–	–	0.080	–	783.417	0.9192	–	–	0.0808	–	783.4170
	0.7945	–	–	–	0.2055	806.253	0.7945	–	–	–	0.2055	806.2532
	–	0.9534	0.046	–	–	764.471	–	0.9534	0.0466	–	–	764.4715
	–	0.9784	–	0.021	–	774.620	–	0.9784	–	0.0216	–	774.6204
	–	0.9122	–	–	0.0878	774.793	–	0.9122	–	–	0.0878	774.7930
	–	–	0.426	0.573	–	4770.800	–	–	0.4267	0.5733	–	4770.800
	–	–	0.080	–	0.9196	963.523	–	–	0.0804	–	0.9196	963.5230
	–	–	–	0.000	0.9997	1001.800	–	–	–	0.0003	0.9997	1001.8000
Every three methods combination	0.3758	0.5605	0.063	–	–	760.831	0.3758	0.5605	0.0636	–	–	760.8311
	0.2109	0.7545	–	0.034	–	773.767	0.2109	0.7545	–	0.0346	–	773.7676
	0.0000	0.9122	–	–	0.0878	774.793	−0.0653	0.9775	–	–	0.0878	774.6327
	0.9003	–	0.072	0.027	–	766.098	0.9003	–	0.0726	0.0271	–	766.098
	0.8018	–	0.084		0.1141	764.410	0.8018	–	0.0841	–	0.1141	764.4101
	0.8880	–		0.076	0.0357	783.132	0.8880	–	–	0.0763	0.0357	783.1323
	–	0.9534	0.046	0.000	–	764.471	–	0.9615	0.0589	−0.0204	–	763.3232
	–	0.8729	0.046	–	0.0811	762.677	–	0.8729	0.046	–	0.0811	762.6773
	–	0.9172	—	0.016	0.0660	773.543	–	0.9172	–	0.0168	0.0660	773.5430
	–	–	0.080	0.000	0.9196	963.523	–	–	0.1401	−0.1055	0.9654	935.2297
Every four methods combination	0.3758	0.5605	0.063	0.000	–	760.831	0.3674	0.5700	0.0642	−0.0015	–	760.8265
	0.3695	0.4893	0.062	–	0.0783	759.160	0.3695	0.4893	0.0628	–	0.0783	759.1604
	0.1506	0.7697	–	0.027	0.0526	773.152	0.1506	0.7697	–	0.0271	0.0526	773.1527
	0.8152	–	0.077	0.011	0.0957	764.135	0.8152	–	0.0777	0.0114	0.0957	764.1353
	–	0.8729	0.046	0.000	0.0811	762.677	–	0.8447	0.0669	−0.0350	0.1234	759.7671
Five methods	0.3695	0.4893	0.062	0.000	0.0783	759.160	0.2629	0.5848	0.0693	−0.0190	0.1021	758.5953

Step 5: Combine every method for a total of 26 combination models using the proposed GA–CM–NNCT method.
Step 6: Compare individual method forecast results with combination method results in three ways, forecast results and analyze the differences among three different combination methods.
Step 7: Use the CPSO algorithm to optimize parameters of BP mode developing CPSO–BP and then use the optimized model to forecast wind speed. Last, the CPSO–BP model is proposed to compare with the proposed individual models and combination models.

The schematic forecasting processes can be seen in Fig. 11. In Fig. 11, Combination 1 represents the average forecasting results of every two method combination, Combination 2 the results of every three method combination, Combination 3 the results of every four method combination and Combination 4 the results of every five method combination.

6.4. Forecast result analysis and comparisons between different models

The three wind turbines forecasting results of different forecasting combination models are compared in Tables 2–7, and SSE is adopted to monitor the forecasting accuracy. Combination models have been categorized into four situations: every two method combination, every three method combination, every four method combination and five method combination. When compared to other combination models, the overall performance of the five method combination model is clearly much better than the performances of the other models. Furthermore, combination models that assign non-zero weight to the combined forecasts have stronger forecasting capacity compared to the five individual forecasts, mainly because the combination models combine advantages of each combined forecast. The other result is that the performance of combination models is not absolutely good when well performing individual forecasts are combined. The results indicate that when the best forecast and the worst forecast are combined, the forecasting result performs better than each of the forecasts when assessed separately. Tables 2 through 4 compare traditional combination models and TCM–NNCT models for three wind turbines. Analyses of the forecasting results indicate that the TCM–NNCT method performs better than the traditional combination model.

Tables 5 through 7 show the comparison results between the GA–CM–NNCT models and the PSO weight-determined combination models of the three turbines where the SSE is adopted to compare the forecasting results. The SSE values indicate that the overall performance of GA–CM–NNCT is undesirable because the SSE values of three wind turbines show high volatility when different models are combined. For wind turbine 1, for six every two method combination models, the GA–CM–NNCT method performs better than the CPSO–CM–NNCT method and only one every three method GA–CM–NNCT model has a better performance than the CPSO–CM–NNCT model. Table 8 shows a detailed comparison between them where 1, 2, 3, 4 and 5 represent the five individual forecasts (ARIMA, ARCH, BP, SVM and HKF) and the first column SSE values belong to the GA–CM–NNCT models and the second column is the CPSO–CM–NNCT models. All of every four combination models and five combination models indicate that the GA–CM–NNCT method has undesirable forecasting performance. Moreover, the comparison results from wind turbine 2 and wind turbine 3 also underline the poor performance of the GA–CM–NNCT method. Meanwhile, with increasing the combined model numbers, more differences appear related to SSE between the GA–CM–NNCT models and the CPSO–CM–NNCT models.

The forecasting results from Tables 5 through 7 show that GA–CM–NNCT has a poor robustness. In contrast, the CPSO–CM–NNCT models always do well in terms of the SSE. These results are really

Table 3
Comparison between traditional combination forecast and no negative constraint combination forecast for wind turbine 2.

