A Novel Ensemble Approach for the Forecasting of Energy Demand Based on the Artificial Bee Colony Algorithm

Jun Hao [1,2], Xiaolei Sun [1,2,*] and Qianqian Feng [1,2]

[1] Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China; 18954210260@163.com (J.H.); fengqianqian18@mails.ucas.ac.cn (Q.F.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[*] Correspondence: xlsun@casipm.ac.cn

check for updates

Abstract: Accurate forecasting of the energy demand is crucial for the rational formulation of energy policies for energy management. In this paper, a novel ensemble forecasting model based on the artificial bee colony (ABC) algorithm for the energy demand was proposed and adopted. The ensemble model forecasts were based on multiple time variables, such as the gross domestic product (GDP), industrial structure, energy structure, technological innovation, urbanization rate, population, consumer price index, and past energy demand. The model was trained and tested using the primary energy demand data collected in China. Seven base models, including the regression-based model and machine learning models, were utilized and compared to verify the superior performance of the ensemble forecasting model proposed herein. The results revealed that (1) the proposed ensemble model is significantly superior to the benchmark prediction models and the simple average ensemble prediction model just in terms of the forecasting accuracy and hypothesis test, (2) the proposed ensemble approach with the ABC algorithm can be employed as a promising framework for energy demand forecasting in terms of the forecasting accuracy and hypothesis test, and (3) the forecasting results obtained for the future energy demand by the ensemble model revealed that the future energy demand of China will maintain a steady growth trend.

Keywords: energy demand; ensemble framework; forecasting; machine learning; artificial bee colony algorithm

1. Introduction

As a strategic supply, energy is an important foundation for the development of an economy and a society [1]. With the growth of socioeconomic and technological advancements in the past few decades, energy demand has increased considerably [2]. Particularly, as the world's largest energy consumer, China has witnessed a three-fold increase in the energy demand throughout a 20-year period. In terms of this increased demand, a sufficient amount of energy is required to satisfy the nationwide demands, and the abundance of energy is of importance to guaranteeing the national energy security and sustainable development [3,4].

The scientific accurate prediction of energy demand is crucial for the rational formulation of energy policies and the basis for ensuring the security of energy supply [5,6] while the forecasting of the energy demand is conducive to prevent energy supply risks, reduce the gap between the energy supply and demand, slow down economic cycle fluctuations, and promote sustainable economic development and social stability. Therefore, forecasting its demand is crucial to the energy management and transportation sectors [7].

However, the forecasting of energy demand faces severe challenges due to the characteristics and properties of energy demand data. Energy demand data constitute a time series with strong nonlinear and chaotic characteristics. It contains a series of dynamic random factors, and it is affected by various factors, e.g., economic development level, industrial structure, and technology innovation [8]. Besides, the annual energy demand data are classified as small sample data, presenting higher requirements on the applicability of the prediction model [9]. In this case, the manner in which an effective forecast model is selected, and the accurate forecasting of energy demand are crucial issues that must be focused on.

To appropriately handle such a situation, a novel ensemble forecasting approach of energy demand based on the artificial bee colony (ABC) algorithm was constructed. The main aim was to seek forecasting models that are liable to train and exhibit high precision. Furthermore, seven forecast models were applied and compared to find the most suitable one, i.e., autoregressive integrated moving average (ARIMA), second exponential smoothing (SES), support vector regression (SVR), back propagation neural network (BPNN), radical basis function neural network (RBFNN), generalized regression neural network (GRNN), and extreme learning machine (ELM) models, respectively. In practice, seldom can a model achieve a good performance under any circumstances [10]. Ensemble learning functions by building and combining multiple machine learning machines to complete learning tasks, which can achieve a higher generalization ability and prediction accuracy [11,12]. Considering the rationality and adaptability of weight allocation, the ABC intelligent algorithm was introduced to integrate the prediction results of the base model. Different forecast models were selected for the combinatorial integration forecast to determine the model that exhibits the best performance.

The motivations and contributions of this paper were as follows. First, in view of the complexity of energy demand forecasting issue, this paper proposes a novel ensemble forecasting framework, which includes factor selection, data preprocessing, forecasting model training and testing process, and forecasting future energy demand. Second, considering the merits of the ABC algorithm in seeking the best integrated solution, this paper introduces the ABC algorithm to integrate the forecasting results of the base model for accurate energy demand forecasting. Third, it is difficult to characterize the impact of exogenous factors on energy demand in a single time series. To this end, this paper considers multiple factors to establish the forecasting model of energy demand. Additionally, this paper predicts the future energy demand of China based on the ensemble model.

The layout of this paper is organized as follows. Section 2 discusses the related studies focusing on energy demand forecasting. Section 3 discusses the formulation of the energy demand forecasting framework and the application of machine learning methods to the problem of energy demand forecasting. Section 4 discusses the empirical results and summary. Section 5 provides the concluding remarks.

2. Literature Review

2.1. Energy Demand Influencing Factors

In the last few decades, energy demand forecasting has turned into an active research area due to its significant impact on the energy security and socioeconomic development of a country [13,14]. A series of prominent research studies have reported on energy demand forecasting, including factor selection [7,15,16] and method determination [13,17].

Owing to the complexity of energy demand forecasting [18,19], it is an extremely difficult task to develop an energy demand forecasting model considering all of the influencing factors. Therefore, studies on the selection of influencing factors are mainly divided into single variables and multivariate variables. Due to the method characteristics, single-variable prediction mainly focuses on the traditional econometric model [20–23] and decomposition-integrated prediction model [24–27]. For example, Tso and Yau [28] examined the effects of socioeconomic indicators on the electricity demand. Adom and Bekoe [29] adopted macro socioeconomic indicators, such as GDP, population, and degree of

urbanization. To build a model to forecast energy demand, Ghanbari et al. [30] collected data on potentially useful factors, such as the GDP, consumer price index (CPI), energy intensity, energy efficiency, and population, and utilized the feature selection technique to eliminate low impact factors and select the most influential ones. Wu and Peng [31] utilized multiple factors, such as economic growth, total population, fixed-asset investments, energy efficiency, and energy structure (ES), to estimate the energy demand of China. The influencing factors were identified (Table 1).

Table 1. Influencing factors of energy demand.

Literatures	Forecast Range	Case Study	Factors
Hribar, Potočnik, Šilc, and Papa [7]	Daily/City	Natural gas demand	Weather, socioeconomic factors
Adom and Bekoe [29]	Yearly/Country	Electrical demand	Macro socioeconomic indicators, population, gross national product, degree of urbanization, industry efficiency, industry value-added
Ghanbari, Kazemi, Mehmanpazir, and Nakhostin [30]	Yearly/Country	Energy demand	Population, GDP, CPI, imports, exports, energy intensity, energy efficiency
Wu and Peng [31]	Yearly/Country	Energy demand	Economic growth, total population, fixed-asset investments, energy efficiency, energy structure, household energy consumption per capita
Yu and Zhu [32]	Yearly/Country	Energy demand	Economic growth, total population, economic structure, urbanization rate, energy structure, energy price
Karadede et al. [33]	Yearly/Country	Natural gas demand	Gross national product, population, growth rate
Angelopoulos et al. [34]	Yearly/Country	Electricity demand	GDP, unemployment rate, population, weather-related criteria, electricity price, energy efficiency criterion
Piltan et al. [35]	Yearly/Country	Energy demand	Number of employees, investment value, gas price, electricity price
Sonmez et al. [36]	Yearly/Country	Transport energy demand	Gross domestic product, population
Yuan et al. [37]	Yearly/Country	Energy demand	Economic level, industrial structure, demographic change, urbanization process, technological progress
He et al. [38]	Yearly/Province	Energy demand	Historical energy consumption, population, GDP growth rate, total GDP, the three major industrial GDP, CPI
He et al. [39]	Yearly/City	Energy demand	Population, GDP growth rate, GDP, the three major industrial GDP, CPI
Forouzanfar et al. [40]	Yearly/Country	Transport energy demand	Population, gross domestic product, number of vehicles
Liao et al. [41]	Yearly/Country	Energy demand	GDP, population, oil price

The factors influencing energy demand forecasting are different due to different energy types and forecasting ranges. Clear differences in terms of the influencing factors among different energy types are observed, such as the electricity demand, natural gas demand, and transportation energy demand. Besides, due to different forecasting ranges, differences in the selection of influencing factors are observed. For example, most of the annual forecasts are based on macroeconomic indicators while the influencing factors of daily forecasts focus on external factors, such as weather and energy prices. According to the range and type of energy forecast, macroeconomic indicators, such as the GDP, population, and industrial structure, for energy demand forecasting are focused herein. As mentioned above, this paper will make full use of expert knowledge and the related literature to identify the influencing factors of energy demand and establish a systematic and reasonable influencing factors list.

2.2. Energy Demand Forecasting Method

Due to the characteristics and properties of energy demand data, energy demand forecasting faces severe challenges and requirements on the applicability of the prediction model. Currently, various models have been used to forecast the energy demand [7,40,42–45], which can be divided into three categories: Statistical, mathematical programming (MP), and computational intelligence (CI). Statistical methods investigate the accumulation, examination, elucidation, presentation, and association of data [46], including linear regression, multiple regression, stepwise regression, and nonparametric

regression. All of these methods have been widely applied in previous studies; these methods are capable of yielding better results for solving linear problems [29,35].

With the emergence of artificial intelligence, various learning technologies, such as the heuristic algorithm [30,40], support vector machine (SVM) [47,48], artificial neural network (ANN) [49], extreme learning machine (ELM) [50], and ensemble techniques, have become quite popular. Actually, real-life problems exhibit nonlinear characteristics during forecasting, especially for energy demand [46]. MP methods provide the best solutions from a set of available alternatives under some constraints. Among these models, nonlinear programming models have been widely used to deal with nonlinear forecasting problems while computational methods have been used for prediction problems where mathematical formulae and the prior known data on the relationship between inputs and outputs are unknown. ANN and SVM are the most popular forecasting techniques for dealing with nonlinear problems. Although each of the forecasting model exhibits its own merits and demerits (Table 2), CI methods have already attracted the most attention due to their high reliability and forecasting accuracy. These CI-based models have been vastly implemented for energy demand forecasting, including SVR, recurrent neural network (RNN), and convolutional neural network (CNN).

Table 2. Comparison of different forecasting models for energy demand.

