Prediction of China's Energy Consumption Based on Robust Principal Component Analysis and PSO-LSSVM Optimized by the Tabu Search Algorithm

Lihui Zhang [1,2], Riletu Ge [1,2,*] and Jianxue Chai [1,2]

[1] School of Economics and Management, North China Electric Power University, Beijing 102206, China; zlh6699@126.com (L.Z.); 1182106023@ncepu.edu.cn (J.C.)

[2] Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Beijing 102206, China

[*] Correspondence: 1172106002@ncepu.edu.cn; Tel.: +86-198-0487-6606

Abstract: China's energy consumption issues are closely associated with global climate issues, and the scale of energy consumption, peak energy consumption, and consumption investment are all the focus of national attention. In order to forecast the amount of energy consumption of China accurately, this article selected GDP, population, industrial structure and energy consumption structure, energy intensity, total imports and exports, fixed asset investment, energy efficiency, urbanization, the level of consumption, and fixed investment in the energy industry as a preliminary set of factors; Secondly, we corrected the traditional principal component analysis (PCA) algorithm from the perspective of eliminating "bad points" and then judged a "bad spot" sample based on signal reconstruction ideas. Based on the above content, we put forward a robust principal component analysis (RPCA) algorithm and chose the first five principal components as main factors affecting energy consumption, including: GDP, population, industrial structure and energy consumption structure, urbanization; Then, we applied the Tabu search (TS) algorithm to the least square to support vector machine (LSSVM) optimized by the particle swarm optimization (PSO) algorithm to forecast China's energy consumption. We collected data from 1996 to 2010 as a training set and from 2010 to 2016 as the test set. For easy comparison, the sample data was input into the LSSVM algorithm and the PSO-LSSVM algorithm at the same time. We used statistical indicators including goodness of fit determination coefficient (R^2), the root means square error (RMSE), and the mean radial error (MRE) to compare the training results of the three forecasting models, which demonstrated that the proposed TS-PSO-LSSVM forecasting model had higher prediction accuracy, generalization ability, and higher training speed. Finally, the TS-PSO-LSSVM forecasting model was applied to forecast the energy consumption of China from 2017 to 2030. According to predictions, we found that China shows a gradual increase in energy consumption trends from 2017 to 2030 and will breakthrough 6000 million tons in 2030. However, the growth rate is gradually tightening and China's energy consumption economy will transfer to a state of diminishing returns around 2026, which guides China to put more emphasis on the field of energy investment.

Keywords: energy consumption forecasting; improved PSO-LSSVM algorithm; Tabu Search; robust principal component analysis

1. Introduction

China is a major energy consumer. Since the first half of 2018, coal, natural gas, petrol, and electricity consumption has been on the rise, among which the highest increase was coal consumption with an increase of 3.1%, and we found thermal power was a major factor of the continued growth of coal consumption. According to "BP Statistical Review of World Energy" [1], China's renewable

energy consumption accounts for 36% of global volume in 2017, among which natural gas consumption accounted for 32.6% of global natural gas consumption. This showed that China's energy consumption has an important role in world energy consumption. At present, China's energy consumption structure upgraded gained initial success, however there is still a distance from the long-term goal to build a low-carbon clean energy consumption structure. Therefore, based on a multitude of influencing factors of energy consumption, it is imperative to apply highly accurate predictive model to forecast energy consumption, and then study more economic information about China's energy consumption. A high-precision energy consumption forecasting model can provide a quantitative basis for decision-making by relevant Chinese institutions, enabling China to better understand energy consumption trends and solve energy-related problems such as pollutant emissions and carbon emissions. In this article, a breakthrough is made in the machine intelligence algorithm. The PSO-LSSVM forecasting model optimized by the Tabu search algorithm is put forward, which avoids PSO-LSSVM falling into local optimum and speeds up the global search at the same time. Finally, the comparison of various forecasting models proves the fact that the accuracy and generalization ability of the TS-PSO-LSSVM forecasting model proposed in this article is higher than the others.

Energy consumption can be influenced by many factors directly or indirectly. Up to now, there are many factors affecting energy consumption have been studied [2–6]. Energy consumption problem is a complex nonlinear problem. So far, scholars have proposed various forecasting models, including grey prediction theory [7–9], multiple regression [10–12], input-output method [13,14], and time-series forecasting models [15,16].

Hsu et al. [7] applied artificial neural network and an improved gray model to predict electricity consumption of Taiwan, and the examples showed that the improved grey prediction model has higher prediction accuracy. Sehgal, V et al. [10] proposed the wavelet-bootstrap-multiple linear regression (WBMLR) predictive model to forecast India Mahathir Power Load Nadi River Basin, and the examples showed that the model owns higher accuracy than the artificial neural networks (ANN), wide area network (WAN), machine learning in *R* (MLR) model. Erdogdu E [16] applied co-integration analysis and auto-regressive integrated moving average (ARIMA) model to predict power load of the Republic of Turkey, and proved the high power load forecasting officially.

Up to now, a multitude of scholars have proposed robustness problem of the principal component analysis (PCA) based on various ways and put forward their own optimized algorithm. Luong et al. [17] considered a new method named online robust principal component analysis (RPCA) for time-varying decomposition problems and proposed a compressive online RPCA algorithm that can combine various information about decomposed vectors via an $n-l_1$ minimization method. Chretien et al. [18] proposed a robust principal component analysis (RPCA) method to build the Low Rank + Sparse models when the used data is corrupted by outliers and applied it to estimate the topology in power grid networks. Sadeghian et al. [19] thought that traditional robust principal component analysis (RPCA) algorithms only focused on output outliers, however, both input and output data can make mistakes in developing soft sensors. They built a robust probabilistic predictive model to overcome this problem by appropriate formulation of noise distributions. Wu et al. [20] proposed a multi-component groups sparse RPCA model to solve the problems under the condition of complex dynamic background and applied alternating direction method of multipliers algorithm to the proposed model. Experiments demonstrated that the proposed method has better performance than others.

Along with the wide application of intelligent algorithms, more and more scholars have proposed various intelligent algorithms in all areas. Least squares support vector machine (LSSVM) prediction algorithm is one of the most widely used and has high accuracy and applicability. Roushangar et al. [21] built three types of models about flow characteristics, flow and bedform characteristics, and sediment characteristics based on the Least Squares Support Vector Machine optimized by Particle Swarm Optimization (PSO-LSSVM) and proved the forecasting model can predict the roughness coefficient precisely. Huan et al. [22] proposed a forecasting model based on integrated empirical mode decomposition (EEMD) and Least Squares Support Vector Machine

(LSSVM) and showed that EEMD-LSSVM model is a better predictor algorithm than wavelet denoising least squares support machine (WD-LSSVM) and traditional LSSVM. Xue [23] optimized the LSSVM by improved particle swarm optimization algorithm (IMPSO-LSSVM) and proposed the combined concrete compressive strength prediction model. Then, he compared IMPSO-LSSVM, PSO-LSSVM, GA-LSSVM (the Least Squares Support Vector Machine optimized by genetic algorithm) and back-propagation neural network to prove the proposed model is an effective tool to forecast concrete compressive strength. Lu et al. [24] presented a new forecasting model based on empirical mode decomposition integrated permutation entropy (EEMD-PE), LSSVM, and gravitational search algorithm (GSA) to overcome the nonlinear prediction of wind power and volatility difficulties and predicted ultra-short-term forecasting of wind power accurately. Zhao et al. [25] used the salp swarm algorithm (SSA) on LSSVM to optimize two machine parameters in LSSVM algorithm, and showed that the forecasting model have higher accuracy than traditional LSSVM, PSO-LSSVM and BP neural network through integrated statistical indicators. Wen et al. [26] proposed the GA-LSSVM prediction model to predict landslide displacement and showed that the model can predict high-precisive consistency between measured displacement and predicted displacement. Liu et al. [27] proposed an improved gravitational search algorithm (AC-GSA) to improve the performance of GSA and optimize LSSVM parameters. They used a novel model to forecast heat rate of a 600 MW supercritical steam turbine unit. Results indicate that the AC-GSA–LSSVM model is a powerful technique to forecast load. Gorjaei et al. [28] applied the LSSVM model to predict liquid flow rate for two-phase flow through wellhead chokes and used particle swarm optimization (PSO) to optimize two parameters of the LSSVM algorithm. The PSO-LSSVM model is excellently consistent with actual measured rates. Results indicated that the PSO-LSSVM model demonstrated better regression precision and generalization capability. Zhang [29] proposed a hybrid model that combines fuzzy clustering (FC), LSSVM, and the wolf pack algorithm (WPA), and used two cases to train and test data. The results proved that the proposed model obtains higher prediction accuracy and stability.

