Research Article

Short-Term Wind Speed Forecasting Using Support Vector Regression Optimized by Cuckoo Optimization Algorithm

Jianzhou Wang, Qingping Zhou, Haiyan Jiang, and Ru Hou

1 School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
2 School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China

Correspondence should be addressed to Qingping Zhou; zhouqp12@lzu.edu.cn

Received 6 June 2014; Revised 5 September 2014; Accepted 16 September 2014

1. Introduction

Energy is an indispensable ingredient of economic growth, but rapid global economic growth has contributed to the ever-increasing demand for energy. However, traditional energy sources, such as coal, oil, and natural gas, not only are facing severe depletion, but also produce various forms of air pollution, which have adverse human health effects [1]. Thus, there is an urgent need to develop renewable resources that aid governments in reducing their reliance on fossil fuels, simultaneously decreasing carbon emissions and improving air quality. Hence, renewable energy sources such as wind power, hydropower, solar energy, geothermal energy, and biofuels have gained more and more attention in recent years [2]. Among these renewable resources, wind power has the characteristics of a mature technique and is relatively cost competitive. Wind power has naturally become the fastest growing energy resource around the world. According to the Global Wind Energy Council (GWEIC), the global cumulative installed capacity reached 282.5 GW in 2012 [3]. Moreover, approximately 10 million MW of wind energy is continuously installed every year. By 2020, 12 percent of the world's power could feasibly come from wind energy. The Chinese wind market shared approximately 27 percent of the global wind energy in 2012. The wind-generated electricity in China reached 100.4 billion kWh in 2012, accounting for 2% of the total electricity consumption.

However, high wind power penetration requires addressing many issues, such as balancing power production and consumption, interconnection standards, and power system stability and reliability [4]. The power grid dispatchers face a significant challenge due to the increased integration of large scale wind power into existing power systems. To address these challenges, dispatchers must forecast wind power production for wind turbine control, preload sharing, and power system management, as well as for determining the power reserves [5]. The relationship between wind speed and wind power is basically cubic; therefore, a small error in the wind speed forecast will lead to a larger error in wind...
power production. Thus, the accurate forecast of short-term wind speed prediction plays a crucial role in wind power forecasting [6].

A number of methods have been developed to forecast wind speed. These methods are usually classified into four categories: (a) physical modeling methods, (b) time series models, (c) artificial intelligence models, and (d) hybrid models. Physical models, such as numerical weather prediction, require various weather data, such as temperature, pressure, and topography information, and solve complex mathematical equations. However, physical modeling approaches are not effective for very short-term predictions. In the time series forecasting methods, mathematical equations are used to predict wind speed based on historical data. The commonly used time series models are the autoregressive moving average model (ARMA), autoregressive integrated moving average model (ARIMA), fractional ARIMA, and seasonal ARIMA [7–9]. These time series methods perform well in situations when the data set has low frequency in nature, such as weekly/yearly patterns, but have difficulties with high-frequency changes and rapid variations of the original data.

With the capability to track complex nonlinearity systems, artificial intelligence techniques have received considerable attention in the wind energy field. An artificial intelligence technique is a generalized term that encompasses artificial neural networks (ANNs) [10, 11], fuzzy logic methods [12, 13], and support vector machines (SVM) [6, 14–21]. Mohandes et al. [16] used the SVM model to forecast the mean daily wind speed and compared the results with the multilayer perceptron (MLP) neural networks. They concluded that the SVM model outperformed the MLP forecasting method. Similar conclusions are found elsewhere [18, 19]. Hybrid forecasting methods, which involve the mixture of several methods or algorithms, have also been proven to obtain better performance. Zhou et al. [21] presented a hybrid method for wind speed prediction based on wavelet analysis and support vector machine. The experimental results show that this method is capable of improving forecasting precision and generalization performance. Salcedo-Sanz et al. [17] successfully used SVM to forecast wind speed at a Spanish wind farm, in which the hyperparameters are tuned by two evolutionary algorithms including the evolutionary programming algorithm (EP) and particle swarm optimization approach (PSO). Because an abundance of redundant information in the wind speed series would decrease the forecasting validity and precision, Cheng and Guo [14] proposed a hybrid forecasting method based on SVM and information granulation. The original data were refined with fuzzy information granulation, and then the refined data are forecasted by the SVM model. Liu et al. [15] presented a hybrid model based on wavelet transform (WT), genetic algorithm (GA), and SVM. The parameters in SVM were chosen by GA. A case study demonstrates that the hybrid method outperforms the comparison models. A hybrid approach [22] based on the ensemble empirical mode decomposition (EEMD) and SVM was proposed to forecast the mean hourly wind speed, which proved to be effective in improving the prediction accuracy.

