Research Article

The Green Photovoltaic Industry Installed Capacity Forecast in China: Based on Grey Relation Analysis, Improved Signal Decomposition Method, and Artificial Bee Colony Algorithm

Hang Liu,1,2 Ming Zeng,1,2 Ting Pan,1,2 Weicheng Chen,1,2 Xiaochun Zhang,1,2 Xianxu Huo,3 and Zhigang Zhang3

1School of Economics and Management, North China Electric Power University, Beijing 102206, China
2State Key Laboratory of Alternate Electrical Power System with Renewable Energy, North China Electric Power University, Beijing 102206, China
3State Grid Tianjin Electric Power Company, Tianjin 300010, China

Correspondence should be addressed to Hang Liu; liuh2018@ncepu.edu.cn and Ming Zeng; zengmingbj@vip.sina.com

Received 19 October 2019; Accepted 19 December 2019; Published 11 February 2020

Academic Editor: Georgios Dounias

Copyright © 2020 Hang Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite photovoltaics being a new type of green energy technology, the output of the photovoltaic industry has been declining year by year since 2018. Thus, China’s photovoltaic industry must adapt to transformations from original extensive growth to the pursuit of high-quality energy. In order to accurately predict the installed capacity of photovoltaics in China, based on an extensive literature review and expert consultation, this paper is the first to construct a set of influencing factors that affect the photovoltaic industry, and we selected the main influencing factors as the predictive model’s input through the grey correlation analysis method. Then, we provide a novel grid investment forecast named the CEEMD-ABC-LSSVM predictive model (a least squares support vector machine algorithm based on complete total empirical mode decomposition and an artificial bee colony algorithm). This algorithm is based on the traditional LSSVM algorithm. The ABC algorithm is used to optimize the parameters, while CEEMD decomposes the original time series to obtain multiple components. While maintaining the data information, the data are expanded and the training is fully carried out. Next, in the empirical analysis, by comparing the prediction results of LSSVM, ABC-LSSVM, and the EMD-ABC-LSSVM algorithm, we demonstrate that the CEEMD-ABC-LSSVM model has strong generalization capabilities and achieves good Chinese PV growth based on the predicted effects of the installed capacity. Finally, the CEEMD-ABC-LSSVM model was used to predict the installed PV capacity in China from 2019 to 2035. We find that China’s installed PV capacity will surpass 4000 GW around 2035. As this installed capacity will increase year by year, China’s PV industry development will maintain steady overall growth.

1. Introduction

In recent years, the development of new energy industries has gradually become a focus of countries all over the world. The rapid development of the economy has become more and more dependent on energy demands, not only in the pursuit of high efficiency, low cost, and stabile energy but also to improve the cleanliness and environmental friendliness of energy. Among many clean energy sources, solar energy is widely distributed with large energy values and has high development potential, representing new growth potential for the world’s economy. According to the National Bureau of Statistics, in 2018, China’s annual power generation reached 6.8 trillion kWh, of which thermal power generation accounted for 75.08%, while photovoltaic power generation was only 0.0539 trillion kWh, accounting for only 0.79%. China’s photovoltaic power generation still has large room for improvement. In the “13th Five-Year Plan” for China’s national power development, by 2020, the installed capacity of photovoltaic power generation is planned to reach 110 million kilowatts [1]. In 2017, the newly installed capacity of photovoltaic power generation in China was 53
million kilowatts, a year-to-year increase of 53.62% [2]. After continuous construction throughout the year, by the end of 2017, the total installed capacity of photovoltaic power generation in China reached 130 million kilowatts, and the “13th Five-Year Plan” was exceeded in advance.

From the perspective of PV installation, China has become the world’s largest PV application market. Except for the occasional impact of the international market, the global PV growth rate has slowed. The remainder of the annual increase rate is above 100%, which indicates that the development center of the photovoltaic market is moving from Europe to emerging markets like China, Japan, and the United States. In 2018, China’s photovoltaic power generation had a new installed capacity of 10.6 million kilowatts, accounting for about 20% of the new installed capacity of the global photovoltaic market, and continues to maintain a good development trend [3]. However, China’s PV industry is greatly affected by the import and export trade situation, and the domestic PV market is still immature. Therefore, it is reasonable to develop China’s PV capacity and installed PV capacity, especially for the spread of centralized and distributed PV installed capacity [4]. Installation that is too rapid will lead to an increase in the phenomenon of “abandonment of light” and a serious overcapacity; a slow installation speed and insufficient installed capacity will lead China to miss the growth forecast of its economic development. Therefore, a high-precision PV installed capacity forecasting model is proposed. Meanwhile, the innovations of this paper are as follows:

1. In order to accurately predict the cumulative installed capacity of photovoltaic power in China, this paper proposes a new support vector forecasting model based on improved empirical mode decomposition and an artificial bee colony algorithm to predict the cumulative installed capacity of photovoltaics in China.

2. Since the initial parameters of the traditional LSSVM model include the regularization parameter $c$ and the width $g$ of the radial basis function, the ABC (artificial bee colony) algorithm can solve the chattering and premature problems of LSSVM parameter optimization and determine the optimal weight and threshold.

3. The traditional LSSVM algorithm cannot fully determine the time–frequency characteristics of the time series data and thus greatly affects the performance of the forecasting model. Then, the accuracy of the LSSVM cannot be ensured. The raw data are separated by CEEMD (complete ensemble empirical mode decomposition), which expands the magnitude of the data to achieve the data requirements of the machine algorithm. Based on the above reasons, this paper applies two different algorithms, CEEMD and ABC, into the LSSVM algorithm.

The main contents of this article are as follows: Section 2 introduces recent related literature; Section 3 introduces the mathematical principle of the signal decomposition model, including the artificial bee colony algorithm and the LSSVM algorithm, as well as the flow of the overall forecasting algorithm; Section 4 concludes the flow chart of China’s cumulative PV installed capacity; Section 5 applies the proposed CEEMD-ABC-LSSVM forecasting model to the prediction of cumulative PV installed capacity in China. By comparing the forecasting results of LSSVM, ABC-LSSVM, and EMD-ABC-LSSVM, the experimental results prove the practicality and accuracy of the proposed model in predicting China’s cumulative PV installed capacity; Section 6 presents several forward-looking conclusions.