In recent years, the Tabu search algorithm is widely used to shorten the computing time of the algorithm [30–33]

Peng et al. [30] added a Tabu search procedure into the framework of path relinking to generate solutions to the job shop scheduling problem (JSP). The results showed that Tabu search/ Path relinking (TS/PR) obtained better performance than the traditional state-of-the-art algorithms for JSP. Escobar et al. [31] proposed a hybrid Granular Tabu Search algorithm to solve the Multi-Depot Vehicle Routing Problem (MDVRP). The results of cases showed that the proposed algorithm solved problems with short computing time and got best solutions. Li et al. [32] applied a hybrid algorithm (HA) to the genetic algorithm (GA) and used Tabu search (TS) at the same time to solve the flexible job shop scheduling problem with the aim to minimize the make span and proved that the proposed method can provide the best solutions. Sicilia et al. [33] presented a novel algorithm to solve the problem of the capillary distribution of goods in major urban areas. The proposed Tabu search algorithm can minimize the wide variety of constraints and complexities and reduced costs, which made problems quickly solved in time.

In order to accurately predict China's energy consumption, this article proposes a native PSO-LSSVM model optimized by the Tabu search algorithm based on robust principal component analysis. The innovations of this article are as follows:

(1) Energy consumption is a macroeconomic issue and is affected by many influencing factors. In order to achieve accurate prediction of China's energy consumption, based on a large amount of literature research, combined with China's energy consumption characteristics, we selected GDP, population, industrial structure, energy consumption structure, energy intensity, total import and export, social fixed Asset investment, energy utilization rate, urbanization rate, household consumption level, and fixed investment in energy industry as the set of initial influencing factors. In this article, the main influencing factors are selected by robust optimized principal component analysis (RPCA) method. Based on the idea of signal reconstruction, the judgment basis of a "bad point" sample is given, which can reduce the difficulty of data collection. The first five influencing factors have the ability to represent the information of other influencing factors. By comparing the

classification results, it is concluded that the RPCA algorithm is significantly better than the traditional PCA, and the classification effect is more accurate. The information of the original sample can be more comprehensively represented. This operation also greatly improves the accuracy of the forecasting mode.

(2) This article innovatively applies the Tabu search algorithm to optimizing the PSO-LSSVM algorithm. The combined forecasting model greatly improves the search ability of parameters and reduces the search time, then, it can avoid local optimal results. The empirical analysis proves that the TS-PSO-LSSVM model has strong generalization ability and can reliably forecast China's energy consumption, and its prediction accuracy is better than PSO-LSSVM and LSSVM.

(3) This article innovatively applies the machine learning algorithm into the hot spot of international research about forecasting China's energy consumption. The traditional methods to forecast energy consumption mainly includes mathematical statistics methods, such as linear regression, time series analysis, gray prediction, etc. These methods all regard energy consumption as a linear problem, which is greatly limited by the choice of influencing factors, so it is difficult to be rational and scientific. However, the machine learning algorithm used in this article can consider more influencing factors, which can turn the energy consumption problem into a nonlinear problem with higher rationality and adaptability.

The main contents of the article are as follows: the second section describes the mathematical principle of robust principal component analysis and PSO-LSSVM optimized by the Tabu search algorithm; the third part proves that the proposed forecasting model has higher prediction accuracy, generalization ability, and higher training speed by compared results with traditional LSSVM and PSO-LSSVM models, and then we apply the model to forecast the energy consumption in China from 2017 to 2030; the fourth part makes forward-looking conclusions according to the results of the RPCA-TS-PSO-LSSVM forecasting model.

2. The Forecasting Model

2.1. Robust Principal Component Analysis

Robust principal component analysis (RPCA) [34] was proposed by John Wright and belongs to the subspace learning model and was improved on the basis of principal component analysis algorithm. The core idea of the RPCA algorithm is to replace the whole data with part of data to reduce the dimension of the original data redundancy. From the perspective of linear algebra, it is to replace the original data with another set of data under the principle of minimizing redundancy and noise. The RPCA calculation steps are as follows:

Based on recently reconfigurable, centralizing a sample data set {x_1, $x_2...x_n$}, $\left\{x_1, x_2,...,x_n\right\}$ $\sum_{i=1}^{n} x_i = 0$, Given that new coordinates after the projection conversion is {w_1, $w_2...w_d$}, then

$$w_{i2} = I, w_i^T w_j = 0 \left(i \neq j\right)$$ (1)

After the coordinates of the new coordinate system portion are discarded to reduce the dimension d' ($d' < d$), the portion of the sample data in the low-dimensional coordinate system in the coordinate system is the projection

$$z_i = \left(z_{i1}; z_{i2} ... z_{id}\right), z_{ij} = w_j^T x_i$$ (2)

Reconstructing x_i based on z_i:

$$\hat{x}_i = \sum_{j=1}^{d'} z_{ij} w_j$$ (3)

Thus, the distance between the original sample points x_i and sample points $\hat{x}_i$ based on the projection reconstruction is

$$\sum_{}^{m}\sum_{}^{d} z_{ij} w_j - x_i = \sum_{}^{m} z_i^T z_i - 2\sum_{}^{m} z_i^T W^T x_i + const \tag{4}$$

Among them, the w_j group is orthogonal, $\sum x_i x_i^T$ is the covariance matrix. We obtain formula (5) from the recent reconstruction.

$$\min_{w} - \mathrm{tr}\left(W^T X X^T W\right)$$

$$s.t. W^T W = I \tag{5}$$

Since the premise of principal component analysis is that the noise contained in the original data is Gaussian, large noise or severe outliers cause the algorithm to fail. Applying robust analysis to principal component analysis can comprehensively consider the redundant information and noise. Robust principal component analysis calculation steps are as follows:

Provided the original data matrix is *D*, we decomposed it into a sparse matrix and low-rank matrix through Robust principal component analysis

$$D = A + E \tag{6}$$

E is the sparse matrix which can be further expressed as:

$$\min_{A,K} rank\left(A\right) + \gamma E_0, s.t. D = A + E \tag{7}$$

Among them, *A* is the low-rank part of *D*, E_0 is the zero norm of the matrix which is the algebraic sum of non-zero elements in the matrix, γ is the weight between rank of matrix *A* and sparse matrix E. Wright [34] et al. proposed to replace the rank of the matrix with a kernel norm and replaced norm with 1-norm, so that the original non-convex problem is converted into a convex function problem:

$$\min_{A,E} D_* + \gamma E_1, s.t. D = A + E \tag{8}$$

Among them, D_* represents the norm of the matrix core which is the algebraic sum of all the eigenvalues, E_1 denotes the matrix norm which is an absolute value of the algebraic sum of all elements in the matrix.

Lin [35] et al. proposed exact augmented Lagrange multipliers (EALM) to solve the Equation (7) using the Augmented Lagrange multipliers (ALM).

Defining $\begin{cases} X = (D, E) \\ f(X) = D_* + \gamma E_1 \\ h(x) = A - E - D \end{cases}$

The Lagrangian function is:

$$L\left(D, E, Y, \mu\right) = D_* + \gamma E_1 + Y, A - D - E + \frac{\mu}{2} A - D - E_F^2 \tag{9}$$

2.2. Least Squares Support Vector Machine (LSSVM)

Given a set of training data samples $(x_i, y_i)^N$, among which $x_i \in R^m$ is the m-dimensional data samples, $y_i \in R$ is a sample output, and the LSSVM optimization algorithm is as follows:

$$\min J = \frac{1}{2} w^T w + \frac{1}{2} r \sum_{i=1}^{N} e_i^2$$

$$s.t. y_j = w^T \varphi\left(x_i\right) + b + e_i, i = 1, 2 \ldots n \tag{10}$$

Among them, $\varphi(x_i): R^m \to R^{mf}$ is the mapping function from original space to high dimensional space, $w \in R^{mf}$ is the weight vector, $e \in R$ is a tolerance, b is the offset, γ is the normalized coefficient.

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

$$L = \frac{1}{2} w^T w + \frac{1}{2} r \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} a_i \left\{ w^T \varphi(x_i) + b + e_i - y_i \right\} \tag{11}$$

Among them, a_i is the Lagrange factor.