Wind speed forecasting based on weather data encounters some difficulties. The accurate measurement of weather data is commonly expensive and unavailable. On the other hand, inaccurate measurements would introduce information to the forecasting model [6, 22, 23]. As a dynamic system, wind speed has a relationship with its past values at any time [24]. Therefore, achieving a low wind speed forecasting error through a relatively simple forecasting model with the historical wind speed data is desired.

Based on the above discussion, this paper proposed a hybrid wind speed forecasting model based on RP and optimized SVM, using its historical data. RP is used to analyze the wind system and to extract the useful information from the wind speed data; the optimized SVM model is then operated to obtain accurate hourly wind speeds. An actual case study is tested in the Shandong province in China.

This paper is organized as follows. Section 2 describes the framework of the study. Section 3 demonstrates the fundamentals of the methods, including the recurrence plot, support vector regression model, and some optimized algorithms. Section 4 presents some chosen performance criteria. Section 5 presents an experimental study to evaluate the proposed models, and a discussion and some conclusions are drawn in Section 6.

2. Framework of the Hybrid Model

The framework of the hybrid model is described in Figure 1. The detailed processes are given as follows.

1. Apply the RP to analyze the predictability of the original wind speed data, which forms the prerequisite conditions for accurate forecasting. Moreover, RP can also extract the optimal information from the original data that is used as the input variables of the optimized SVR models.

2. The SVR models, in which parameters are optimized by GA, PSO, and COA, are established for multi-step-ahead wind speed prediction.

3. Compare the forecasts and the actual wind speed. Meanwhile, a two-tailed hypothesis test is performed to evaluate the stability of the forecasting methods.

3. Methodology

In this section, the fundamentals of the recurrence plot (RP) and support vector regression (SVR) are introduced. To obtain the best performance of the SVR method, three optimized algorithms, including genetic algorithm (GA), particle swarm optimization (PSO), and cuckoo optimization algorithm (COA), are used to select the optimal parameters of SVR.

3.1. Recurrence Plot

Recurrence plot (RP) can identify whether the wind speed series are generated by a nonlinear deterministic process and select the input set of SVR. Recurrence is a basic property of the dynamic system. This property can be visualized and analyzed using the powerful tool of recurrence plot (RP). Eckmann et al. [25] first introduced
RP and used it to detect nonlinearities and chaotic dynamics in experimental signals in physics. The RP technique has been used recently to identify structural changes and hidden patterns in data or detect similarities in patterns across the time series in many fields, such as economics [26], physiology [27], and energy markets [27]. In this paper, the RP technique will be applied to analyze the predictability of the wind time series.

In the process of producing an RP, the original one-dimensional data are first mapped to a higher dimensional reconstructed phase space. Letting \( \{x_i\}_{i=1}^{N} \) be the wind speed time series, the reconstructed vector in the phase space can be represented as

\[
X_i = (x_i, x_{i+1}, \ldots, x_{i+(m-1)\tau}), \quad i = 1, 2, \ldots, N - (m - 1)\tau, \tag{1}
\]

where \( m \) is the embedding dimension and \( \tau \) is the embedding delay.

Recurrence is defined as the degree of closeness between two different elements in the phase space. RP visualizes the behavior of recurrence via a graphical representation of the distance matrix:

\[
R_{i,j} = \Theta(\varepsilon - \|X_i - X_j\|), \quad i, j = 1, 2, \ldots, N - (m - 1)\tau, \tag{2}
\]

where \( \varepsilon \) is a predefined threshold, \( \| \cdot \| \) is the usual Euclidean norm, and \( \Theta(\cdot) \) is the Heaviside function. RP can identify three types of dynamic systems, as shown in Figure 2. (1) Periodic systems are characterized by diagonal parallel lines with the same periodic distance. Data with this RP indicates that the data are periodic or quasiperiodic. (2) Chaotic systems are characterized by diametric lines with the irregular distances which have been cut. Data with this RP reveals that the data often has abrupt changes. (3) Stochastic systems are characterized by many individual dots that indicate their distribution is quite irregular. An example of such an RP is that of a random time series.

3.2. Support Vector Regression Model. The support vector machine (SVM), proposed by Vapnik and his coworkers [28], has been successfully used for pattern recognition,
Figure 2: Recurrence plots of periodic series, chaotic series, and stochastic series.
classification, and regression. SVM is based on the statistical learning theory, the Vapnik-Chervonenkis dimension theory, and the structural risk minimization (SRM) principle. Support vector regression (SVR), also developed by Vapnik and coworkers [29], is an extension of SVM. Compared with SVM, SVR minimizes the generalized error bound instead of minimizing the observed training error [30].