Models	Trend Features		Forecast Period		Number of Variables	
	Linear	Nonlinear	Long	Short	Multiple	Single
Regression-based	✓			✓	✓	
ARIMA	✓		✓			✓
SVM		✓		✓	✓	
ANN		✓		✓	✓	
ELM		✓		✓	✓	

Note: symbol ✓ represents the superiority of forecasting performance.

Time series analysis utilizes regression methods to establish various function equations of time series for trend prediction. Some classical time series analysis models, such as the exponential smoothing (ES) method [51] and ARIMA [52], have been widely used in the fields of social sciences. However, in the face of the large number of nonlinear complex time series prevalent in social and economic phenomena, the performance of a conventional mathematical statistics prediction method is not sufficiently good. On the contrary, machine learning models exhibit good performance for dealing with nonlinear problems, especially complex and nonlinear timing data, such as SVMs, neural networks, and various derivation methods. However, machine learning models have strict requirements on the data structure. When the amount of data is small, over-fitting may occur for the establishment of prediction models, leading to a weak generalization ability. In this case, computational learning theory results revealed that ensemble forecasting methods exhibit good prediction performance.

Ensemble learning functions by building and combining multiple machine learning machines to complete learning tasks, which can achieve a higher generalization ability and prediction accuracy [11,12]. As Bates and Granger [10] pointed out, a linear combination of the base models would yield a smaller forecasting variance than a single forecasting model. Since the linear weighting ensemble method cannot adjust the weights of single models adaptively, the adaptability of the overall prediction model is relatively poor [53]. Galicia et al. adopted the weighted least square method to forecast electricity demand [54]. Liu et al. [55] constructed an ensemble forecasting model combining the grey model, ARIMA, and second-order polynomial regression model with particle swarm optimization to predict CO_2 emission in China. Wang et al. [56], Zhu and Wei [57], and Qu et al. [58] adopted a genetic algorithm, particle swarm optimization algorithm, and bat algorithm, respectively, for the ensemble forecasting model. The artificial bee colony (ABC) algorithm combines the advantages of local deep search and global-wide search and exhibits good performance for seeking the best integrated solution, and the ABC algorithm is simple to operate and easy to implement [59].

In this paper, a predictive analysis framework based on the ABC algorithm is proposed in this paper for small sample, nonlinear, and chaotic time series. Seven base models, including SES, ARIMA, SVR, BPNN, GRNN, RBFNN, and ELM, are used for energy demand forecasting, and the ABC algorithm is implemented to combine the base model predicted results.

3. Forecasting Framework and Methods

In this section, the framework and methods of energy demand forecasting are mainly discussed. First, the ensemble framework of energy demand forecasting is introduced, and the applied base models are described. Finally, crucial information on the ABC ensemble algorithm is provided.

3.1. Ensemble Framework of Energy Demand Forecasting

To better forecast energy demand under multi-factor disturbance, this paper presents the ensemble framework of energy demand forecasting (Figure 1) and the pseudo-code of the ensemble forecasting model, as shown in Appendix B Table A3. The ensemble framework is divided into four stages: (1) Influence index selection, (2) data preprocessing, (3) forecasting model training and testing, and (4) future energy demand forecasting.
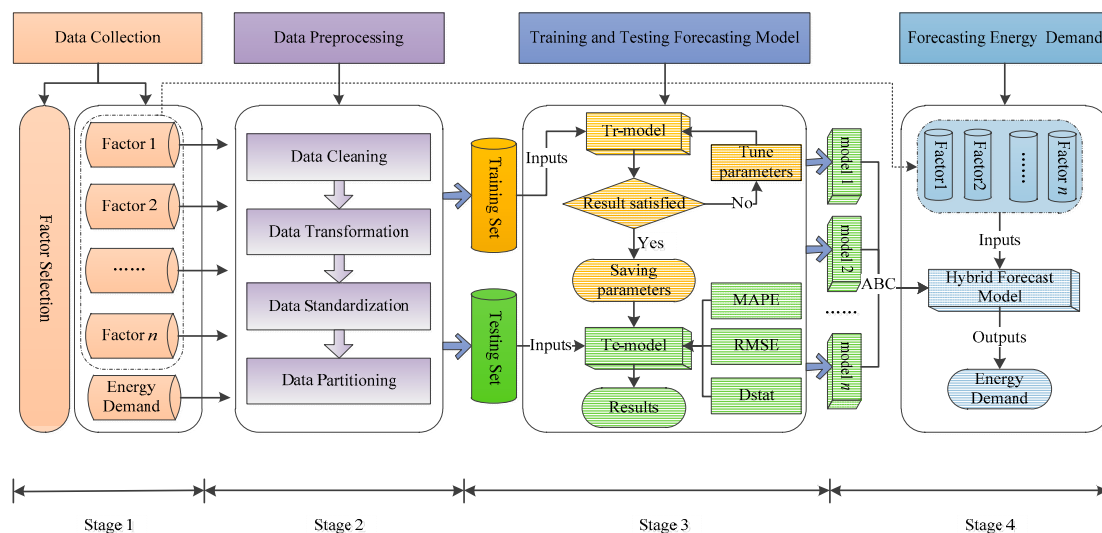

Figure 1. Ensemble framework of energy demand forecast.

3.1.1. Factor Selection

Energy systems are complex nonlinear systems, and energy demand is affected by various socioeconomic factors, such as the GDP, CPI, economic structure, and energy efficiency. To more accurately forecast China's future energy demand, it is crucial to systematically analyze the influencing factors of energy demand. By reviewing the literature, this paper selected the indicators that affect energy demand (Table 3).

In addition, correlations, r, between the energy demand and selected factors are listed, and p is used to test the hypothesis that there is no relationship between the observed phenomena (null hypothesis). All of the selected factors pass the correlation test under the condition of $\alpha = 95\%$. The urbanization rate is one of the strongest factors of energy demand. Furthermore, a significant negative correlation between technological progress and energy demand is observed (Figure 2). In fact, the correlation plot (Figure 2) revealed that the ES exhibits a weak positive correlation with energy demand.

No.	Factors	r	p
1	Gross domestic product (GDP)	0.965	0.0000
2	Industrial structure (IS)	0.887	0.0000
3	Energy structure (ES)	0.328	0.0387
4	Technological innovation (TI)	−0.697	0.0000
5	Urbanization rate (UR)	0.978	0.0000
6	Population (Pop)	0.873	0.0000
7	Consumer price index (CPI)	0.959	0.0000
8	Energy demand (ED)	1.000	——

Note: Correlation in *r* is considered significant if the off-diagonal elements of *p* are less than the significance level (0.05).
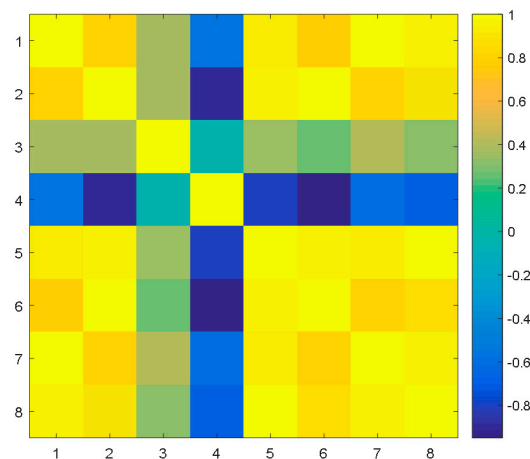
Figure 2. Correlation plot between the energy demand and selected factors.

3.1.2. Data Preprocessing

Actually, the collected datasets are extremely sensitive to various differences (such as data frequency, noise, original units, magnitude, and missing values). These differences or errors in the raw data might result in a considerable deviation in the forecasted results. Therefore, it is crucial to preprocess the data to ensure reliability during knowledge mining. Generally, data preprocessing involves the following processes.

- Data cleaning: Data cleaning is mainly utilized to fill in missing values, remove noise, monitor outliers, and deal with differences in datasets.
- Data transformation: The data transformation process involves several methods, including the transformation of multiple files into a unified available data format as well as feature extraction.
- Data standardization: Due to the different dimensions of data, large differences in the magnitude of collected data exist, often leading to large deviations in data analysis. Therefore, it is crucial to conduct the standardized processing of data to eliminate effects of dimensions and magnitude. The following Formula (1) is adopted in this paper:

$$x_{ik} = \frac{X_{ik} - \min(X_i)}{\max(X_i) - \min(X_i)}, \tag{1}$$

where x_{ik} represents the data of the *k-th* element corresponding to the *i-th* influencing factor after standardization, and X_i represents the sequence corresponding to the *i-th* influencing factor.

- Data partitioning: To train and test the prediction model, observation values need to be divided into training sets and test sets. The data from the training set are used to train the prediction model

to obtain optimal parameters, and then the data from the test set are used to test the generalization performance of the model.

The processed data are presented in the Appendix A Table A2.

3.1.3. Forecasting Model Training and Testing

At this stage, the collected data are mainly used to train and test the forecasting model. The model is trained by training set data. If the training results meet the requirements, the model parameters are saved; otherwise, the model parameters with the best prediction accuracy and generalization performance are retrained until they are obtained.

3.1.4. Forecasting Future Energy Demand

To predict the future energy demand, the input data (influencing factors) need to be predicted. For the time series data of influencing factors, the trend extrapolation method for forecasting is adopted (Figure 3). The future energy demand data can be obtained by inserting the data into the integrated model.
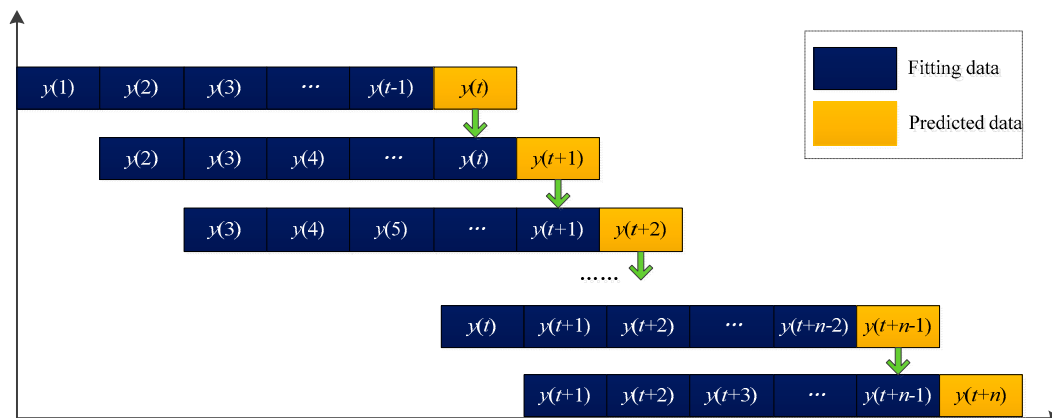

Figure 3. The trend extrapolation method of energy demand influencing factors.