The KKT conditions, L derivative can be obtained

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \to w = \displaystyle\sum_{i=1}^{N} \alpha_i \varphi(x_i) \\[2mm] \dfrac{\partial L}{\partial b} = 0 \to \displaystyle\sum_{i=1}^{N} \alpha_i = 0 \\[2mm] \dfrac{\partial L}{\partial e_i} = 0 \to \alpha_i = \gamma e_i, i = 1,2\ldots n \\[2mm] \dfrac{\partial L}{\partial \alpha_i} = 0 \to w^T \varphi(x_i) + b + e_i - y_i = 0, i = 1,2\ldots n \end{cases} \tag{12}$$

After canceling w and e, we obtain the matrix equation:

$$\begin{bmatrix} 0 & I_v^T \\ I_{v\Omega} + \dfrac{I}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{13}$$

Among them, I is the identity matrix, $I_v = [I,\ldots,I]$, $\alpha = [\alpha_1, \alpha_2 \ldots, \alpha_n]^T$, $\Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$

The optimal decision function is:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b \tag{14}$$

2.3. PSO-LSSVM Optimized by the Tabu Search Algorithm

2.3.1. Particle Swarm Optimization Algorithm

The particle swarm optimization algorithm designs a set of particles to mimic the flock of birds searching for food in the defined domain, with each particle corresponding to a solution. Particle size of the group is N, and the position of the particle is x_i. The optimal position it experienced is the optimal solution, which is recorded as the individual extreme value $pBest_i$. Therefore, the optimal position of the population is recorded as $gBest$. The particle swarm optimization problem can be calculated as follows:

$$\begin{aligned} \min f(\mathrm{x}) &= f(x_1, x_2 \ldots x_n) \\ \mathrm{x} \in \mathrm{S} &= \left\{ x = (x_1, x_2 \ldots x_n) \mid a_i < x_i < b_i, i = 1,2\ldots n \right\} \end{aligned} \tag{15}$$

Equation (15) will continue to be optimized by Equation (16)

$$v_i = \omega_1 \times v_i + \eta_1 \times rand(\ ) \times (\text{p}Best_i - x_i) + \eta_2 \times rand() \times (\text{g}Best - x_i)$$
$$x_i = x_i + v_i$$

(16)

Among them, v_i is flight speed, $rand()$ is a random number between 0 and 1, ω_1, η_1, η_2 are learning factors.
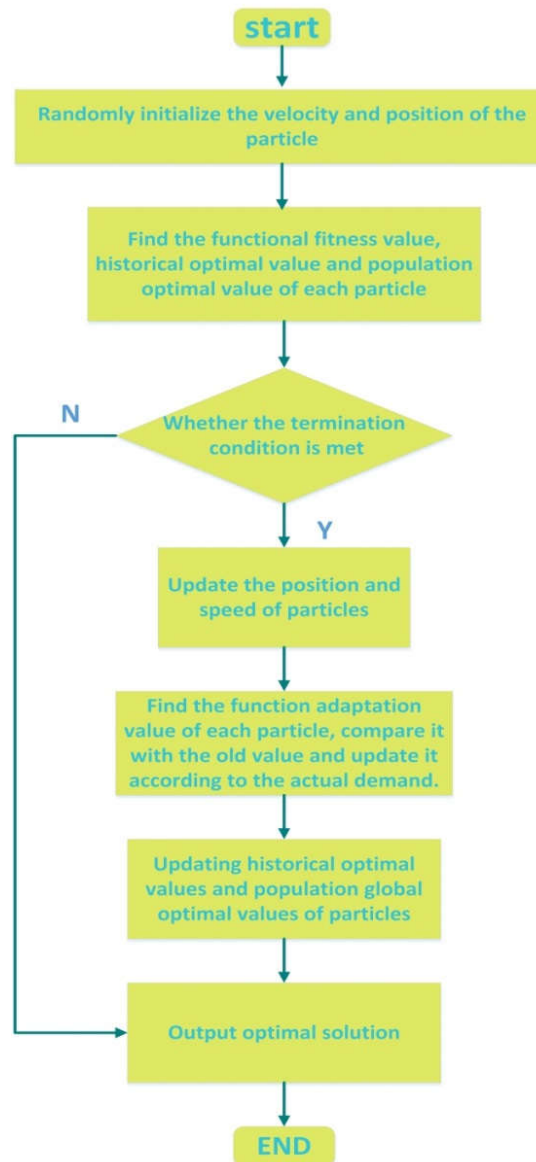
PSO algorithm flow is shown in Figure 1.

Figure 1. Flow chart of particle swarm optimization (PSO) algorithm.

2.3.2. Particle Swarm Optimization Algorithm Optimized by Tabu Search (TS-PSO)

The traditional PSO algorithm has the advantages of easy operation and simple parameters, but it also has the disadvantages of single population, premature convergence, and easily falling into local optimal solutions. This article solved the shortcomings of traditional particle swarms by applying Tabu search tables to store the optimal and worst particles. Tabu Search is an algorithm proposed by Glover in 1986 [28]. After each search is completed, the optimal solution is marked to prevent it from falling into the local optimum. At present, many scholars have further studied it and introduced TS into the PSO algorithm.

The flow of the TS-PSO algorithm is as follows:

1. Initializing the speed and position of the particle.
2. The initial particle swarm is divided into two subgroups, and the class-free scaled network model is used to calculate the fitness value of each particle, then the better particles of the two subgroups are compared.
3. Updating the particle velocity and position according to Formula (16).
4. Assume that after R iterations, the worst and best particle adaptation values are substantially unchanged. The position of the optimal particle is stored as the current best position P_j in the local contraindication table (Table 1), and the position of the worst particle is stored in Table 2 as the current worst position G_j.
5. Repeat steps 3 and 4.
6. Determine whether the termination condition is met, and exit if it is satisfied, otherwise return to the second step.

2.4. Least Squares Support Vector Machine Optimized by the TS-PSO Algorithm

The parameter of regularization and the width of the radial basis function should be determined before using the Least Squares Support Vector Machine with RBF kernel function to forecast energy consumption. In this article, the TS-PSO algorithm was used to optimize the parameters of LSSVM. The steps are as follows:

(1) Perform TS-PSO algorithm steps 1–6.
(2) Assign the optimized parameters to Least Squares Support Vector Machine for constructing the forecasting model.

The flowchart of Least Squares Support Vector Machine Optimized by TS-PSO Algorithm (TS-PSO-LSSVM) is shown in Figure 2.

According to the above analysis, the difference between the parameter settings of TS-PSO-LSSVM and PSO-LSSVM, LSSVM, and other traditional forecasting models can be clarified, as shown in Table 1.

Table 1. Comparison of parameter settings for various models.

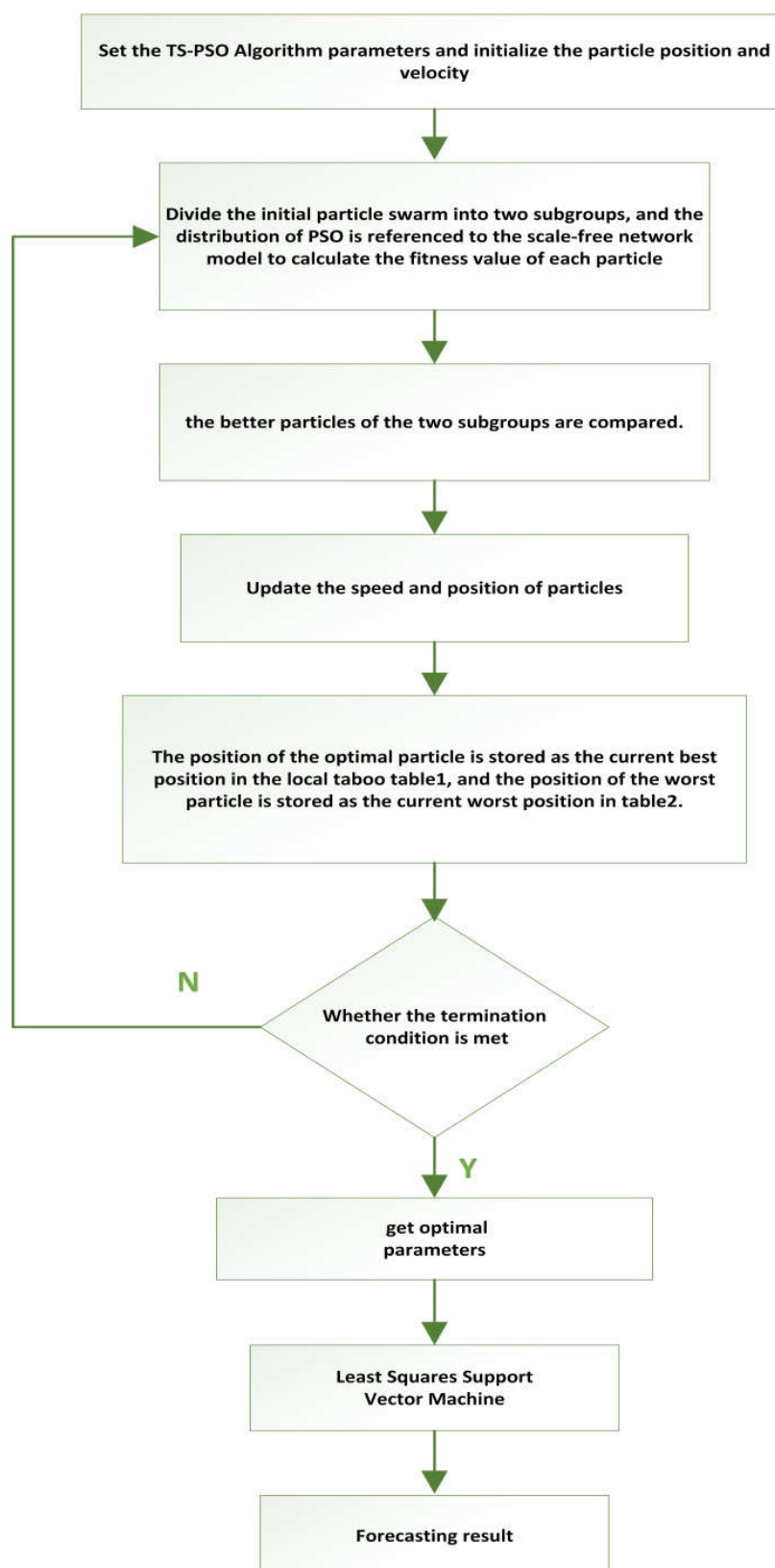| Parameter	LSSVM	PSO-LSSVM	TS-PSO-LSSVM
Regularization parameter c	✓	✓	✓
Radial basis kernel function parameter g	✓	✓	✓
Bird group inertia factor ω		✓	✓
Learning factor c_1, c_2		✓	✓
Maximum speed v_{max}		✓	✓
The maximum number of iterations R		✓	✓
Taboo search step size t			✓

Figure 2. The flow chart of Least Squares Support Vector Machine Optimized by TS-PSO Algorithm.