Weight	Traditional combination forecast methods						TCM–NNCT forecast methods					
Methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two methods combination	1.0000	0.0000	–	–	–	669.8995	1.8703	−0.8703	–	–	–	669.2118
	0.9669	–	0.0331	–	–	662.6500	0.9669	–	0.0331	–	–	662.6500
	0.9800	–	–	0.0200	–	668.0040	0.9800	–	–	0.0200	–	668.0040
	0.9038	–	–	–	0.0962	667.9189	0.9038	–	–	–	0.0962	667.9189
	–	0.9615	0.0385	–	–	662.4506	–	0.9615	0.0385	–	–	662.4506
	–	0.9748	–	0.0252	–	669.326	–	0.9748	–	0.0252	–	669.326
	–	0.8765	–	–	0.1235	668.9395	–	0.8765	–	–	0.1235	668.9395
	–	–	0.3055	0.6945	–	4846.0000	–	–	0.3055	0.6945	–	4846.0000
	–	–	0.0628	–	0.9372	815.5107	–	–	0.0628	–	0.9372	815.5107
	–	–	–	0.0066	0.9934	842.4698	–	–	–	0.0066	0.9934	842.4698
Every three methods combination	0.3344	0.6290	0.0366	–	–	662.3733	0.3344	0.6290	0.0366	–	–	662.3733
	0.9800	0.0000	–	0.0200	–	668.0040	1.3638	−0.3818	–	0.0180	–	667.8903
	0.9038	0.0000	–	–	0.0962	667.9189	1.2382	−0.3259	–	–	0.0876	667.8383
	0.9669	–	0.0331	0.0000	–	662.6500	0.9709	–	0.0378	−0.0087	–	662.4390
	0.7953	–	0.0249	–	0.1966	656.3324	0.7953	–	0.0249	–	0.1966	656.3324
	0.9071	–	–	–	0.0771	666.8147	0.9071	–	–	–	0.0771	666.8147
	–	0.9615	0.0385	0.0000	–	662.4506	–	0.9647	0.0424	−0.0071	–	662.3089
	–	0.8425	0.0381	–	0.1194	659.2258	–	0.8425	0.0381	–	0.1194	659.2258
	–	0.8820	–	0.0193	0.0988	667.2914	–	0.8820	–	0.0193	0.0988	667.2914
	–	–	0.0628	0.0000	0.9372	815.5107	–	–	0.1021	−0.0755	0.9734	800.7597
Every four methods combination	0.3344	0.6290	0.0366	0.0000	–	662.3733	0.3874	0.5798	0.0406	−0.0078	–	662.2063
	0.0000	0.8425	0.0381	–	0.1194	659.2258	−1.0498	1.8566	0.0439	–	0.1492	658.6644
	0.8567	0.0492	–	0.0160	0.0782	666.8131	0.8567	0.0492	–	0.0160	0.0782	666.8131
	0.8673	–	0.0334	0.0000	0.0993	660.5414	0.8505	–	0.0448	−0.0208	0.1255	659.4799
	–	0.8425	0.0381	0.0000	0.1194	659.2258	–	0.8259	0.0498	−0.0215	0.1458	658.0842
Five methods	0.0000	0.8425	0.0381	0.0000	0.1194	659.2258	−1.1916	1.9759	0.0572	−0.0231	0.1815	657.3665

Table 4
Comparison between traditional combination forecast and TCM–NNCT forecast for wind turbine 3.

Weight	Traditional combination forecast methods						TCM–NNCT forecast methods					
Methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two methods combination	0.0028	0.9972	–	–	–	589.3653	0.0028	0.9972	–	–	–	589.3653
	0.9524	–	0.0476	–	–	580.5047	0.9524	–	0.0476	–	–	580.5047
	0.9896	–	–	0.0104	–	592.5905	0.9896	–	–	0.0104	–	592.5905
	0.9389	–	–	–	0.0611	592.4256	0.9389	–	–	–	0.0611	592.4256
	–	0.9408	0.0592	–	–	569.2336	–	0.9408	0.0592	–	–	569.2336
	–	0.9849	–	0.0151	–	588.1506	–	0.9849	–	0.0151	–	588.1506
	–	0.9959	–	–	0.0041	589.3624	–	0.9959	–	–	0.0041	589.3624
	–	–	0.5112	0.4888	–	4652.1000	–	–	0.5112	0.4888	–	4652.1000
	–	–	0.1094	–	0.8906	691.3006	–	–	0.1094	–	0.8906	691.3006
	–	–	–	0.0167	0.9833	764.7916	–	–	–	0.0167	0.9833	764.7916
Every three methods combination	0.0000	0.9408	0.0592	–	–	569.2336	−1.6631	2.5814	0.0817	–	–	561.5899
	0.0000	0.9849	–	0.0151	–	588.1506	−0.1184	1.1027	–	0.0158	–	588.0993
	0.0118	0.9836	–	–	0.0046	589.3619	0.0118	0.9836	–	–	0.0046	589.3619
	0.9524	–	0.0476	0.0000	—	580.5047	0.9652	–	0.0663	−0.0315	–	577.2484
	0.8237	–	0.0527	–	0.1236	577.6428	0.8237	–	0.0527	–	0.1236	577.6428
	0.9370	–	–	0.0088	0.0542	592.0243	0.9370	–	–	0.0088	0.0542	592.0243
	–	0.9408	0.0592	0.0000	–	569.2336	–	0.9559	0.0816	−0.0375	–	564.6322
	–	0.8778	0.0610	–	0.0612	568.5835	–	0.8778	0.0610	–	0.0612	568.5835
	–	0.9849	–	0.0151	0.0000	588.1506	–	0.9954	–	0.0155	−0.0108	588.1304
	–	–	0.1094	0.0000	0.8906	691.3006	–	–	0.1671	−0.1001	0.9330	660.4313
Every four methods combination	0.0000	0.9408	0.0592	0.0000	–	569.2336	−1.9589	2.8924	0.1142	−0.0477	–	554.3720
	0.0000	0.8778	0.0610	–	0.0612	568.5835	−1.6412	2.5479	0.0817	–	0.0116	561.5681
	0.0000	0.9849	–	0.0151	0.0000	588.1506	-0.1618	1.1639	–	0.0166	-0.0187	588.0455
	0.8237	–	0.0527	0.0000	0.1236	577.6428	0.7634	–	0.0853	-0.0493	0.2007	570.7511
	–	0.8778	0.0610	0.0000	0.0612	568.5835	–	0.8187	0.0924	−0.0489	0.1379	561.7638
Five methods	0.0000	0.8778	0.0610	0.0000	0.0612	568.5835	−1.8279	2.6718	0.1192	−0.0547	0.0915	553.1544

related to the different information sharing mechanisms between GA and CPSO; chromosomes share information among each other when the whole population moves in the GA, but in the CPSO algorithm, only the global best particles contain all of the information about the other particles. CPSO can search best individuals without too much effect from some bad particles and has a faster convergence speed. As a consequence, CPSO–CM–NNCT shows a desirable robustness compared to GA–CM–NNCT.