3.2. Base Models

The time series of energy demand exhibit nonlinear and chaotic characteristics. It is affected by various factors, such as the economic development level, technology innovation, and total population. Meanwhile, the annual energy demand belongs to the small sample, presenting higher requirements on the performance of the prediction model. Hence, the traditional econometric models and emerging machine learning models were selected as base models in this section. The detailed information about the selected base is as follows.

3.2.1. Autoregressive Integrated Moving Average

The ARIMA model is a combination of the autoregressive moving average model with differencing. It can be described as ARIMA (p, d, q), where p and q represent the lagged values of autoregressive and moving average sections, respectively [60]. d denotes the number of differences in the time series. Generally, the ARIMA model can be expressed by the following formulations:

$$X_t = \theta_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \tag{2}$$

where X_t denotes the actual value at time period t; φ_i ($i = 1, 2, \ldots, p$) and θ_j ($j = 1, 2, \ldots, q$) are parameters to be estimated; and ε are the random errors, which are assumed to be independently and identically distributed with a mean of zero and a constant variance of δ^2.

3.2.2. Second Exponential Smoothing

Exponential smoothing is a classical time-series prediction method with a small sample, which is widely used due to its high operability. SES eliminates the random fluctuation in the historical statistical sequence and finds the main trend of its development. It is suitable for simple time series analysis and short- and medium-term prediction. Figure 4 shows the flow chart of the SES model.
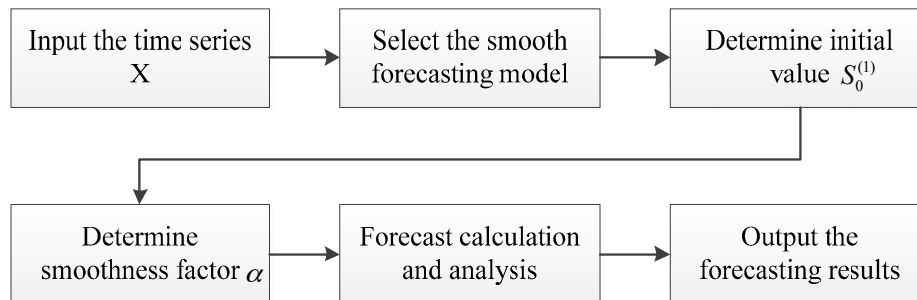

Figure 4. Flow chart of the SES model.

Assume that $X_0, X_1, \cdots, X_n$ are the observations of the time series, and $S_t^{(1)}, S_t^{(2)}$ are the first and second exponential smoothing values of the observation, respectively:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)}, \tag{3}$$

$$S_t^{(1)} = \alpha X_t + (1-\alpha)S_{t-1}^{(1)}. \tag{4}$$

The prediction formula is expressed by Equation (5):

$$\hat{X}_{t+T} = a_t + b_t T, \tag{5}$$

$$a_t = 2S_t^{(1)} - S_t^{(2)}, \tag{6}$$

$$b_t = \frac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}), \tag{7}$$

where α is the smoothing factor, and T is the lead time of the forecast.

3.2.3. Support Vector Machine

SVM developed by Cortes C and Vapnik V is the most popular, robust, and widely used intelligence method [61]. SVM exhibits a good performance for dealing with small samples, as well as high-dimensional nonlinear problems. Assume that there is a training sample set $\{(x_i, y_i), i = 1, 2, \cdots, l\}$, where $x_i(x_i \in R^d)$ with d features, and $y_i \in R$ represents the dependent variables associated with each input. The SVM model tries to formulate a mapping $f(x) : R^d \to R$, where $f(x)$ is expressed as follows:

$$f(x) = w^T x + b, \tag{8}$$

where w and b are the model parameters to be solved.

It is imperative to adopt a transformation function, $\phi : R^d \rightarrow R^s (s > d)$, to all points in the input space for capturing nonlinear data features. The modified mapping function is expressed as follows:

$$f(x) = w\phi(x) + b. \tag{9}$$

The ε linear insensitive loss function is defined as follows:

$$L(f(x), y, \varepsilon) = \begin{cases} 0 & , |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & , |y - f(x)| > \varepsilon \end{cases}, \tag{10}$$

where $f(x)$ denotes the forecasting value, and y represents the observed values.

Formally, the SVR model can be expressed as a convex optimal problem as follows:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} (\varepsilon_i + \varepsilon_i^*)$$
$$s.t. \begin{cases} y_i - w\phi(x) - b \leq \varepsilon_i + \varepsilon_i^*, i = 1, 2, \cdots, l \\ -y_i + w\phi(x) + b \leq \varepsilon_i + \varepsilon_i^*, i = 1, 2, \cdots, l \\ \varepsilon_i \geq 0, \varepsilon_i^* \geq 0 \end{cases}, \tag{11}$$

where ε_i, ε_i^* are the slack variables, C is the penalty factor, and ε is the error requirement of the regression function. The above problem in Equation (10) can also be expressed in its dual form as follows:

$$\max\left[-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) - \sum_{i=1}^{l} (a_i + a_i^*)\varepsilon + \sum_{i=1}^{l} (a_i - a_i^*)y_i\right]$$
$$s.t. \begin{cases} \sum_{i=1}^{l} (a_i - a_i^*) = 0 \\ 0 \leq a_i \leq C \\ 0 \leq a_i^* \leq C \end{cases}, \tag{12}$$

where $K(x_i, x_j) = \Phi(x_i)\Phi(x_j)$ is the kernel function.

By setting the optimal solution of the above equation as $a = [a_1, a_2, \cdots, a_l]$ and $a^* = [a_1^*, a_2^*, \cdots, a_l^*]$, Equations (13) and (14) can be obtained:

$$w^* = \sum_{i=1}^{l} (a_i - a_i^*)\Phi(x_i), \tag{13}$$

$$b^* = \frac{1}{N_{nsv}}\{\sum_{0<a_i<C}^{l} [y_i - \sum_{x_i \in SV}^{l} (a_i - a_i^*)K(x_i, x_j) - \varepsilon] + \sum_{0<a_i<C}^{l} [y_i - \sum_{x_i \in SV}^{l} (a_j - a_j^*)K(x_i, x_j) - \varepsilon]\}, \tag{14}$$

where N_{nsv} is the number of support vectors.

Therefore, the regression function is shown in Equation (15):

$$\begin{aligned} f(x) &= w^*\phi(x) + b^* \\ &= \sum_{i=1}^{l} (a_i - a_i^*)\Phi(x_i)\phi(x) + b^* \\ &= \sum_{i=1}^{l} (a_i - a_i^*)K(x_i, x) + b^* \end{aligned}. \tag{15}$$

3.2.4. Artificial Neural Networks

ANNs has been a research hotspot in the field of artificial intelligence since the 1980s. The ANN abstracts the neural network of a human brain from the perspective of information processing, builds some simple models, and forms different networks according to different connection modes. A neural network is an operational model comprising a large number of nodes (or neurons) connected to each other. Each node represents a specific type of an output function called an activation function. The connection between each node represents a weighted value of the signal passing through the connection, known as weight, which is equivalent to the memory of the ANN. The output of the network varies according to the connection mode, weight value, and excitation function [62,63]. In this paper, three models of BPNN, RBFNN, and GRNN, respectively, are mainly introduced.

3.2.5. Extreme Learning Machine

ELM is a simple efficient algorithm that does not require the tuning of parameters and exhibits extremely rapid learning speeds [64]. The network of an ELM training model is a single-hidden-layer feed-forward neural network structure (Figure 5).
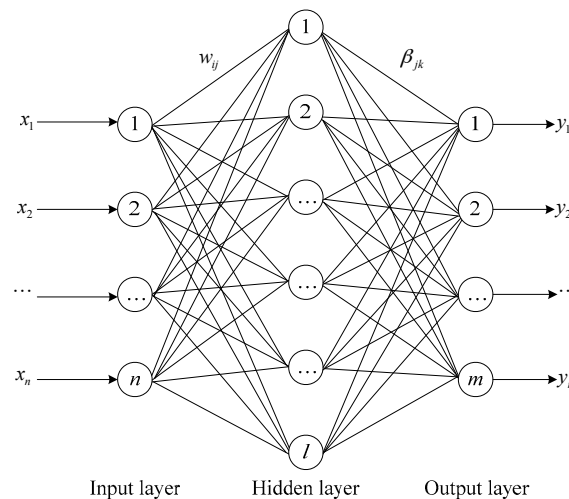

Figure 5. Structure of the single-hidden-layer feed-forward neural network.

Take N data samples $(x_i, y_i), i = 1, 2, \cdots, N$ as an example, where the input data are $x_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,N}]$ and the output data are $y_i = [y_{i,1}, y_{i,2}, \cdots, y_{i,M}]$. The ELM algorithm is as follows:

Assume that single-layer feed-forward neural networks (SLFNs) with L hidden layer nodes and activation function $g(a, b, x)$ can fit N data samples with non-errors, indicating that the presence of the parameters $a_i, b_i, \beta_i, i = 1, 2, \cdots, L$ makes Equation (16) true:

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i g(a_i, b_i, x_j) = y_j, j = 1, 2, \cdots, N. \tag{16}$$

Equation (16) can be represented by a matrix as follows:

$$H\beta = Y \Leftrightarrow \begin{bmatrix} g(a_1, b_1, x_1) & \cdots & g(a_L, b_L, x_L) \\ \vdots & \ddots & \vdots \\ g(a_1, b_1, x_N) & \cdots & g(a_L, b_L, x_N) \end{bmatrix}_{N \times L} * \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times M} = \begin{bmatrix} Y_1^T \\ \vdots \\ Y_N^T \end{bmatrix}_{N \times M}, \tag{17}$$

where H is the output matrix of the hidden layer, the *i*-th column represents the output of $(x_1, \cdots, x_N)$ at the *i*-th hidden layer node, and the *j*-th row represents the output of $(x_1, \cdots, x_N)$ on all hidden layer nodes.