To gain the optimal generalization ability, the basic principle of SVR is to map the original data into a high-dimensional feature space through nonlinear mapping [31]. Supposing the training data are \(\{(x_1, y_1), \ldots, (x_n, y_n)\}\), the regression formula can be expressed as

\[
 f(x) = \sum_{i=1}^{D} \omega_i \Phi_i(x) + b, \quad \Phi_i : R^n \rightarrow F, \quad \omega_i \in F, \quad b \in R.
\]

(3)

Here \(\{\omega_i\}_{i=1}^{D}\) are weights estimated from the training data, \(b\) is the threshold value, and \(\{\Phi_i\}_{i=1}^{D}\) are nonlinear mapping functions which map the sample datasets to high-dimensional feature space \(F\). Based on the SRM principle, the weights \(\{\omega_i\}_{i=1}^{D}\) can be obtained from the sample data by minimizing the following quadratic programming problem:

\[
 \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-),
\]

subject to

\[
 |y_i - \omega \cdot \Phi(x_i) - b| \leq \varepsilon + \xi_i, \quad \xi_i^+, \xi_i^- \geq 0, \quad i = 1, 2, \ldots, n,
\]

(5)

where constant \(C\), also called the penalties factor, is greater than zero and determines the tradeoff by minimizing the training error and minimizing the model complexity. \(\xi_i\) and \(\xi_i^+\) are the slack variables. \(\varepsilon(\cdot)\) is the \(\varepsilon\)-intensive loss function and is defined as follows:

\[
 \varepsilon(y_i) = \begin{cases} 
 0, & |f(x_i) - y_i| < \varepsilon, \\
 |f(x_i) - y_i| - \varepsilon, & \text{otherwise}.
\end{cases}
\]

(6)

By solving the optimization problem, the estimation function can be obtained as follows:

\[
 f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^+) K(x_i, x_i) + b,
\]

(7)

subject to

\[
 \sum_{i=1}^{n} (\alpha_i - \alpha_i^+) = 0, \quad 0 \leq \alpha_i, \quad \alpha_i^+ \leq C,
\]

(8)

where \(K(x_i, x_j)\) is a kernel function that represents the inner product in the \(D\)-dimensional feature space. SVR is characterized by the use of “kernel trick” to apply linear classification techniques to nonlinear classification problems. Consider

\[
 K(x_i, x_j) = \sum_{i=1}^{D} \Phi_i(x_i) \cdot \Phi(x_j).
\]

(9)

The typical kernel functions include Gaussian Radial Basis Functions (RBF), polynomial kernel functions, Gaussian kernel functions, and sigmoid kernel functions. In this paper, the RBF kernel function is used due to its computational simplicity and its robustness in dealing with nonlinear data. The RBF is defined mathematically as

\[
 K(x_i, x_j) = e^{-(x_i - x_j)^2/d^2},
\]

(10)

where \(d\) is the width of RBF.

3.3. Optimized Algorithms. Evolutionary algorithms and heuristic optimization algorithms have played a significant part in addressing practical mathematics and engineering problems in recent years [6, 8]. Some optimization algorithms used in this study are described as follows.

3.3.1. Genetic Algorithm. The genetic algorithm (GA) proposed by Holland [32] is inspired by the living natural evolution procedure. GA is a global optimization method that mimics the biological evolution mechanisms, including survival of the fittest, crossover, and mutation. It performs well in complex optimization problems given its simplicity and robustness, and it has been successfully applied in various forecasting fields. To effectively search for the optimal solution, attention should be focused on the selection of population size, crossover rate, and mutation rate. More details of GA can be found elsewhere [33]. The basic optimization procedure is described as follows.

Step 1. Randomly initialize a group of chromosomes. The chromosomes correspond to the solutions in an optimization problem.

Step 2. Evaluate the fitness of the candidate chromosomes. The chromosomes with smaller fitness are more likely to survive.

Step 3. Produce a new generation according to the proportionate. The chromosomes with higher fitness are more likely to be selected.

Step 4. Perform a crossover or mutation operation to generate new candidate chromosomes.

Step 5. Return the solution if the optimal solution has been achieved or return to Step 2 until a certain number of iterations has been reached.

3.3.2. Particle Swarm Optimization Algorithm. PSO, first proposed by Kennedy and Eberhart [34], is an efficient evolutionary computation technique. In the PSO algorithm, a swarm of particles keep updating around a search space, and an optimal solution will emerge at the end. Suppose the search space is \(m\)-dimensional space and the number of particles is \(n\); then the \(i\)th particle can be defined as an \(m\)-dimensional vector \(X_i = (x_{i1}, x_{i2}, \ldots, x_{im})\) \((i = 1, 2, \ldots, n)\) and the update velocity of this particle is \(V_i = (v_{i1}, v_{i2}, \ldots, v_{im})\) \((i = 1, 2, \ldots, n)\). The mechanism of the algorithm is described as follows.
Step 1. Randomly initialize a population of \(n\) particles with positions and velocities. The positions represent the candidate solutions in the optimized problems.