Table 5
Comparison between TCM–NNCT and CPSO–CM–NNCT forecast for wind turbine 1.

Weight	TCM–NNCT forecast methods						CPSO–CM–NNCT forecast methods					
Methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two methods combination	−0.0646	1.0646	–	–	–	776.7376	−0.0451	1.0429	–	–	–	776.6334
	0.9098	–	0.0902	–	–	768.2096	0.8805	–	0.0960	–	–	754.4853
	0.9192	–	–	0.0808	–	783.4170	0.9182	–	–	0.0772	–	782.9502
	0.7945	–	–	–	0.2055	806.2532	0.8207	–	–	–	0.1736	805.7316
	–	0.9534	0.0466	–	–	764.4715	–	0.9435	0.0493	–	–	763.2360
	–	0.9784	–	0.0216	–	774.6204	–	0.9785	–	0.0219	–	774.6175
	–	0.9122	–	–	0.0878	774.7930	–	0.9015	–	–	0.1013	774.6589
	–	–	0.4267	0.5733	—	4770.800	–	–	0.6444	0.1733	—	4200.7000
	–	–	0.0804	—	0.9196	963.5230	–	–	0.0595	—	0.9780	934.9232
	–	–	–	0.0003	0.9997	1001.8000	–	–	–	0.0238	1.0248	951.7287
Every three methods combination	0.3758	0.5605	0.0636	–	–	760.8311	0.7191	0.1736	0.0870	–	–	754.0186
	0.2109	0.7545	–	0.0346	–	773.7676	0.2156	0.7494	–	0.0343	–	773.755
	−0.0653	0.9775	–	–	0.0878	774.6327	−0.1078	1.0012	–	–	0.1113	774.2953
	0.9003	–	0.0726	0.0271	–	766.098	0.8829	–	0.1139	−0.02506	–	753.3239
	0.8018	–	0.0841	–	0.1141	764.4101	0.9283	–	0.1000	–	−0.0557	753.9838
	0.8880	–	–	0.0763	0.0357	783.1323	0.9046	–	–	0.0759	0.0157	782.9106
	–	0.9615	0.0589	−0.0204	–	763.3232	–	0.9492	0.0850	−0.0522	–	758.4536
	–	0.8729	0.046	–	0.0811	762.6773	–	0.8857	0.0475	–	0.0632	762.4588
	–	0.9172	–	0.0168	0.0660	773.5430	–	0.8995	–	0.0187	0.0864	773.1726
	–	–	0.1401	−0.1055	0.9654	935.2297	–	–	0.1049	−0.0672	0.9865	927.0130
Every four methods combination	0.3674	0.5700	0.0642	−0.0015	–	760.8265	0.6201	0.2832	0.1049	−0.0337	–	752.1970
	0.3695	0.4893	0.0628	–	0.0783	759.1604	0.7889	0.1410	0.0920	–	−0.0462	753.6907
	0.1506	0.7697	–	0.0271	0.0526	773.1527	0.1101	0.7968	–	0.0257	0.0707	772.9955
	0.8152	–	0.0777	0.0114	0.0957	764.1353	0.9190	–	0.1154	−0.0234	−0.0424	753.0423
	–	0.8447	0.0669	−0.0350	0.1234	759.7671	–	0.8793	0.0843	−0.0540	0.0767	757.3179
Five methods	0.2629	0.5848	0.0693	−0.0190	0.1021	758.5953	0.6543	0.2643	0.1062	−0.0321	−0.0195	752.1422

Table 6
Comparison between TCM–NNCT and CPSO–CM–NNCT forecast for wind turbine 2.

Weight	No negative constraint combination forecast methods						CPSO–CM–NNCT forecast methods					
methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two Methods combination	1.8703	−0.8703	–	–	–	669.2118	1.6571	−0.6475		–	–	667.0852
	0.9669	–	0.0331	–	–	662.6500	0.9757	–	0.0301	–	–	661.9144
	0.9800	–	–	0.0200	–	668.0040	0.9852	–	–	0.9852	–	663.8944
	0.9038	–	–	–	0.0962	667.9189	0.7983	–	–	–	0.2236	660.0311
	–	0.9615	0.0385	–	–	662.4506	–	0.9703	0.0355	–	–	661.6981
	–	0.9748	–	0.0252	–	669.326	–	0.9806	–	0.0347	—	664.1885
	–	0.8765	–	–	0.1235	668.9395	–	0.7687	–	–	0.2555	658.9409
	–	–	0.3055	0.6945	–	4846.0000	–	–	0.5467	0.2616	–	4259.5000
	–	–	0.0628	–	0.9372	815.5107	–	–	0.0276	–	1.0270	760.2322
	–	–	–	0.0066	0.9934	842.4698	–	–		0.0302	1.0325	760.8294
Every three Methods combination	0.3344	0.6290	0.0366	—	—	662.3733	0.7052	0.2687	0.0319	–	–	661.7237
	1.3638	−0.3818	–	0.0180	—	667.8903	0.7021	0.2818	–	0.0303	–	663.8373
	1.2382	−0.3259	–	–	0.0876	667.8383	−0.873	1.6037	–	–	0.2965	658.5336
	0.9709	–	0.0378	−0.0087	–	662.4390	0.9755	–	0.0287	0.0021	–	661.9065
	0.7953	–	0.0249	–	0.1966	656.3324	0.7953	–	0.0249	–	0.1966	656.3324
	0.9071	–	–	–	0.0771	666.8147	0.7932	–	–	0.0236	0.2072	657.6233
	–	0.9647	0.0424	−0.0071	–	662.3089	–	0.9698	0.0323	0.0049	–	661.6571
	–	0.8425	0.0381	—	0.1194	659.2258	–	0.7697	0.0284	–	0.2203	654.1479
	–	0.8820	–	0.0193	0.0988	667.2914	–	0.7665	–	–	0.2324	655.6523
	–	–	0.1021	−0.0755	0.9734	800.7597	–	–	0.0192	0.0127	1.0254	759.9511
Every four methods combination	0.3874	0.5798	0.0406	−0.0078	–	662.2063	0.3054	0.6661	0.0307	0.0046	–	661.5973
	−1.0498	1.8566	0.0439	–	0.1492	658.6644	1.0995	−0.2732	0.0219	–	0.1678	657.3232
	0.8567	0.0492	–	0.0160	0.0782	666.8131	0.6295	0.1332	–	0.0276	0.2356	657.3814
	0.8505	–	–	−0.0208	0.1255	659.4799	0.7950	–	–	0.0024	0.1967	656.3224
	–	0.8259	0.0498	−0.0215	0.1458	658.0842	–	0.7690	0.0254	0.0045	0.2204	654.1127

The behavior of the GA–CM–NNCT models shows that with an increasing number of combined models, the forecasting accuracies and the robustness of the GA–CM–NNCT models fall under the measure of SSE, especially for wind turbine 2 and wind turbine 3.