2. Literature Review

In recent years, many scholars have studied the photovoltaic industry’s development system, which is mainly composed of two major parts—the industrial main body and the industrial environment [5]. Among them, the main body of the industry refers to the main production factors within the photovoltaic industry. The industrial environment mainly includes factors such as the level of social and economic development, the relevant policy system of the photovoltaic industry, market demand, and industrial capital investment, which affect the development of the industry. According to the system framework, the photovoltaic industry’s development system is divided into two subsystems: the photovoltaic industry subsystem and the industrial development environment subsystem. Xiao et al. [6] argued that in the photovoltaic (PV) technology value chain, improving technology is an important factor in reducing PV costs. Investigating the policies of Germany and China proves that the deployment support and investment equations of PV part manufacturing plants affect the PV industry’s important factors. In [7], Adaramola studied the technical and economic performance of an 80kW solar photovoltaic grid-connected power generation system. In the case of solar photovoltaic power generation costs and global average solar radiation, the proportion of solar photovoltaic power generation and its average energy costs were obtained. Finally, it is economically feasible to develop solar photovoltaic systems for grid-connected power generation in northeastern Nigeria. Perez et al. [8] proposed a model predictive control (MPC) method to manage the energy generated by grid-connected photovoltaic (PV) power plants with energy storage in real time to optimize their economic benefits. Jafarzadeh et al. [9] believes that the randomness of solar energy resources is one of the factors affecting their large-scale diffusion in the power system. The system for solar power plant modeling uses fuzzy logic and the predicted fuzzy system uses type-1 and interval-2 Takagi-Sugeno–Kang (TSK) fuzzy systems. The results of this example show that the type-2 TSK model provides the best performance index based on the type-2 prefactor and clear results.

PV installed capacity prediction is a complex nonlinear solution problem. So far, many scholars have proposed various prediction models, such as grey theory [10, 11],...
multiple regression [12, 13], and time series [14, 15]. In recent years, various intelligent algorithms and signal decomposition models have also been widely used in installed capacity prediction [16–25]. Yuan et al. [10] defined the approximate real starting point for a steady-state growth trend sequence with random perturbations and calculated the generalized grey relation analysis (GRA) on this basis. The main reasons for the expansion of the photovoltaic industry were analyzed. It was found that the global installed capacity of photovoltaics is the main factor affecting the development of the photovoltaic industry. Shi et al. [11] proposed a periodic relationship model using the horizontal and vertical coordinates of the curve for the traditional grey correlation model. Ultimately, this model only refers to the wave period but is independent of it. The swing can show positive and negative correlations. Wolff et al. [12] evaluated multiple numerical weather prediction (NWP) parameters to improve their ability to improve photovoltaic power generation prediction capabilities. The importance of functionality is determined by a random forest algorithm. Furthermore, by using support vector regression, the random forest and linear regression models performed PV power predictions to test the resulting highest ranked features. Raghoebarsing and Reinders [13] outlined the current state of Suriname’s energy sector and photovoltaic (PV) systems and examined the role that PV systems can play in the country’s future energy transition. According to the government and nongovernment agencies that have planned photovoltaic projects, it is predicted that future PV systems will generate 0.8 TWh of electricity to meet the power demands of 2027. Sun et al. [14] proposed a new predictive model for multistep wind speed prediction (WSF). A two-stage signal decomposition method combining set empirical mode decomposition (EEMD) and variational mode decomposition (VMD) was used to decompose empirical wind speed data. Further, the parameters in the proposed wavelet neural network (WNN) model were optimized to improve the prediction performance, which was realized by a hybrid backtracking search, and its effectiveness was illustrated. Pavlyshenko [15] studied the use of machine learning models in sales forecast analysis and proved that when there is a small amount of historical data in a specific sales time series, this effect can be used to make sales forecasts. The results prove that machine learning can improve sales time. This shows the performance of predictive models of sequence prediction.

Signal decomposition and machine intelligence algorithms have been widely used in the field of hybrid prediction. Hu et al. [16] proposed a hybrid prediction method that includes EWT (empirical wavelet transform), CSA (coupled simulated annealing), and LSSVM (least squares support vector machine) to improve the accuracy of short-term wind speed predictions. The proposed hybrid model has been shown to predict the average half-hour wind speed series obtained from a windmill farm in northwestern China. Wang et al. [17] proposed a variational mode decomposition (VMD) based on phase space factoring (PSR) and the genetic algorithm. The results show that Wang et al.’s model is superior to all other comparison models, which proves that this model has superior multistep performance. Li et al. [18] developed four time series prediction techniques, including the metabolic grey model (MGD), the autoregressive integrated moving average (ARIMA), the grey model (GM)-ARIMA, and the nonlinear metabolic grey model (NMGM). A more accurate prediction of installed capacity can help to develop a better energy sustainability strategy. Tsaur and Lin [19] defined a solar building as a building-affiliated photovoltaic (BAPV) system and proposed a technology acceptance model (TAM) to predict and explain the public acceptance of BAPV. Wu et al. [20] considered the output radiation, cloud cover, and other weather conditions as the main influencing factors affecting PV production, considering the short-term predictions of operational plans, exchange sources, and peak load matching. Five types of prediction modules have been developed, including ARIMA, SVM, ANN, ANFIS, and a combined model using genetic algorithms. The prediction results show that the combined model has higher precision and efficiency.

LSSVM is a practical machine learning method that has been widely used in many fields. In order to ensure performance, it is crucial to choose the right technology to obtain the optimized hyper parameters of the LSSVM algorithm. Wu and Peng [21] proposed a bat algorithm (BA) optimized least squares support vector machine (LSSVM) hybrid model to improve the prediction of wind energy production. In order to effectively select the input of the LSSVM, the parameters in the LSSVM are optimized by the BA to verify the learning ability and generalization ability of the LSSVM. The results of the example show that the accuracy of the prediction model proposed in the paper can be improved by about 20%. Xiang et al. [22] proposed a decomposition-based artificial bee colony algorithm (ABC) to deal with multiobjective optimization problems (MAOPs). Their example shows that this new algorithm has good convergence and diversity. The quality of its solution is high, and its speed is fast. Zhao et al. [23] used the least squares support vector machine (LSSVM) method to establish the relationship between the reliability index and design variables and used the artificial bee colony (ABC) algorithm for reliability optimization. The numerical results show that this method not only satisfies the design accuracy but also improves optimization of the reliability parameters, such as the mechanical parameters, ground stress, and internal pressure of the rock mass. In [24], Xiao et al. proposed a fault diagnosis method based on heterogeneous information fusion in order to accurately determine fault type based on obtained information. The fusion feature was then introduced into the artificial bee colony algorithm- (ABC-) optimized least squares support vector machine (LSSVM). Simulation experiments show that the proposed method has higher diagnostic accuracy. Mustaffa and Yusof [25] used an enhanced artificial bee colony (EABC) to obtain the ideal values for the hyper parameters of the LSSVM, namely, the regularization parameters, the gamma parameters, and the kernel parameters, $\sigma^2$. Then, the LSSVM was used as the forecasting model. This model was applied to the prediction of financial time series data, and the simulation results ensure the accuracy of parameter selection.
3. Principles of Forecasting Model