2.5. The Forecasting Model Based on Robust Principal Component Analysis and Least Squares Support Vector Machine Optimized by TS-PSO Algorithm (RPCA-TS-PSO-LSSVM)

Energy consumption is influenced by a multitude of direct or indirect factors. We firstly selected the GDP, population, industrial structure and energy consumption structure, and urbanization as the

main affecting factors according to RPCA, which can improve data availability and forecasting accuracy. Then, we proposed the RPCA-TS-PSO-LSSVM forecasting model to forecast the energy consumption of China. The prediction process is shown in Figure 3:
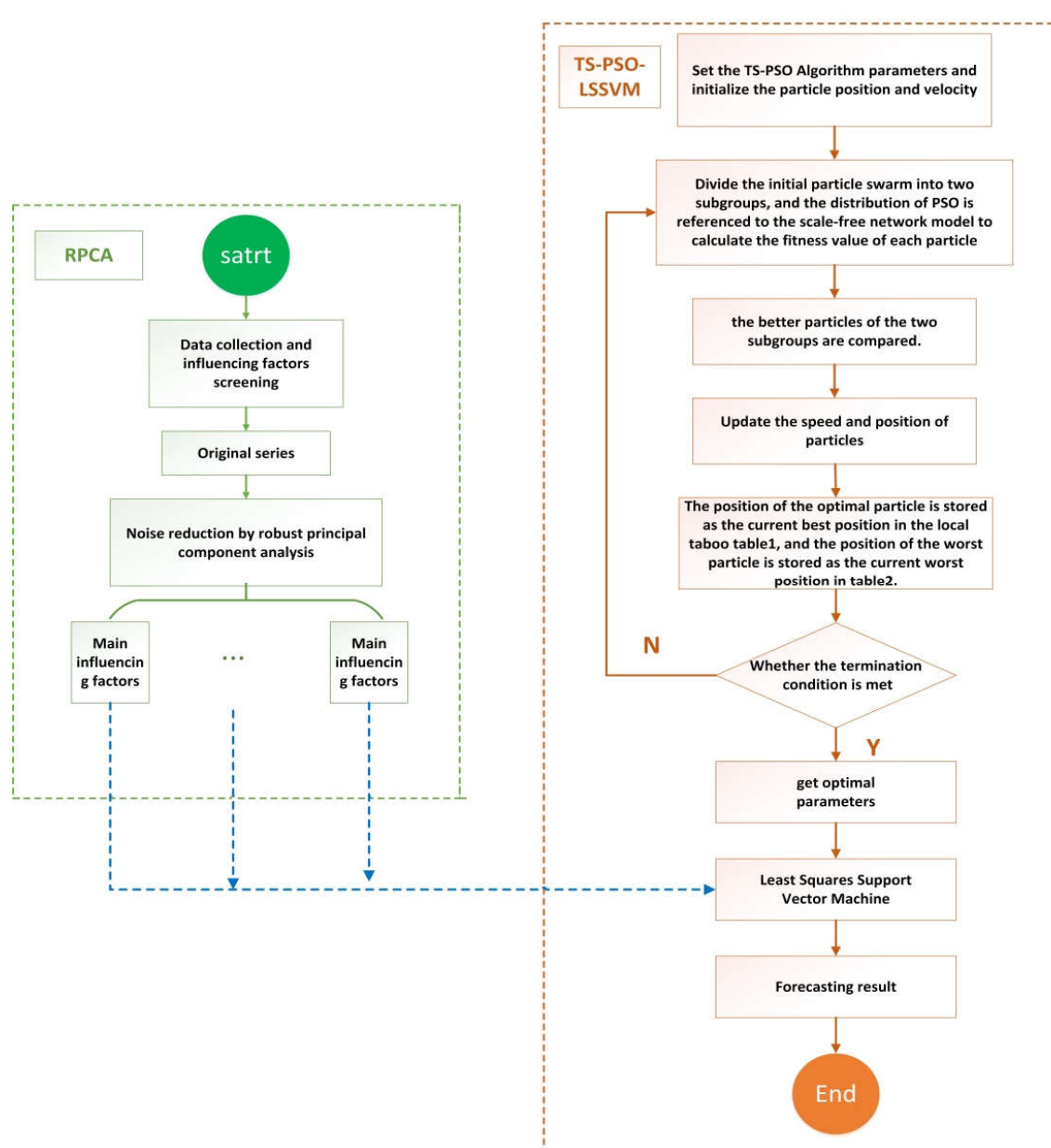

Figure 3. The forecasting process based on the RPCA-TS-PSO-LSSVM (Robust Principal Component Analysis and PSO-LSSVM Optimized by the Tabu Search) model.

3. Empirical Analysis

3.1. Screening of Influencing Factors for Model Input

According to the literature study and China Statistical Yearbook, we selected GDP, population, industrial structure and energy consumption structure, energy intensity, total imports and exports, fixed asset investment, energy efficiency, urbanization, the level of consumption, and fixed investment in the energy industry as a set to input into the RPCA model to achieve hierarchical clustering.

3.2. Data Preprocessing

Because the scale of the experimental test data is not uniform, directly classifying according to the RPCA algorithm will result in unsatisfactory classification. The obtained raw data must be preprocessed to eliminate the non-uniformity of dimension and scale. The preprocessing method adopted in this article is a normalization method commonly used in clustering algorithms to scale the data to a small range. According to the meaning of the diagonal element representation of the covariance matrix, the data is averaged, and the difference in the degree of change of each variable can be reflected by the diagonal elements of the covariance matrix, and the original data is averaged and the variables are inter-variable. Meanwhile, the relevance of the data still exists.

$$Z_{ij} = \frac{x_{ij}}{u}(i = 1, 2 \ldots m; j = 1, 2 \ldots m) \tag{17}$$

$$u = \frac{1}{n}\sum_{i=1}^{n}x_i \tag{18}$$

Among them, u is the mean, x_{ij} is the normalized sample data.

3.3. Hierarchical Clustering according to RPCA

In this article, the cumulative contribution rate of the main components of the two-dimensional data was compared and is shown in Table 2.

Table 2. Comparison between robust principal component analysis (RPCA) and principal component analysis (PCA).

Instructions	The First Two-Dimensional Principal Component Contribution Rate
PCA	92.7%
RPCA	99.8%

It can be concluded from the above analysis that the traditional PCA algorithm can't effectively classify the test data, indicating that the correlation between the samples is not high, and the first two principal elements selected are not representative. The RPCA algorithm is obviously superior to traditional PCA, and the classification effect is more accurate. The comprehensive preservation of the original sample information can comprehensively represent the main information of the sample and solve the information loss problem.

After RPCA clustering, the contribution rates of affecting factors are shown in Table 3.

Table 3. The contribution rates of affecting factors.

Factors	Cumulative Contribution Rate	Individual Variance Contribution Rate
GDP	99.85%	38.39%
population	61.46%	30.23%
Industrial structure	31.23%	15.44%
Energy consumption structure	15.79%	10.61%
Urbanization rate	5.18%	3.13%
Energy intensity	2.05%	0.41%
Total import and export	1.64%	0.39%
Investment in fixed assets	1.25%	0.36%
Energy efficiency	0.89%	0.31%
Energy industry fixed investment	0.58%	0.30%

level of consumption	0.28%	0.28%

According to the above, a scree plot of the influencing factors is shown in Figure 4.