While the TCM–NNCT, GA–CM–NNCT and CPSO–CM–NNCT methods all allow combination weights to be negative, the CPSO–CM–NNCT models clearly have better performance than the TCM–NNCT and GA–CM–NNCT models mainly because the

Table 7
Comparison between TCM–NNCT and CPSO–CM–NNCT forecast for wind turbine 3.

Weight	TCM–NNCT forecast methods						CPSO–CM–NNCT forecast methods					
Methods	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$	ARIMA	ARCH(1)	BP	SVM	KF	SSE $(m/s)^2$
Every two methods combination	0.0028	0.9972	–	–	–	589.3653	−0.4517	1.4635	–	–	–	587.0454
	0.9524	–	0.0476	–	–	580.5047	0.9468	–	0.0494	–	–	580.1937
	0.9896	–	–	0.0104	–	592.5905	0.9914	–	–	0.0118	–	592.3679
	0.9389	–	–	–	0.0611	592.4256	0.8362	–	–	–	0.1784	590.3405
	–	0.9408	0.0592	–	–	569.2336	–	0.9432	0.0584	–	–	569.1790
	–	0.9849	–	0.0151	–	588.1506	–	0.9913	–	0.0193	–	585.7104
	–	0.9959	–	–	0.0041	589.3624	–	0.8649	–		0.1538	585.5750
	–	–	0.5112	0.4888	–	4652.1000	–	–	0.6594	0.1315		3720.6
	–	–	0.1094	–	0.8906	691.3006	–	–	0.0729	–	0.9866	630.0726
	–	–	–	0.0167	0.9833	764.7916	–	–	–	0.0267	1.048	655.6800
Every three methods combination	−1.6631	2.5814	0.0817	–	–	561.5899	−1.0637	2.0000	0.0707		–	561.5052
	−0.1184	1.1027	–	0.0158	–	588.0993	−0.8555	1.8459	–	0.0266	–	583.8929
	0.0118	0.9836	–	–	0.0046	589.3619	−0.5627	1.4197	–	–	0.1661	584.6861
	0.9652	–	0.0663	−0.0315	–	577.2484	0.9531	–	0.0803	−0.0454	–	574.8269
	0.8237	–	0.0527	–	0.1236	577.6428	0.7538	–	0.0512	–	0.2054	576.5940
	0.9370	–	–	0.0088	0.0542	592.0243	0.8251	–	–	0.0119	0.1788	589.6275
	–	0.9559	0.0816	−0.0375	–	564.6322	–	0.9494	0.0891	−0.0449	–	563.9439
	–	0.8778	0.0610	–	0.0612	568.5835	–	0.8054	0.0582	–	0.1475	567.2917
	—	0.9954	—	0.0155	-0.0108	588.1304	–	0.8555	–	0.0183	0.1461	583.8640
	—	—	0.1671	−0.1001	0.9330	660.4313	–	–	0.1084	−0.0523	0.9946	622.9879
Every four methods combination	−1.9589	2.8924	0.1142	−0.0477	–	554.3720	−1.0576	2.0000	0.1004	−0.0437	–	556.5371
	−1.6412	2.5479	0.0817	–	0.0116	561.5681	−1.2509	2.0000	0.0724	–	0.1993	558.1153
	−0.1618	1.1639	–	0.0166	−0.0187	588.0455	−0.9298	1.7745	–	0.0256	0.1573	581.6059
	0.7634	–	0.0853	−0.0493	0.2007	570.7511	0.7459	–	0.0838	−0.0478	0.2209	570.6775
	–	0.8187	0.0924	−0.0489	0.1379	561.7638	–	0.7968	0.0901	-0.0468	0.1637	561.6292
Five methods	−1.8279	2.6718	0.1192	−0.0547	0.0915	553.1544	−1.2585	2.0000	0.1038	−0.0460	0.2143	552.6342

Table 8
performances of every two methods combination models and every three method combination models between GA–CM–NNCT models and CPSO–CM–NNCT models for wind turbure1.

	GA–CM–NNCT performs better than CPSO–CM–NNCT			CPSO–CM–NNCT perform better than GA–CM–NNCT		
	Method	SSE $(m/s)^2$		Method	SSE $(m/s)^2$	
Every two methods combination	1,2	773.8911	776.6334	1,3	784.0531	754.4853
	2,4	763.5455	774.6175	1,4	842.1843	782.9502
	2,5	773.4923	774.6589	1,5	831.9923	805.7316
	3,4	2102.000	4200.700	2,3	763.5703	763.2360
	3,5	926.0932	934.9232	–	–	–
	4,5	812.7742	951.7287	–	–	–
Every three methods combination	1,2,4	767.7142	773.755	1,2,3	758.2240	773.755
	–	–	–	1,2,5	776.0847	774.2953
	–	–	–	2,3,4	787.1289	758.4536
	–	–	–	2,4,5	776.7883	762.4583
	–	–	–	3,4,5	945.3544	927.0130
	–	–	–	1,3,4	755.4224	753.3239
	–	–	–	1,3,5	760.8816	753.9838
	–	–	–	1,4,5	786.2697	782.9106

random particles have stronger search ability and can thus find better weight values. The weight assignment of combination models is shown in Tables 2 through 7, where the larger weights are assigned to the better forecasts and the sum of the weights is 1 or converges to 1 with probability.