3.1. LSSVM. Given a set of training data samples \((x_i, y_i)\)\(^N\), among which \(x_i \in \mathbb{R}^m\) is the m-dimensional data samples and \(y_i \in \mathbb{R}\) is a sample output, the LSSVM optimization algorithm is as follows [26]:

\[
\min \quad J = \frac{1}{2} w^T w + \frac{1}{2} r \sum_{i=1}^{N} e_i^2,
\]

s.t. \(y_j = w^T \phi(x_j) + b + e_j, \quad i = 1, 2, \ldots, n,\)

where \(\phi(x_j): \mathbb{R}^m \rightarrow \mathbb{R}^m\) is the mapping function from the original space to a high dimensional space, \(w \in \mathbb{R}^m\) is the weight vector, \(e \in \mathbb{R}\) is the tolerance, \(b\) is the offset, and \(y\) is the normalized coefficient.

According to the objective function and constraints, we establish the Lagrange function as

\[
L = \frac{1}{2} w^T w + \frac{1}{2} r \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} a_i \{w^T \phi(x_i) + b + e_i - y_i\},
\]

where \(a_i\) is the Lagrange factor.

For the KKT conditions, the \(L\) derivative can be obtained as

\[
\begin{align*}
\frac{\partial L}{\partial w} &= 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N} a_i \phi(x_i), \\
\frac{\partial L}{\partial b} &= 0 \quad \Rightarrow \quad \sum_{i=1}^{N} a_i = 0, \\
\frac{\partial L}{\partial e_i} &= 0 \quad \Rightarrow \quad a_i = \gamma e_i, \quad i = 1, 2, \ldots, n, \\
\frac{\partial L}{\partial a_i} &= 0 \quad \Rightarrow \quad w^T \phi(x_i) + b + e_i - y_i = 0, \quad i = 1, 2, \ldots, n.
\end{align*}
\]

After cancelling \(w\) and \(e\), we obtain the following matrix equation:

\[
\begin{bmatrix}
0 & 1^T \\
1 & \Omega + I
\end{bmatrix}
\begin{bmatrix}
b \\
y
\end{bmatrix} = \begin{bmatrix}
0 \\
y
\end{bmatrix},
\]

where \(I\) is the identity matrix, \(1_x = [1, \ldots, 1]\), \(a = [a_1, a_2, \ldots, a_N]^T\), and \(\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)\).

The optimal decision function is

\[
y(x) = \sum_{i=1}^{N} a_i K(x_i, x_j) + b.
\]

3.2. Artificial Bee Colony Algorithm. A swarm of bees can adapt to most living environments on the planet. We divided the bee colony into different types of work (such as collecting honey, breeding, and defense) to ensure a division of labor and cooperation. At the same time, there is a special “information transmission mechanism” between bees [27, 28]. The bees communicate information to inform other bees around them and distribute work. The ABC algorithm is a swarm intelligence optimization algorithm designed by simulating the process of bees constantly searching for excellent honey sources.

The ABC algorithm is a highly efficient and simple intelligent algorithm that is suitable for many function optimization problems. The honey source position in the algorithm represents a feasible solution for the target optimization function. The optimal honey source represents the optimal solution of the function optimization problem. The corresponding relationship between honey bee behavior and the optimization function is shown in Table 1.

In the ABC algorithm, the honey bee collecting process is divided into four stages: the initialization stage, the hired bee stage, observing the bee, and the detection bee stage [29].

(1) Initialization phase: the initial honey source location is randomly generated within the feasible solution domain. These locations represent the initial source of honey (a feasible solution) and calculate their fitness values.

(2) Hired bee stage: hired bees search the neighborhood at the current honey source location and record better honey source information. The recorded honey source information is then transmitted to the observation bee through the “swing dance.” When the search for the honey source is completed and a better source of honey cannot be found, the hired bee gives up the honey source and turns into a detecting bee.

(3) Observing the bee: this stage entails observing the bee waiting in the “swinging dance area” and receiving the honey source information transmitted by the employed bee. After receiving the information, the observation bee decides whether to choose to follow the hired bee to the honey source. If it chooses to follow, the observation bee turns into a hired bee.

(4) Detection bee stage: if the employer is unable to find a better source after multiple searches, the hired bee will abandon the current honey source and convert into a detection bee. The scouting bee then begins to randomly search for a better source of honey. When a better source of honey is found, the scout is converted back to a hired bee.

During the entire search process, the bees constantly change between these three different roles until the end of the algorithm. Thus, during the operation of the algorithm, the three bees constantly change roles, and the possibility of the algorithm finding the global optimal solution is greatly improved. At the same time, the speed of the calculation method is also improved. The bees adjust their respective divisions of labor through the transformation of their roles and complete the transformation of their roles through the sharing of information. The transformation diagram is shown in Figure 1.
Table 1: Correspondence between honey bee behavior and the optimization function.

<table>
<thead>
<tr>
<th>Bee colony</th>
<th>Optimization function problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honey source location</td>
<td>Possible solution to the optimization problem</td>
</tr>
<tr>
<td>Bee collecting process</td>
<td>Feasible solution fitness value</td>
</tr>
<tr>
<td>Optimal honey source</td>
<td>Function optimization speed</td>
</tr>
</tbody>
</table>

Figure 1: Biological principle of the artificial bee colony algorithm.

The ABC algorithm is a very simple and easy-to-calculate intelligent algorithm. The specific steps for its implementation are as follows [23].