Step 2. Evaluate the fitness value for each particle.

Step 3. Calculate the position of the best fitness value for each particle from its historical movement. The best position of the \(i\)th particle is denoted as \(P_i = (p_{i1}, p_{i2}, \ldots, p_{in})\).

Step 4. Evaluate the position of the best fitness value for all particles. The best position for all particles is expressed as \(P^* = (P_{g1}, P_{g2}, \ldots, P_{gn})\).

Step 5. Update the speed and the position of particles using the following:

\[
v_{i}^{k+1} = \omega \cdot v_{i}^{k} + c_1 \cdot r_1 \cdot \left( p_{i}^{k} - x_{i}^{k} \right) + c_2 \cdot r_2 \cdot \left( P_{g}^{k} - x_{i}^{k} \right), (11)
\]

\[
x_{i}^{k+1} = x_{i}^{k} + v_{i}^{k+1}, (12)
\]

where \(\omega\) is the inertia weight, \(r_1\) and \(r_2\) are two uniformly distributed random variables in the interval from 0 to 1, and \(c_1\) and \(c_2\) are the personal learning and global learning coefficients, respectively. The first term on the right side of (II) makes the particles move around the search space; the second term makes the particles stay close to their personal best positions, and the last term, known as the social collaboration, has the advantage of quickly approaching the global optimal value.

Step 6. If the terminated condition is not satisfied, return to Step 2. The terminated condition is usually a predefined iteration value or fitness value.

3.3.3. Cuckoo Optimization Algorithm. The cuckoo optimization algorithm (COA) is a metaheuristic optimization method first proposed by Yang and Deb in 2009 [35]. The basic idea of this new algorithm was based on the obligate brood parasitic behavior of the cuckoo and its method in egg laying and breeding. In nature, many animals search for food in a quasirandom manner. In fact, an animal’s path of searching for food is effectively a random walk because the next move is based on the current location and the transition probability to the next location. Choosing the direction depends on a probability which can be modeled mathematically. A recent study presents that *Drosophila melanogaster*, a fruit fly, wander in their territory using a series of straight flight paths punctuated by sudden 90° turns, leading to a Levy flight-style intermittent scale-free search pattern.

The basic COA is defined by the endeavor to survive among cuckoos. In the survival competition, the cuckoo makes two main operations, including a random search based on the probability of finding an alien egg of a host bird and a direct search based on Levy flights. With the combination of the two operations, COA is more effective than other optimization methods for nonconvex and complex optimization problems [36]. In the COA method, each nest represents a candidate solution and a number of nests are used for finding the appropriate solution of the optimization problem. The main steps of the COA method are described as follows.

Step 1. Initialize the population. Suppose the number of host nests is \(N_p\) and the optimization problem has \(N\) parameters. Then, a population of \(N_p\) nests is denoted as \(X_i = [P_{i1}, P_{i2}, \ldots, P_{in}] (i = 1, 2, \ldots, N_p)\).

Step 2. Update the new solution via Levy flights. The best path of the Levy flights is estimated by Mantegna’s algorithm [37]. The new solution of each nest is calculated using the following:

\[
X_i^{new} = X_{best} + \alpha \cdot \text{rand}_1 \cdot \Delta X_i^{new}, \\
\Delta X_i^{new} = v \cdot \frac{\sigma_x(\beta)}{\sigma_y(\beta)} \cdot (X_{best} - \text{Gbest}), (13)
\]

where \(X_{best}\) is the best solution for the individual, \(\text{Gbest}\) is the best solution for the total population, \(\alpha > 0\) is the iteration step size, and \(\text{rand}_1\) is a uniformly distributed random number in the interval from 0 to 1. The variable \(v\) is defined as \(\text{rand}_1 / \text{rand}_2 \cdot \text{rand}_3\) and \(\text{rand}_4\) having normal distributions with the standard deviations \(\sigma_x(\beta)\) and \(\sigma_y(\beta)\) given as follows:

\[
\sigma_x(\beta) = \left[ \frac{\sin(\pi \beta/2)}{\Gamma((1+\beta)/2) \beta^{1/\beta}} \right]^{1/\beta}, \\
\sigma_y(\beta) = 1, (14)
\]

where \(\beta\) is the distribution factor between 0.3 and 1.99 and \(\Gamma()\) is the gamma function.