In most cases, the number of the hidden layer nodes, L, is less than the number of samples, N, making it difficult for the constructed neural network to approximate Equation (16). By using E to represent the error between the output of the training sample and the actual output, Equation (18) is obtained:

$$E = \left[e_1^T, \cdots, e_1^T \right]_{N \times M}^T. \tag{18}$$

Combining Equations (17) and (18), Equation (17) can be expressed as Equations (19)–(21):

$$H\beta = T, \tag{19}$$

$$\sum_{j=1}^{N} \| t_j - y_j \| < \varepsilon, \tag{20}$$

$$\min_{\beta} \| H\beta - T \|, \tag{21}$$

$$\hat{\beta} = H^+ T, \tag{22}$$

where ε indicates any small number, which can ensure t_j can approach y_j infinitely. $\hat{\beta}$ is the optimal solution of Equation (22), and H^+ is the Moore–Penrose generalized inverse of the hidden layer output matrix, H.

3.3. Artificial Bee Colony Ensemble Algorithm

The ABC algorithm is a random search meta-heuristic global optimization algorithm, which divides worker bees into three categories: Employed bees, onlooker bees, and scout bees [65]. Employed bees and onlooker bees are used for mining honey sources while scout bees search for new honey sources. The location of nectar sources is the result of the optimization problem. First, a group of initial solutions (honey source) are randomly generated from the feasible domain, and all employed bees conduct a domain search. Second, the scout bees randomly search the honey source and record the current best solution. Then, the onlooker bees choose honey sources according to some strategies and conduct a random search to record the best solution. Finally, the honey source that falls into the local optimal point is replaced by the randomly generated honey source, and the global optimal solution is finally obtained after several iterations [66]. Figure 6 shows the schematic of the ABC algorithm.

To better demonstrate the algorithm flow, this paper gives the pseudo-code of the ABC algorithm in the Appendix B Table A4.
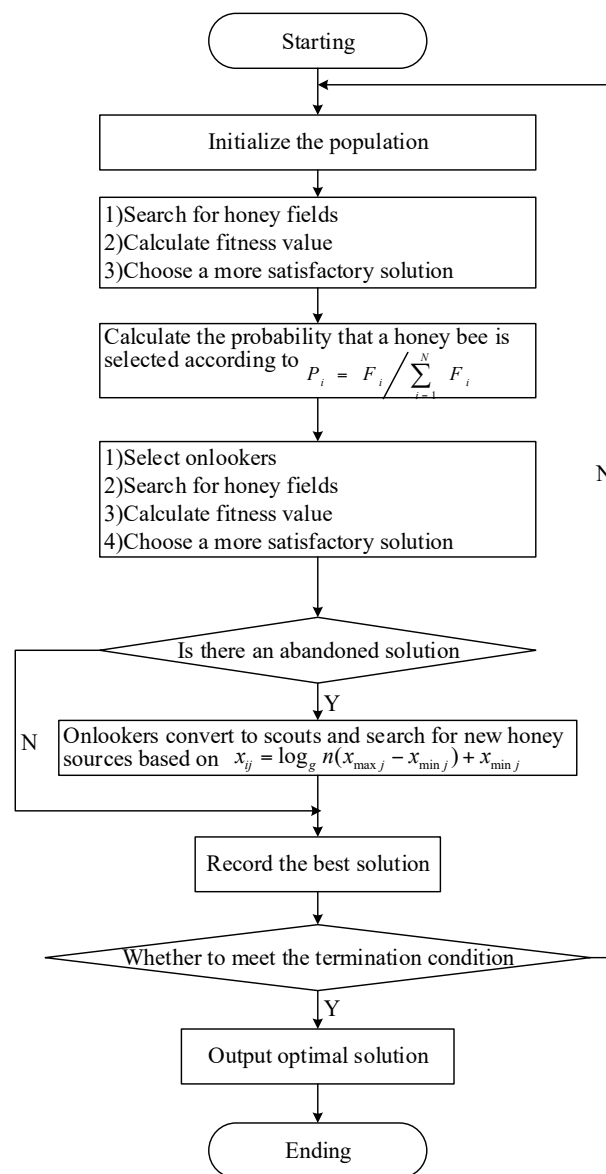
Figure 6. Schematic of the artificial bee colony algorithm.

4. Empirical Analysis

4.1. Datasets

In this paper, an energy demand series was investigated for forecasting analysis. The primary energy consumption/demand data of China from 1978 to 2017 were collected from the Wind database, BP statistical review of world energy (https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html), and the China National Bureau of Statistics database (http://www.stats.gov.cn/). All the original data are listed in the Appendix A Table A1. Figure 7 shows the raw demand series and energy demand increase rate. Meanwhile, factors, such as the GDP, industrial structure, energy structure, technological innovation, urbanization rate, population, and CPI, were selected as exogenous variables. All of the series are yearly data from 1978 to 2017 including 40 observations. For the training and testing model, the series wre partitioned into two parts. The observations from 1978 to 2012 were regarded as the training samples, and the remainder were testing samples.
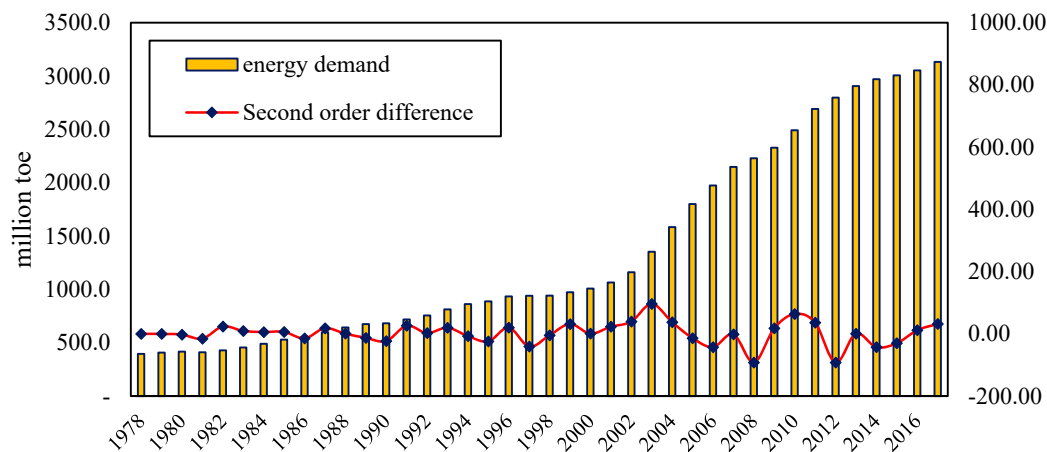
Figure 7. Primary energy demand time series from 1978 to 2017.

4.2. Error Metric and Statistic Test

The chosen error metric should reflect the forecasting performance of the models from different aspects. In view of this, three indicators were adopted herein: Root mean square error (RMSE) and mean absolute percentage error (MAPE). These indicators have been widely utilized in recent years:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2},\tag{23}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|,\tag{24}$$

where n is the size of the predictions, and y_t and $\hat{y}_t$ represent the observed value and predicted value, respectively.

Besides, to test the difference in the model prediction performance from the statistical aspect, the Diebold–Mariano (DM) statistical test was introduced to determine whether the prediction accuracy of model A is significantly better than that of model B. The null hypothesis of DM is that the prediction accuracy of model A is no greater than that of model B, that is, the prediction error $e_{A,t} = x_t - \hat{x}_{A,t}$ of model A is greater than or equal to the prediction error $e_{B,t} = x_t - \hat{x}_{B,t}$ of model B [67]. Accordingly, DM test statistics can be given by Equations (25)–(27):

$$DM = \frac{\frac{1}{n}\sum_{t=1}^{n}g_t}{\sqrt{(\gamma_0 + 2\sum_{l=1}^{\infty}\gamma_l)/n}} \sim N(0,1),\tag{25}$$

$$g_t = e_{A,t}^2 - e_{B,t}^2,\tag{26}$$

$$\gamma_l = \text{cov}(g_t, g_{t-l}).\tag{27}$$

The unilateral test of DM statistics can effectively identify the superiority of model A over model B according to the DM statistics and its p value.

4.3. Parameter Settings

To test the superiority of the proposed framework, six forecasting base models and one simple average ensemble model were established and applied as benchmarks. First, the ARIMA and SES models that exhibit better performance for small-sample nonlinear time series were selected.

Then, five emerging machine learning models, i.e., SVR, BPNN, GRNN, RBFNN, and ELM models, were adopted.

For the ARIMA model, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are widely used to determine the orders of the autoregressive and moving averages process [68]. ACF measures the correlation between time-series and its lags while PACF are the correlation coefficients between time-series and its lags without the influence of members in between. The lags of the autoregressive process are usually determined by the PACF diagram while the lag value of the moving averages process is determined by the ACF diagram. In this paper, parameters p and q were all set to be 5 according to the ACF and PACF diagrams. As the second-order difference of the original sequence is stable, d was set as 2. As for the SES model, the smoothing factor α was set to 0.7.

For artificial neural networks, it is widely accepted that a feed-forward network with one hidden layer and enough neurons in the hidden layers can fit any finite input-output mapping problem [64]. Therefore, the BPNN was set as a standard three-layer neural network, including an input layer, one hidden layer, and output layer [67]. The number of hidden layer nodes was set to 20 according to Zhao et al. [67] as a small number of hidden nodes results in an inaccurate fitting and too much results in local optimums. There are four layers, i.e., input layer, pattern layer, summation layer, and output layer, respectively, in the GRNN model. The number of nodes in the input layer is equal to the dimensions of the input vector. The spread of the radial basis function in the RBF model was set to 1, and the number of input neurons was equal to the number of columns in the data matrix.

Also according to Zhao et al. [67], Godarzi et al. [69], and Yu et al. [70], the kernel function was set to the Gaussian kernel function, and C and gamma were respectively set as iqr(Y)/1.349 and 1, where iqr(Y) is the interquartile range of the processed target series. Besides, the number of neurons in the hidden layer was equal to the number of training samples in the ELM model. The sigmoid function was selected as the activation function.

For the ABC ensemble algorithm, the size of the bee colony was set to 100, and the maximum number of loop iterations was 1000. All of the mentioned forecasting models were run 20 times by using MATLAB R2016a software: (2016a, MathWorks, Natick, MA, USA). The average results are shown in the following sections.

