A Novel Adaptive Intelligent Ensemble Model for Forecasting Primary Energy Demand

Wenting Zhao 1, Juanjuan Zhao 1,2,*, Xilong Yao 1, Zhixin Jin 1,3 and Pan Wang 2

1 School of Economic and Management, Taiyuan University of Technology, Jinzhong 030600, China; zhaowenting0094@link.tyut.edu.cn (W.Z.); xilongyao@163.com (X.Y.); jzxin88@163.com (Z.J.)
2 School of Information and Computer, Taiyuan University of Technology, Jinzhong 030600, China; wp_wangpan_wp@163.com
3 Shanxi Coking Coal Group Co. Ltd., Taiyuan 030024, China
* Correspondence: zhaojuanjuan@tyut.edu.cn; Tel.: +86-186-3666-7123

Received: 12 March 2019; Accepted: 2 April 2019; Published: 8 April 2019

Abstract: Effectively forecasting energy demand and energy structure helps energy planning departments formulate energy development plans and react to the opportunities and challenges in changing energy demands. In view of the fact that the rolling grey model (RGM) can weaken the randomness of small samples and better present their characteristics, as well as support vector regression (SVR) having good generalization, we propose an ensemble model based on RGM and SVR. Then, the inertia weight of particle swarm optimization (PSO) is adjusted to improve the global search ability of PSO, and the improved PSO algorithm (APSO) is used to assign the adaptive weight to the ensemble model. Finally, in order to solve the problem of accurately predicting the time-series of primary energy consumption, an adaptive inertial weight ensemble model (APSO-RGM-SVR) based on RGM and SVR is constructed. The proposed model can show higher prediction accuracy and better generalization in theory. Experimental results also revealed outperformance of APSO-RGM-SVR compared to single models and unoptimized ensemble models by about 85% and 32%, respectively. In addition, this paper used this new model to forecast China’s primary energy demand and energy structure.

Keywords: forecasting; primary energy demand; ensemble model; improved particle swarm optimization algorithm

1. Introduction

Ensuring adequate energy supply is an important material basis for economic construction and development in most countries of the world. Especially for developing countries which are in the process of transformation and economic restructuring, energy supply and demand and structural adjustment are the keys to their transformation [1]. Estimating energy demand will not only affect energy market allocation [2], energy policy adjustment [3], economic development [4], carbon emission prediction [5], and other related energy planning research, it can also serve as the basis for building energy planning models [6]. Therefore, accurately forecasting energy demand can provide effective advice for energy policy-makers on energy development planning and policy formulation. It also helps to avoid imbalances in energy supply and demand and ensures energy security in the region and worldwide. In addition, adjustment and transformation of the energy structure directly affect environmental protection and governance in the region and worldwide [7]. In view of this, accurately forecasting primary energy demand and energy structure is important for energy planning.

Many methods have been presented in the field of prediction. In short, energy forecasting models fall into three categories: time-series models, soft-computing technologies, and the ensemble of
models [8,9]. Time-series models are the most widely used in energy prediction, including regression models, the autoregressive integrated moving average model (ARIMA), and the grey model (GM), which are commonly used for energy demand forecasting [10,11]. The construction of regression models requires rigorous assumptions, which leads to poor generalization. In recent years, ARIMA and GM have performed better in the forecasting of small samples [11,12]. With the development of artificial intelligence (AI), soft-computing technology has gradually been applied to energy demand and power load forecasting [13]. This kind of method has high prediction accuracy, but most of the models have strict requirements on training samples. For remedying the shortcomings of single forecasting models, scholars began to develop ensemble models and applied them to various types of energy consumption and load forecasting research. Ensemble models extract the characteristics of every single model for improving forecast precision [14,15]. At present, the single-model features of most ensemble models are extracted by fixed weights. In comparison with the single model, the precision of prediction has been improved. However, for the ensemble forecasting model with a fixed weight, the feature extraction of a single model cannot be adjusted in real-time, resulting in poor generalization ability. In view of the above, although many forecasting methods have been put forward, more accurate and optimized forecasting of energy demand remains a key issue for energy management departments, because forecasting energy demand is the basis of energy planning and policy adjustment for a region or even a country. There are not many statistical data on primary energy consumption. Especially for primary electricity consumption, only about a dozen years of data are available, which is clearly a small sample. In addition, the change in energy consumption is volatile. The use of ARIMA [12], the first-order single variable grey model (GM (1,1)) [16], and rolling grey model (RGM) [17] in time-series models perform better for forecasting small-size samples [8,9]. In addition, support vector regression (SVR) in artificial intelligence methods has good generalization [18]. Therefore, considering the characteristics of primary energy consumption, this paper combines ARIMA, GM (1,1), and RGM with SVR to construct ensemble models. Then, we improve the global search ability of particle swarm optimization (PSO) by improving the inertia weight, and then optimize the ensemble model with the improved adaptive inertia weight particle swarm optimization algorithm (APSO). Thus, an adaptive variable weight ensemble forecasting model (APSO-RGM-SVR) is innovatively proposed to accurately forecast primary energy demand and energy structure. Accurately forecasting primary energy demand can provide reference and evidence for energy–environment–economic planning and policy-making. It can also help to maintain the balance of energy supply and demand, and ensure energy security and stability. The contribution of this paper mainly includes two aspects:

1. The APSO algorithm based on adaptive inertia weight is proposed. The inertial weight of PSO is an important factor affecting global search ability. In this paper, an adaptive adjustment method of inertia weight was used to improve the PSO algorithm. A “momentum” factor was added to PSO to adjust the adaptive inertia weight of the particles so as to improve the global search ability of PSO.

2. An adaptive variable weight ensemble forecasting model (APSO-RGM-SVR) is proposed to predict primary energy demand and energy structure. This paper combines the GM (1,1), RGM, and ARIMA in time-series model with SVR in machine learning method, and constructs GM (1,1)-SVR, RGM-SVR, and ARIMA-SVR ensemble models based on the standard deviation weight method. Through comparative experiments, we found RGM-SVR had higher precision. Then, the RGM-SVR model with higher accuracy was optimized by the APSO algorithm, and the APSO-RGM-SVR model was obtained. The new method proposed in this paper was used to forecast China’s primary energy demand and structure. The empirical results showed that the new method had higher accuracy and better generalization.

The remaining paper is structured as follows. Section 2 reviews the related work. Section 3 presents the single forecasting models involved in this paper, including GM (1,1), RGM, ARIMA, and SVR. Section 4 presents ensemble models, including ensemble models with the standard deviation
method and the adaptive inertial weight ensemble forecasting model. Section 5 presents the results. Section 6 presents the conclusion.

2. Literature Review

Forecasting energy demand has always been the focus of authors in the study of energy. The key to improving the forecasting accuracy is to construct a reasonable and appropriate energy demand forecasting model according to different types of forecasting objects and targets [8]. Over the past decade, the energy consumption index has shown a globally increasing trend, and its actual situation is more complex. More new technologies will be developed for energy forecasting.

2.1. Time-Series Forecasting Model

The time-series forecasting model is a kind of traditional forecasting method which was put forward early. The popular methods include regression model [9,19], ARIMA [12], and the grey theory model [16,17]. The regression model can forecast the changes in the dependent variable by establishing the regression equation through the correlation between the dependent variable and the independent variable. The regression model was the earliest method applied to energy management [19,20]. Bianco et al. (2009) established a regression model to predict power consumption based on data of electricity consumption [20]. The construction of the regression model uses strict assumptions, which leads to poor generalization [9,12].

The ARIMA is a forecasting method that was developed by Box and Jenkins [19]. The principle of ARIMA is to treat the data sequence formed by the prediction object over time as a random sequence, and then establish a model to express the sequence [12,19]. This method is often used in energy consumption and price forecasting [21,22]. Contreras et al. (2003) used ARIMA to predict the power price in the Spanish mainland and the California market [21]. Abdel-Aal et al. (2014) forecasted power load using ARIMA [22]. Building the ARIMA model does not require the changes in random variables to be considered directly, but this method is not suitable for non-linear forecasting [23].

In time-series model research, the grey model has been more widely used in recent decades. It was proposed by Julong Deng in 1982 [16]. It can use small sample data to establish the grey differential equation and describe the development trend of things for a long time [16]. Prediction methods based on grey theory are often used to predict energy consumption [24,25]. Zhao and Guo (2016) used the RGM to predict power load [24]. Tsai et al. (2017) used three grey models to forecast the growth trend of renewable energy [25]. However, GM (1,1) has low prediction accuracy for samples with complex backgrounds and seasonal variation characteristics, and it is not good at long-term prediction, although the model has the advantages of a simple model structure and high accuracy for short-term predictions. At present, scholars generally optimize the grey prediction model or combine it with other methods to improve the accuracy and generalization of GM [26].