Forecasting accuracy measurements of the three wind turbines are computed and compared in this paper, and the results are presented in Table 9 where 1, 2, 3, 4 and 5 represent the five individual forecasts (ARIMA, ARCH, BP, SVM and HKF), and 6, 7, 8 and 9 represent the four combination methods (traditional combination, TCM–NNCT, CPSO–CM–NNCT and GA–CM–NNCT). The results show that the time series methods (ARIMA and ARCH) are superior to the others. On the other hand, the results of the neural network methods (BP neural networks and SVM) are not ideal when they are applied to hourly wind speed forecasting in the Chengde area as the forecasting trend has opposite features to the original series (Fig. 5). The regulation of the combination models are displayed in Table 9 and the four forecasting error measures adopted clearly show that, on average, combination forecasting methods are much better than individual forecasts. According to the comparisons in Table 9, among the four combination methods, the TCM–NNCT method is superior to the traditional combination method but inferior to the CPSO–CM–NNCT method, and the GA–CM–NNCT method shows high volatility in forecasting accuracies. An important conclusion from Table 9 is that the comparison results show differences under different accuracy measures. For example, when the BP model and SVM model are compared, the BP model performs better than the SVM model

Table 9
Comparison of individual method forecast accuracies of three wind turbines.

	Method	Wind turbine 1				Wind turbine 2				Wind turbine 3			
		MAE (m/s)	MSE (m/s)	MAPE (%)	SSE $(m/s)^2$	MAE (m/s)	MSE (m/s)	MAPE (%)	SSE $(m/s)^2$	MAE (m/s)	MSE $(m/s)^2$	MAPE (%)	SSE $(m/s)^2$
Five individual methods	1	1.327	0.119	17.24	819.336	1.219	0.107	14.61	669.903	1.177	0.101	16.67	593.165
	2	1.293	0.116	17.11	776.890	1.219	0.1080	14.53	672.388	1.167	0.1012	16.35	589.366
	3	3.1421	0.326	69.41	5973.34	3.146	0.320	62.93	5910.96	2.890	0.303	61.40	5307.93
	4	3.999	0.307	58.59	5436.85	3.7636	0.3015	51.77	903.274	4.089	0.316	61.00	5751.17
	5	1.529	0.131	19.79	1001.82	1.395	0.121	16.47	842.665	1.404	0.115	18.54	766.232
	Average	2.258	0.199	36.42	2801.60	2.149	0.191	32.06	1799.80	2.145	0.187	34.792	2601.600
Every two method combination	6	1.585	0.136	22.14	1218.50	1.500	0.128	19.05	1117.30	1.985	0.163	27.63	1601.70
	7	1.585	0.136	22.14	1218.50	1.5009	0.128	19.01	1117.30	1.465	0.121	21.27	1021.00
	8	1.549	0.133	21.12	1152.00	1.4690	0.125	18.47	1041.80	1.404	0.117	20.06	909.673
	9	1.454	0.126	18.49	937.36	1.471	0.125	18.52	1051.36	1.410	0.117	20.18	918.356
	Average	1.544	0.133	20.97	1131.59	1.4854	0.1269	18.76	1081.94	1.566	0.129	22.28	1112.68
Every three method combination	6	1.317	0.116	17.87	788.724	1.2404	0.1085	15.12	679.278	1.824	0.1513	24.54	1319.10
	7	1.316	0.116	17.86	785.764	1.4128	0.1205	17.57	677.327	1.190	0.100	16.88	586.774
	8	1.314	0.116	17.75	781.338	1.2371	0.1078	15.33	669.126	1.175	0.100	16.97	580.922
	9	1.642	0.117	21.93	789.802	1.237	0.1080	15.20	673.72	1.410	0.101	20.18	918.36
	Average	1.48	0.120	19.88	787.54	1.270	0.110	15.61	674.48	1.400	0.110	19.82	873.65
Every four method combination	6	1.293	0.115	17.68	763.991	1.2449	0.1083	15.40	661.636	1.165	0.099	16.83	574.439
	7	1.293	0.115	17.68	763.408	1.2248	0.1072	14.99	661.049	1.158	0.099	16.57	567.300
	8	1.2953	0.114	17.45	757.853	1.2298	0.1071	15.25	660.948	1.158	0.099	16.73	565.714
	9	1.308	0.116	17.74	774.119	1.237	0.108	15.42	588.665	1.181	0.101	16.95	602.79
	Average	1.301	0.116	17.67	767.94	1.235	0.108	15.32	624.938	1.171	0.100	16.83	585.971
Five method combination	6	1.294	0.114	17.75	759.158	1.2251	0.1070	15.06	659.226	1.161	0.099	16.87	568.583
	7	1.292	0.114	17.73	758.595	1.2248	0.1068	15.04	657.366	1.158	0.099	16.57	562.041
	8	1.294	0.114	17.44	752.142	1.2245	0.1066	15.24	654.010	1.1512	0.098	16.62	552.634
	9	1.299	0.115	17.07	767.698	1.241	0.1078	14.68	669.618	1.161	0.100	16.18	576.816
	Average	1.296	0.115	17.36	762.165	1.233	0.107	14.89	663.242	1.159	0.098	16.43	568.951
	CPSO–BP	4.317	0.341	58.01	7003.00	4.1610	0.3255	51.54	6102.90	3.907	0.1980	51.26	5115.50S

based on MAE, but based on MSE, MAPE or SSE, the SVM model performs better than the BP model. For wind turbine 1, the MAPE values for every four model combination using combination methods 6, 7, 8 and 9 are 17.68%, 17.68%, 17.45% and 17.74%, respectively, and for the five model combination using combination methods 6, 7, 8 and 9 the MAPE values are 17.75%, 17.73%, 17.44% and 17.07%, respectively. Moreover, on average, the measure for MAPE for the four forecast combination (17.67%) is higher than the MAPE value for the five forecast combination (17.36%). The MAPE along with the other three measures (MAE, MSE and SSE) indicate that the five forecast combination (under combination methods 6, 7, 8 and 9) performs better than the four forecast combination on average. For wind turbine 1, in terms of MAE, MSE and MAPE, every two methods combined by the GA–CM–NNCT method perform better than the CPSO–CM–NNCT models on average. However, with increasing the combined model numbers, the GA–CM–NNCT models are inferior to the CPSO–CM–NNCT models. An interesting result in Table 9 for wind turbine 2 and wind turbine 3 shows that some of GA–CM–NNCT models have better performance in terms of MAE, MSE and MAPE on average compared to CPSO–CM–NNCT although they have undesirable performance for SSE. Even so, SSE is a metric used to measure the fluctuation of forecasting values, so a higher SSE value means a more unstable model. In this way, GA–CM–NNCT models demonstrate high volatility compared to the CPSO–CM–NNCT models. However, for a grid-connected wind power system, high volatility models could be lethal for grid security. Thus, adopting CPSO–CM–NNCT models is advisable in wind speed forecasting. Generally speaking, as SSE is regarded as an objective function to determine the weights of combination forecasting models, the information in Table 9 indicates that the more combination models, the better the forecasting abilities. Additionally, with increasing numbers of individual forecasts, the forecasting accuracy improvement of combination models is not obvious. Thus, compared with individual forecasts, the forecasting accuracy improvement of combination models is encouraging.