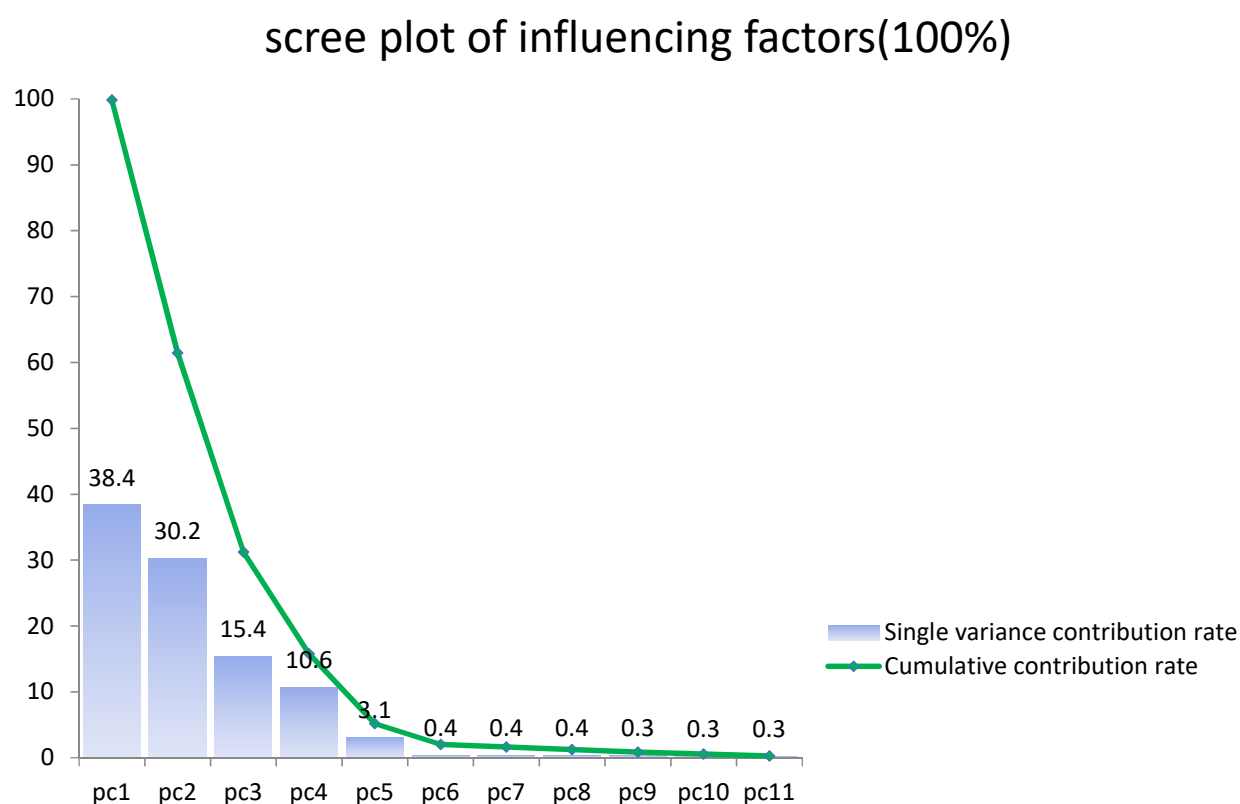## scree plot of influencing factors(100%)

Figure 4. Scree plot of the influencing factors.

According to the RPCA analysis results, GDP, population, industrial structure, energy consumption structure, and urbanization rate are the main components of energy consumption which are representative of the sample information. Based on the China National Statistical Yearbook, the data of the above influencing factors of 1996 to 2016 are normalized and shown in Table 4.

Table 4. The main factors' normalized results.

Year	GDP	Population	Industrial Structure	Energy Consumption Structure	Urbanization Rate
1996	0.011186	0.044472	0.057085	0.050384	0.034561
1997	0.012416	0.044922	0.05424	0.048944	0.035611
1998	0.01327	0.045334	0.054791	0.048602	0.036672
1999	0.014106	0.045706	0.052221	0.048396	0.037755
2000	0.01562	0.046054	0.054699	0.046956	0.038848
2001	0.017268	0.046375	0.042584	0.046614	0.040169
2002	0.018959	0.046676	0.045338	0.046956	0.041609
2003	0.021405	0.046957	0.053139	0.048122	0.043071
2004	0.025208	0.047233	0.04754	0.048122	0.044544
2005	0.029177	0.047512	0.046347	0.04963	0.046038
2006	0.03418	0.047778	0.045613	0.04963	0.047499
2007	0.042092	0.048011	0.04598	0.049698	0.048939
2008	0.049767	0.048256	0.044604	0.049013	0.05039

2009	0.054373	0.048491	0.047999	0.049081	0.05182
2010	0.064334	0.048724	0.05268	0.047436	0.053346
2011	0.076214	0.048958	0.047724	0.048122	0.054765
2012	0.084168	0.049201	0.045797	0.046956	0.056183
2013	0.092716	0.049444	0.044512	0.046202	0.057569
2014	0.100306	0.049702	0.043869	0.044968	0.058911
2015	0.107328	0.049949	0.038913	0.043666	0.060222
2016	0.115906	0.050243	0.034325	0.042501	0.061477

### 3.4. Forecasting Energy Consumption in China Based on TS-PSO-LSSVM Model

We used the outputs of the RPCA analysis as the input for three types of forecasting models. In the models, the data of 1996–2009 were used as the training set and the data of 2010–2016 were used as the test set. In order to verify that the TS-PSO-LSSVM model has high prediction accuracy, we also inputted the sample data into the traditional LSSVM and PSO-LSSVM algorithms. The forecasting results are shown in Figure 5.
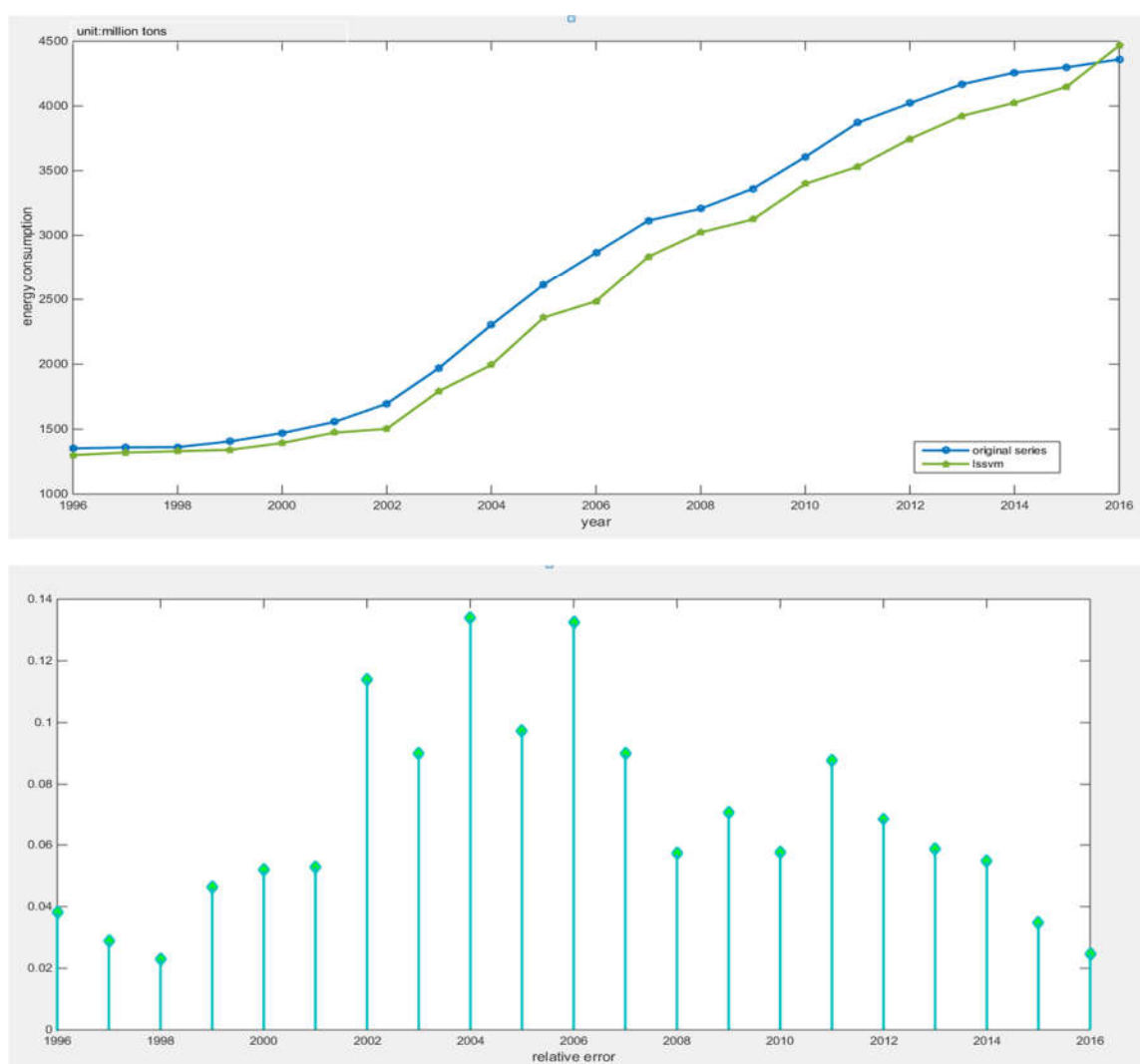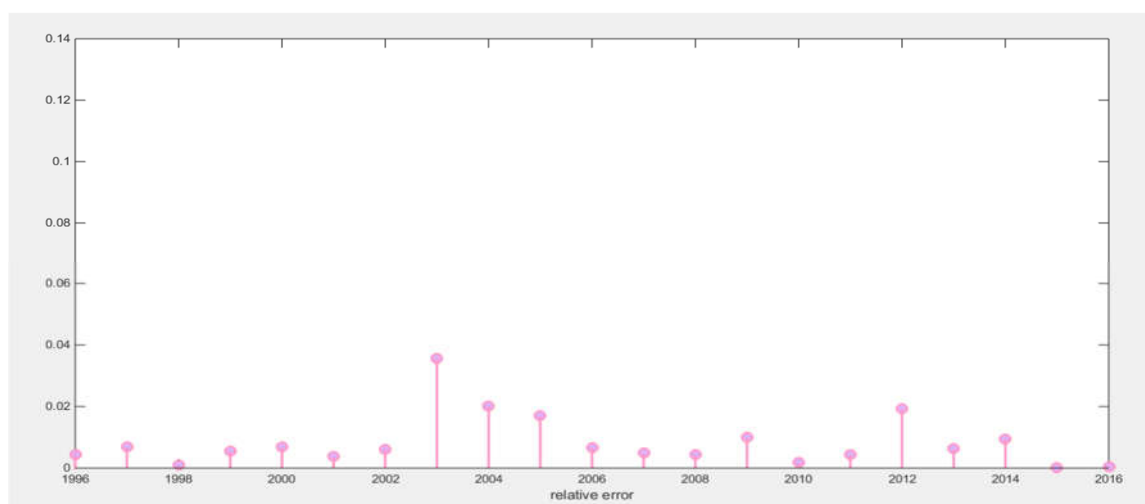