2.2. Soft-Computing Technology

With the progress of AI, soft-computing technology based on AI has gradually entered the field of energy management [27]. Compared with traditional time-series prediction methods, soft-computing technology can cope with more complex changes and has a higher fitting accuracy [28].

The artificial neural network (ANN) has been a hotspot since the 1980s [29]. It imitates human brain neurons from the perspective of information processing, and forms different networks according to different connection modes, and establishes models. In later years, ANN has been gradually applied to predict energy consumption and power load [29,30]. Ekonomou (2010) used ANN to predict energy demand [29]. Azadeh et al. (2013) proposed ANN based on multi-layer perceptron (MLP) training and to test data to forecast Iran’s energy consumption [30].

Deep learning was proposed by Hinton et al. in 2006 [31]. As the arrival of the era of big data, deep learning has been greatly developed. Scholars have proposed several time-series prediction methods based on in-depth learning, including Long Short-Term Memory (LSTM) [32,33], the Recurrent Neural
Network (RNN) [34,35], and other prediction methods based on deep learning [36,37], which are used to forecast photovoltaic power generation, energy demand, and power load [36,38–41]. Liu et al. (2016) established the Gated Recurrent Unit prediction model (GRU) and forecasted China’s primary energy demand [40]. Lee and Kim (2019) used ANN, DNN, and LSTM to forecast photovoltaic power combined with meteorological information [41]. However, ANN and time-series prediction models based on deep learning have better performance in the training of large samples. Compared with ANN and deep learning, SVR has a good generalization ability and low-computational complexity.

The SVR model improves the generalization of learning machines by seeking structural risk minimization, so as to obtain good statistical rules in the case of fewer statistical samples [18]. This method has been widely used in financial market forecasting [42], weather forecasting [43], energy consumption and power load forecasting [44,45], etc. Jain et al. (2014) used SVR to construct sensor-based forecasting models to forecast energy consumption in New York buildings [44]. Chen et al. (2017) constructed SVR to forecast the power load based on the demand response [45]. The SVR model structure is simple, and the prediction accuracy is high. However, the model has better performance for short-term prediction. At present, more scholars choose to optimize SVR or combine it with other models to improve its generalization [46,47]. The applicability, advantages, and disadvantages of some single prediction models are listed in Table 1.

Table 1. Summary of several models commonly used for energy prediction.

<table>
<thead>
<tr>
<th>Models</th>
<th>Feature</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Applied to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>Establishing the regression equation through the correlation between variables.</td>
<td>The correlation degree between the factors can be analyzed.</td>
<td>Poor generalization and low accuracy.</td>
<td>Finding the relationship between energy consumption and economic growth [4]. forecasting long-term electricity consumption [20]. Forecasting the next-day electricity price in the Spanish mainland [21]. forecasting monthly electricity consumption in Eastern Saudi Arabia [22]. Forecasting annual power load in Shanghai, China [14]. forecasting growth trends in renewable energy [25]. Forecasting long-term energy demand in Greece [29]. forecasting Iran’s future monthly energy consumption [30]. Forecasting photovoltaic power generation [33,35]. Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
<tr>
<td>ARIMA</td>
<td>Treating the data sequence formed by the prediction object over time as a random sequence, and then establishing a model to express the sequence.</td>
<td>The model is simple and only requires endogenous variables.</td>
<td>Requirements are stable time-series data or are stable after differentiation, essentially only captures linear relationships, not non-linear relationships. Cannot consider the relationship between factors; medium- and long-term prediction error is relatively large. More model parameters; learning time is too long; large sample size required for model training.</td>
<td>Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
<tr>
<td>Grey</td>
<td>Using small sample data to establish a grey differential equation and describing the development trend of things over a long time.</td>
<td>Less sample size required; short-term prediction accuracy is higher; less model parameters.</td>
<td>Can fully approximate complex non-linear relationships; parallel distributed processing; highly robust and fault tolerant. More model parameters; learning time is too long; large sample size required for model training.</td>
<td>Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
<tr>
<td>ANN</td>
<td>A highly complex non-linear dynamic learning system; suitable for handling inaccurate and fuzzy information; processing problems under various factors and conditions.</td>
<td>Can fully approximate complex non-linear relationships; parallel distributed processing; highly robust and fault tolerant.</td>
<td>The training time of the model is a little long.</td>
<td>Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long-term save input; identifying useful information.</td>
<td>Suitable for dealing with problems highly related to time-series; solves long sequence dependency problems.</td>
<td>The training time of the model is a little long.</td>
<td>Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
<tr>
<td>SVR</td>
<td>Application of Support Vector in Regression Function; maps data to a high-dimensional feature space through a non-linear mapping and performs regression in this space.</td>
<td>Small samples; simplifies regression problems; high flexibility.</td>
<td>With the increase of sample size, the time complexity of model training will increase.</td>
<td>Forecasting buildings’ energy consumption in New York [14]. forecasting the electrical load of four typical office buildings [43].</td>
</tr>
</tbody>
</table>

2.3. Optimization Algorithm

The intelligent optimization algorithm is an optimization algorithm with global optimization performance and strong versatility that is suitable for parallel computing [48]. In 1995, Kennedy and Eberhart [49] proposed the PSO algorithm, which represents significant progress in the research of the meta-heuristic optimization algorithm. The PSO algorithm starts from a random solution and finds the optimal solution through iteration. This algorithm has fewer parameters that need to be adjusted and the structure is simple. Thus, it is widely used in parameter optimization, engineering application, and stochastic optimization [50]. However, PSO also has the potential to fall into a local optimum.
With the continuous evolution of optimization algorithms and the increasing complexity of optimization problems, more and more scholars are choosing to improve PSO for better application [51,52].

In recent years, some new swarm intelligence optimization algorithms have been proposed, including gray wolf optimization (GWO), moth-flame optimization (MFO), the whale optimization algorithm (WOA), etc. The GWO algorithm is an optimization method that imitates the hunting behavior of gray wolves [53]. The GWO algorithm has the characteristics of simple structure and good global performance, but its local search ability is poor [54]. The MFO algorithm is an optimization algorithm for simulating the flight characteristics of moths, which was proposed by Mirjalili in 2015 [55]. Although MFO has fast convergence speed and high accuracy, this method also has the problem of being easily trapped in local extreme points [56]. In 2016, the WOA was proposed by Mirjalili [57], and it is often used to solve large-scale, complex optimization problems [58]. Easily falling into local extremum is the common characteristic of most optimization models, and this represents a challenge and a goal for scholars trying to improve optimization models.

2.4. Ensemble-Based Methods

Energy management is affected by energy policy, population, consumption, and economic development, and the change in energy consumption is unstable. The single forecasting model can only extract some features of energy consumption time-series, and it is difficult to obtain a more perfect prediction effect. In later years, many scholars have chosen to adopt an ensemble model approach. More specifically, multiple single-prediction models have been established, and then these single models are combined. This combination can extract the advantages of each method and achieve more accurate prediction results [15,59,60]. Li et al. (2017) forecasted power demand based on a hybrid approach [59]. Xiao et al. (2018) combined with the grouping method of data processing selection set to construct an ensemble prediction model, which is mainly used to forecast the non-linear changes in energy consumption [60].

The ensemble model extracts the advantages of each individual model and has a higher prediction accuracy. Therefore, for constructing an adaptive ensemble model with stronger generalization ability and higher forecasting accuracy, determining the combined weight of every single model is the focus of the ensemble model [61,62]. The linear combination of predictive models can improve the predictive power of a single model, but the flexibility in assigning weights is relatively poor, thus affecting the adaptability of the ensemble model [63,64]. Liu et al. (2014) adjusted the weight of the ensemble model by PSO and used the optimized model to forecast carbon emissions [63]. Li et al. (2018) successfully developed an ensemble prediction model for short-term wind speed prediction using the variable weighted combination theory [64]. Thus, while improving the prediction accuracy of the model, the adaptability and flexibility of the enhanced model are also issues that deserve attention from scholars [65].