4.4. Forecasting Error and Statistical Test

In this section, according to the error metric criteria (i.e., RMSE and MAPE), the forecasting error of out-of-sample data for eight benchmark forecasting models and the established model is shown. Then, the *DM* test was conducted for testing the significance level of the difference between any two models.

Figure 8a plots the actual values of the energy demand (line) and the forecasted values (bar) of each model, and Figure 8b shows the absolute errors of each forecasting model. From the diagram of the prediction fitting curve (Figure 8a), the benchmark models and the proposed ensemble forecasting model are effective, reflecting that the selected models are rational. In addition, the out-of-sample prediction performance of the GRNN and RBFNN models is not very good. Such a conclusion can also be verified from Figure 8b. The absolute errors of GRNN and RBFNN fluctuate considerably. Compared to all benchmark models, the ensemble model proposed herein exhibits the smallest fluctuation of the absolute prediction error (Figure 8b).
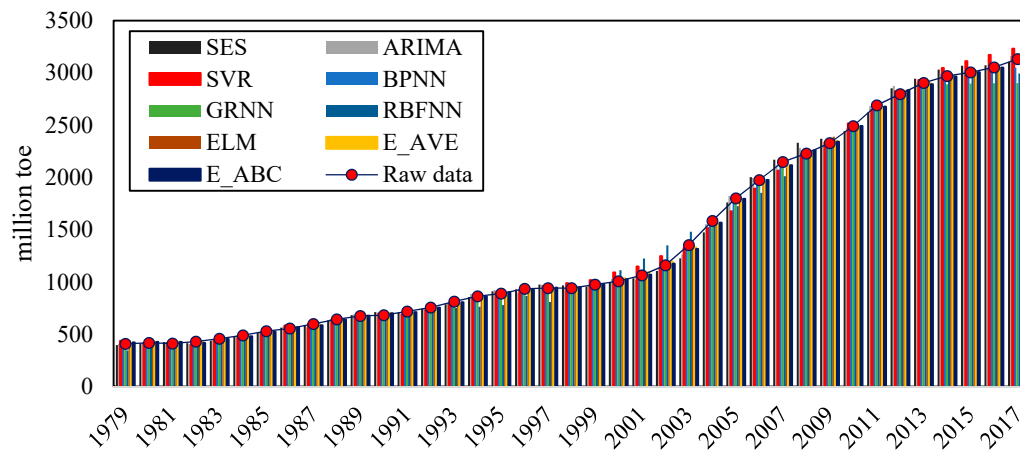
(**a**)
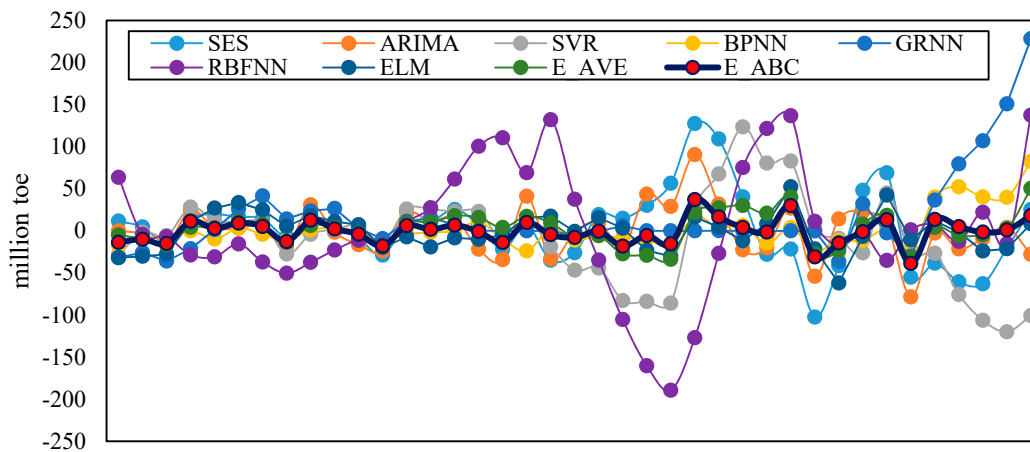
(**b**)

Figure 8. Forecasting results (**a**) and absolute errors (**b**) of each forecasting model.

For the in-depth analysis of the prediction performance, two evaluation indexes were calculated. Table 4 summarizes the performance of each model, including RMSE and MAPE, as well as the mean value of the statistics.

Table 4. Forecasting accuracy of all models.

Model	RMSE	MAPE
SES	46.68	1.41%
ARIMA	18.31	0.51%
SVR	101.90	3.31%
BPNN	56.42	1.76%
GRNN	152.15	4.62%
RBFNN	70.28	1.51%
ELM	16.57	0.45%
E_AVE	25.52	0.51%
E_ABC	**9.46**	**0.21**%

Note: The boldface value represents the best performance among the seven models in terms of RMSE and MAPE.

First, by the comparison of the base prediction models, including SES, ARIMA, SVR, and ANN, the ELM model exhibits the best prediction performance probably because the parameters do not need

to be adjusted in the training process for the use of the ELM algorithm, and the unique optimal solution can be obtained by setting the number of neurons in the hidden layer. In addition, the ELM algorithm exhibits advantages of a rapid learning speed and good generalization performance. For the remaining base models, ARIMA exhibits better performance than SES, SVR, BPNN, GRNN, and RBFNN on the part of RMSE and MAPE, reflecting that ARIMA can forecast small-sample time series.

Once the ensemble prediction model is considered, the prediction model (the ensemble model with the ABC algorithm, E_ABC) proposed herein exhibits the best prediction performance. This model exhibits the lowest RMSE and MAPE. The comparison of the two integrated models revealed that E_ABC exhibits a better performance than E_AVE with respect to the criteria of RMSE and MAPE. For example, the RMSE values for E_ABC and E_AVE are 9.46 and 25.52, respectively. The RMSE of model E_ABC is less than half that of model E_AVE. Meanwhile, the MAPE values for E_ABC and E_AVE are 0.21% and 0.51%, respectively. The MAPE of E_AVE is considerably greater than that of E_ABC.

Besides, the prediction performance of the simple average ensemble model (E_AVE) is worse than that of some base prediction models (i.e., ELM and ARIMA model), possibly because the simple average ensemble method ignores the correlation between base models, and the number of base models is relatively small. The ensemble model with the ABC algorithm (E_ABC) makes use of the prediction results of each base model to train the model and to minimize the absolute error. The optimal training results can be obtained without considering the correlation between base models.

Finally, the results revealed that the integrated model E_ABC exhibits the best performance due to its lowest RMSE and MAPE. Hence, the integrated model E_ABC is powerful for energy demand forecasting.

Table 5 summarizes the empirical results from the DM test, as well as the *p* values of the relevant statistics between any two models. For explanation purposes, in Table 5, the *p* value in row 2, column 2 is 0.0081 and less than 0.01, indicating that the test rejects the null hypothesis (i.e., there is a significant difference between the forecasted results of ARIMA and SES) at a 99% confidence level.

Table 5. DM statistical test result.

	SES	ARIMA	SVR	BPNN	GRNN	RBFNN	ELM	E_AVE
ARIMA	0.00811							
SVR	0.86552	0.99625						
BPNN	0.00757	0.11478	0.00005					
GRNN	0.69070	0.88984	0.45013	0.95854				
RBFNN	0.99407	0.99977	0.97985	0.99991	0.93608			
ELM	0.00782	0.12005	0.00031	0.55476	0.06843	0.00012		
E_AVE	0.00283	0.02631	0.00005	0.27089	0.05017	0.00005	0.16362	
E_ABC	0.00071 ***	0.00123 ***	0.00005 ***	0.09451 *	0.04439 **	0.00005 ***	0.01202 **	0.01955 **

Note that (*), (**), and (***) represent rejection of the null hypothesis at confidence levels of 90%, 95%, and 99%, respectively.

By focusing on the DM test in Table 5, in the last row, the *p* values represent the statistical test results of the proposed ensemble model (E_ABC) and other benchmark models. The *p* values are all less than 0.1, indicating that there is a significant difference between the forecasting results of the integrated model E_ABC and those of eight benchmark models at a 90% confidence level. For example, the *p* values in row 9, columns 2, 3, 4, and 7 are less than 0.01; hence, the proposed model is better than SES, ARIMA, SVR, and RBFNN at a 99% confidence level. Considering the forecasting accuracy of all models in Table 5, the forecasting accuracy of the integrated model E_ABC is considerably better than that of the above-mentioned eight models from the statistical perspective. Generally, the integrated model E_ABC is confirmed to exhibit good performance for energy demand forecasting according to the error metric and DM statistic test.

Except for the forecasting accuracy comparison with the benchmark models listed in Tables 4 and 5, this paper compares the experimental results with other related works [30–33,35]. As presented in Equation (24), the value of MAPE eliminates the magnitude of the predicted target and can directly reflect the prediction accuracy of the model. Table 6 presents the MAPE value of the various energy demand forecasting experiments (including different forecasting target and data set). The MAPE of the E-ABC model is the lowest compared with the others.

Table 6. MAPE comparison with other forecasting models.

Article	Model	Time Range	Target/Country	MAPE (%)
[30]	Cooperative-ACO-GA Cooperative-ACO-GA Cooperative-ACO-GA	1971–2007	Oil/Turkey Natural gas/Turkey Electricity/Turkey	0.72 1.18 0.90
[31]	BAG-SA	1990–2015	Energy/China	0.28
[32]	PSO-GA-linear PSO-GA-exponential PSO-GA-quadratic	1980–2009	Energy/China	0.45 0.95 0.29
[33]	BGA BGA-SA	2001–2014	Natural gas/Turkey	1.88 1.43
[35]	PSO Real coding GA	1980–2003	electricity consumption/Turkey	5.24 5.93
Present study	E_ABC	1978–2017	Energy/China	0.21

Note: GA indicates Genetic Algorithm; PSO denotes Particle Swarm Optimization; ACO stands for Ant Colony Optimization; BAG denotes Bat Algorithm, Gaussian Perturbations; SA indicates Simulated Annealing; BGA indicates Breeder Genetic Algorithm.

4.5. Future Energy Demand Forecasting Results

In this section, the integrated model E_ABC with a good prediction performance was applied to forecast the future energy demand of China. However, the out-of-sample information was required to forecast the energy demand from 2018 to 2022.