Moreover, the CPSO–BP model is proposed to compare the existing forecasts. The forecasting results indicate that this model does not obtain better performance compared to the combination models for all wind turbines and its performance is not consistent in forecasting wind speed for the three wind turbines under four proposed measures (MAE, MSE, MAPE and SSE). Three metrics, MAE, MSE and SSE, indicate that the CPSO–BP model does not perform well compared to the BP model for wind turbine 1 and wind turbine 2; however, the MAPE values of wind turbine 1 is 58.01% and of wind turbine 2 is 51.54%, which are superior to the BP models. Moreover, comparing the SVM models for wind turbine 1 and wind turbine 2 shows that the CPSO–BP models demonstrate better performance in terms of MAPE values. For wind turbine 3, the PSO–BP model performs better than the SVM model for all evaluation criteria and demonstrates better performance than the BP model in terms of MSE, MAPE and SSE. According to the obtained results, although the CPSO–BP model can sometimes improve the forecasting accuracies, its prediction performance varies with data sets and has less robustness compared to the combination models. In conclusion, combination models are the winners both in forecasting accuracies and robustness.

The combination forecasting errors by the no negative constraint method among the 26 combination models of the three wind turbines are shown in Table 10. The forecasting errors of every two combination model, every three combination model, every four combination model and the five combination model from wind turbine 1 to wind turbine 3 provide the following conclusions:

Table 10
Comparison of forecast accuracies of three wind turbines for different combination models (TCM–NNCT).

	Methods	Wind turbine 1			Wind turbine 2			Wind turbine 3		
		MAE $(m/s)^2$	MSE (m/s)	MAPE (%)	MAE (m/s)	MSE (m/s)	MAPE (%)	MAE (m/s)	MSE (m/s)	MAPE (%)
Every two methods combination	1,2	1.2936	0.1161	17.11	1.2224	0.1078	14.72	1.1674	0.1011	16.35
	1,3	1.2973	0.1155	17.96	1.2300	0.1073	15.13	1.1732	0.1003	17.10
	1,4	1.2952	0.1166	17.48	1.2210	0.1073	14.74	1.1778	0.1015	16.77
	1,5	1.3182	0.1183	17.35	1.2155	0.1077	14.55	1.1768	0.1014	16.62
	2,3	1.3003	0.1152	17.75	1.2280	0.1072	15.09	1.1604	0.0995	16.86
	2,4	1.2956	0.1160	17.29	1.2217	0.1078	14.69	1.1677	0.1010	16.47
	2,5	1.2921	0.1160	17.11	1.2136	0.1078	14.46	1.1675	0.1012	16.35
	3,4	3.7115	0.2878	58.98	3.6666	0.2901	53.08	3.7252	0.2842	58.76
	3,5	1.5237	0.2059	12.93	1.3949	0.1190	17.13	1.3393	0.1096	18.80
	4,5	1.5298	0.1319	19.79	1.3950	0.1209	16.47	1.4023	0.1152	18.58
	Average	1.5857	0.1439	21.38	1.5009	0.1283	19.01	1.4658	0.1215	21.27
Every three methods combination	1,2,3	1.2936	0.1149	17.77	1.2285	0.1072	15.10	1.1491	0.0987	16.57
	1,2,4	1.2918	0.1159	17.28	2.1553	0.1744	28.06	1.1670	0.1010	16.44
	1,2,5	1.2932	0.1160	17.14	1.9981	0.1622	25.67	1.1676	0.1012	16.36
	1,3,4	1.2934	0.1153	17.94	1.2308	0.1072	15.13	1.1751	0.1001	16.99
	1,3,5	1.2932	0.1152	17.89	1.2308	0.1071	15.10	1.1745	0.1001	17.09
	1,4,5	1.2941	0.1166	17.44	1.2177	0.1076	14.67	1.1773	0.1014	16.71
	2,3,4	1.2998	0.1151	17.73	1.2281	0.1072	15.09	1.1600	0.0990	16.70
	2,3,5	1.2987	0.1151	17.73	1.2251	0.1070	15.06	1.1618	0.0994	16.87
	2,4,5	1.2929	0.1159	17.23	1.2161	0.1076	14.59	1.1674	0.1010	16.48
	3,4,5	1.5150	0.1274	20.47	1.3976	0.1179	17.23	1.4023	0.1071	18.58
	Average	1.3166	0.1167	17.86	1.4128	0.1205	17.57	1.1902	0.1009	16.88
Every four methods combination	1,2,3,4	1.2938	0.1149	17.77	1.2289	0.1072	15.10	1.1479	0.0981	16.36
	1,2,3,5	1.2924	0.1148	17.75	1.2228	0.1069	15.03	1.1495	0.0987	16.59
	1,2,4,5	1.2910	0.1159	17.24	1.2175	0.1076	14.66	1.1663	0.1010	16.44
	1,3,4,5	1.2920	0.1152	17.89	1.3279	0.1130	17.10	1.1698	0.0995	16.83
	2,3,4,5	1.2999	0.1148	17.72	1.2275	0.1069	15.09	1.1575	0.0988	16.62
	Average	1.2938	0.1151	17.67	1.2449	0.1083	15.40	1.1582	0.0992	16.57
Five methods	1,2,3,4,5	1.2944	0.1148	17.73	1.2251	0.1070	15.06	1.1502	0.0980	16.37

1) The forecasting results of combination models vary depending on the evaluation measures. For example, the forecasting results of wind turbine 1 show that the ARIMA and ARCH (1) combination model perform better than the ARIMA and BP combination model for MAE and MAPE, but it is inferior to the ARIMA and BP combination model for the MSE measure. For wind turbine 2, the ARIMA and ARCH combination model is not as good as the ARIMA and BP combination model for the MSE measure, but it is better than the ARIMA and BP combination model when using the MAE and MAPE measures.
2) On average, the more individual forecasts that are combined, the higher the forecasting accuracy. For wind turbine 1, the average values of MAE, MSE and MAPE for every three method combination are 1.3166m/s, 0.1167m/s and 17.86%, respectively, which are better than the same values for every four method combination but lower than the values for every two method combination 1.5857 m/s, 0.1439 m/s and 21.38%. For wind turbine 2 and wind turbine 3, every three method combination model performs worse than every four method combination model, but better than every two method combination model. In Table 8, the forecasting results for wind turbine 2 and wind turbine 3 show that the five forecast combination models are the best models among all of the combination models; however, for wind turbine 1, the five forecast combination model is not superior to every four method combination model according to the MAE and MAPE values.
3) The forecasting accuracies of two well-performed forecast combination models may not always be better than the combination of a good model and a bad model based on different error measures. For wind turbine 1, the ARCH and BP model are better than the SVM model, but the forecasting results of the ARCH and BP combination model are inferior to the ARCH and SVM combination model, according to the measures of MAE and MAPE.