(**a**)
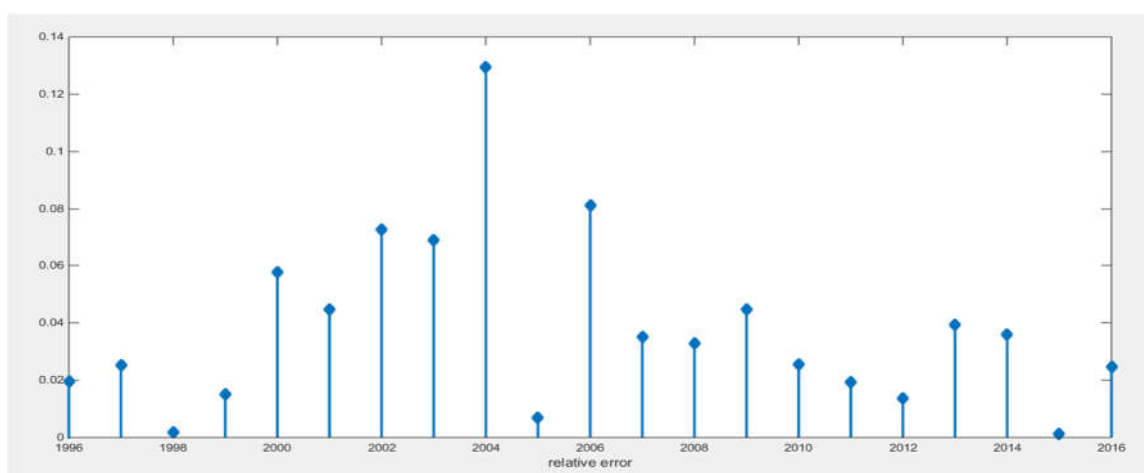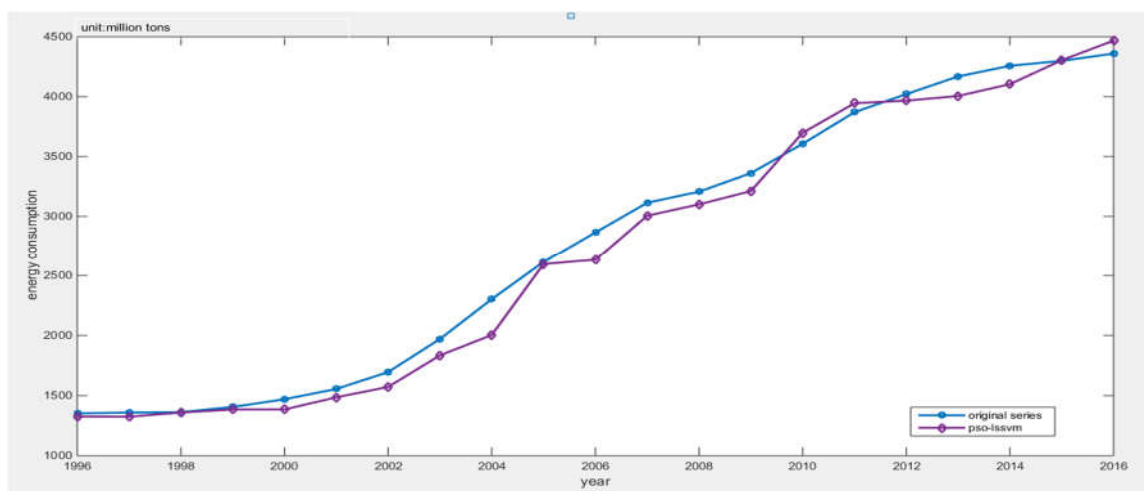
(**b**)
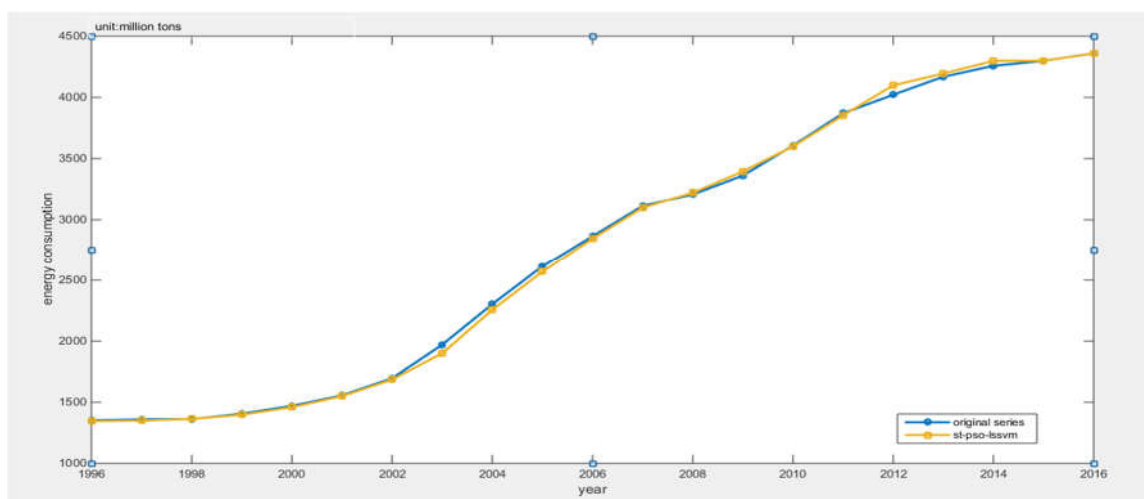
(**c**)

Figure 5. Comparison of the three forecasting models: (**a**) forecasting results of LSSVM model; (**b**) forecasting result of PSO-LSSVM model; (**c**) forecasting result of TS-PSO-LSSVM model.

The correlation error (RE) of the three forecasting results are shown in Table 5.

Table 5. Correlation error (RE) of three forecasting results.

Year	Actual Data	LSSVM		PSO-LSSVM		TS-PSO-LSSVM	
		Forecasting Results	RE	Forecasting Results	RE	Forecasting Results	RE
1996	1351.92	1300.23	0.0382	1325.64	0.0194	1345.85	0.0045
1997	1359.09	1319.84	0.0289	1324.96	0.0251	1349.9	0.0068
1998	1361.84	1330.45	0.0230	1359.32	0.0019	1363.2	0.0010
1999	1405.69	1340.64	0.0463	1384.56	0.0150	1397.83	0.0056
2000	1469.64	1393.42	0.0519	1384.98	0.0576	1459.32	0.0070
2001	1555.47	1473.23	0.0529	1485.96	0.0447	1549.69	0.0037
2002	1695.77	1502.34	0.1141	1572.24	0.0728	1685.33	0.0062
2003	1970.83	1793.4	0.0900	1834.59	0.0691	1900.43	0.0357
2004	2302.81	1994.32	0.1340	2004.55	0.1295	2256.43	0.0201
2005	2613.69	2359.32	0.0973	2595.83	0.0068	2569.43	0.0169
2006	2864.67	2485.33	0.1324	2632.13	0.0812	2845.33	0.0068
2007	3114.42	2834.55	0.0899	3005.44	0.0350	3099.45	0.0048
2008	3206.11	3022.34	0.0573	3100.45	0.0330	3220.5	0.0045
2009	3361.26	3123.58	0.0707	3211.31	0.0446	3394.56	0.0099
2010	3606.48	3398.52	0.0577	3698.34	0.0255	3600.02	0.0018
2011	3870.43	3530.96	0.0877	3945.22	0.0193	3853.42	0.0044
2012	4021.38	3745.62	0.0686	3966.58	0.0136	4098.55	0.0192
2013	4169.13	3924.45	0.0587	4005.64	0.0392	4194.99	0.0062
2014	4258.06	4024.55	0.0548	4104.54	0.0361	4298.54	0.0095
2015	4299.05	4149.64	0.0348	4304.65	0.0013	4299.95	0.0002
2016	4360	4466.93	0.0245	4467.87	0.0247	4358.43	0.0004

In order to objectively compare the accuracy of three models, statistical indicators including RMSE, r^2, and MRE were adopted in the article and the index calculation formula was as follows:

$$\text{rmse} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{q}_i - q_i\right)^2} \tag{19}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{q}_i - \overset{-}{q}_i\right)^2}{\sum_{i=1}^{n}\left(q_i - \overset{-}{q}_i\right)^2} \tag{20}$$

$$\text{mre} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{q}_i - q_i\right|}{q_i}\times 100\% \tag{21}$$

where q_i is a real value, $\hat{q}_i$ is a predicted value, $\overline{q}_i$ is a sample mean, n is a sample number.