Step 3. Discover the alien egg and randomize the nests. Suppose the probability of a host bird to discover an alien egg in its nest is \(P_a\). The action of this discovery also creates a new solution, which is defined as the following:

\[
X_i^{dis} = X_{best} + K \cdot \Delta X_i^{dis}, \\
\Delta X_i^{dis} = \text{rand}_2 \cdot \left[ \text{rand}_1 (X_{best}) - \text{rand}_2 (X_{best}) \right], (15)
\]

where \(\text{rand}_2\) is a uniformly distributed random number and ranges from 0 to 1. The variables \(\text{rand}_1(X_{best})\) and \(\text{rand}_2(X_{best})\) are the random perturbations of nest positions in \(X_{best}\). Parameter \(K\) is determined based on \(P_a\). Consider

\[
K = \begin{cases} 
1, & \text{if } \text{rand}_2 < P_a, \\
0, & \text{otherwise}. 
\end{cases} (16)
\]

Step 4. Stop the algorithm when the maximum number of iterations is reached.

4. Model Performance Evaluation

To evaluate the performance of the optimized SVR models, MAE, RMSE, MAPE, \(R^2\), and IA (see (17) through (21)) are used to measure the forecasting accuracy. MAE and RMSE
are employed to measure the absolute error of the actual data and the forecasts. Including the quadratic term in the RMSE makes it more sensitive to excessive values than MAE. Moreover, the relative error is also evaluated using MAPE. $R^2$ and IA measure the overall similarity between the actual and the forecasted wind speed. They both range from 0 to 1. For a perfect forecasting model, MAE, RMSE, and MAPE take values of 0, and the values of $R^2$ and IA are equal to 1. Let $A_t$ and $P_t$ be the actual and the forecasting values at time-point $t$, respectively; $\bar{A}$ and $\bar{P}$ denote the mean of the actual and the forecasting values, respectively. Then, these performance metrics are defined as follows.

(i) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |A_t - P_t|,$$

(ii) Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (A_t - P_t)^2}.$$  \hspace{1cm} (18)

(iii) Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \frac{|A_t - P_t|}{|A_t|} \times 100\%.$$ \hspace{1cm} (19)

(iv) Coefficient of determination ($R^2$):

$$R^2 = \frac{\sum_{t=1}^{N} (P_t - \bar{P})(A_t - \bar{A})^2}{\sum_{t=1}^{N} (P_t - \bar{P})^2 \sum_{t=1}^{N} (A_t - \bar{A})^2}.$$ \hspace{1cm} (20)

(v) Index of agreement (IA):

$$\text{IA} = 1 - \frac{\sum_{t=1}^{N} (P_t - A_t)^2}{\sum_{t=1}^{N} (|P_t - \bar{A}| + |A_t - \bar{A}|)^2}.$$ \hspace{1cm} (21)

5. Case Studies

The hybrid methods are applied to wind speed prediction at the Shandong wind farm in China. The prediction horizon is 15 min (one-step-ahead), 30 min (two-step-ahead), 45 min (three-step-ahead), and 1 h ahead (four-step-ahead). In this paper, the following forecasting approaches are performed:

(i) forecasting based on SVR and COA, that is, the COA-SVR method,

(ii) forecasting based on SVR and PSO, that is, the PSO-SVR method,

(iii) forecasting based on SVR and GA, that is, the GA-SVR method.

5.1. Available Data. The wind speed data used in this paper are collected from a Shandong wind farm in China. The studied time range covers September 23rd to October 22nd, 2012. The data are sampled in 15 min time intervals, so there are 96 data points for one day. The total dataset includes 2880 wind speed samples; the first 2592 samples are used to build the models, and the remaining 288 samples will be used to validate the forecasting performance. Figure 3 shows the variation trend of the wind speed data. Shandong province had a total installed capacity of 6980.5 MW in 2013, ranking fourth in China. Given the considerable wind power potential around the coastal areas and islands as well as those wind farms located in electricity demand centers [38], Shandong wind farms have received more and more attention in China. Thus, accurate wind speed forecasting is crucial for planning and operating a power system for the city’s sustainable development.

5.2. Predictability of Wind Speed. Figure 2 shows severe fluctuations in the wind speed data, while no hallmark of periodicity is demonstrated. However, previous researchers have concluded that the fluctuation comes from the stochastic or chaotic nature of a nonlinear system [39, 40]. To study the predictability of wind speed, RP is performed on the original wind speed data. Based on the derived results in Section 5.3, a correlation analysis is performed to choose proper input sets.

The embedding dimension and the embedding delay of the wind speed series must be acquired first to reconstruct the phase space. The false nearest neighbor method and the mutual information method are used to calculate these two parameters. The embedding dimension and the embedding delay are 3 and 2, respectively. The RP of wind speed time series is shown in Figure 4 [41]. It can be concluded that the short-term erratic distribution of recurrence points is representative of the strong stochastic nature of the wind speed series with mimic predictability. White ribbons in the RP correspond to transitive processes in the system. The dynamic transition process is representative of both the seasonality and no stationary of the wind speed time series.