1. Initialize the population: when the algorithm starts running, the number of bee colonies is np and the number of hired bees is SN. Randomly generate NP honey sources, which are usually set as SN = NP/2. The feasible domain dimension of the algorithm is D. The maximum number of iterations of the algorithm is maxcycle, and the maximum number of stagnations is limited. Within the feasible domain of the objective function, we randomly generate the hired bees’ SN honey source locations according to formula (6), which is recorded as \( X = \{ X^i_j | i = 1, 2, \ldots, SN; \ j = 1, 2, \ldots, D \} \), and calculate the fitness of the honey source fitness according to formula (7):

\[
\text{new}_i^j = X^i_j + \text{rand}(0,1)(X^i_{\max} - X^i_{\min}), \quad (6)
\]

\[
\text{fitness} = \begin{cases} 
1 & \frac{1}{1 + \text{fitness}}, \\
1 + \text{abs}(\text{fitness}), 
\end{cases} \quad (7)
\]

where \( \text{rand}(0,1) \) refers to a random number representing a uniform distribution between \( (0, 1) \). D is the dimension of the optimized space, the value of \( k \) is selected from \( D[1, 2, \ldots, SN] \), and \( k \neq i, X^j_i \neq X^j_k \). maxcycle is used to specify the maximum number of iterations of the algorithm. limit is used to specify the maximum number of times a hired bee stays in a single source of honey.

2. Hired bee search phase: this stage involves hiring a bee to conduct a honey source search to find the candidate a new honey source formula according to formula (8). The new honey source location will update according to formula (9):

\[
V_{ij} = X_{ij} + q_{ij}(X_{ij} - X_{kj}), \quad (8)
\]

\[
V_{ij} = \begin{cases} 
V_{ij}, & \text{fitness}(V_{ij}) > \text{fitness}(X_{ij}), \\
X_{ij}, & \text{fitness}(V_{ij}) \leq \text{fitness}(X_{ij}), 
\end{cases} \quad (9)
\]

where \( q_{ij} \) is a random number between \([-1, 1]\), \( k \in (1, 2, \ldots, SN) \) and \( X_{ij} \neq X_{kj} \). The old and new honey sources are selected by the greedy rule of formula (9); that is, the fitness values of the new and old honey sources is compared according to their fitness, and then a larger value of honey is taken. If the fitness value fitness of the new honey source \( V_{ij} \) is better than the old honey source \( X_{ij} \), we use the new honey source \( V_{ij} \) to replace the location of the old honey source, \( X_{ij} \). Otherwise, we save honey source location \( X_{ij} \), which is still the current honey source location. At the same time, the number of stagnating bees in the honey source is increased by one.

3. Observation bee search phase: at this stage, the observation bee selects the honey source according to the roulette method based on the honey source information obtained by the hired bee to perform the honey source search. If we choose to follow the hired bee, we search for the new honey source according to the method of step 2. The probability of the observation phase is as follows:

\[
P_i = \frac{\text{fitness}(X_{ij})}{\sum_{k=1}^{SN} \text{fitness}(X_{kj})}, \quad (10)
\]

4. Detection bee search stage: each time the number of the stagnation of the honey source changes, a judgment is made. If this number reaches the threshold limit and the bee still does not find a better source, the bee gives up the honey source of the current search. The bee then turns into a scouting bee and begins a random search until a new, better source of honey is found.

5. Optimal solution judgment: in this stage, we judge whether the stop criterion is reached and then output the optimal honey source position, fitness value, and other parameters; otherwise, we continue to run step 2.

3.3. LSSVM Optimized by ABC. Since the initial parameters of the traditional LSSVM model include the regularization parameter \( c \) and the width \( g \) of the radial basis function, the ABC algorithm can solve the chattering and premature problems of the LSSVM parameter optimization and determine the optimal weight and threshold. In the ABC algorithm, the fitness
function in the ABC algorithm is calculated; the initial population size and the maximum evolution algebra maxgen are set, and the genetic operations, such as selection, improved crossover, and mutation, are performed on the individuals in the population. Finally, for the global excellent fitness, we use optimal fitness to obtain the regularization parameter \( c \) and the width \( g \) of the radial basis function. The flow chart of LSSVM optimized by ABC is shown in Figure 2.

3.4. Basic Principle of CEEMD. In 1998, Huang et al. [27] proposed Empirical Mode Decomposition (EMD). This analysis method can decompose any complex signal according to different time-frequency scales and then generate a series of intrinsic mode function components (IMFs) from different frequencies (from a high frequency to a low frequency).

Huang et al. proposed three hypotheses:

1. Any complex signal consists of several \( M \) components
2. The number of extreme points of each IMF component is consistent with the number of zeros, and the envelope signal diagram is symmetrical around the time axis
3. Any signal can be decomposed into several \( M \) components at any time, and IMF full component superposition reconstruction can completely recover the signal.

In 2010, Yeh et al. [29] made some improvements to the EMD algorithm. By adding positive and negative paired auxiliary noise, the residual auxiliary noise in the reconstructed signal can be eliminated and the computational efficiency can also be improved. This method is called complete ensemble empirical mode decomposition (CEEMD).

The steps of the CEEMD algorithm are as follows:

1. Firstly, fixed-intensity white noise is added into the original signal; then, the noise-added signal is subjected to EMD decomposition to obtain an IMF component, and \( n \) different kinds of white noise are continuously decomposed for corresponding times. Then, the obtained components are averaged:

   \[
   \text{IMF}_1 = \frac{1}{n} \sum_{i=1}^{n} e_i [x + \epsilon \omega_i],
   \]

   where \( \text{IMF}_1 \) is the first-order of IMF component, \( n \) is the number of times the noise is added, \( e_i \) indicates the \( i \)-th component of the EMD decomposition, \( \epsilon \) indicates the proportion of noise added, and \( \omega_i \) is the white noise added.

2. Next, we calculate the residual by subtracting the first-order IMF component:

   \[
   r_i = x - \text{IMF}_1.
   \]

3. Then, we decompose \( r_i + \epsilon e_i \omega_i \), \( i = 1, 2, \ldots, n \) and continue to iteratively decompose until the component needed to satisfy \( \text{IMF}_1 \) emerges. Then, we obtain the overall average of the whole \( \text{IMF}_1 \). In this way, the second-order of the original signal \( \text{IMF}_2 \) is obtained, which is

\[
\text{IMF}_2 = \frac{1}{n} \sum_{i=1}^{n} e_i [r_i + \epsilon e_i \omega_i].
\]

4. Calculate the residual of the \( k \)-th order:

\[
r_k = r_{k-1} - \text{IMF} \text{ and then extract the } r_k + \epsilon e_k [\omega_i] \text{ first-order IMF}
\]

\[
\text{IMF}_{k+1} = \frac{1}{n} \sum_{i=1}^{n} e_i [r_i + \epsilon e_i \omega_i].
\]

Figure 2: The flow chart of LSSVM optimized by the ABC algorithm.
component = 1, 2, ..., N. Then, take their overall average to obtain the \( k + 1 \)-th-order component of the original signal \( \text{IMF}_{k+1} \), which is

\[
\text{IMFF}_{k+1} = \frac{1}{n} \sum_{i=1}^{n} e_i [r_k + \varepsilon e_k [\omega_j]].
\] (14)

(5) Continue iterative screening and stop when the number of extreme values of the residual does not exceed two:

\[
R = x - \sum_{k=1}^{K} \text{IMF}_k,
\] (15)

Figure 3: Flow chart of China’s PV cumulative installed capacity based on the CEEMD-ABC-LSSVM forecasting model.
where \( r \) is the residual value and \( k \) is the number of IMF components.