According to the 13th Five Year Plan proposed by the State Council, the economic growth rate will be greater than 6.5% by 2020, which is authoritative and representative to a certain extent. Thus, in this study, a GDP growth rate of 6.5% was used to calculate the GDP from 2018 to 2022. With respect to the other influencing factors, all of them belong to small samples; hence, the trend extrapolation method typically performs better than the others. In view of this, we applied the trend extrapolation and rolling regression methods to predict the data of each influencing factor for the next 5 years in this paper (Table 7).

Table 7. Prediction results of each input variable under ARIMA.

Year	GDP (10^8 Yuan)	IS (/)	ES (/)	TI (/)	UR (%)	POP (10^4)	CPI (/)
2018	900,309	0.9267	25.34	0.5281	59.71	139,991	24,602.38
2019	958,829	0.9321	25.47	0.4343	60.75	140,971	26,353.20
2020	1,021,153	0.9363	25.38	0.3322	61.88	142,022	28,016.92
2021	1,087,528	0.9428	25.11	0.2173	63.02	143,149	29,666.45
2022	1,158,217	0.9519	24.72	0.0875	64.15	144,347	31,267.32

Table 8 summarizes the forecasting results and growth rate for the proposed ensemble model (E_ABC) in this study, and Figure 9 summarizes the growth rate and energy demand, including the actual and forecasted data. The forecasting results revealed that the energy demand of China will maintain a steady growth trend. By 2022, the energy demand of China will reach 3429 million tons of the standard coal equivalent, corresponding to an increase of 9.48% compared to the energy demand

in 2017 and an average annual growth rate of 1.897%. With respect to the growth rate of the future energy demand, the forecasted results revealed a steady fluctuation trend: The growth rate increases to 3.02% in 2018 and then decreases to 1.38% from 2019 to 2022. From Figure 9, the growth rate of the energy demand is clearly leveling off gradually.

Table 8. Forecasting results of the future energy demand by the E_ABC model.

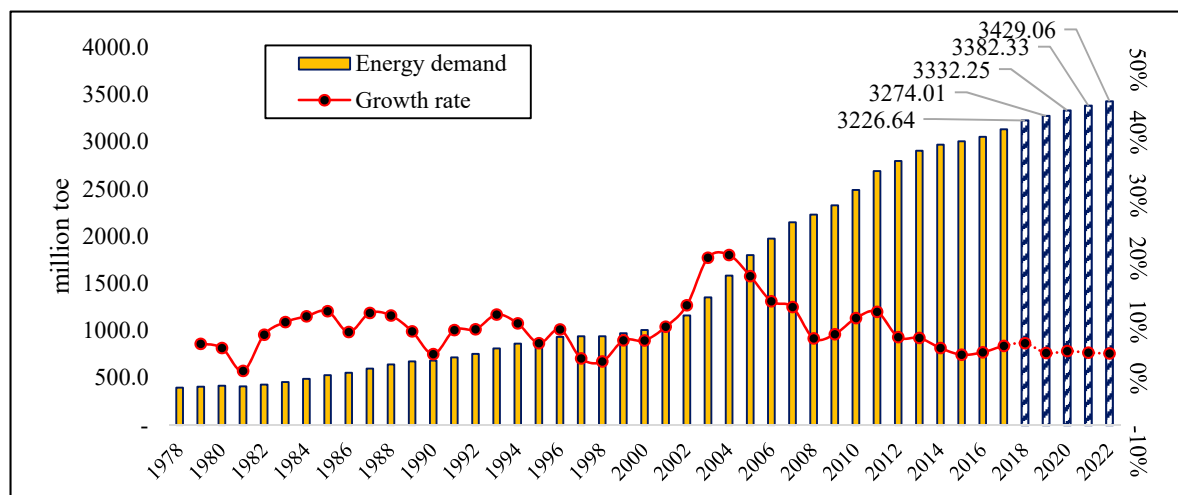| Year	Energy Demand (Million Toe)	Increase Rate
2017	3132	–
2018	3226	3.02%
2019	3274	1.47%
2020	3332	1.78%
2021	3382	1.50%
2022	3429	1.38%
Total	–	9.48%

Figure 9. Forecasting results and the growth rate of the future energy demand.

4.6. Discussion

By comparing the forecasting performance of the two categories models (single model and ensemble model), the ensemble models were verified to perform the best in forecasting energy demand, since the ensemble model achieves the highest forecasting accuracy in terms of MAPE and RMSE, which further supports that the ensemble model can be chosen as a powerful algorithm in forecasting energy demand.

When two ensemble models were compared, the results indicated that the prediction model (E_ABC) proposed herein exhibits the best prediction performance, which exhibited the lowest RMSE and MAPE. Meanwhile, the MAPE values for E_ABC and E_AVE are 0.21% and 0.51%, respectively. The MAPE of E_AVE is considerably greater than that of E_ABC. The ensemble model with the ABC algorithm (E_ABC) makes use of the prediction results of each base model to train the model and to minimize the absolute error.

Furthermore, the forecasting accuracy of the integrated model E_ABC is considerably better than that of the above-mentioned eight models from the statistical perspective. Generally, the integrated model E_ABC is confirmed to exhibit a good performance for energy demand forecasting according to the error metric and DM statistic test.

In summary, according to the above experiments analyzed in Sections 4.4 and 4.5, three main results can be drawn as follows: (1) The ensemble model established in this study is significantly superior to some other benchmark prediction models just in terms of the forecasting accuracy and

hypothesis test; (2) the proposed ensemble approach with the ABC algorithm can be employed as a promising framework for energy demand forecasting in terms of the forecasting accuracy and hypothesis test; and (3) the results for the forecasting of the future energy demand by the ensemble model revealed that the energy demand of China will maintain a steady growth trend.

5. Conclusions and Further Research

The scientific accurate prediction of energy demand is crucial for the rational formulation of energy policies, and also for the basis of ensuring the security of the energy supply. However, it is difficult to accurately predict energy demand because of the complex relationship between the energy demand and a series of influencing factors. In this context, a novel ensemble forecasting framework was proposed, which offers a complement to both traditional and data-driven forecasting techniques.

In this approach, the artificial bee colony algorithm (ABC), a type of a meta-heuristic global optimization algorithm with a random search, was adopted to integrate the monolithic forecast models. As the ABC algorithm combines the advantages of a local deep search and global-wide search, an ensemble model with better predictive performance can be obtained. In the empirical study, the performance of the proposed integrated model was better than those of the benchmark machine learning models (i.e., SES, ARIMA, SVR, ANN, and ELM) and the simple average integrated prediction model in terms of the evaluation criteria (i.e., RMSE, MAPE, and statistic test), indicating that the proposed approach can be used as a promising tool for energy demand forecasting. Besides, the results obtained for the forecasting of the future energy demand by the ensemble model (E_ABC) revealed that the energy demand of China will maintain a steady growth trend. By 2022, the energy demand of China will reach 3429 million tons of the standard coal equivalent, corresponding to an increase of 9.48% compared to the energy demand in 2017 and an average annual growth rate of 1.897%.

Although the proposed integrated model exhibits good prediction performance, there is still considerable room for improvement. Various factors are well known to affect the energy demand, not only those selected herein. If those factors can be considered in the proposed integrated model, the prediction performance may be considerably better. In addition, energy demand forecasting faces more challenges due to the data frequency and data type. Hence, the use of text mining technology to expand data types and broaden the data frequency has become the focus of our future research.

Appendix A

Table A1. Total energy demand and related factors of China between 1978 and 2017.

Year	GDP	IS	ES	TI	UR	POP	CPI	ED
Unit	10^8 Yuan	/	/	/	%	10^4	/	Million Toe
1978	3645.00	0.72	25.90	15.68	17.92	96,259.00	184.00	396.6
1979	4062.00	0.69	25.10	14.42	19.99	97,542.00	208.00	408.2
1980	4545.00	0.70	23.80	13.26	19.39	98,705.00	238.00	417.4

Table A1. *Cont.*

Year	GDP	IS	ES	TI	UR	POP	CPI	ED
Unit	10^8 Yuan	/	/	/	%	10^4	/	Million Toe
1981	4892.00	0.68	22.80	12.15	20.16	100,100.00	264.00	411.6
1982	5323.00	0.67	21.40	11.66	21.13	101,700.00	288.00	429.5
1983	5963.00	0.67	20.50	11.07	21.62	103,000.00	316.00	456.9
1984	7208.00	0.68	19.80	9.84	23.01	104,400.00	361.00	490.2
1985	9098.90	0.72	19.30	8.43	23.71	105,851.00	446.00	529.9
1986	10,376.20	0.73	19.50	7.79	24.52	107,507.00	497.00	555.3
1987	12,174.60	0.73	19.10	7.12	25.32	109,300.00	565.00	598.8
1988	15,180.40	0.74	19.10	6.13	25.81	111,026.00	714.00	643.1
1989	17,179.70	0.75	19.20	5.64	26.21	112,704.00	788.00	674.6
1990	18,872.90	0.73	18.70	5.23	26.41	114,333.00	833.00	683.2
1991	22,005.60	0.75	19.10	4.72	26.94	115,823.00	932.00	718.0
1992	27,194.50	0.78	19.40	4.01	27.46	117,171.00	1116.00	755.6
1993	35,673.20	0.80	20.10	3.25	27.99	119,517.00	1393.00	812.7
1994	48,637.50	0.80	19.30	2.52	28.51	119,850.00	1833.00	862.7
1995	61,339.90	0.80	19.30	2.14	29.04	121,121.00	1355.00	888.8
1996	71,813.60	0.80	20.50	1.88	30.48	122,389.00	2789.00	935.1
1997	79,715.00	0.82	22.20	1.70	31.91	123,626.00	3002.00	940.6
1998	85,195.50	0.82	22.60	1.60	33.35	124,761.00	3159.00	941.6
1999	90,564.40	0.84	23.50	1.55	34.78	125,786.00	3346.00	974.3
2000	100,280.10	0.85	24.40	1.45	36.22	126,743.00	3632.00	1007.9
2001	110,863.10	0.86	24.20	1.36	37.66	127,627.00	3869.00	1064.6
2002	121,717.40	0.86	24.70	1.31	39.09	128,453.00	4106.00	1161.0
2003	137,422.00	0.87	23.70	1.34	40.53	129,227.00	4411.00	1353.5
2004	161,840.20	0.87	23.80	1.32	41.76	129,988.00	4925.00	1583.8
2005	187,318.90	0.88	22.40	1.26	42.99	130,756.00	5463.00	1800.4
2006	219,438.50	0.89	22.20	1.18	44.34	131,448.00	6138.00	1974.7
2007	270,232.30	0.89	22.10	1.04	45.89	132,129.00	7081.00	2147.8
2008	319,515.50	0.89	22.00	0.91	46.99	132,802.00	8183.00	2229.0
2009	349,081.40	0.90	19.90	0.88	48.34	133,450.00	9514.00	2328.1
2010	413,030.30	0.90	21.40	0.79	49.95	134,091.00	10,919.00	2491.1
2011	489,300.60	0.90	21.40	0.71	51.27	134,735.00	13,134.00	2690.3
2012	540,367.40	0.90	21.80	0.74	52.57	135,404.00	14,699.00	2797.4
2013	595,244.40	0.91	22.40	0.70	53.73	136,072.00	16,190.00	2905.3
2014	643,974.00	0.91	23.10	0.66	54.77	136,782.00	17,778.00	2970.6
2015	689,052.10	0.92	24.20	0.62	56.10	137,462.00	19,397.00	3005.9
2016	744,127.20	0.92	24.70	0.59	57.35	138,271.00	21,228.00	3053.0
2017	827,121.70	0.92	25.12	0.54	58.52	139,008.00	22,902.00	3132