5.3. Selection of SVR Inputs. It is concluded above that the wind speed has mimic predictability. Therefore, the mimic
predictability of the wind speed series should be taken into account in selecting the inputs for the SVR model. Moreover, because the input set of the forecasting model is the lagged wind speed, it should be as close as possible to the forecasted time to compensate for the no stationary of the system. From Section 5.2, it is known that the embedding dimension \((m)\) and the embedding delay \((\tau)\) are 3 and 2, respectively. According to (1), the element of reconstructed phase space is four-dimensional. The input set is 4, and thus, the lagged wind speed series is the element in the reconstructed phase space:

\[
X_i = (x_i, x_{i+1}, \ldots, x_{i+4}), \quad i = 1, 2, \ldots, N - 4.
\]  

(22)

5.4. Selection of SVR Parameters. As mentioned previously, the performance of SVR modeling depends on its parameters. The optimization algorithms including COA, PSO, and GA are developed to tune crucial parameters of SVM via minimizing training errors and validation errors. In this paper, we choose RBF as the kernel function. The crucial parameters that need to be optimized are the penalties factor \((C)\), which determines the tradeoff cost between minimizing the training error and minimizing the SVM model complexity, and gamma of the kernel function RBF \((\gamma)\), which defines the nonlinear mapping from the input space to the high-dimensional feature space. The corresponding parameters of the optimized algorithms are shown in Table 1.

The process to search the SVR parameters is described as the following.

**Step 1.** Set the range of the penalties factor \((C)\) and gamma of the kernel function RBF \((\gamma)\). In this paper, \(C\) ranges from 0.1 to 100, and \(\gamma\) ranges from 0.1 to 10.

**Step 2.** Initialize the parameters of each optimization algorithm according to Table 1. Note that the dimension of chromosomes of GA, particles of PSO, and population of COA are all equal to 2.

**Step 3.** Perform the optimization processes according to Section 3.3.

Figures 5, 6, 7, and 8 show the fitness curves for one-step- to four-step-ahead predictions of all three algorithms. It can be seen that COA has a stationary process of finding
Table 1: Parameters of the employed optimization algorithms.

<table>
<thead>
<tr>
<th></th>
<th>COA</th>
<th>PSO</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum iteration</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Number of nests ($N_p$)</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Discovery probability ($P_d$)</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Best $C$ and $g$ values of multi-step-ahead predictions tuned by optimized algorithm.

<table>
<thead>
<tr>
<th>Forecasting horizons</th>
<th>COA</th>
<th>PSO</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-step-ahead</td>
<td>9.986</td>
<td>1.351</td>
<td>1.911</td>
</tr>
<tr>
<td>Two-step-ahead</td>
<td>7.629</td>
<td>0.100</td>
<td>0.232</td>
</tr>
<tr>
<td>Three-step-ahead</td>
<td>9.873</td>
<td>0.100</td>
<td>2.490</td>
</tr>
<tr>
<td>Four-step-ahead</td>
<td>8.660</td>
<td>0.084</td>
<td>1.081</td>
</tr>
</tbody>
</table>

Figure 8: Fitness curves of four-step-ahead predictions by the optimization algorithms.

the minimal fitness, while GA and PSO show fairly high fluctuations of the fitness curve. COA is more likely to reach the minimal fitness and to maintain that status, while GA detects the minimal fitness and quickly loses it. PSO fails to find the best fitness, and overfitting may be one reason for the failure [42]. The restricted parameters of these optimization algorithms form another possible reason. To quickly obtain the forecasting results in the very short-term wind speed prediction, the parameters of these optimization algorithms are restricted. For example, if the global learning coefficient ($c_2$) of PSO is small or large, the running time increases. The experiment illustrates the superiority of COA over both PSO and GA.

The best parameters (in Table 2) for multi-step-ahead predictions, penalties factor ($C$), and gamma of kernel function RBF ($g$) are selected out via the three optimized algorithms.

5.5. Comparison of the Forecasting Results. The SVR models with the optimal parameters are then used to forecast the wind speed data. A randomly chosen SVR model with $C$ of 2 and $g$ of 1, denoted as the SVR method, is run for comparison with the optimized SVR models based on five criteria including MAE, MAPE, RMSE, $R^2$, and IA, as shown in Table 3. The forecasting results of one-step-, two-step-, three-step-, and four-step-ahead forecasts are shown in Figures 9, 10, 11, and 12.

The following can be seen from Figures 9–12 and Table 3.

(1) The hybrid methods have satisfactory performances in the wind speed predictions because the forecasting curves are very close to the actual wind speed curve. The subplot in the north-west corner of each figure shows that the forecasting curve of COA-SVR is closer to the actual wind speed data, especially for the two- to four-step-ahead predictions.