(6) The original signal recovered from each IMF component and the residual sequence can be expressed as

\[
x = R + \sum_{k=1}^{K} \text{IMF}_k.
\]

Compared with EMD, CEEMD not only solves the modal aliasing problem perfectly but also has higher computational efficiency and complete signal decomposition. Therefore, the CEEMD method is more suitable for filtering denoising.

After the signal is decomposed by CEEMD, several IMF components with high to low frequency are obtained from small to large, including \( \text{IMF}_1, \text{IMF}_2, \ldots, \text{IMF}_k, \text{IMF}_{k+1} \). The high-frequency IMF component with a small order generally contains more noise, and the noise of the low-frequency IMF component with a large order has little effect. There must be a component in the middle of the IMF component \( \text{IMF}_k \) as the demarcation point. The components between \( \text{IMF}_{k+1} \) to \( \text{IMF}_k \) refer to the signal dominant zone, and the other is the dominant zone of noise. In 2007, Boundraa and Cexus [30] proposed a method based on continuous mean square error to determine the demarcation point for \( \text{IMF}_k \). The discriminant formula of this method is as follows:

\[
e_k = \frac{1}{n} \sum_{i=1}^{n} (y_k(t_i) - y_{k+1}(t_i))^2 = \frac{1}{n} \sum_{i=1}^{n} \text{IMF}_k(t_i)^2,
\]

where \( y_k(t_i) \) is the selected signal reconstructed from the \( k \)-th layer to the residual, \( y_{k+1}(t_i) \) is the selected signal reconstructed from the \( k + 1 \)-th layer to the residual, \( e_k \) is the energy, and \( n \) is the signal length.

4. The Forecasting Model of CEEMD-ABC-LSSVM

Due to the late development of China’s PV market, the historical data of the installed capacity are insufficient, and it is difficult to realize the data requirements of intelligent algorithms. Therefore, it is very easy to create homogeneity among the influencing factors of the original data as well as different scale information characteristics for the input data. The traditional LSSVM algorithm cannot fully determine the time-frequency characteristics of the time series data, which greatly affects the performance of the forecasting model. Thus, the accuracy of the LSSVM cannot be ensured. The raw data are separated by CEEMD, which expands the magnitude of the data to achieve the data requirements of the machine algorithm. Based on the above reasons, this paper applies two different algorithms (CEEMD and ABC) into the LSSVM algorithm and proposes a Chinese PV installed capacity forecasting model called the CEEMD-ABC-LSSVM combined forecasting model.

The forecasting steps of the overall algorithm are as follows:

(1) Time series data decomposition. Use the CEEMD algorithm to decompose \( x(t) \) and then obtain \( n \) IMF components and one residual \( r_n \).

(2) Construct training and test sample sets. In the IMF component, the input and output of each component training sample set and test sample set are constructed.

(3) Construct ABC-LSSVM training and forecasting models for each component. In the ABC algorithm, the fitness function in the ABC algorithm is calculated; the initial population size \( NP \) and the maximum evolution algebra \( maxcycle \) are thus set.

(4) Perform steps 1–5 of the ABC algorithm to determine the global optimal fitness. Use the optimal fitness to obtain the optimal honey source position (i.e., the optimal solution of the regularized parameter \( c \) and the width \( g \) of the radial basis function).

(5) Assign the optimized parameters to the least squares support vector machine to construct the forecasting model.

(6) Output the forecasting results for each IMF component and residual \( r_n \). The predicted results are then summed to obtain the final forecast results of the cumulative installed capacity of photovoltaics in China. The forecasting flow chart is shown in Figure 3.

5. Empirical Analysis

5.1. Data Processing. Based on [1–3, 5–9], this paper selects the industry prosperity index (the annual average of the electric power production and supply industry), the cost of power generation (the average value of the photovoltaic on-grid electricity price for centralized and distributed general projects), and the domestic PV Output value (100 million), PV module production, foreign market demand, domestic market demand, international PV trade policy easing, PV tariff, GDP, population, total social electricity consumption,
Table 3: Index normalized values of the main factors affecting the PV installed capacity in China.