7. Conclusions

As one of the most promising renewable energy sources, wind energy performs a vital role in routine production. As wind speed directly influences wind energy generation, forecasting wind speed is important. To understand and successfully implement wind speed forecasting, many studies have been developed. However, as these studies are not always suitable to forecast various wind speed time series, the forecasting accuracy must be improved for different wind speed time series. This study reviews the existing wind speed forecasting approaches, provides a detailed presentation about the combined approaches for wind speed forecasting and then proposes no negative constraint theory (NNCT) and artificial intelligence-based combination models to predict future wind speed. To illustrate the effectiveness of the proposed combination models, wind speed data of three wind turbines in the Chengde region of China were simulated, and the simulation results indicate that the proposed combination methods are more effective than traditional combination forecasts. The traditional combination models, TCM–NNCT, CPSO–CM–NNCT and GA–CM–NNCT models were compared to each other, and the CPSO–CM–NNCT was found to be superior to the TCM–NNCT while the GA–CM–NNCT showed high volatility. Otherwise, CPSO–BP is adopted to compare forecasting results of combination models, even though it shows lower forecasting accuracy and poor robustness. While there is no convincing evidence for negative weight, the forecasting results indicate that if weights are allowed to be negative, they are, in fact, better than nothing. Although the comparison results vary according to the monitored accuracy measures, the performances of the combination models are encouraging. Combinations result in a deterioration of the forecasting accuracy due to statistical measures that occur when

comparing combination models. Nonetheless, the performance after the proposed combining forecasts to increase the robustness of the models is encouraging. All of the combination models demonstrate better performance than do the individual forecast models. Accordingly, the overall proposed combination forecasting result is a significant and worthwhile improvement. Therefore, the proposed combined models have integrated different models' advantages and are very suitable for large wind farms in China.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 71171102).

References

[1] Harmsen J, Patel MK. The impact of copper scarcity on the efficiency of 2050 global renewable energy scenarios. Energy 2013;50:62–73.
[2] Khatib H. IEA world energy outlook 2010—a comment. Energy Policy 2011;39:2507–11.
[3] Li Danny HW, Liu Y, Joseph C. Zero energy buildings and sustainable development implications – a review. Energy 2013;54:1–10.
[4] Yong K, Kim J, Lee J, Ryu M, Lee J. An assessment of wind energy potential at the demonstration offshore wind farm in Korea. Energy 2012;46:555–63.
[5] ⟨http://www.sgcc.com.cn/shouye/tbxw/279179.shtml⟩.
[6] ⟨http://www.51baogao.cn/xinnengyuan/200901fenglifadian.shtml⟩.
[7] Guo Z, Zhao J, Zhang W, Wang J. A corrected hybrid app roach for win d speed forecasting in Hexi Corridor of China. Energy 2011;36(16):68–1679.
[8] Esen Hikmet, Mustafa Inalli, Abdulkadir Sengur, Esen Mehmet. Performance prediction of a ground-coupled heat pump system using artificial neural networks. Expert Syst Appl 2008;35:1940–8.
[9] Esen Hikmet, Mustafa Inalli, Abdulkadir Sengur, Esen Mehmet. Forecasting of a ground-coupled heat pump performance using neural networks with statistical data weighting pre-processing. Int J Therm Sci 2008;47:431–41.
[10] Esen Hikmet, Inalli Mustafa, Sengur Abdulkadir, Esen Mehmet. Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system. Energy Build 2008:1074–83.
[11] Esen Hikmet, Ozgen Filiz, Esen Mehmet, Sengur Abdulkadir. Artificial neural network and wavelet neural network approaches for modeling of a solar air heater. Expert Syst Appl 2009;36:11240–8.
[12] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13:915–20.
[13] Esen Hikmet, Inalli Mustafa, Sengur Abdulkadir, Esen Mehmet. Modeling a ground-coupled heat pump system by a support vector machine. Renew Energy 2008;33:1814–23.
[14] Esen Hikmet, Ozgen Filiz, Esen Mehmet, Sengur Abdulkadir. Modelling of a new solar air heater through least-squares support vector machines. Expert Syst Appl 2009;36:10673–82.
[15] Guo Z, Wu J, Lu H, Wang J. A case study on a hybrid wind speed forecasting method using BP neural network. Knowl.-Based Syst 2011;24:1048–56.
[16] Bates JM, Granger CWJ. The combination of forecasts. Oper Res Q 1969;20:451–68.
[17] Wang J, Liang-dong Liu, Zuo-yi Wang. The status and development of the combination forecast method. Forecast 1997;6:37–8.
[18] Chen H. The validity of the theory and its application of combination forecast methods. Beijing: Science Press; 2008.
[19] 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.
[20] Cassola F, Burlando M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. Appl Energy 2012;99:154–66.
[21] Liu H, Shi J, Erdem E. Prediction of wind speed time series using modified Taylor Kriging method. Energy 2010;35:4870–9.
[22] González-Mingueza Carlos, Muñoz-Gutiérrez Francisco. Wind prediction using Weather Research Forecasting model (WRF): a case study in Peru. Energy Convers Manag 2014;81:363–73.
[23] Ren Chao, Anb Ning, Wang Jianzhou, Li Lian, Hu Bin, Shang Duo. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl-Based Syst 2014;56:226–39.
[24] Su Zhongyue, Wang Jianzhou, Lu Haiyan, Zhao Ge. A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. Energy Convers Manag 2014;85:443–52.
[25] Zhao W, Wang J, Lu H. Combining forecasts of electricity consumption in China with time-varying weights updated by a high-order Markov chain model. Omega 2014;45:80–91.
[26] Bouzgou H, Benoudjit N. Multiple architecture system for wind speed prediction. Appl Energy 2011;88:2463–71.
[27] Sánchez Ismael. Adaptive combination of forecasts with application to wind energy. Int J Forecast 2008;24:679–93.
[28] Lei C, Ran L. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. In: Proceedings of the third international conference on electric utility Deregulation and Restructuring and Power Technologies (DRPT); 2008. P. 2525–9.
[29] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37:241–9.
[30] Zhou H, Jiang JX, Huang M. Short-term wind power prediction based on statistical clustering. In: Proceedings of the IEEE power and energy society general meeting; 2011. P. 1–7.
[31] Pourmousavi Kani SA, Ardehali MM. Very short-term wind speed prediction: a new artificial neural network—Markov chain model. Energy Convers Manag 2011;52:738–45.
[32] Xingpei L, Yibing L, Weidong X. Wind speed prediction based on genetic neural network. In: Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications (ICIEA); 2009. P. 2448–51.
[33] Hui T, Dongxiao N. Combining simulate anneal algorithm with support vector regression to forecast wind speed. In: Proceedings of the second IITA International Conference on Geoscience and Remote Sensing (IITA-GRS); 2010. P. 92–4.
[34] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerodyn 2008;96:2348–62.
[35] Zhang W, Deng Y. Short-term wind speed prediction based on combination model.
[36] Ma Y, Tang X. Multi-objective optimal combination forecast model. Stat Res 1997;4:45–8.
[37] Pan D, Liu H, Li Y. Wind farms wind speed forecast Optimization Model based on time series analysis and Kalman filtering algorithms. Power Syst Technol 2008;32:82–6.
[38] Gao Y, Li X. Chaos particle swarm optimization algorithm. Comput Sci Technol 2004;31(7):8.
[39] James K, Russell E. Particle swarm optimization. IEEE 1995;95:1942–8.
[40] Liu B, Wang L, Jin Y. an effective pso-based memetic algorithm for flow shop scheduling. IEEE Trans Syst Man Cybern—Part B: Cybern 2007;37(1).
[41] Huang Han-Xiong, Li Jiong-Cheng, Xiao Cheng-Long. A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. Expert Syst Appl 2015;42:146–55.
[42] Homayounia Seyed Mahdi, Tang Sai Hong, Motlagh Omid. A genetic algorithm for optimization of integrated scheduling of cranes, vehicles, and storage platforms at automated container terminals. J Comput Appl Math J Comput Appl Math 2014;270:545–56.
[43] Box G, Jenkins G. Time series analysis. Forecasting and control. San Francisco, CA: Holden-Day; 1970.
[44] Pai P, Lin C. A hybrid ARIMA and support vector machines model in stock price forecast. Int J Manag Sci Omega 2005;33:497–505.
[45] Engle Robert F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 1982;50:987–1007.
[46] Yi D. Data analysis and eviews application. Beijing, China: Renmin University Press; 2008.
[47] Yang S, Li H. Financial crisis warning model based on BP neural network. Syst Eng—Theory Pract 2005;1:12–18, 26.
[48] Vapnik Vladimir N. An overview of statistical learning theory. IEEE Trans Neural Netw 1999;10:988–99.
[49] Quine Brendan M. A derivative-free implementation of the extended Kalman filter. Automatica 2006;42:1927–34.
[50] Lu Jun J. Development of Chengde wind energy on carbon-free. Coal Technol 2010;29:178–80.