Figure 9: One-step-ahead forecasting results.
(2) The forecasting approaches are effective at forecasting the nonstationary wind speed series, and all forecasting models display satisfactory performance. Taking the four-step-ahead prediction as an example, the MAE errors of the general SVR, GA-SVR, PSO-SVR, and COA-SVR are 1.2602, 1.2616, 1.2761, and 1.2386, respectively.

(3) The performance of the optimized SVR models is much better than that of the SVR model in which $C$ and $γ$ take values of 2 and 1, respectively. The COA-SVR model is the best among the optimized SVR models.

(4) When comparing the performance of the COA-SVR model with that of the SVR model, the optimized algorithm part has improved the performance of the general SVR model. The improved percentage of the RMSE from one-step to four-step is 1.98%, 1.33%, 0.80%, and 0.98%, respectively. The improved percentage of the MAPE for one-step, three-step, and four-step is 0.29%, 0.37%, and 0.63%, respectively. The improved percentage of $R^2$ from one-step to four-step is 0.2%, 0.24%, 0.30%, and 0.4%, respectively.

(5) When comparing the performance of the COA-SVR model with that of the GA-SVR model, the COA part has improved the performance more than the GA part. The improved percentage of the RMSE from one-step to four-step is 0.71%, 4.53%, 1.99%, and 1.25%, respectively. The improved percentage of the MAPE from one-step to four-step is 0.16%, 0.42%, 1.20%, and 0.79%, respectively. The improved percentage of $R^2$ from one-step to four-step is 0.08%, 1.07%, 0.7%, and 0.53%, respectively.

(6) When comparing the performance of the COA-SVR model with that of the PSO-SVR model, the COA part has improved the performance more than the PSO part. The improved percentage of the RMSE from one-step to four-step is 2.82%, 5.42%, 5.36%, and 2.48%, respectively. The improved percentage of the MAPE from one-step to four-step is 0.54%, 0.71%, 2.60%, and 1.12%, respectively. The improved percentage of $R^2$ from one-step to four-step is 0.31%, 1.17%, 1.94%, and 0.94%, respectively.

5.6. Hypothesis Test. The aim of using the hypothesis test is to answer whether the COA-SVR model is statistically robust. If the COA-SVR method is statistically robust, we
### Table 3: Performance evaluation for one-step to four-step-ahead predictions.

<table>
<thead>
<tr>
<th>Predictions</th>
<th>Methods</th>
<th>MAE (m/s)</th>
<th>RMSE (m/s)</th>
<th>MAPE (%)</th>
<th>$R^2$ (—)</th>
<th>IA (—)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-step-ahead prediction</strong></td>
<td>SVR</td>
<td>0.6911</td>
<td>0.9300</td>
<td>9.6650</td>
<td>0.9424</td>
<td>0.9846</td>
</tr>
<tr>
<td></td>
<td>GA-SVR</td>
<td>0.6860</td>
<td>0.9184</td>
<td>9.6524</td>
<td>0.9435</td>
<td>0.9851</td>
</tr>
<tr>
<td></td>
<td>PSO-SVR</td>
<td>0.6951</td>
<td>0.9376</td>
<td>9.6896</td>
<td>0.9414</td>
<td>0.9844</td>
</tr>
<tr>
<td></td>
<td>COA-SVR</td>
<td>0.6836</td>
<td>0.9119</td>
<td>9.6731</td>
<td>0.9443</td>
<td>0.9853</td>
</tr>
<tr>
<td><strong>Two-step-ahead prediction</strong></td>
<td>SVR</td>
<td>1.0101</td>
<td>1.3598</td>
<td>14.6636</td>
<td>0.8779</td>
<td>0.9657</td>
</tr>
<tr>
<td></td>
<td>GA-SVR</td>
<td>1.0418</td>
<td>1.4028</td>
<td>14.7376</td>
<td>0.8706</td>
<td>0.9630</td>
</tr>
<tr>
<td></td>
<td>PSO-SVR</td>
<td>1.0472</td>
<td>1.4148</td>
<td>14.7800</td>
<td>0.8697</td>
<td>0.9618</td>
</tr>
<tr>
<td></td>
<td>COA-SVR</td>
<td>1.0051</td>
<td>1.3420</td>
<td>14.6764</td>
<td>0.8800</td>
<td>0.9673</td>
</tr>
<tr>
<td><strong>Three-step-ahead prediction</strong></td>
<td>SVR</td>
<td>1.1608</td>
<td>1.5941</td>
<td>17.3654</td>
<td>0.8318</td>
<td>0.9521</td>
</tr>
<tr>
<td></td>
<td>GA-SVR</td>
<td>1.1814</td>
<td>1.6129</td>
<td>17.5094</td>
<td>0.8285</td>
<td>0.9502</td>
</tr>
<tr>
<td></td>
<td>PSO-SVR</td>
<td>1.2290</td>
<td>1.6662</td>
<td>17.7517</td>
<td>0.8181</td>
<td>0.9460</td>
</tr>
<tr>
<td></td>
<td>COA-SVR</td>
<td>1.1515</td>
<td>1.5815</td>
<td>17.3018</td>
<td>0.8343</td>
<td>0.9539</td>
</tr>
<tr>
<td><strong>Four-step-ahead prediction</strong></td>
<td>SVR</td>
<td>1.2602</td>
<td>1.7247</td>
<td>18.9360</td>
<td>0.8043</td>
<td>0.9423</td>
</tr>
<tr>
<td></td>
<td>GA-SVR</td>
<td>1.2616</td>
<td>1.7293</td>
<td>18.9662</td>
<td>0.8032</td>
<td>0.9421</td>
</tr>
<tr>
<td></td>
<td>PSO-SVR</td>
<td>1.2761</td>
<td>1.7502</td>
<td>19.0296</td>
<td>0.7999</td>
<td>0.9398</td>
</tr>
<tr>
<td></td>
<td>COA-SVR</td>
<td>1.2386</td>
<td>1.7079</td>
<td>18.8183</td>
<td>0.8075</td>
<td>0.9451</td>
</tr>
</tbody>
</table>