3. Methodology

The GM (1,1) [16], RGM [17], and ARIMA [12] in a time-series model have better performance in forecasting small-sized samples [8,9]. In addition, SVR in an AI method has good generalization ability [18]. In view of the small sample size of primary energy annual data, we combined ARIMA, GM (1,1), and RGM with SVR to construct three ensemble forecasting models based on standard deviation. Then, the global search ability of PSO was enhanced by improving the inertial weight, and the APSO algorithm was used to optimize the ensemble model with higher accuracy. Thus, an adaptive ensemble forecasting model with high accuracy and good generalization was constructed. The overall framework of this study is shown in Figure 1. In this section, we introduce the modeling principles and methods. We first introduce the modeling principles and processes of the four models: GM (1,1), RGM, ARIMA, and SVR in detail.
3.1. GM (1,1)

The basic idea of GM is that the original sequence \(x^{(0)}\) is composed of the original data, and the sequence \(x^{(1)}\) is generated by the method of cumulative generation \[19\]. GM can weaken the randomness of the original data and make them show more obvious characteristics. GM is a differential equation model that is used to generate a transformed sequence \(x^{(1)}\).

GM (1,1) represents the first-order differential equation model with one variable. Specific steps are as follows:

Step 1. Assume \(x^{(0)}\) is the original sequence:

\[
x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}
\]  

(1)

where \(n\) is the number of years observed. Perform an accumulation of the original sequence \(x^{(0)}\), recorded as the first-order accumulated generating sequence (1-AGO), and generate a sequence:

\[
x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\}
\]  

(2)

where,

\[
x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \ldots, n
\]  

(3)

Step 2. Assume \(x^{(1)}\) is continuously differentiable and satisfies the first-order linear differential equation:

\[
\begin{cases}
\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \\
x^{(1)}(t)|_{t=0} = x^{(0)}(1)
\end{cases}
\]  

(4)

where \(a\) and \(u\) are pending parameters.

Step 3. Discretize Equation (4), and the differential equation becomes a difference equation of the form

\[
x^{(0)}(t) + ax^{(1)}(t) = u
\]  

(5)

where \(z^{(1)}(t) = \{z^{(1)}(1), z^{(1)}(2), \ldots, z^{(1)}(n)\}\) is a sequence of background values constructed by \(x^{(1)}(t)\), namely,

\[
z^{(1)}(t) = \lambda x^{(1)}(t-1) + (1 - \lambda)x^{(1)}(t), t = 2, 3, \ldots, n, \lambda \in [0, 1]
\]  

(6)

where \(\lambda = 0.5\).

Step 4. Solve for the parameters \(a\) and \(u\) in Equation (4) using the least squares method:

\[
\hat{a} = [a, u]^T = (B^T, B)^{-1}B^T Y_N
\]  

(7)
where

$$B = \begin{bmatrix} -z^{(1)(2)} & 1 \\ -z^{(1)(3)} & 1 \\ \vdots & \vdots \\ -z^{(1)(n)} & 1 \end{bmatrix}, \quad Y_N = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

(8)

Step 5. Establish a prediction formula. Solve Equation (8) after finding $a$ and $u$:

$$\hat{x}^{(1)}(t) = \left( x^{(0)}(1) - \frac{u}{a} \right) e^{-a(t-1)} + \frac{u}{a}$$

(9)

Perform a subtractive generation on Equation (9) to get the prediction formula:

$$\hat{x}^{(0)}(t) = \left( x^{(0)}(1) - \frac{u}{a} \right) (1 - e^a) + e^{-a(t-1)}$$

(10)

3.2. RGM

In theory, GM (1,1) is a continuous-time function that can continue from the initial value to any moment in the future. As time goes by, some disturbances in the future may affect the system. The farther away the future is, the larger the grey interval of the predicted value. Therefore, for GM (1,1), the meaningful prediction data are only a few data points after $x^{(0)}(n)$, and other data can only represent the planning data under the premise that the current conditions are stable. For improving the precision of the predicted data, new information has to be constantly added while making full use of the known information [17].

Therefore, an improved prediction method is proposed: GM (1,1) is established based on the known series, a grey value is predicted, and then the predicted value is added to the known series to form a new sequence of information. GM (1,1) is established based on the known series, a grey value is predicted once more, and then the predicted value is added to the known series again to form a new sequence of information. Whenever a new datum is added, GM (1,1) is created. Because the previous data are more and more unable to reflect the new situation, each time a new datum is added, the oldest datum is eliminated. Therefore, the dimensions of the sequence are unchanged, and then GM (1,1) is constructed to complete the prediction target. For each step of prediction, the parameters are corrected once, the model is improved, and the predicted values are generated dynamically. This prediction method is called the “grey model with the rolling mechanism” (RGM). Specific steps are as follows:

Step 1. Construct GM (1,1) by using the original sequence $x^{(0)}_0 = (x^{(0)}(1), x^{(0)}(0), \cdots, x^{(0)}(n))$.

Step 2. Forecast a new datum as $\bar{x}(n+1)$, add it to the original data $x^{(0)}_0$, and remove the oldest original datum $x^{(0)}(1)$ to keep the original data dimension unchanged, the new sequence is recorded as $x^{(0)}_1$.

Step 3. Establish GM (1,1) with $x^{(0)}_1 = (x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n), \bar{x}_1(n+1))$.

Step 4. Forecast the next value, denoted as $\bar{x}_2(n+1)$. Add it to the original datum $x^{(0)}_1$ and remove one of the oldest raw datum $x^{(0)}(2)$ to form a new sequence: $x^{(0)}_2$.

Remove the old and add the new, then add them one by one, until the prediction target is achieved, or a certain accuracy requirement is met. At each prediction step, the parameters are modified, and the model is improved constantly.

3.3. ARIMA

The digital characteristics of a non-stationary sequence, such as the mean, variance, and covariance, change with time. In 1970, Box and Jenkins [19] proposed a time-series analysis method based on stochastic theory. The basic models include three types [23]: the autoregressive model (AR ($p$)), the moving average model (MA ($q$)), and the ARIMA. The ARIMA carries out the disturbance analysis
by using AR, single integer, and MA. It describes the change in the time-series by the extrapolation mechanism, and comprehensively takes into account the previous data, current data, and error of the prediction variable, to improve the prediction accuracy.

For a single sequence $Y_t$, the non-stationary sequence can be translated into a stationary sequence $X_t$ using the $d$-order difference so that the stationary sequence $X_t$ is fitted to autoregressive moving average model (ARMA $(p, q)$).

The ARMA $(p, q)$ consists of AR $(p)$ and MA $(q)$. The AR $(p)$ is an autoregressive model with order $p$.

\[ X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + u_t \]  

where $\phi_1, \cdots, \phi_p$ are parameters, $c$ is a constant, and the random variable $u_t$ is a white-noise sequence.

MA $(q)$ is a moving average model of order $q$ with the expression:

\[ X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \]

where $\theta_1, \cdots, \theta_q$ are parameters, $\mu$ is the expected value of $X_t$ (usually assumed to be 0), $\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-q}$ are the error terms of the white-noise sequence.

The expression of the ARMA $(p, q)$ model is:

\[ X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \]  

The ARMA $(p, q)$ is changed into ARIMA $(p, d, q)$ by $d$-order differential transformation. In ARIMA $(p, d, q)$, the parameter $p$ is the order of AR $(p)$, $d$ is the difference, and $q$ is the order of MA $(q)$.

The construction of ARIMA $(p, d, q)$ and the determination of the parameters are summarized as follows: (i) First of all, it should be examined whether the time-series is stable. A smooth time-series is the basic precondition for constructing this model. For unstable time-series, the $d$-order difference should be made before the model is constructed to make it stationary. (ii) Then the main parameters of the model need to be determined. The $d$-value can be obtained by the difference, and then the $q$-value and the $p$-value can be obtained by the autocorrelation function and the partial correlation function. (iii) After the model parameters are determined, the saliency of the model is tested to determine the rationality of each parameter and to verify whether the model’s settings conform to the specifications. (iv) The model is trained with the sample, the sample data are fitted, and ARIMA $(p, d, q)$ is found.

### 3.4. SVR

The basic principle of SVR is to map the input factor $x$ to the high-dimensional feature space through the non-linear mapping $\varphi$, and then linear regression is performed in this high-dimensional space to find the function $f(x)$ [44]. The modeling steps are as follows.

Step 1. Given a set of numbers $G = \{(x_i, y_i)\}_{i=1}^n$, with $x_i$ the input factor, $y_i$ the expected value, and $n$ is the amount of data, then the expression is found as:

\[ f(x_i) = \omega^T \varphi(x_i) + b \]  

where, $\omega$ is the weight and $b$ is the bias.

Step 2. By means of the principle of structural risk minimization, the parameters $\omega$ and $b$ are estimated:

\[
\begin{align*}
\min f &= \frac{1}{2}||w||^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
\text{s.t.} & \quad y_i - f(x_i) \leq \epsilon + \xi_i \\
& \quad f(x_i) - y_i \leq \epsilon + \xi_i^* \\
& \quad \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2 \ldots n
\end{align*}
\]
where $\|w\|_2$ is a confidence risk, $C$ is a penalty parameter, $\varepsilon$ is an insensitive coefficient, and $\xi^*_i$ and $\xi_i$ are slack variables.