<table>
<thead>
<tr>
<th>Year</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>2.0579</td>
<td>0.9343</td>
<td>0.8701</td>
<td>0.6439</td>
<td>1.2586</td>
<td>0.9906</td>
<td>0.4649</td>
<td>1.4387</td>
<td>0.4439</td>
<td>0.6331</td>
</tr>
<tr>
<td>1991</td>
<td>1.9200</td>
<td>0.8502</td>
<td>0.7381</td>
<td>0.6395</td>
<td>1.2586</td>
<td>0.9778</td>
<td>0.3971</td>
<td>1.2915</td>
<td>0.4437</td>
<td>0.6327</td>
</tr>
<tr>
<td>1992</td>
<td>1.6221</td>
<td>0.8225</td>
<td>0.6863</td>
<td>0.6351</td>
<td>1.2586</td>
<td>0.9565</td>
<td>0.3358</td>
<td>1.4085</td>
<td>0.4429</td>
<td>0.6315</td>
</tr>
<tr>
<td>1993</td>
<td>1.2800</td>
<td>0.8086</td>
<td>0.5929</td>
<td>0.6308</td>
<td>1.2586</td>
<td>0.9217</td>
<td>0.2746</td>
<td>1.4405</td>
<td>0.4344</td>
<td>0.6308</td>
</tr>
<tr>
<td>1994</td>
<td>1.2304</td>
<td>0.7143</td>
<td>0.5562</td>
<td>0.6264</td>
<td>1.2586</td>
<td>0.8685</td>
<td>0.2139</td>
<td>1.4631</td>
<td>0.4320</td>
<td>0.6292</td>
</tr>
<tr>
<td>1995</td>
<td>1.1476</td>
<td>0.6727</td>
<td>0.4911</td>
<td>0.6220</td>
<td>1.2586</td>
<td>0.7735</td>
<td>0.1974</td>
<td>1.4674</td>
<td>0.4273</td>
<td>0.6277</td>
</tr>
<tr>
<td>1996</td>
<td>1.0593</td>
<td>0.5728</td>
<td>0.4713</td>
<td>0.6176</td>
<td>1.2586</td>
<td>0.7411</td>
<td>0.1561</td>
<td>1.1266</td>
<td>0.4186</td>
<td>0.6246</td>
</tr>
<tr>
<td>1997</td>
<td>1.0207</td>
<td>0.5589</td>
<td>0.4553</td>
<td>0.6132</td>
<td>0.9842</td>
<td>0.7411</td>
<td>0.0422</td>
<td>1.1459</td>
<td>0.4121</td>
<td>0.6258</td>
</tr>
<tr>
<td>1998</td>
<td>0.9545</td>
<td>0.3620</td>
<td>0.4515</td>
<td>0.6088</td>
<td>0.4353</td>
<td>0.7186</td>
<td>0.0094</td>
<td>1.1266</td>
<td>0.4186</td>
<td>0.6246</td>
</tr>
<tr>
<td>1999</td>
<td>0.5628</td>
<td>0.3121</td>
<td>0.4440</td>
<td>0.6044</td>
<td>0.1609</td>
<td>0.6966</td>
<td>0.0560</td>
<td>1.1207</td>
<td>0.3903</td>
<td>0.6231</td>
</tr>
<tr>
<td>2000</td>
<td>0.4635</td>
<td>0.2788</td>
<td>0.4308</td>
<td>0.6000</td>
<td>0.1136</td>
<td>0.6567</td>
<td>0.0996</td>
<td>1.1042</td>
<td>0.3851</td>
<td>0.6227</td>
</tr>
<tr>
<td>2001</td>
<td>0.1489</td>
<td>0.2705</td>
<td>0.4223</td>
<td>0.5956</td>
<td>0.1136</td>
<td>0.6133</td>
<td>0.1398</td>
<td>1.0023</td>
<td>0.3774</td>
<td>0.6188</td>
</tr>
<tr>
<td>2002</td>
<td>0.7608</td>
<td>0.2510</td>
<td>0.4110</td>
<td>0.5604</td>
<td>0.1136</td>
<td>0.5688</td>
<td>0.1773</td>
<td>0.9419</td>
<td>0.3697</td>
<td>0.6111</td>
</tr>
<tr>
<td>2003</td>
<td>0.7150</td>
<td>0.2483</td>
<td>0.3987</td>
<td>0.5165</td>
<td>0.1609</td>
<td>0.5043</td>
<td>0.2125</td>
<td>0.8493</td>
<td>0.3620</td>
<td>0.5996</td>
</tr>
<tr>
<td>2004</td>
<td>0.7023</td>
<td>0.1817</td>
<td>0.3874</td>
<td>0.4725</td>
<td>0.4353</td>
<td>0.4042</td>
<td>0.2471</td>
<td>0.6836</td>
<td>0.3542</td>
<td>0.5957</td>
</tr>
<tr>
<td>2005</td>
<td>0.6775</td>
<td>0.1623</td>
<td>0.3573</td>
<td>0.4286</td>
<td>0.7098</td>
<td>0.2997</td>
<td>0.2821</td>
<td>0.5745</td>
<td>0.3388</td>
<td>0.5803</td>
</tr>
<tr>
<td>2006</td>
<td>0.7652</td>
<td>0.1207</td>
<td>0.3375</td>
<td>0.3846</td>
<td>0.7098</td>
<td>0.1679</td>
<td>0.3154</td>
<td>0.3983</td>
<td>0.3259</td>
<td>0.5457</td>
</tr>
<tr>
<td>2007</td>
<td>0.8921</td>
<td>0.0929</td>
<td>0.3290</td>
<td>0.3407</td>
<td>0.7098</td>
<td>0.0405</td>
<td>0.3445</td>
<td>0.1757</td>
<td>0.3156</td>
<td>0.4611</td>
</tr>
<tr>
<td>2008</td>
<td>0.7183</td>
<td>0.0929</td>
<td>0.3205</td>
<td>0.2967</td>
<td>1.2113</td>
<td>0.2426</td>
<td>0.3751</td>
<td>0.0694</td>
<td>0.3105</td>
<td>0.3341</td>
</tr>
<tr>
<td>2009</td>
<td>0.6273</td>
<td>0.0122</td>
<td>0.3026</td>
<td>0.2528</td>
<td>1.2113</td>
<td>0.3639</td>
<td>0.4046</td>
<td>0.0843</td>
<td>0.3028</td>
<td>0.3071</td>
</tr>
<tr>
<td>2010</td>
<td>0.9384</td>
<td>0.0042</td>
<td>0.2752</td>
<td>0.1649</td>
<td>1.2113</td>
<td>0.6262</td>
<td>0.4337</td>
<td>0.2104</td>
<td>0.2873</td>
<td>0.0084</td>
</tr>
<tr>
<td>2011</td>
<td>0.6824</td>
<td>0.0097</td>
<td>0.2253</td>
<td>0.2746</td>
<td>0.9369</td>
<td>0.9391</td>
<td>0.4630</td>
<td>0.4411</td>
<td>0.2770</td>
<td>0.5240</td>
</tr>
<tr>
<td>2012</td>
<td>0.8143</td>
<td>0.0097</td>
<td>0.1310</td>
<td>0.3625</td>
<td>0.9369</td>
<td>1.1486</td>
<td>0.4935</td>
<td>0.5641</td>
<td>0.2513</td>
<td>0.5394</td>
</tr>
<tr>
<td>2013</td>
<td>0.9710</td>
<td>0.0064</td>
<td>0.0462</td>
<td>0.5383</td>
<td>1.2113</td>
<td>1.3737</td>
<td>0.5238</td>
<td>0.7318</td>
<td>0.2127</td>
<td>0.8472</td>
</tr>
<tr>
<td>2014</td>
<td>0.7922</td>
<td>0.0083</td>
<td>0.9126</td>
<td>0.9339</td>
<td>1.2113</td>
<td>1.5736</td>
<td>0.5561</td>
<td>0.8242</td>
<td>0.1992</td>
<td>0.9319</td>
</tr>
<tr>
<td>2015</td>
<td>0.8579</td>
<td>0.0072</td>
<td>1.3321</td>
<td>1.1976</td>
<td>1.2113</td>
<td>1.7585</td>
<td>0.5871</td>
<td>0.8370</td>
<td>0.5647</td>
<td>1.3244</td>
</tr>
<tr>
<td>2016</td>
<td>0.8071</td>
<td>0.0067</td>
<td>2.0061</td>
<td>1.6810</td>
<td>1.2113</td>
<td>1.9822</td>
<td>0.6238</td>
<td>1.1047</td>
<td>1.2597</td>
<td>2.3133</td>
</tr>
<tr>
<td>2017</td>
<td>0.7464</td>
<td>0.0028</td>
<td>2.7980</td>
<td>2.6918</td>
<td>1.2113</td>
<td>2.3248</td>
<td>0.6574</td>
<td>1.1869</td>
<td>2.5982</td>
<td>2.4557</td>
</tr>
<tr>
<td>2018</td>
<td>0.7017</td>
<td>0.0027</td>
<td>3.0827</td>
<td>3.1753</td>
<td>1.2113</td>
<td>0.6974</td>
<td>5.0188</td>
<td>1.5039</td>
<td>4.1246</td>
<td>2.6173</td>
</tr>
</tbody>
</table>