Table 4: Two-tailed hypothesis test results.

<table>
<thead>
<tr>
<th>Compare groups</th>
<th>Levene’s test for equality of variances</th>
<th>$t$-test for quality of means</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>t</td>
<td>Sig.</td>
</tr>
<tr>
<td>$\mu_0$ and $\mu_1$</td>
<td>Equal variances assumed</td>
<td>0.188</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
<td>0.188</td>
</tr>
<tr>
<td>$\mu_0$ and $\mu_2$</td>
<td>Equal variances assumed</td>
<td>0.483</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
<td>0.483</td>
</tr>
<tr>
<td>$\mu_0$ and $\mu_3$</td>
<td>Equal variances assumed</td>
<td>0.638</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
<td>0.638</td>
</tr>
<tr>
<td>$\mu_0$ and $\mu_4$</td>
<td>Equal variances assumed</td>
<td>1.092</td>
<td>0.297</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td></td>
<td>1.092</td>
</tr>
</tbody>
</table>

6. Discussion and Conclusions

Accurate short-term wind speed prediction is crucial for improving the stability of grid-connected wind power generation units and avoiding disadvantageous impacts on the electric networks. This paper proposed a hybrid model based on recurrence plot and optimized support vector regression model for wind speed prediction.

Wind is caused by the complex interaction of meteorological conditions, which makes wind a complex dynamic process and extremely difficult to forecast. Analyzing the predictability of wind speed is a primary step to build a proper model.

Recurrence plot (RP), which performs excellently in studying dynamic systems, is thus chosen to unravel the available wind speed. The results of RP reveal that wind speed data here has short-term predictability over the short-term timescale.

With the capability of modeling complex nonlinearity systems, support vector regression (SVR) models are employed to forecast the wind speed data at the Shandong wind farm in China. The forecast horizons are 15 min (one-step-ahead), 30 min (two-step-ahead), 45 min (three-step-ahead), and 1 h ahead (four-step-ahead). The forecasting accuracy of SVR
greatly depends on its parameters \((C, g)\) and its input signals. Therefore, proper selection of inputs and parameters is essential for satisfactory SVR performance. The elements of the reconstructed phase space of the RP with the biggest predictability are naturally chosen as the input sets. Once the input sets are fixed, the optimum parameters of SVR are selected by some optimized algorithms, including genetic algorithm (GA), particle swarm optimization (PSO), and cuckoo optimization algorithm (COA).

Finally, the optimized approaches are then evaluated according to some chosen performance criteria. The forecasting results show that the available wind data has short-term predictability over the short-term time scale; the performance of optimized SVR models is better than that of the SVR model without parameter selection. COA-SVR is superior among the optimized SVR models, especially in dealing with the jumping samples. Furthermore, a two-tailed hypothesis is performed for checking the robustness of the COA-SVR model. With 95% confidence, the COA-SVR method is statistically robust in multi-step-ahead predictions.

**Highlights**

(i) Recurrence plot (RP) is a new and powerful tool to explore wind system.

(ii) Short-term predictability of wind speed on a short-term time scale is proved by RP.

(iii) The optimized SVR techniques are highly satisfactory.

(iv) The COA-SVR model is statistically robust, especially in addressing jumping samples.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgment**

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

**References**