Step 3. For convenience, Equation (15) is transformed into a dual problem, and SVR function is obtained:

$$f(x_i) = \sum_{i=1}^{n} (\alpha_i - \alpha^*_i) K(X_i, X) + b$$

where $\alpha_i$ and $\alpha^*_i$ are support vector parameters, and $K(X_i, X)$ is an inner product function.

Step 4. Since the radial basis function (RBF) is a good general kernel and can achieve non-linear projection [66], according to the Mercer condition, the kernel function is defined, and RBF is selected:

$$K(X_i, X) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$$

By substituting Equation (17) into Equation (16), we obtain the following equivalent Equation (18):

$$f(x_i) = \sum_{i=1}^{n} \alpha_i \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) + b$$

where $\sigma$ is a hyper-parameter of the RBF kernel that defines the feature-length scale of the similarity between learning samples, and $x$ is the center of the kernel function.

Equation (18) is operated to obtain parameters $\alpha_i$ and $b$, thereby obtaining an energy demand prediction model.

3.5. Model Evaluation Criteria

Three criteria are used to evaluate the performance of a forecasting algorithm: Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Mean Square Percent Error (MPSE) which are defined as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i| = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|e_i|}{x_i} = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{x_i - \hat{x}_i}{x_i}\right|$$

$$MSPE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (\frac{e_i}{x_i})^2} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (\frac{x_i - \hat{x}_i}{x_i})^2}$$

where, $e_i$ represents the error values, $x_i$ represents the sample data, $\hat{x}_i$ represents the prediction data, and $n$ is the sample size.

The accuracy examination of each forecasting algorithm is determined using the appropriate tests.

4. Construction of Ensemble Forecasting Model

In the process of constructing the ensemble forecasting model, the most important thing is to extract the effective forecasting results of the single forecasting model, that is, to determine the weight of each single forecasting model. In this paper, GM (1,1), RGM, and ARIMA are combined with SVR, respectively, and the weights of the ensemble prediction model are determined based on the standard deviation method to forecast China’s primary energy demand. Then, we propose adding the momentum factor into PSO to improve the global search ability of PSO, and compare with PSO, GWO, MFO, and WOA to test the optimization accuracy of the improved PSO (APSO). Finally, the cost function is constructed by the sum of the absolute values of the prediction errors, and RGM-SVR is optimized by the improved PSO to achieve the best combination; thus, the adaptive weighted ensemble prediction model (APSO-RGM-SVR) is constructed to forecast China’s primary energy demand.
4.1. Construction of the Ensemble Model Based on the Standard Deviation Weight Method

In this section, we describe three ensemble models: GM (1,1)-SVR, RGM-SVR, and ARIMA-SVR. The relevant parameters of each single model are estimated by model training and calculation. Then we calculate the weights of each model based on the standard deviation method to combine the single models based on the calculated weights.

4.1.1. Standard Deviation Method

The standard deviation weight method is a relatively common weight calculation method. The calculation formula is as follows:

Assume the standard deviations of the prediction methods are \( s_1 \) and \( s_2 \), and

\[
s = s_1 + s_2
\]  

(22)

where, \( s_1 \) and \( s_2 \) are the standard deviations of the prediction error of each model.

Then, the weights are

\[
\omega_i = \frac{s - s_i}{s}, \quad i = 1, 2
\]  

(23)

4.1.2. GM (1,1)-SVR

The modeling procedure for the GM (1,1)-SVR combined prediction model is shown in Figure 2. The modeling steps are as follows:

Step 1. GM (1,1) and SVR are constructed, respectively. According to Step 4 of Section 3.1., the parameters \( a \) and \( u \) of GM (1,1) are calculated as shown in Table 2. The Lagrangian multiplier method is used to transform SVR into a “dual problem”, and then the parameters are solved using sequential minimal optimization (SMO). The parameters of SVR have been tested repeatedly. It is shown that the fitting result is optimal when \( C = 10000 \) and \( \sigma^2 = 0.01 \). The parameters \( \alpha \) and \( b \) of the SVR are estimated, as shown in Table 3.

Step 2. The results of GM (1,1) and SVR are assigned weights according to the standard deviation method, and the weight distribution is as shown in Table 4.

Step 3. The accuracy of the GM (1,1)-SVR is evaluated by MAE, MAPE, and MSPE.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>-0.056</td>
<td>-0.047</td>
<td>-0.055</td>
<td>-0.153</td>
<td>-0.096</td>
</tr>
<tr>
<td>( u )</td>
<td>2.656</td>
<td>1.966</td>
<td>4.528</td>
<td>6.285</td>
<td>1.875</td>
</tr>
</tbody>
</table>

Table 2. Parameter estimation of GM (1,1).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-2.1974</td>
<td>0.9999</td>
<td>-1.9356</td>
<td>-8.1981</td>
<td>-2.4216</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-6.8117</td>
<td>-8.0748</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>-4.5862</td>
<td>0.9999</td>
<td>-7.3687</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>-1.4721</td>
<td>0.9999</td>
</tr>
<tr>
<td>( b )</td>
<td>0.5148</td>
<td>0.4971</td>
<td>0.5021</td>
<td>0.5119</td>
<td>0.4878</td>
</tr>
</tbody>
</table>

Table 3. Parameter estimation of SVR.
Table 4. Weight distribution of GM (1,1)-SVR.

<table>
<thead>
<tr>
<th>Weight</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_{GM(1,1)} )</td>
<td>0.1201</td>
<td>0.2437</td>
<td>0.1871</td>
<td>0.1606</td>
<td>0.5512</td>
</tr>
<tr>
<td>( \omega_{SVR} )</td>
<td>0.8799</td>
<td>0.7563</td>
<td>0.8129</td>
<td>0.8394</td>
<td>0.4488</td>
</tr>
</tbody>
</table>

Figure 2. The procedure of GM (1,1)-SVR.

4.1.3. RGM-SVR

The description of the RGM-SVR combined prediction model is shown in Figure 3. The steps are as follows:

Step 1. RGM and SVR are constructed, respectively. Take China’s primary energy consumption from 2005 to 2012 as a training set, which is brought into the model for parameter estimation. The training samples are unchanged, therefore, the estimated parameters of the SVR model are unchanged. The parameter estimation results of the RGM model are shown in Table 5. The parameters of SVR have been tested repeatedly. \( C = 10000 \), \( \sigma^2 = 0.01 \). The parameters \( \alpha \) and \( b \) of the SVR are estimated as shown in Table 3.

Step 2. The prediction results of RGM and SVR are assigned weights according to the standard deviation method. The weight distribution is shown in Table 6.

Step 3. The accuracy of RGM-SVR is evaluated by MAE, MAPE, and MSPE.
### Table 4. Weight distribution of GM (1,1)-SVR.

<table>
<thead>
<tr>
<th>Weight</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{GM}$</td>
<td>0.1201</td>
<td>0.2437</td>
<td>0.1871</td>
<td>0.1606</td>
<td>0.5512</td>
</tr>
<tr>
<td>$\omega_{SVR}$</td>
<td>0.8799</td>
<td>0.7563</td>
<td>0.8129</td>
<td>0.8394</td>
<td>0.4488</td>
</tr>
</tbody>
</table>

4.1.3. RGM-SVR

The description of the RGM-SVR combined prediction model is shown in Figure 3. The steps are as follows:

- **Step 1.** RGM and SVR are constructed, respectively. Take China’s primary energy consumption from 2005 to 2012 as a training set, which is brought into the model for parameter estimation. The training samples are unchanged, therefore, the estimated parameters of the SVR model are unchanged. The parameter estimation results of the RGM model are shown in Table 5. The parameters of SVR have been tested repeatedly. $C = 10000$, $\sigma = 0.01$. The parameters $\alpha_i$ and $b$ of the SVR are estimated as shown in Table 3.

- **Step 2.** The prediction results of RGM and SVR are assigned weights according to the standard deviation method. The weight distribution is shown in Table 6.

- **Step 3.** The accuracy of RGM-SVR is evaluated by MAE, MAPE, and MSPE.