The Results of CCEMD signal decomposition

Figure 4: The results of CEEMD signal decomposition.
photovoltaic industry investment, photovoltaic power generation, macro policy support, and photovoltaic power generation technology as preliminary indicators affecting China’s PV installed capacity. Among these indicators, relaxation of international PV trade policy, macro policy support, and photovoltaic power technology maturity are qualitative indicators. Through the Delphi method, experts score the data. Since China’s PV installed capacity is a complex nonlinear problem with many influencing factors, the relationship between each index and PV installed capacity is vague. Therefore, this paper uses the grey correlation method to define the distance between the influencing factors and the installed capacity of PV, as well as its initial impact. The effective collection of indicator sets can effectively improve the speed and accuracy of the forecasting model. In addition, based on the same data test set, Matlab R2014a is used for programming. This test was performed on a computer with an Intel Core i5-6200U, 8G memory, and Windows 10 Professional Edition as a support.

5.1. Normalization of Indicators. If any data are missing for some years, in order to eliminate the dimensional difference between the data of different metrics and to ensure the availability of data, the data should first be rationally adjusted and supplemented. Normalize all the corrected data and use the $Z$-score data normalization method to standardize the $N$ sets of data in the data sets of the $M$ indicators:

$$Z_{mn} = \frac{x_{mn} - \bar{x}_n}{\sigma} \quad (m = 1, 2, \ldots, j, \ n = 1, 2, \ldots, n),$$

(18)

where $Z_{mn}$ is the normalized data, $x_{mn}$ is the corrected data, $\bar{x}_n$ is the average of $x_{mn}$, and $\sigma$ is the standard deviation of $x_{mn}$.

Since the cost of photovoltaic power generation is a negative index, in order to achieve the forwardization of data, the inverse index forward processing method is adopted:

$$x_{nm} = \frac{Z_{mn} - \max}{\max - \min}$$

(19)

where $x_{nm}$ refers to the data after forwardization, max is the maximum value of the normalized data, and min refers to the minimum value of the normalized data.

5.1.2. Grey Relational Analysis. In this paper, the normalized data for 1990–2018, which provide the preliminary influencing indicator for China’s PV installed capacity forecast, are used as the input for the GRA. Then, the degree between each impact index and China’s PV installed capacity can be calculated, as shown in Table 2.

Exclude the indicators below 0.5500. Then, the domestic market demand, PV tariffs, macro policy support, PV industry investment, and photovoltaic power technology maturity are eliminated, and the main factors affecting China’s PV installed capacity can be collected. The main factors include total social electricity consumption, GDP, power generation cost, population, photovoltaic power generation, PV module production, industry prosperity index, domestic PV output value, international PV trade policy easing, and foreign market demand. The normalized index values of the main factors affecting China’s PV installed capacity (1990–2018) are shown in Table 3.

5.2. PV Installed Capacity Forecasting Based on the CEEMD-ABC-LSSVM Model. The original historical load sequence is decomposed by CEEMD. Input the signal sequence of China’s historical PV installed capacity and then load the CEEMD model with the whole forecasting algorithm. With this process, we obtain 4 IMFs and 1 residual. The decomposition results are shown in Figure 4.

It can be seen from Figure 4 that the time series of China’s PV installed capacity has clear multiscale features, and the four IMF components present information for high and low variances with different fluctuation scales. Among them, the frequency of IMF1 is relatively high, which reflects the random noise information of the original time series data. Meanwhile, the residual frequency is low and the change is stable. The trend information reflecting the timing of China’s PV installed capacity reflects the overall change characteristics of China’s PV installed capacity. The number of IMFs obtained by CEEMD decomposition is related to the characteristics of the original data time series itself. Based on the nature of these IMF components, we can understand the practical significance of CEEMD decomposition. The original Chinese PV installed capacity sequence and each of the IMF components decomposed from the original data sequence are independent of each other, and the respective IMF components are mutually orthogonal. The last item obtained by decomposition removes the remainder of the original sequence from all IMF components. Indeed, this item represents the average trend of the original Chinese PV installed capacity change.

The influential factor indicator set as well as each IMF component and residual are used as the inputs for the training set and the test set of the ABC-LSSVM model. In this way, the predictive model is trained. After many training cycles, the accuracy of the test set forecasting result increases significantly. The main adaptive and nonadaptive parameters of ABC-LSSVM multiple training cycles are shown in Table 4.

We reconstructed the forecasting series to obtain the forecasting results and relative errors of the components and residuals of the time series decomposition of China’s PV installed capacity since 1990, as shown in Figure 5.

Based on an analysis of Figure 5, we find that the proposed CEEMD-ABC-LSSVM forecasting model has high forecasting accuracy for each IMF and residual prediction sequence of the PV installed capacity, and the

---

**Table 4: Main parameters of the ABC-LSSVM model.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (number of bees)</td>
<td>(5000, 10000)</td>
</tr>
<tr>
<td>limit (maximum number of iterations)</td>
<td>100</td>
</tr>
<tr>
<td>maxcycle (maximum number of stagnation)</td>
<td>50</td>
</tr>
<tr>
<td>c (regularization parameter)</td>
<td>0.97</td>
</tr>
<tr>
<td>g (width of radial basis function)</td>
<td>0.49</td>
</tr>
<tr>
<td>Global optimal fitness</td>
<td>1.94</td>
</tr>
</tbody>
</table>

---

5.2. PV Installed Capacity Forecasting Based on the CEEMD-ABC-LSSVM Model. The original historical load sequence is decomposed by CEEMD. Input the signal sequence of China’s historical PV installed capacity and then load the CEEMD model with the whole forecasting algorithm. With this process, we obtain 4 IMFs and 1 residual. The decomposition results are shown in Figure 4.