The calculation results of objectively three forecasting models are compared in Table 6.

Table 6. The calculation results of the three models.

Model	RMSE (100%)	r^2 (100%)	MRE
LSSVM	21.22	99.48	6.73
PSO-LSSVM	12.27	99.62	3.79
TS-PSO-LSSVM	3.087	99.97	0.83

Furthermore, boxplots of the results are shown in Figure 6.
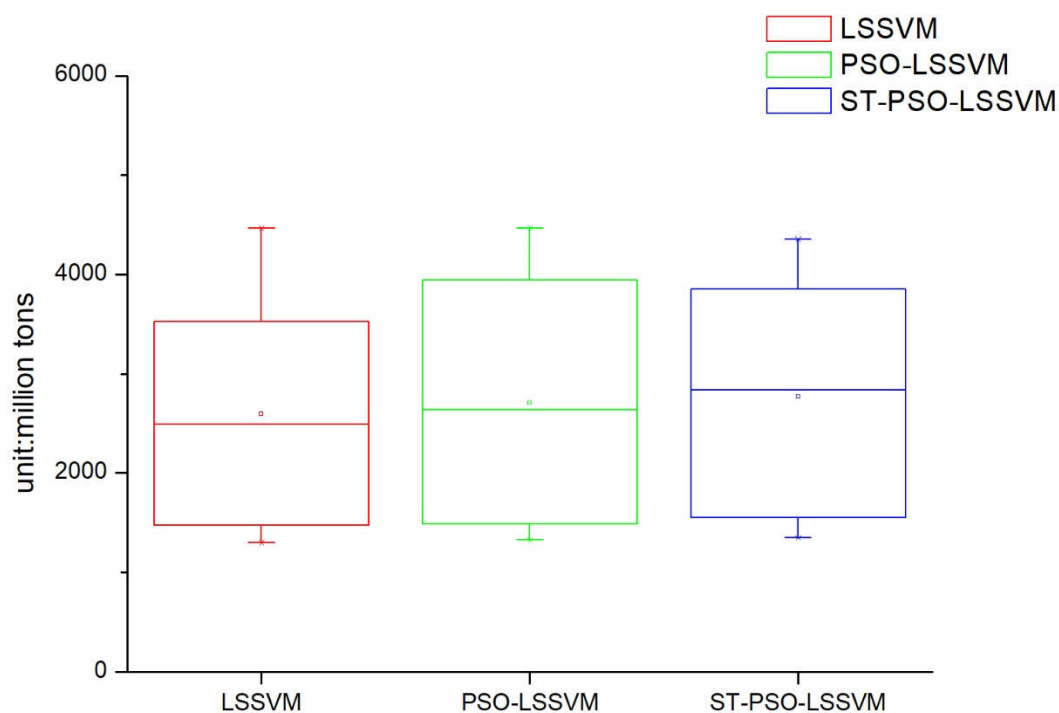
Figure 6. Boxplots of the three forecasting models' results.

Through the analysis of the results in Figures 5 and 6 and Tables 5 and 6, we can find that the TS-PSO-LSSVM forecasting model has higher prediction accuracy from different directions. Based on relative error, the TS-PSO-LSSVM forecasting model not only maintains a relatively low relative error, but also has a small dispersion between relative errors, which shows a high degree of stability. From the perspective of mathematical statistics, we find that TS-PSO-LSSVM< PSO-LSSVM <LSSVM from the perspective of RMSE, which indicates the TS-PSO-LSSVM model in this article has the best performance from the perspective of degree of dispersion. The prediction results have higher robustness, which can also be proved by the boxplots shown in Figure 6; TS-PSO-LSSVM> PSO-LSSVM > LSSVM from the perspective of r^2 indicator; TS-PSO-LSSVM < PSO-LSSVM < LSSVM from the perspective of MRE indicator, these two indicators jointly illustrate that the proposed TS-PSO-LSSVM model has higher prediction accuracy.

Another important indicator for comparing the superiority of machine learning algorithms is the training time of the model. The shorter the training time, the higher the calculation speed, and the superiority of the algorithm can be reflected when a large amount of data is encountered. This article compared the training time of the three forecasting models, as shown in Table 7.

Table 7. Training times of the three forecasting models.

Forecasting Model	Training Time(s)	Total Time(s)
TS-PSO-LSSVM	45 s	64 s
PSO-LSSVM	52 s	66 s
LSSVM	64 s	71 s

By comparison, we can find that TS-PSO-LSSVM algorithm proposed in this article reduces the number of repetitive process executions of the PSO algorithm which needs select the optimal position and the poor position, and takes up less resources. Thus, the algorithm enables faster training speeds. At the same time, the accuracy comparison of the forecasting model has also proved that the forecasting model proposed in this article will not fall into the local optimum. Therefore, the superiority of the proposed combination algorithm is proved from the perspective of both accuracy and operation speed.

3.5. Forecasting Results

We applied GM (1, 1) (grey prediction theory) to forecast the GDP, population, industrial structure, energy consumption structure, and urbanization rate from 2017 to 2030, which was used as the input data of the TS-PSO-LSSVM forecasting model. Finally, we calculated the amount of energy consumption from 2017 to 2030, which is shown in Figure 7.
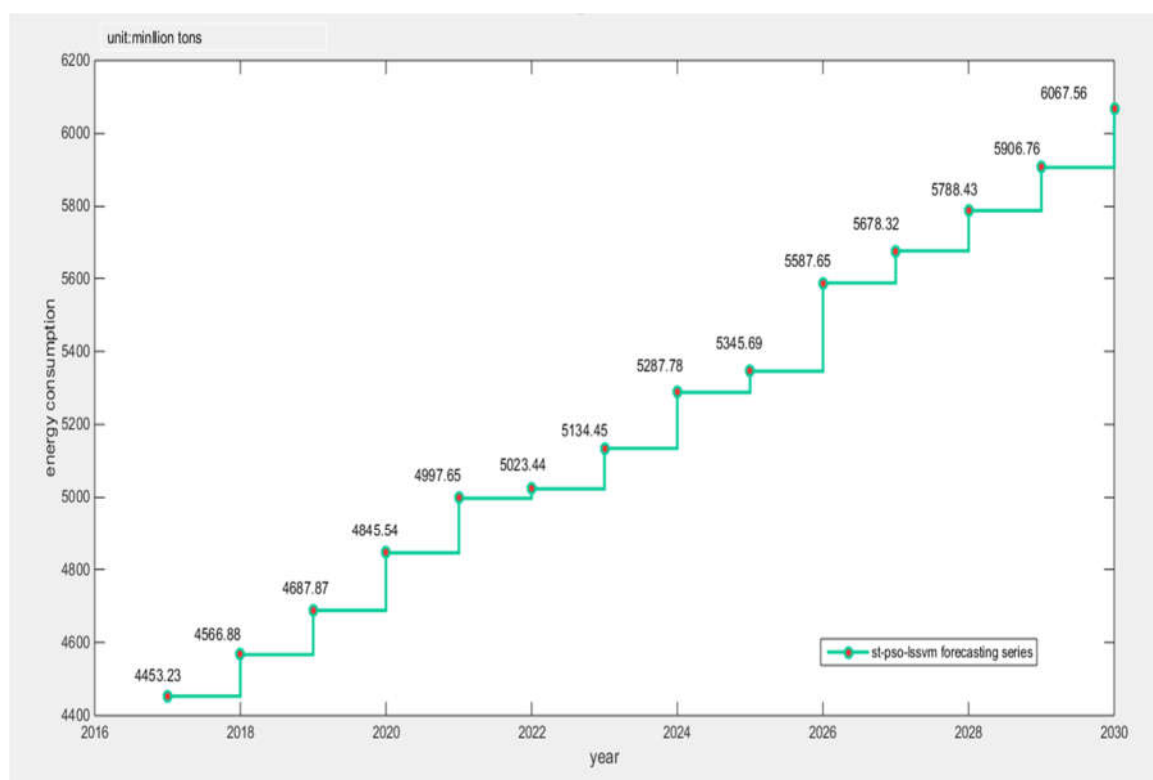
Figure 7. The forecasting results of energy consumption from 2017 to 2030 in China.