### Table 5. Parameter estimation of RGM.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$-0.056$</td>
<td>$-0.047$</td>
<td>$-0.055$</td>
<td>$-0.153$</td>
<td>$-0.096$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$0.055$</td>
<td>$-0.045$</td>
<td>$-0.057$</td>
<td>$-0.152$</td>
<td>$-0.094$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$0.058$</td>
<td>$-0.047$</td>
<td>$-0.060$</td>
<td>$-0.152$</td>
<td>$-0.092$</td>
</tr>
<tr>
<td>$a_4$</td>
<td>$0.057$</td>
<td>$-0.047$</td>
<td>$-0.059$</td>
<td>$-0.153$</td>
<td>$-0.093$</td>
</tr>
<tr>
<td>$u_1$</td>
<td>$2.656$</td>
<td>$1.966$</td>
<td>$4.528$</td>
<td>$6.285$</td>
<td>$1.876$</td>
</tr>
<tr>
<td>$u_2$</td>
<td>$2.816$</td>
<td>$2.073$</td>
<td>$4.719$</td>
<td>$7.327$</td>
<td>$2.087$</td>
</tr>
<tr>
<td>$u_3$</td>
<td>$2.940$</td>
<td>$2.139$</td>
<td>$4.906$</td>
<td>$8.456$</td>
<td>$2.325$</td>
</tr>
<tr>
<td>$u_4$</td>
<td>$3.121$</td>
<td>$2.250$</td>
<td>$5.239$</td>
<td>$9.807$</td>
<td>$2.522$</td>
</tr>
</tbody>
</table>

### Table 6. Weight distribution of RGM-SVR.

<table>
<thead>
<tr>
<th>Weight</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{RGM}$</td>
<td>0.2201</td>
<td>0.2407</td>
<td>0.1809</td>
<td>0.2501</td>
<td>0.4748</td>
</tr>
<tr>
<td>$\omega_{SVR}$</td>
<td>0.7799</td>
<td>0.7593</td>
<td>0.8191</td>
<td>0.7599</td>
<td>0.5254</td>
</tr>
</tbody>
</table>

---

**Figure 3.** The procedure of RGM-SVR.
4.1.4. ARIMA-SVR

The description of the ARIMA-SVR model is shown in Figure 4. The steps are as follows.

Step 1. Establish an ARIMA \((p, d, q)\) model for China’s primary energy demand forecast and examine whether the input samples are stationary time-series.

Step 2. Perform a time-series difference \(d\) on the sample. Begin with the first-order difference and observe the sequence. After the \(d\)-order difference, the sequence tends to be stationary, and the difference order is the parameter \(d\) of ARIMA.

Step 3. Estimate the parameters \(p\) and \(q\). Check and observe the autocorrelation and partial correlation graphs of the stationary time series to obtain \(q\) and \(p\), and so that the parameters of ARIMA are determined which are shown in Table 7.

Step 4. Assign weights to ARIMA and SVR by the standard deviation method, and the weight distribution is as shown in Table 8.

Step 5. The accuracy of ARIMA-SVR is evaluated by MAE, MAPE, and MSPE.

![Figure 4. The procedure of ARIMA–SVR.][1]

Table 7. Parameter estimation of ARIMA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(d)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(q)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

[1]: https://example.com/figure4.png
4.2. Construction of the Adaptive Weight Ensemble Forecasting Model

Linear and non-linear combinations are the most commonly used methods for ensemble forecasting models. However, when the prediction accuracy of the predicted system is unstable, the fixed-weight combination cannot be adjusted according to the changes in the predicted system in time, and thus, the generalization is poor. The adaptive variable weight combination can adjust the weight of each individual prediction model in time, thus improving the overall prediction accuracy and generalization of the ensemble model. Therefore, this section introduces the method of improving the inertia weight of PSO (APSO). Then, APSO is compared with PSO, GWO, MFO, and WOA to test the performance of the optimization algorithms. Finally, APSO is used to optimize the ensemble model to construct the adaptive weight ensemble prediction model (APSO-RGM-SVR).

4.2.1. PSO

PSO is a new evolutionary algorithm with easy implementation, high precision, and fast convergence. Therefore, PSO is suitable for solving practical problems.

The algorithm principle is as follows.

PSO abstracts the potential solution of the optimization problem into a particle in the D-dimensional search space. Each particle has a fitness value set according to the objective function, and the particle adjusts its position \( s \) according to the flight speed \( v \). Assume the search space has \( P \) particles composing a population \( s = (s_1, s_2, \ldots, s_i, \ldots, s_P) \), where the position of the \( i \)th particle \( (1 \leq i \leq P) \) is a \( D \)-dimensional vector. Then the position change rate of particle \( i \) at time \( t \) is \( v_t^i = (v_{t1}^i, v_{t2}^i, \ldots, v_{tD}^i) \), and the current optimal position of particle \( i \) is represented as \( p_t^i = (p_{t1}^i, p_{t2}^i, \ldots, p_{tD}^i) \), which is often denoted by \( p_{best}^i \). The optimal position of the population from the particle swarm search to time \( t \) is denoted by \( p_{tg1} = (p_{tg1}^1, p_{tg1}^2, \ldots, p_{tg1}^D) \), or by \( g_{best} \). The position of the particle is updated by the following formula:

\[
\begin{align*}
  v_{t+1}^{id} &= \omega v_{t}^{id} + c_1 r_1 (p_{t}^{id} - x_{t}^{id}) + c_2 r_2 (p_{tg}^{id} - x_{t}^{id}) \\
  s_{t+1}^{id} &= s_{t}^{id} + v_{t+1}^{id}
\end{align*}
\]

where \( \omega \) is the inertia weight, \( d = 1, 2, \ldots, D \) is the particle dimension, \( t \) is the current iteration number, \( v_{t}^{id} \in [v_{min}, v_{max}] \) is the particle velocity, the acceleration factors \( c_1 \) and \( c_2 \) are non-negative constants, and \( r_1 \) and \( r_2 \) are random numbers distributed in the interval \((0, 1)\).

4.2.2. APSO

The main factor affecting the search capability of particle swarm is the inertia weight. For improving the adaptive ability of PSO, an adaptive adjustment method based on inertia weight is proposed, which is called APSO. The algorithm is briefly described as follows:

Define the particle population diversity function

\[
F_d(t) = \frac{f_{min}(\alpha(t))}{f_{min}(\alpha(t)) + f_{max}(\alpha(t))}
\]

where

\[
\begin{align*}
  f_{min}(\alpha(t)) &= \text{Min}(f(\alpha_i(t))) \\
  f_{max}(\alpha(t)) &= \text{Max}(f(\alpha_i(t)))
\end{align*}
\]
where \( f(a_i(t)) \) is the fitness value of the \( i \)th particle \((i = 1, 2, \ldots, P_i)\), and \( f_{\min}(a(t)) \) and \( f_{\max}(a(t)) \) are the minimum and maximum values of the particle fitness at time \( t \), respectively.

The diversity function \( F_d(t) \) can be used to describe the motion characteristics of the particle, and the non-linear function \( \delta(t) \) is defined accordingly for the adaptive adjustment of the inertia weight:

\[
\delta(t) = e^{(F_d(t) - L)^{-1}}
\]

where \( L \geq 2 \) is the initialization constant.

The adjustment rules for defining particle inertia weight adaptation are

\[
v_{i}^{t+1} = \beta v_{i}^{t} + \eta d_{i}^{t}
\]

\[
\omega_{i}^{t+1} = e^{-(\delta(t)\epsilon_{i}^{t} + \epsilon_{i}^{t+1})^{-1}}
\]

where \( v_{i}^{t} \) is the velocity of the particle at time \( t \), \( \beta \) is the “momentum” hyper-parameter, that is usually set to 0.9, \( \eta \) is the learning rate, \( d_{i}^{t} \) is the gradient of the weight, \( \xi > 0 \) is a custom slack variable used to adjust the global search ability of the ion, and \( \epsilon_{i}^{t} \) is the distance between the particle \( i \) and the global optimal particle, defined as

\[
\epsilon_{i}^{t} = e^{\frac{|b_{i} - g_{b}|^2}{\xi}}
\]

4.2.3. Contrast Test of Optimization Algorithms

The PSO algorithm has fewer parameters that need to be adjusted and the structure is simple. However, PSO also has the potential to fall into local optimum. In recent years, some new swarm intelligence optimization algorithms have been proposed, including GWO, MFO, and WOA. These algorithms have high accuracy, but the problem of falling into local extremum is the common characteristic of most optimization models. Therefore, we improved PSO, which has fewer parameters and a simpler structure, to improve the global search ability of the optimization algorithm. In order to test the performance of the improved optimization algorithm, we conducted a comparative test.

The optimization performance of APSO was tested in a 30-dimensional search space. The population size was set to 30, the maximum number of iterations was 1000, and the benchmark functions are \( F_1 - F_6 \). The standard deviation comparison results after 40 independent experiments of GWO, MFO, WOA, POS, and APSO are shown in Table 9.