It can be seen from Figure 4 that the time series of China’s PV installed capacity has clear multiscale features, and the four IMF components present information for high and low variances with different fluctuation scales. Among them, the frequency of IMF1 is relatively high, which reflects the random noise information of the original time series data. Meanwhile, the residual frequency is low and the change is stable. The trend information reflecting the timing of China’s PV installed capacity reflects the overall change characteristics of China’s PV installed capacity. The number of IMFs obtained by CEEMD decomposition is related to the characteristics of the original data time series itself. Based on the nature of these IMF components, we can understand the practical significance of CEEMD decomposition. The original Chinese PV installed capacity sequence and each of the IMF components decomposed from the original data sequence are independent of each other, and the respective IMF components are mutually orthogonal. The last item obtained by decomposition removes the remainder of the original sequence from all IMF components. Indeed, this item represents the average trend of the original Chinese PV installed capacity change.

The influential factor indicator set as well as each IMF component and residual are used as the inputs for the training set and the test set of the ABC-LSSVM model. In this way, the predictive model is trained. After many training cycles, the accuracy of the test set forecasting result increases significantly. The main adaptive and nonadaptive parameters of ABC-LSSVM multiple training cycles are shown in Table 4.

We reconstructed the forecasting series to obtain the forecasting results and relative errors of the components and residuals of the time series decomposition of China’s PV installed capacity since 1990, as shown in Figure 5.

Based on an analysis of Figure 5, we find that the proposed CEEMD-ABC-LSSVM forecasting model has high forecasting accuracy for each IMF and residual prediction sequence of the PV installed capacity, and the
Figure 5: Continued.
Figure 5: The predicted results of all training and testing sets: (a) IMF₁; (b) IMF₂; (c) IMF₃; (d) IMF₄; (e) the residual.

Figure 6: Comparison of the forecasting results for the four models.
5.3. Model Comparison Analysis. In order to compare the superiori-ty of the proposed model, this paper also uses data for the main factors affecting China’s PV installed capacity in 2019–2040, such as LSSVM, ABC-LSSVM, EMD-ABC-LSSVM, and CEEMD-ABC-LSSVM. Algorithm training was performed on the four forecasting models, and the degree of fitting for the four forecasting models is shown in Figure 6.

In order to objectively compare the accuracy of a variety of models, common statistical indicators, including RMSE, $R^2$, and MRE, are adopted, and the index calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{q}_i - q_i)^2},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{q}_i - \hat{q})^2}{\sum_{i=1}^{n} (q_i - \bar{q})^2},$$

$$MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{q}_i - q_i}{q_i} \right| \times 100\%.$$  

where $\hat{q}_i$ is a predicted value, $\bar{q}_i$ is a sample mean, and $n$ is a sample number.

The calculation results for the four models are compared and shown in Table 5.

<table>
<thead>
<tr>
<th>Forecasting model</th>
<th>RMSE</th>
<th>$R^2$ (100%)</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSSVM</td>
<td>5.15</td>
<td>95.09</td>
<td>87.54</td>
</tr>
<tr>
<td>ABC-LSSVM</td>
<td>5.26</td>
<td>97.49</td>
<td>90.76</td>
</tr>
<tr>
<td>EMD-ABC-LSSVM</td>
<td>1.95</td>
<td>99.51</td>
<td>41.46</td>
</tr>
<tr>
<td>CEEMD-ABC-LSSVM</td>
<td>0.63</td>
<td>99.84</td>
<td>28.82</td>
</tr>
</tbody>
</table>

5.4. Prediction Results. We applied GM (1, 1) to predict the total electricity consumption, GDP, power generation cost, population, photovoltaic power generation, PV module production, industry prosperity index, domestic PV output value, international PV trade policy easing, and foreign market demand from 2019 to 2035, which was used as the input data for the CEEMD-ABC-LSSVM forecasting model. Finally, we examined China’s cumulative photovoltaic installed capacity from 2019 to 2035, as shown in Figure 7.
6. Conclusion

This article combined the CEEMD algorithm with the ABC-LSSVM algorithm to construct a CEEMD-ABC-LSSVM forecasting model for China’s PV installed capacity. After selecting the main influential factors, we used data from 1990 to 2010 as the training set for the CEEMD-ABC-LSSVM, EMD-ABC-LSSVM, ABC-LSSVM, and LSSVM forecasting models and data from 2011 to 2018 as the test set. Then, we compared the results from the forecasted results of all training and testing sets. Finally, we applied the CEEMD-ABC-LSSVM forecasting model to forecast China’s future PV installed capacity in 2019–2035. We find that China’s PV installed capacity will surpass 4000GW around 2035. And, its installed capacity will increase year by year. Moreover, China’s PV industry development will maintain a steady overall growth rate (compound annual growth rate). Our final conclusions and policy recommendations are as follows:

(1) From 2019 to 2035, forecasts of China’s photovoltaic industry show a gradual upward trend, except for growth rate. From the perspective of market certainty, although the government has a corresponding PV installed capacity index for every year, internal and external uncertainty factors are faced by the photovoltaic industry. At the same time, international PV market volatility also produces continuous market uncertainty.

(2) China’s PV installed capacity will be transferred to a scale effect stage around 2030. At that time, excessive PV industry investment will yield sustained GDP growth. In the future, China will continue to develop renewable energy technologies. Meanwhile, China must pursue better opportunities in the energy investment field and continue to reduce its pollutants and carbon emissions.

(3) China must pay more attention to the development of energy consumption terminal power to successfully realize its energy conservation and emission reduction responsibilities to the international community. As China’s photovoltaic power generation industry gradually enters the era of affordable Internet access, the cost of photovoltaics on grid electricity will experience a gradual reduction. Indeed, China’s PV industry still has significant room for future development. China’s PV installed capacity will gradually become the main force among the global PV installed capacity.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors’ Contributions

All the authors were involved in the data collection and preprocessing phase, model constructing, empirical research, result analysis and discussion, and manuscript preparation. All the authors have approved the submitted manuscript.

Acknowledgments

The completion of this paper has been helped by many teachers and classmates. We would like to express our gratitude to them for their help and guidance. This research was funded by the National Natural Science Foundation of China, grant number (71601078), the Social Science Fund of...
Beijing, grant number (16GLC070), and the State Grid Corporation Science and Technology Project, grant number (SGTJD00DWJS1800015).

References