4. Conclusions

This article combined the Tabu search algorithm with the PSO-LSSVM algorithm to construct the TS-PSO-LSSVM forecasting model for prediction of China's energy consumption. Because the study of energy consumption is a complex issue and energy consumption is influenced by a multitude of factors, we adopted RPCA to select the main factors of GDP, population, industrial structure and energy consumption structure, urbanization, energy intensity, total imports and exports, fixed asset investment, energy efficiency, energy industry fixed investment, household

consumption, and level of noise reduction. The main influential factors can contain information about other factors, while reducing the complexity of the studied factors. Compared with the traditional PCA, RPCA was proved to be better for generalizing information. After selecting the five main influential factors, we used data from 1996 to 2010 as the training set for the TS-PSO-LSSVM, PSO-LSSVM, and LSSVM forecasting models, and data from 2011 to 2016 as the test set. Then, we compared the results from forecasting of the test set form the perspective of both accuracy and operation speed. Finally, we applied the RPCA-TS-PSO-LSSVM forecasting model to forecast the future energy consumption of China in 2017–2030. We found that China's energy consumption will break through 5000 million tons in 2020, and energy consumption will increase year by year, eventually reaching 6000 million tons in 2030. Our final conclusions and policy recommendations are as follows:

(1) From 2018 to 2030, China's energy consumption shows a gradual upward trend, but the growth rate is gradually tightening. From the perspective of technological progress, this forecast proves that China's energy efficiency will increase year by year.

(2) China's energy consumption economy will transfer into the stage of diminishing returns around 2026. At that time, excessive energy investment will not bring about sustained GDP growth. Therefore, China should give priority to improving energy efficiency in the future, and continue to develop renewable energy technologies. Meanwhile, China needs to look for better opportunities in the energy investment field and continue to reduce pollutants and carbon emissions.

(3) China should pay more attention to the development of energy consumption terminal power for energy conservation and emissions reduction of the international community. At the same time, in the field of energy investment, China must attach importance to the development of the Belt and Road and grasp the opportunities of international cooperation in energy-based society. According to the forecasting results, the energy strategy of the next ten years must be carefully planned.

The forecasting results and corresponding conclusions drawn in this article have laid a strong foundation for our future research, especially the research hotspots of China's carbon emission related to energy consumption and China's investment ability in Belt and Road countries.

References

1. BP Statistical Review of World Energy. Available online: https://www.bp.com/content/dam/bp/en/corporate/pdf/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf (accessed on 30 June 2018).
2. Liang, Z. Effects of different stages of the energy consumption of urbanization factors. *J. Shanghai Univ. Financ. Econ.* **2010**, *5*, 89–96.
3. Xu, S.; He, Z.; Long, R. Factors that influence carbon emissions due to energy consumption in China: Decomposition analysis using LMDI. *Appl. Energy* **2014**, *127*, 182–193.
4. Omri, A. CO2 emissions, energy consumption and economic growth nexus in MENA countries: Evidence from simultaneous equations models. *Energy Econ.* **2013**, *40*, 657–664.
5. Poumanyvong, P.; Kaneko, S. Does urbanization lead to less energy use and lower CO2 emissions? A cross-country analysis. *Ecol. Econ.* **2010**, *70*, 434–444.
6. Zhou, X.; Zhang, J.; Li, J. Industrial structural transformation and carbon dioxide emissions in China. *Energy Policy* **2013**, *57*, 43–51.

7. Meng, F.; Liu, Y.; Liu, L.; Li, Y.; Wang, F. Studies on Mathematical Models of Wet Adhesion and Lifetime Prediction of Organic Coating/Steel by Grey System Theory. *Materials* **2017**, *10*, 715.

8. Lee, Y.; Tong, L. Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Convers. Manag.* **2011**, *52*, 147–152.

9. Lin, C.; Liou, F.; Huang, C. Grey forecasting model for CO2 emissions: A Taiwan study. *Adv. Mater.* **2011**, *88*, 3816–3820.

10. Ghaedi, M.; Rahimi, M.; Ghaedi, A.M.; Tyagi, I.; Agarwal, S.; Gupt, V.K. Application of least squares support vector regression and linear multiple regression for modeling removal of methyl orange onto tin oxide nanoparticles loaded on activated carbon and activated carbon prepared from Pistacia atlantica wood. *J. Colloid Interface Sci.* **2016**, *461*, 425–434.

11. Sehgal, V.; Tiwari, M.K.; Chatterjee, C. Wavelet Bootstrap Multiple Linear Regression Based Hybrid Modeling. *Water Resour. Manag.* **2014**, *10*, 2793–2811.

12. Cogoljevic, D.; Gavrilovic, M.; Roganovic, M.; Matic, I.; Piljan, I. Analyzing of consumer price index influence on inflation by multiple linear regression. *Phys. A Stat. Mech. Appl.* **2018**, *505*, 941–944.

13. Yi, M.; Qiao, M.; Ying, F.; Okada, N.; Tsai, H.T. A scenario analysis of energy requirements and energy intensity for China's rapidly developing society in the year 2020. *Technol. Forecast. Soc. Chang.* **2006**, *73*, 405–421.

14. Zhu, Y.; Zheng, W.; Li, P.; Wang, L.; Zou, X. Simulation on China's Economy and Prediction on Energy Consumption and Carbon Emission under Optimal Growth Path. *Acta Geogr. Sin.* **2009**, *64*, 935–944.

15. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175.

16. Erdogdu, E. Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. *Energy Policy* **2007**, *35*, 1129–1146.

17. Luong, H.V.; Deligiannis, N.; Seiler, J.; Forchhammer, S.; Kaup, A. Compressive Online Robust Principal Component Analysis via n-l1 Minimization. *IEEE Trans. Image Process.* **2018**, *27*, 4314–4329.

18. Chrétien, S.; Clarkson, P.; Garcia, M.S. Application of Robust PCA with a structured outlier matrix to topology estimation in power grids. *Int. J. Electr. Power Energy Syst.* **2018**, *100*, 559–564.

19. Sadeghian, A.; Wu, O.; Huang, B. Robust probabilistic principal component analysis based process modeling: Dealing with simultaneous contamination of both input and output data. *J. Process. Control* **2018**, *67*, 94–111.

20. Wu, M.; Sun, Y.; Hang, R.; Liu, Q.; Liu, G. Multi-component group sparse RPCA model for motion object detection under complex dynamic background. *Neurocomputing* **2018**, *314*, 120–131.

21. Roushangar, K.; Saghebian, S.M.; Mouaze, D. Predicting characteristics of dune bedforms using PSO-LSSVM. *Int. J. Sediment Res.* **2017**, *32*, 515–526.

22. Huan, J.; Cao, W.; Qin, Y. Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. *Comput. Electron. Agric.* **2018**, *150*, 257–265.

23. Xue, X. Evaluation of concrete compressive strength based on an improved PSO-LSSVM model. *Comput. Concr.* **2018**, *21*, 501–511.

24. Lu, P.; Ye, L.; Su, B.; Zhang, C.; Zhao, Y.; Teng, J. A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA. *Energies* **2018**, *11*, 697.

25. Zhao, H.; Huang, G.; Yan, N. Forecasting energy-related CO2 emissions employing a novel SSA-LSSVM model: Considering structural factors in China. *Energies* **2018**, *11*, 781.

26. Wen, T.; Tang, H.; Wang, Y.; Lin, C.; Xiong, C. Landslide displacement prediction using the GA-LSSVM model and time series analysis: A case study of Three Gorges Reservoir, China. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 2181–2198.

27. Liu, C.; Niu, P.; Li, G.; You, X.; Ma, Y.; Zhang, W. A Hybrid Heat Rate Forecasting Model Using Optimized LSSVM Based on Improved GSA. *Neural Process. Lett.* **2017**, *45*, 299–318.

28. Gorjaei, R.G.; Songolzadeh, R.; Torkaman, M.; Safari, M.; Zargar, G. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. *J. Nat. Gas Sci. Eng.* **2015**, *24*, 228–237

29. Zhang, X. Short-Term Load Forecasting for Electric Bus Charging Stations Based on Fuzzy Clustering and Least Squares Support Vector Machine Optimized by Wolf Pack Algorithm. *Energies* **2018**, *11*, 1449.

30. Peng, B.; Lü, Z.; Cheng, T.C.E. A tabu search/path relinking algorithm to solve the job shop scheduling problem. *Comput. Oper. Res.* **2015**, *53*, 154–164.

31. Escobar, J.W.; Linfati, R.; Toth, P.; Baldoquin, M.G. A hybrid Granular Tabu Search algorithm for the Multi-Depot Vehicle Routing Problem. *J. Heuristics* **2014**, *20*, 483–509.

32. Li, X.; Gao, L. An effective hybrid genetic algorithm and tabu search for flexible job shop scheduling problem. *Int. J. Prod. Econ.* **2016**, *174*, 93–110.

33. Sicilia, J.A.; Quemada, C.; Royo, B.; Escuín, D. An optimization algorithm for solving the rich vehicle routing problem based on Variable Neighborhood Search and Tabu Search metaheuristic *J. Comput. Appl. Math.* **2016**, *291*, 468–477.

34. Wright, J.; Ma, Y.; Candes, E.J.; Li, X. Robust Principal Component Analysis? *J. ACM* **2011**, *58*, 11.

35. Lin, Z.; Chen, M.; Ma, Y. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. Available online: https://arxiv.org/pdf/1009.5055.pdf (accessed on 2 June 2018).