<table>
<thead>
<tr>
<th>Function</th>
<th>GWO</th>
<th>MFO</th>
<th>WOA</th>
<th>POS</th>
<th>APSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_1 )</td>
<td>8.24E - 007</td>
<td>3.34E + 001</td>
<td>1.07E - 006</td>
<td>2.32E + 001</td>
<td>1.00E - 010</td>
</tr>
<tr>
<td>( F_2 )</td>
<td>8.47E - 001</td>
<td>4.22E - 001</td>
<td>1.26E - 001</td>
<td>1.95E + 005</td>
<td>2.63E - 001</td>
</tr>
<tr>
<td>( F_3 )</td>
<td>7.67E - 004</td>
<td>2.80E - 003</td>
<td>3.85E - 004</td>
<td>7.97E - 002</td>
<td>5.27E - 005</td>
</tr>
<tr>
<td>( F_4 )</td>
<td>1.03E + 003</td>
<td>6.45E + 002</td>
<td>2.75E + 002</td>
<td>3.94E + 002</td>
<td>2.39E + 001</td>
</tr>
<tr>
<td>( F_5 )</td>
<td>5.69E + 000</td>
<td>0.79E + 000</td>
<td>0.12E + 000</td>
<td>2.98E + 001</td>
<td>0.04E + 000</td>
</tr>
<tr>
<td>( F_6 )</td>
<td>6.45E + 003</td>
<td>6.50E + 002</td>
<td>3.78E - 003</td>
<td>3.98E - 001</td>
<td>8.25E - 009</td>
</tr>
</tbody>
</table>

It can be seen from Table 10 that for six benchmark functions, the optimization accuracy of APSO for most functions was many orders of magnitude higher than that of PSO, and the accuracy of APSO was better than that of traditional GWO, MFO, and WOA. In the process of solving continuous single-mode functions, APSO had the highest accuracy for solving function \( F_1 \). For the optimization of function \( F_3 \), APSO was slightly better than other algorithms. For the multimodal functions \( F_4, F_5, \) and \( F_6 \) with multiple local extremums, the optimization accuracy of APSO was significantly better than other algorithms.
Table 10. The benchmark functions.

<table>
<thead>
<tr>
<th>Function</th>
<th>Formula</th>
<th>Dimension</th>
<th>Range</th>
<th>Theoretical Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwefel 2.21</td>
<td>$F_1 = \max_{1 \leq i \leq D}</td>
<td>x_i</td>
<td>$, $1 \leq i \leq D$</td>
<td>30</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>$F_2 = \sum_{i=1}^{\text{Dim}} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]$</td>
<td>30</td>
<td>$[-30, 30)$</td>
<td>0</td>
</tr>
<tr>
<td>Quartic</td>
<td>$F_3 = \sum_{i=1}^\text{Dim} i \cdot x_i^4 + \text{random}(0, 1)$</td>
<td>30</td>
<td>$[-1.28, 1.28]$</td>
<td>0</td>
</tr>
<tr>
<td>Schwefel 2.26</td>
<td>$F_4 = \sum_{i=1}^{\text{Dim}} -x_i \sin(\sqrt{x_i})$</td>
<td>30</td>
<td>$[-500, 500]$</td>
<td>0</td>
</tr>
<tr>
<td>Rastrigin</td>
<td>$F_5 = \sum_{i=1}^{\text{Dim}} [x_i^2 - 10 \cos(2\pi x_i) + 10]$</td>
<td>30</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Griewank</td>
<td>$F_6 = \frac{1}{4000} \sum_{i=1}^{\text{Dim}} x_i^2 - \prod_{i=1}^{\text{Dim}} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$</td>
<td>30</td>
<td>$[-600, 600]$</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2.4. Construction of APSO-RGM-SVR Based on Adaptive Inertia Weight

Based on the advantages of various forecasting models, considering the characteristics of the small sample size and long forecasting target of our paper, we combined RGM with SVR. Then, the combined weights of the two prediction models were determined by APSO. Finally, an adaptive optimal weight ensemble prediction model was established, which is shown in Figure 5. The involved steps are as follows.

Step 1. Based on SVR and RGM, a model for China’s primary energy demand forecasting is constructed and brought into the sample training set. The parameters of SVR and RGM required to determine the model are shown in Tables 3 and 5.

Step 2. The sum of the absolute values of the prediction errors is used as the cost function of the adaptive weight. The expression is

$$\min Q = \sum_{i=1}^{n} |e_i|$$

where $e_i$ represents the error values, and $n$ is the sample size.

Step 3. The optimal weighted model is obtained by the APSO based on the adaptive inertia weight, and thus, an optimal ensemble forecasting model is constructed. The optimal weight is shown in Table 11:

$$x_i(\text{APSO–RGM–SVR}) = \omega_i x_i(\text{RGM}) + (1 - \omega_i) x_i(\text{SVR})$$

Step 4. The prediction accuracy of the APSO-RGM-SVR prediction model is evaluated by MAE, MAPE, and MSPE.

Table 11. Weight distribution of APSO-RGM-SVR.

<table>
<thead>
<tr>
<th>Weight</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{\text{APSO–RGM}}$</td>
<td>0.2928</td>
<td>0.3252</td>
<td>0.1281</td>
<td>0.5436</td>
<td>0.0981</td>
</tr>
<tr>
<td>$\omega_{\text{APSO–SVR}}$</td>
<td>0.7072</td>
<td>0.6748</td>
<td>0.8719</td>
<td>0.4564</td>
<td>0.9019</td>
</tr>
</tbody>
</table>
4.2.4. Construction of APSO-RGM-SVR Based on Adaptive Inertia Weight

Based on the advantages of various forecasting models, considering the characteristics of the small sample size and long forecasting target of our paper, we combined RGM with SVR. Then, the combined weights of the two prediction models were determined by APSO. Finally, an adaptive optimal weight ensemble prediction model was established, which is shown in Figure 5. The involved steps are as follows.

Step 1. Based on SVR and RGM, a model for China’s primary energy demand forecasting is constructed and brought into the sample training set. The parameters of SVR and RGM required to determine the model are shown in Tables 3 and 5.

Step 2. The sum of the absolute values of the prediction errors is used as the cost function of the adaptive weight. The expression is

\[
\min_{i} \sum_{i} Qe_i = \sum_{i} |\hat{y}_i - y_i|
\]

where \(ie\) represents the error values, and \(n\) is the sample size.

Step 3. The optimal weighted model is obtained by the APSO based on the adaptive inertia weight, and thus, an optimal ensemble forecasting model is constructed. The optimal weight is shown in Table 11:

\[
\omega_{APSO-RGM-SVR} = \omega_{RGM} + \omega_{SVR} = \omega_{SVR} - \omega_{RGM}
\]

Step 4. The prediction accuracy of the APSO-RGM-SVR prediction model is evaluated by MAE, MAPE, and MSPE.

Figure 5. The procedure of APSO-RGM–SVR model.

5. Results

The previous section introduced the data sources, described the detailed steps for constructing the four ensemble models, and estimated the parameters of each model. The calculation steps for applying the model for prediction are similar to those used to evaluate the fitting, and thus are not described here. This section introduces the dataset and presents the fitting results and evaluation of primary energy consumption by the single forecasting models, unoptimized ensemble models, and optimized ensemble models. Then, the forecast results of the future primary energy demand and primary energy structure are presented.

5.1. Data

The data used to construct China’s primary energy demand and energy structure prediction models came from the National Bureau of Statistics of China (http://www.stats.gov.cn/tjsj/). We selected 12 years (2005–2016) of annual consumption data as a dataset, including China’s total annual energy consumption, coal, crude oil, natural gas, and annual consumption of water, wind, and nuclear power, as shown in Table 12. Seventy percent of the dataset (2005–2012) was used as a model training set, and 30% of the sample (2013–2016) was used as the model test set.

Figure 6 depicts the annual consumption of coal, crude oil, natural gas, water, wind, and nuclear power as well as China’s total annual energy consumption. The growth rate of energy consumption is shown in Figure 7. From 2005 to 2016, China’s total energy consumption showed an overall growth trend. During the period 2005–2011, primary energy demand showed rapid growth. For promoting economic development, the demand for primary energy increased steadily, especially the demand for coal. Oil and natural gas are affected by supply, and their growth rates were small. Under the influence of environmental policies, the total primary energy consumption growth slowed from 2011. The primary energy structure began to change significantly, and coal consumption growth began to slow down and decline. The proportion of clean energy began to increase.
Table 12. The energy consumption of China from 2005–2016 (units: million ton coal equivalent (mtce)).