Note: Gross Domestic Product (GDP); Industrial Structure (IS); Energy Structure (ES); Technological Innovation (TI); Urbanization rate (UR); Population (Pop); Consumer Price Index (CPI); Energy Demand (ED); Ton Oil Equivalent (toe).

Table A2. The preprocessed data.

	Year	GDP	IS	ES	TI	UR	POP	CPI	ED
	1978	0.00	0.20	1.00	1.00	0.00	0.00	0.00	0.00
	1979	0.00	0.08	0.89	0.92	0.05	0.03	0.00	0.00
	1980	0.00	0.12	0.71	0.84	0.04	0.06	0.00	0.01
	1981	0.00	0.06	0.57	0.77	0.06	0.09	0.00	0.01
	1982	0.00	0.00	0.38	0.73	0.08	0.13	0.00	0.01
	1983	0.00	0.01	0.25	0.70	0.09	0.16	0.01	0.02
	1984	0.00	0.05	0.15	0.61	0.13	0.19	0.01	0.03
	1985	0.01	0.19	0.08	0.52	0.14	0.22	0.01	0.05
	1986	0.01	0.24	0.11	0.48	0.16	0.26	0.01	0.06
	1987	0.01	0.25	0.06	0.43	0.18	0.31	0.02	0.07
	1988	0.01	0.30	0.06	0.37	0.19	0.35	0.02	0.09
	1989	0.02	0.32	0.07	0.34	0.20	0.38	0.03	0.10
	1990	0.02	0.24	0.00	0.31	0.21	0.42	0.03	0.10
	1991	0.02	0.34	0.06	0.28	0.22	0.46	0.03	0.12
	1992	0.03	0.45	0.10	0.23	0.23	0.49	0.04	0.13
Training Set	1993	0.04	0.53	0.19	0.18	0.25	0.54	0.05	0.15
	1994	0.05	0.52	0.08	0.13	0.26	0.55	0.07	0.17
	1995	0.07	0.52	0.08	0.11	0.27	0.58	0.05	0.18
	1996	0.08	0.53	0.25	0.09	0.31	0.61	0.11	0.20
	1997	0.09	0.59	0.49	0.08	0.34	0.64	0.12	0.20
	1998	0.10	0.61	0.54	0.07	0.38	0.67	0.13	0.20
	1999	0.11	0.66	0.67	0.07	0.42	0.69	0.14	0.21
	2000	0.12	0.71	0.79	0.06	0.45	0.71	0.15	0.22
	2001	0.13	0.74	0.76	0.05	0.49	0.73	0.16	0.24
	2002	0.14	0.76	0.83	0.05	0.52	0.75	0.17	0.28
	2003	0.16	0.80	0.69	0.05	0.56	0.77	0.19	0.35
	2004	0.19	0.78	0.71	0.05	0.59	0.79	0.21	0.43
	2005	0.22	0.82	0.51	0.05	0.62	0.81	0.23	0.51
	2006	0.26	0.86	0.49	0.04	0.65	0.82	0.26	0.58
	2007	0.32	0.88	0.47	0.03	0.69	0.84	0.30	0.64
	2008	0.38	0.88	0.46	0.02	0.72	0.85	0.35	0.67
	2009	0.42	0.89	0.17	0.02	0.75	0.87	0.41	0.71
	2010	0.50	0.90	0.38	0.02	0.79	0.88	0.47	0.77
	2011	0.59	0.90	0.38	0.01	0.82	0.90	0.57	0.84
	2012	0.65	0.90	0.43	0.01	0.85	0.92	0.64	0.88
	2013	0.72	0.93	0.51	0.01	0.88	0.93	0.70	0.92
	2014	0.78	0.96	0.61	0.01	0.91	0.95	0.77	0.94
Testing Set	2015	0.83	0.97	0.76	0.01	0.94	0.96	0.85	0.95
	2016	0.90	0.98	0.83	0.00	0.97	0.98	0.93	0.97
	2017	1.00	1.00	0.89	0.00	1.00	1.00	1.00	1.00

Appendix B

Table A3. Pseudo-code of the ensemble forecasting model.

Pseudo-Code of the Ensemble Forecasting Model	
Input:	training data set $D = \{X,Y\}$, the number of base model T
Output:	ensemble forecasting model H
Pseudo-code	
1:	**Step 1: training the base model**
2:	For t = 1 to T do
3:	Training the forecasting model t with training data set
4:	Get the fitting value of model t,
5:	End
6:	**Step2: find the optimal weights for based models**
7:	Initialize the weights of based models $\{w_1, w_2, \ldots, w_T\}$
8:	Find the optimal base model weights by minimizing the error function J with ABC:
9:	$J = \sum \sum_{t=1}^{T} w_t \hat{Y}_t - Y$
10:	**Step3 get the ensemble forecasting model *H*:**
11:	Make predictions using the final model: $\hat{Y} = \sum_{t=1}^{T} w^* \hat{Y}_t$

Table A4. Pseudo-code of the ensemble forecasting model.

Pseudo-Code of the Artificial Bee Colony Algorithm	

Input:	int N; // the size of the bee colony
	int M; // the maximum number of cycles
	int L; // the control parameter to abandon the nectar source

Output:	set X^*; // the optimal solution set
Pseudo-code	
1:	initialize the population of solutions
2:	evaluate the population
3:	cycle = 1
4:	Repeat
5:	send employed bees
6:	calculate the probability of the current nectar sources
7:	send onlooker
8:	send scout
9:	memorize the best nectar sources position achieved currently
10:	cycle = cycle + 1
11:	until cycle = M

References

1. Le, T.-H.; Nguyen, C.P. Is energy security a driver for economic growth? Evidence from a global sample. *Energy Policy* **2019**, *129*, 436–451. [CrossRef]
2. Sharimakin, A.; Glass, A.J.; Saal, D.S.; Glass, K. Dynamic multilevel modelling of industrial energy demand in Europe. *Energy Econ.* **2018**, *74*, 120–130. [CrossRef]
3. Ji, Q.; Zhang, H.; Zhang, D. The impact of OPEC on East Asian oil import security: A multidimensional analysis. *Energy Policy* **2019**, *126*, 99–107. [CrossRef]
4. Cohen, E. Development of Israel's natural gas resources: Political, security, and economic dimensions. *Resour. Policy* **2018**, *57*, 137–146. [CrossRef]
5. Wang, Q.; Li, S.; Li, R. Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. *Energy* **2018**, *161*, 821–831. [CrossRef]

6.　Ji, Q.; Bouri, E.; Roubaud, D.; Kristoufek, L. Information interdependence among energy, cryptocurrency and major commodity markets. *Energy Econ.* **2019**, *81*, 1042–1055. [CrossRef]

7.　Hribar, R.; Potočnik, P.; Šilc, J.; Papa, G. A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy* **2019**, *167*, 511–522. [CrossRef]

8.　He, Y.; Lin, B. Forecasting China's total energy demand and its structure using ADL-MIDAS model. *Energy* **2018**, *151*, 420–429. [CrossRef]

9.　Wang, Z.-X.; Li, Q.; Pei, L.-L. Grey forecasting method of quarterly hydropower production in China based on a data grouping approach. *Appl. Math. Model.* **2017**, *51*, 302–316. [CrossRef]

10.　Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *OR* **1969**, *20*, 451–468. [CrossRef]

11.　Samuels, J.D.; Sekkel, R.M. Model Confidence Sets and forecast combination. *Int. J. Forecast.* **2017**, *33*, 48–60. [CrossRef]

12.　Hsiao, C.; Wan, S.K. Is there an optimal forecast combination? *J. Econ.* **2014**, *178*, 294–309. [CrossRef]

13.　Bedi, J.; Toshniwal, D. Deep learning framework to forecast electricity demand. *Appl. Energy* **2019**, *238*, 1312–1326. [CrossRef]

14.　Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]

15.　Wei, L.; Li, G.; Zhu, X.; Sun, X.; Li, J. Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures. *Energy Econ.* **2019**, *80*, 452–460. [CrossRef]

16.　Kim, K.; Hur, J. Weighting Factor Selection of the Ensemble Model for Improving Forecast Accuracy of Photovoltaic Generating Resources. *Energies* **2019**, *12*, 3315. [CrossRef]

17.　Cui, J.; Yu, R.; Zhao, D.; Yang, J.; Ge, W.; Zhou, X. Intelligent load pattern modeling and denoising using improved variational mode decomposition for various calendar periods. *Appl. Energy* **2019**, *247*, 480–491. [CrossRef]

18.　Ji, Q.; Zhang, D. How much does financial development contribute to renewable energy growth and upgrading of energy structure in China? *Energy Policy* **2019**, *128*, 114–124. [CrossRef]

19.　Ji, Q.; Li, J.; Sun, X. Measuring the interdependence between investor sentiment and crude oil returns: New evidence from the CFTC's disaggregated reports. *Financ. Res. Lett.* **2019**, *30*, 420–425. [CrossRef]

20.　Ma, Y.-R.; Ji, Q.; Pan, J. Oil financialization and volatility forecast: Evidence from multidimensional predictors. *J. Forecast.* **2019**, *38*, 564–581.