Update 1 of 2

Renewable and Sustainable Energy Reviews

Volume 73, Issue , June 2017, Page 276

DOI: https://doi.org/10.1016/j.rser.2017.01.149

Renewable and Sustainable Energy Reviews 73 (2017) 276

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Corrigendum

Corrigendum to "Combined forecasting models for wind energy forecasting: A case study in China" [Renew. Sustain. Energy Rev. 44 (2015) 271–288]

Ling Xiao[a], Jianzhou Wang[b], Yao Dong[a], Jie Wu[a]

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

In the above published article, the following errors needs to be corrected:

(1) In the Section 3, the formula (8) "SSE = $\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "SSE = $\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(2) In the Section 3.1, the formula (10) "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(3) In the Section 3.3, the formula (11) "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(4) In the Section 3.4.1, the formula (15) "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "min $\boldsymbol{J=L^TEL}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(5) We have made a mistake about the reference of Table 1 in our study, which should be referred to the paper titled: "A review of combined approaches for prediction of short-term wind speed and power" - i.e., reference [19] rather than reference [17].

The authors would like to apologize for any inconvenience these errors may have caused to anyone in the reading or understanding of this review paper.

DOI of original article: http://dx.doi.org/10.1016/j.rser.2014.12.012
E-mail address: 13519619146@126.com (J. Wang).

http://dx.doi.org/10.1016/j.rser.2017.01.149

Available online 27 January 2017

Update 2 of 2

Renewable and Sustainable Energy Reviews

Volume 74, Issue , July 2017, Page 1407

DOI: https://doi.org/10.1016/j.rser.2016.12.091

Renewable and Sustainable Energy Reviews 74 (2017) 1407

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Corrigendum

Corrigendum to "Combined forecasting models for wind energy forecasting: A case study in China" [Renew. Sustain. Energy Rev. 44 (2015) 271–288]

Ling Xiao[a], Jianzhou Wang[b,*], Yao Dong[a], Jie Wu[a]

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

In the above published article, the following errors needs to be corrected:

(1) In the Section 3, the formula (8) "SSE = $\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "SSE = $\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(2) In the Section 3.1, the formula (10) "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(3) In the Section 3.3, the formula (11) "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(4) In the Section 3.4.1, the formula (15) "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it}$" should be corrected as "$\min \boldsymbol{J}=\boldsymbol{L}^T\boldsymbol{E}\boldsymbol{L}=\sum_{t=1}^{T}\sum_{j=1}^{m}\sum_{i=1}^{m} l_i l_j e_{it} e_{jt}$"

(5) We have made a mistake about the reference of Table 1 in our study, which should be referred to the paper titled: "A review of combined approaches for prediction of short-term wind speed and power" – i.e., Ref. [19] rather than Ref. [17].

The authors would like to apologize for any inconvenience these errors may have caused to anyone in the reading or understanding of this review paper.

DOI of original article: http://dx.doi.org/10.1016/j.rser.2014.12.012
* Corresponding author.
E-mail address: wjzdufe@gmail.com (J. Wang).

http://dx.doi.org/10.1016/j.rser.2016.12.091

Available online 30 December 2016
1364-0321/