<table>
<thead>
<tr>
<th>Year</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Water, Wind, and Nuclear Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2613.69</td>
<td>1892.31</td>
<td>465.24</td>
<td>62.73</td>
<td>193.41</td>
</tr>
<tr>
<td>2006</td>
<td>2864.67</td>
<td>2074.02</td>
<td>501.32</td>
<td>77.35</td>
<td>211.99</td>
</tr>
<tr>
<td>2007</td>
<td>3114.42</td>
<td>2257.95</td>
<td>529.45</td>
<td>93.43</td>
<td>233.58</td>
</tr>
<tr>
<td>2008</td>
<td>3206.11</td>
<td>2292.37</td>
<td>535.42</td>
<td>109.01</td>
<td>269.31</td>
</tr>
<tr>
<td>2009</td>
<td>3361.26</td>
<td>2406.66</td>
<td>551.25</td>
<td>117.64</td>
<td>285.71</td>
</tr>
<tr>
<td>2010</td>
<td>3506.11</td>
<td>2495.68</td>
<td>627.53</td>
<td>144.26</td>
<td>325.12</td>
</tr>
<tr>
<td>2011</td>
<td>3870.43</td>
<td>2717.04</td>
<td>650.23</td>
<td>178.04</td>
<td>325.12</td>
</tr>
<tr>
<td>2012</td>
<td>4021.38</td>
<td>2754.65</td>
<td>683.63</td>
<td>193.03</td>
<td>390.07</td>
</tr>
<tr>
<td>2013</td>
<td>4169.13</td>
<td>2809.99</td>
<td>712.92</td>
<td>220.96</td>
<td>425.25</td>
</tr>
<tr>
<td>2014</td>
<td>4258.06</td>
<td>2793.29</td>
<td>740.90</td>
<td>242.71</td>
<td>481.16</td>
</tr>
<tr>
<td>2015</td>
<td>4299.05</td>
<td>2738.49</td>
<td>786.73</td>
<td>253.64</td>
<td>520.19</td>
</tr>
<tr>
<td>2016</td>
<td>4360.00</td>
<td>2703.20</td>
<td>797.88</td>
<td>279.04</td>
<td>579.88</td>
</tr>
</tbody>
</table>

Figure 6. Energy consumption in China from 2005–2016.

Figure 7. The energy consumption growth rate in China.
5.2. Model Adaptability

The experimental data came from the total annual consumption of primary energy in China and the annual consumption of four types of primary energy for the years 2005–2016. The dataset was divided into two parts: data from 2005–2012 was used as model data, and data from 2013–2016 was used as a test dataset. The prediction results of the prediction model including GM (1,1), RGM, ARIMA, SVR, GM (1,1)-SVR, RGM-SVR, ARIMA-SVR, and APSO-RGM-SVR are shown in Table 13. Then, the accuracy was evaluated by MAE, MAPE, and MSPE. As far as a single model is concerned, fitting results of SVR are better in the model training phase, followed by RGM and GM (1,1), while SVR had the highest prediction accuracy in the test phase. The optimization of the rolling mechanism makes the accuracy of RGM higher than GM (1,1). The ARIMA showed very poor prediction effect in both the training phase and the prediction phase. The ensemble model had better prediction accuracy than the single models. Compared with other unoptimized ensemble models, the RGM-SVR model had better fitting and prediction effects both in training and testing. However, the adaptive weight ensemble model APSO-RGM-SVR based on the APSO proposed in this study had higher prediction accuracy and stronger adaptability.

<table>
<thead>
<tr>
<th>Method</th>
<th>Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Testing</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>MAPE</td>
<td>MSPE</td>
</tr>
<tr>
<td>PSO-RGM (1,1)-SVR</td>
<td>38.765</td>
<td>0.01174</td>
<td>0.00510</td>
</tr>
<tr>
<td>RGM (1,1)-SVR</td>
<td>41.123</td>
<td>0.01263</td>
<td>0.00548</td>
</tr>
<tr>
<td>GM (1,1)-SVR</td>
<td>42.503</td>
<td>0.01306</td>
<td>0.00570</td>
</tr>
<tr>
<td>ARIMA-SVR</td>
<td>54.939</td>
<td>0.01716</td>
<td>0.00767</td>
</tr>
<tr>
<td>SVR</td>
<td>52.555</td>
<td>0.01617</td>
<td>0.00747</td>
</tr>
<tr>
<td>RGM (1,1)</td>
<td>28.310</td>
<td>0.00840</td>
<td>0.00384</td>
</tr>
<tr>
<td>GM (1,1)</td>
<td>28.310</td>
<td>0.00840</td>
<td>0.00384</td>
</tr>
<tr>
<td>ARIMA</td>
<td>89.969</td>
<td>0.02818</td>
<td>0.01148</td>
</tr>
</tbody>
</table>

Table 13. Model evaluation results.
Table 13. Cont.

<table>
<thead>
<tr>
<th>Method</th>
<th>Natural gas</th>
<th>Hydro, Nuclear and Wind Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>MAPE</td>
</tr>
<tr>
<td>PSO-RGM (1,1)-SVR</td>
<td>1.034</td>
<td>0.00837</td>
</tr>
<tr>
<td>RGM (1,1)-SVR</td>
<td>1.034</td>
<td>0.00837</td>
</tr>
<tr>
<td>GM (1,1)-SVR</td>
<td>1.042</td>
<td>0.00830</td>
</tr>
<tr>
<td>ARIMA-SVR</td>
<td>12.977</td>
<td>0.01023</td>
</tr>
<tr>
<td>SVR</td>
<td>0.622</td>
<td>0.00549</td>
</tr>
<tr>
<td>RGM (1,1)</td>
<td>3.190</td>
<td>0.02351</td>
</tr>
<tr>
<td>GM (1,1)</td>
<td>3.188</td>
<td>0.02348</td>
</tr>
<tr>
<td>ARIMA</td>
<td>2.283</td>
<td>0.01663</td>
</tr>
</tbody>
</table>

The best evaluation results for each type of energy are indicated in bold in Table 13. Because the variation of consumption of each type of energy was different, the characteristics extracted by each prediction method during training were also different. From the prediction results of the total annual energy consumption, the RGM and GM (1,1) had the best fitting effect in the training phase. Because the modeling steps of RGM and GM (1,1) were the same in the training phase, the fitting results of the two models in the phase were basically the same. The results of the three evaluations, MAE, MAPE, and MSPE were 28.310, 0.00840, and 0.00384, respectively, which were the lowest values, followed by those of APSO-RGM-SVR. In the test phase, the values of the three evaluations of APSO-RGM-SVR were 26.252, 0.00616, and 0.00352, respectively, which were the lowest values. The fitting and prediction results of annual coal consumption showed that the fitting errors of RGM and GM (1,1) were the smallest in the training phase, and the MAE, MAPE, and MSPE of the two models were 29.404, 0.01207, and 0.00516, respectively, i.e., they were approximately the same. In the test phase, the APSO-RGM-SVR model had the highest accuracy. The three error evaluations results of the APSO-RGM-SVR model were the lowest compared with other methods: 14066, 0.00505, and 0.00320. For the training and test results of annual oil consumption, SVR had the highest fitting degree to the training data. The results of the three error evaluations were 4.345, 0.00700, and 0.00651, respectively. In the test phase, the accuracy of APSO-RGM-SVR was still higher than that of the other models, and the results of the three error evaluations were 5.956, 0.00785, and 0.00448, respectively. For the results of the annual consumption of natural gas, the results of the MAE and MAPE evaluations in the training phase were analyzed, and SVR had the smallest error of 0.622 and 0.00549, respectively, indicating that the SVR model had the highest fitting degree. However, the analysis results of MSPE, APSO-RGM-SVR, and RGM-SVR error evaluations were the lowest—0.00512. Most noteworthy, the MSPE value was low, indicating that the fitting curve was relatively stable, and the abnormality was low. For a fitted curve with a large MAPE value, if it is relatively stable, its MSPE value is low. Therefore, the MSPE value cannot fully reflect the accuracy of the curve fitting. In evaluating the prediction model, it should be combined with MAE, MAPE, and MSPE to give a comprehensive evaluation. In the test phase, the three error evaluations of APSO-RGM-SVR—2.092, 0.01259, and 0.00995—were the lowest, showing an optimal prediction effect. For the training and test results of water, wind, and nuclear power consumption, in the training phase, it can be seen from the values of MAE and MAPE (7.912 and...
0.02546, respectively) that the SVR errors were the smallest, indicating that the SVR model had the highest fitting degree. However, regarding the value of MSPE, the error evaluations of RGM and GM (1,1) (both were 0.01333) were the lowest. In the test phase, the APSO-RGM-SVR model had the best prediction capacity, and the three error evaluations were the lowest—10.947, 0.02024, and 0.01232, respectively.