21.　Berk, I.; Ediger, V.Ş. Forecasting the coal production: Hubbert curve application on Turkey's lignite fields. *Resour. Policy* **2016**, *50*, 193–203. [CrossRef]

22.　Mehmanpazir, F.; Khalili-Damghani, K.; Hafezalkotob, A. Modeling steel supply and demand functions using logarithmic multiple regression analysis (case study: Steel industry in Iran). *Resour. Policy* **2019**, *63*, 101409. [CrossRef]

23.　Wang, X.; Lei, Y.; Ge, J.; Wu, S. Production forecast of China's rare earths based on the Generalized Weng model and policy recommendations. *Resour. Policy* **2015**, *43*, 11–18. [CrossRef]

24.　Wang, C.; Zhang, H.; Fan, W.; Ma, P. A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction. *Energy* **2017**, *138*, 977–990. [CrossRef]

25.　Yu, L.; Wang, S.; Lai, K.K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ.* **2008**, *30*, 2623–2635. [CrossRef]

26.　Tang, L.; Wu, Y.; Yu, L.A. A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting. *Energy* **2018**, *157*, 526–538. [CrossRef]

27.　Kim, S.; Lee, G.; Kwon, G.-Y.; Kim, D.-I.; Shin, Y.-J. Deep Learning Based on Multi-Decomposition for Short-Term Load Forecasting. *Energies* **2018**, *11*, 3433. [CrossRef]

28.　Tso, G.K.F.; Yau, K.K.W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* **2007**, *32*, 1761–1768. [CrossRef]

29.　Adom, P.K.; Bekoe, W. Conditional dynamic forecast of electrical energy consumption requirements in Ghana by 2020: A comparison of ARDL and PAM. *Energy* **2012**, *44*, 367–380. [CrossRef]

30.　Ghanbari, A.; Kazemi, S.M.R.; Mehmanpazir, F.; Nakhostin, M.M. A Cooperative Ant Colony Optimization-Genetic Algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowl.-Based Syst.* **2013**, *39*, 194–206. [CrossRef]

31.　Wu, Q.; Peng, C. A hybrid BAG-SA optimal approach to estimate energy demand of China. *Energy* **2017**, *120*, 985–995. [CrossRef]

32. Yu, S.-W.; Zhu, K.-J. A hybrid procedure for energy demand forecasting in China. *Energy* **2012**, *37*, 396–404. [CrossRef]

33. Karadede, Y.; Ozdemir, G.; Aydemir, E. Breeder hybrid algorithm approach for natural gas demand forecasting model. *Energy* **2017**, *141*, 1269–1284. [CrossRef]

34. Angelopoulos, D.; Siskos, Y.; Psarras, J. Disaggregating time series on multiple criteria for robust forecasting: The case of long-term electricity demand in Greece. *Eur. J. Oper. Res.* **2019**, *275*, 252–265. [CrossRef]

35. Piltan, M.; Shiri, H.; Ghaderi, S.F. Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms. *Energy Convers. Manag.* **2012**, *58*, 1–9. [CrossRef]

36. Sonmez, M.; Akgüngör, A.P.; Bektaş, S. Estimating transportation energy demand in Turkey using the artificial bee colony algorithm. *Energy* **2017**, *122*, 301–310. [CrossRef]

37. Yuan, X.-C.; Sun, X.; Zhao, W.; Mi, Z.; Wang, B.; Wei, Y.-M. Forecasting China's regional energy demand by 2030: A Bayesian approach. *Resour. Conserv. Recycl.* **2017**, *127*, 85–95. [CrossRef]

38. He, Y.; Zheng, Y.; Xu, Q. Forecasting energy consumption in Anhui province of China through two Box-Cox transformation quantile regression probability density methods. *Measurement* **2019**, *136*, 579–593. [CrossRef]

39. He, Y.X.; Liu, Y.Y.; Xia, T.; Zhou, B. Estimation of demand response to energy price signals in energy consumption behaviour in Beijing, China. *Energy Convers. Manag.* **2014**, *80*, 429–435. [CrossRef]

40. Forouzanfar, M.; Doustmohammadi, A.; Hasanzadeh, S.; Shakouri, G.H. Transport energy demand forecast using multi-level genetic programming. *Appl. Energy* **2012**, *91*, 496–503. [CrossRef]

41. Liao, H.; Cai, J.-W.; Yang, D.-W.; Wei, Y.-M. Why did the historical energy forecasting succeed or fail? A case study on IEA's projection. *Technol. Forecast. Soc. Chang.* **2016**, *107*, 90–96. [CrossRef]

42. Ahmad, T.; Chen, H. Deep learning for multi-scale smart energy forecasting. *Energy* **2019**, *175*, 98–112. [CrossRef]

43. Schaer, O.; Kourentzes, N.; Fildes, R. Demand forecasting with user-generated online information. *Int. J. Forecast.* **2019**, *35*, 197–212. [CrossRef]

44. Xu, S.; Chan, H.K.; Zhang, T. Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *122*, 169–180. [CrossRef]

45. Ünler, A. Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025. *Energy Policy* **2008**, *36*, 1937–1944. [CrossRef]

46. Debnath, K.B.; Mourshed, M. Forecasting methods in energy planning models. *Renew. Sustain. Energy Rev.* **2018**, *88*, 297–325. [CrossRef]

47. Yuan, X.; Chen, C.; Yuan, Y.; Huang, Y.; Tan, Q. Short-term wind power prediction based on LSSVM–GSA model. *Energy Convers. Manag.* **2015**, *101*, 393–401. [CrossRef]

48. Zendehboudi, A.; Baseer, M.A.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. *J. Clean. Prod.* **2018**, *199*, 272–285. [CrossRef]

49. Wei, D.; Wang, J.; Ni, K.; Tang, G. Research and Application of a Novel Hybrid Model Based on a Deep Neural Network Combined with Fuzzy Time Series for Energy Forecasting. *Energies* **2019**, *12*, 3588. [CrossRef]

50. Law, A.; Ghosh, A. Multi-label classification using a cascade of stacked autoencoder and extreme learning machines. *Neurocomputing* **2019**, *358*, 222–234. [CrossRef]

51. Bergmeir, C.; Hyndman, R.J.; Benitez, J.M. Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *Int. J. Forecast.* **2016**, *32*, 303–312. [CrossRef]

52. Han, X.; Li, R. Comparison of Forecasting Energy Consumption in East Africa Using the MGM, NMGM, MGM-ARIMA, and NMGM-ARIMA Model. *Energies* **2019**, *12*, 3278. [CrossRef]

53. Zhao, W.; Zhao, J.; Yao, X.; Jin, Z.; Wang, P. A Novel Adaptive Intelligent Ensemble Model for Forecasting Primary Energy Demand. *Energies* **2019**, *12*, 1347. [CrossRef]

54. Galicia, A.; Talavera-Llames, R.; Troncoso, A.; Koprinska, I.; Martínez-Álvarez, F. Multi-step forecasting for big data time series based on ensemble learning. *Knowl.-Based Syst.* **2019**, *163*, 830–841. [CrossRef]

55. Liu, L.; Zong, H.; Zhao, E.; Chen, C.; Wang, J. Can China realize its carbon emission reduction goal in 2020: From the perspective of thermal power development. *Appl. Energy* **2014**, *124*, 199–212. [CrossRef]

56. Wang, J.-J.; Wang, J.-Z.; Zhang, Z.-G.; Guo, S.-P. Stock index forecasting based on a hybrid model. *Omega* **2012**, *40*, 758–766. [CrossRef]

57. Zhu, B.; Wei, Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega* **2013**, *41*, 517–524. [CrossRef]

58. Qu, Z.; Zhang, K.; Mao, W.; Wang, J.; Liu, C.; Zhang, W. Research and application of ensemble forecasting based on a novel multi-objective optimization algorithm for wind-speed forecasting. *Energy Convers. Manag.* **2017**, *154*, 440–454. [CrossRef]

59. Xue, Y.; Jiang, J.; Zhao, B.; Ma, T. A self-adaptive artificial bee colony algorithm based on global best for global optimization. *Soft Comput.* **2017**, *22*, 2935–2952. [CrossRef]

60. Xu, X.; Wei, Z.; Ji, Q.; Wang, C.; Gao, G. Global renewable energy development: Influencing factors, trend predictions and countermeasures. *Resour. Policy* **2019**, *63*, 101470. [CrossRef]

61. Shi, Y.; Tian, Y.; Kou, G.; Peng, Y.; Li, J. *Optimization Based Data Mining Theory and Applications*; Springer: London, UK, 2011.

62. Wang, C.; Zhang, H.; Fan, W.; Fan, X. A new wind power prediction method based on chaotic theory and Bernstein Neural Network. *Energy* **2016**, *117*, 259–271. [CrossRef]

63. Qureshi, A.S.; Khan, A.; Zameer, A.; Usman, A. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* **2017**, *58*, 742–755. [CrossRef]

64. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

65. Pérez, C.J.; Vega-Rodríguez, M.A.; Reder, K.; Flörke, M. A Multi-Objective Artificial Bee Colony-based optimization approach to design water quality monitoring networks in river basins. *J. Clean. Prod.* **2017**, *166*, 579–589. [CrossRef]

66. Lin, Y.; Jia, H.; Yang, Y.; Tian, G.; Tao, F.; Ling, L. An improved artificial bee colony for facility location allocation problem of end-of-life vehicles recovery network. *J. Clean. Prod.* **2018**, *205*, 134–144. [CrossRef]

67. Zhao, Y.; Li, J.; Yu, L. A deep learning ensemble approach for crude oil price forecasting. *Energy Econ.* **2017**, *66*, 9–16. [CrossRef]

68. Peimankar, A.; Weddell, S.J.; Jalal, T.; Lapthorn, A.C. Multi-objective ensemble forecasting with an application to power transformers. *Appl. Soft Comput.* **2018**, *68*, 233–248. [CrossRef]

69. Godarzi, A.A.; Amiri, R.M.; Talaei, A.; Jamasb, T. Predicting oil price movements: A dynamic Artificial Neural Network approach. *Energy Policy* **2014**, *68*, 371–382. [CrossRef]

70. Yu, L.; Zhao, Y.; Tang, L. A compressed sensing based AI learning paradigm for crude oil price forecasting. *Energy Econ.* **2014**, *46*, 236–245. [CrossRef]