The fitting and prediction curves for primary energy consumption are shown in Figure 8a–e. In the training phase, ARIMA had the worst fitting effect, and the fitting effects of the GM (1,1), RGM, SVR, and APSO-RGM-SVR model were ideal. However, in the test phase, the ensemble model showed its advantages, especially the APSO-RGM-SVR model, which had a high fitting degree to the real data.

![Graph showing fitting and prediction curves for primary energy consumption](image-url)
Figure 8. The simulation and forecasting results for (a) the total energy consumption; (b) coal; (c) oil; (d) natural gas; and (e) hydro, nuclear, and wind power.
5.3. Prediction Results

Using the sample information, the APSO-RGM-SVR model was selected to forecast China’s total annual energy demand and annual demand for the four types of primary energy for the year 2020 and 2025. The prediction results are shown in Table 14. The total primary energy demand was predicted to increase to 4465.41 mtce in 2020 and 4825.43 mtce in 2025. Compared with 2016, the demand for coal was predicted to decrease. It was estimated that the demand for coal will be 2637.23 and 2203.86 mtce in 2020 and 2025, respectively. The total demand for oil is growing slowly. It was estimated that the demand for oil will be 922.15 and 989.19 mtce by 2020 and 2025, respectively. In the future, the demand for clean energy was predicted to drastically increase. It is estimated that the demand for natural gas will be 409.53 mtce in 2020 and 540.64 mtce in 2025. Water, wind, and nuclear power were predicted to receive more and more attention in the future energy market. The demand for 2020 was predicted to be 687.50 mtce, and the demand for 2025 was predicted to be 1091.74 mtce.

<table>
<thead>
<tr>
<th>Year</th>
<th>The Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Hydro, Nuclear, and Wind Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>4656.41</td>
<td>2637.23</td>
<td>922.15</td>
<td>409.53</td>
<td>687.50</td>
</tr>
<tr>
<td>2025</td>
<td>4825.43</td>
<td>2203.86</td>
<td>989.19</td>
<td>540.64</td>
<td>1091.74</td>
</tr>
</tbody>
</table>

Figure 9a shows the prediction results of energy for the years 2017–2025 and the integration and comparison of the energy demand structure with the energy structure, for the years 2005–2017. Figure 9b shows the ratio of coal, oil, natural gas, water, wind, and nuclear power for the years 2005 and 2025. As can be seen from Figure 9, China’s total primary energy demand has grown steadily. The demand for natural gas and water, wind, and nuclear power has risen sharply. In 2020, the ratio of natural gas was predicted to be about 8.8%. In 2025, the ratio of natural gas was expected to reach 11.2%. It was estimated that the ratio of water, wind, and nuclear power will reach 14.8% and 22.6% in 2020 and 2025, respectively. Coal and oil that cause pollution to the environment were predicted to gradually decrease. The ratio of coal will fall to 56.6% in 2020 and 45.7% in 2025. It was estimated that oil will account for 19.8% in 2020 and around 20.5% in 2025, with a very small increase.
Figure 9. (a) Prediction results of primary energy demand forecast; (b) prediction results of primary energy ratio.

Figure 10 shows the growth rates of annual demand for primary energy for the next nine years. Table 15 shows the average annual growth rate of primary energy demand. The results show that the total primary energy demand was predicted to grow steadily in the future, with an average annual growth rate of 1.14%. The annual growth rate of coal demand is negative, i.e., it is declining year by year with an average annual rate of 2.2%. Although the annual demand for oil has increased yearly, the growth rate has shown a downward trend, and the average annual growth rate is 2.42%. It is expected that natural gas and clean energy such as water, wind, and nuclear power will be fully developed. Although the growth rate is predicted to decline slightly after 2020, the annual demand is still predicted to showing a very high growth trend. The average annual growth rates of the two types of primary energy demand were predicted to be 7.43% and 7.7%.
Table 15. Average annual growth rate (%).

<table>
<thead>
<tr>
<th>Energy</th>
<th>Total Energy Consumption</th>
<th>Coal</th>
<th>Oil</th>
<th>Natural Gas</th>
<th>Hydro, Nuclear, and Wind Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average annual growth rate</td>
<td>1.14</td>
<td>−2.2</td>
<td>2.42</td>
<td>7.43</td>
<td>7.70</td>
</tr>
</tbody>
</table>

6. Conclusions

Accurate forecasting of energy demand can help decision departments to develop reasonable policies and plans. At the same time, by forecasting energy demand and structure, the connections between energy consumption and economic, social, and environmental protection can be established to guide the involved departments to adjust the energy structure and industrial layout in a targeted manner. The purpose of this study was to develop an ensemble forecasting technology based on a time-series model and artificial intelligence, together with APSO to determine the optimal combination weight. The ensemble forecasting technology was optimized to predict China’s primary energy demand and primary energy structure more accurately. Through the research results, we obtained the following three important findings:

1. The improved PSO algorithm with the “momentum” factor added had better global search capabilities.

2. By comparing a variety of single-prediction models and fixed-weight ensemble models, the prediction results of the APSO-RGM-SVR model proposed in this study showed the smallest error, and the prediction accuracy of energy demand was significantly improved. The ensemble model (APSO-RGM-SVR), which was optimized by the APSO algorithm, effectively combined the characteristics of a time-series model and artificial intelligence, and could adjust the optimal weight at any time with the change of samples. The new method proposed in this paper is a prediction method with higher accuracy and better generalization.

3. The APSO-RGM-SVR method was used to forecast China’s primary energy consumption from 2017 to 2025. Prediction results indicate that China’s energy demand will continue to rise. By 2020, China’s primary energy demand was predicted to reach about 4656.41 mtce, and the proportions of coal, oil, natural gas, and hydropower and nuclear power were predicted to be 56.6%, 19.8%, 8.8%, and 14.8%, respectively. By 2025, China’s primary energy demand was predicted to reach about 4656.41 mtce, and the proportion of coal, oil, natural gas, and hydro, nuclear, and wind power were predicted to be 45.7%, 20.5%, 11.2%, and 22.6%, respectively.

According to the evaluation results of the model training situation and predictive ability, it can be considered that the prediction model APSO-RGM-SVR proposed in this paper can be used for future short-term and medium-term primary energy demand forecasting and primary energy structure prediction in China, and it can produce better prediction results. At the same time, this method can also be used for the same types of data training and prediction in other regions and countries. However, whether the model proposed in this study is applicable to the prediction of large samples remains to be tested. This is also a direction that we need to explore further.

Author Contributions: W.Z. conceived and designed the experiments, and wrote the paper; J.Z. and Z.J. analyzed the data; X.Y. modified the manuscript; P.W. performed the experiments; and all authors read and approved the final manuscript.

Funding: This study was funded by the National Natural Science Foundation of China [grant number 41401655], the Qualified Personnel Foundation of Taiyuan University of Technology [grant number 2013W005], Program for the Top Young Academic Leaders of Higher Learning Institutions of Shanxi [grant number 163080127-S], and Program for the Philosophy and Social Sciences Research of Higher Learning Institutions of Shanxi [grant number 2016311], and the Postgraduate Education Innovation Project of Shanxi [grant number 2018BY044]. The APC was funded by [grant number 2018BY044].

Acknowledgments: The authors would like to thank the support from the project grants: the National Natural Science Foundation of China [grant number 41401655], the Qualified Personnel Foundation of Taiyuan University
of Technology [grant number 2013W005], Program for the Top Young Academic Leaders of Higher Learning Institutions of Shanxi [grant number 163080127-S], Program for the Philosophy and Social Sciences Research of Higher Learning Institutions of Shanxi [grant number 2016311], and the Postgraduate Education Innovation Project of Shanxi [grant number 2018BY044].

Conflicts of Interest: The authors declare no conflict of interest.

References


15. Yuan, C.Q.; Liu, S.F.; Fang, Z.G. Comparison of China’s primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. *Energy* 2016, 100, 384–390. [CrossRef]


23. Wang, Q.; Li, S.; Li, R. China’s dependency on foreign oil will exceed 80% by 2030: Developing a novel NMGM-ARIMA to forecast China’s foreign oil dependence from two dimensions. *Energy* 2018, 163, 151–167. [CrossRef]


40. Liu, B.; Fu, C.; Bielefield, A.; Liu, Y. Forecasting of Chinese Primary Energy Consumption in 2021 with GRU Artificial Neural Network. *Energies* 2017, 10, 1453. [CrossRef]


59. Li, W.; Yang, X.; Li, H.; Su, L. Hybrid Forecasting Approach Based on GRNN Neural Network and SVR Machine for Electricity Demand Forecasting. *Energies* 2017, 10, 44. [CrossRef]


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


