Tech Science Press

# Research on Prediction Methods of Energy Consumption Data

**Ning Chen[1], Naernaer Xialihaer[2, 3] and Weiliang Kong[3] and Jiping Ren[2,3,*]**

[1]Xinjiang Research Institute of Building Sciences (Co., Ltd.), Urumqi, 830000, China
[2]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, 100083, China
[3]Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing, 100083, China
[*]Corresponding Author: Jiping Ren. Email: renjiping@ustb.edu.cn

**Abstract:** This paper analyzes the energy consumption situation in Beijing, based on the comparison of common energy consumption prediction methods. Here we use multiple linear regression analysis, grey prediction, BP neural net-work prediction, grey BP neural network prediction combined method, LSTM long-term and short-term memory network model prediction method. Firstly, before constructing the model, the whole model is explained theoretically. The advantages and disadvantages of each model are analyzed before the modeling, and the corresponding advantages and disadvantages of these models are pointed out. Finally, these models are used to construct the Beijing energy forecasting model, and some years are selected as test samples to test the prediction accuracy. Finally, all models were used to predict the development trend of Beijing's total energy consumption from 2018 to 2019, and the relevant energy-saving opinions were given.

**Keywords:** Energy consumption; multiple linear regression; grey prediction; LSTM

## 1 Introduction

Forecasting energy consumption may make the government more convenient and more accurate analysis and evaluation of the city's energy consumption and regulation. By analyzing the overall state of energy consumption, energy consumption can be reduced, and it is determined through actual data analysis and research to determine why there is such a certain amount of energy consumption to determine the factors that contribute to the favorable and unfavorable factors of energy consumption. The next step is to reduce energy consumption.

Beijing is a city with relatively scarce energy resources. The dependence on energy is very high. And with the continued development of the economy, the contradiction between energy supply and demand in Beijing will become more and more prominent. Therefore, how to study Beijing's energy consumption de-mand and predict the future energy consumption of Beijing is imminent. Based on the data of the total energy consumption and composition of Beijing in the past 38 years, this paper analyzes the energy classification and analyzes and models the energy consumption types of key energy-using enterprises in Beijing. The total energy consumption is a combination of many complicated processes, and there are many kinds of energy consumption. Generally, the total energy consumption is divided into 39 types, and together with the total energy, there are more than 40 kinds. Its characteristics are many variables, and the relationship is more complicated.

To sum up, this article will explore Beijing's energy consumption patterns and their energy consumption characteristics, domestic and international energy consumption and energy efficiency assessment methods. Late stage, the data law of energy consumption in power, coal, coke, natural gas and liquefied petroleum gas will be analyzed to study energy consumption prediction. The energy consumption

prediction model predicts the energy consumption of the target time period by using regression prediction, neural network and 3 other algorithms in deep learning, and studies the energy consumption data analysis method of the city. The research results also reflect the energy saving direction of the city. There will be corresponding help [1].

## 2 Related Work

Researchers such as Zhang [2] analyzed the error and accuracy of the energy consumption prediction models of several metallurgical enterprises, such as neural network prediction model, genetic algorithm and BP neural network prediction model, discussed the advantages and disadvantages of various methods, and summarized some under construction. The problem you need to pay attention to when modeling.

Ding et al. [3] proposed a short-term energy consumption prediction model based on genetic optimization decision tree. The genetic algorithm was used to optimize the subtree generation process of the gradient decision tree, and the prediction result became this iteration. The evaluation criteria met the predicted goals.

Wang [4] used the method of multiple linear regression model to study the forecast of total output value of Jiangsu construction industry. Based on the data of construction industry in Jiangsu Province from 2000 to 2016, he analyzed a multiple regression model, including wages and electricity. Interpretive variables such as consumption and current assets are important for the province's overall GDP. In terms of inspection, MPE and MAPE (percentage of absolute value of error) were used to explain the percentage, and finally the results of multiple linear regression have better prediction results.

Foreign research scholar use intelligent artificial neural networks to simulate the solar potential of the Souss-Massa region in Morocco, paying special attention to the application of artificial neural networks, especially in the field of renewable energy, especially the prediction of meteorological data such as solar radiation. For this reason, a multi-layer perceptron-based model was developed to predict the evolution of global solar radiation in the Souss-Massa region (southwest Morocco) [5]. In this study, researchers used a large database of a wide range of periods. The database contains a set of metrology and geographic data such as: latitude, longitude, months of the year, average temperature, sunshine duration, relative humidity, and average monthly global solar radiation. After testing a number of models, an appropriate model was given that gave the minimum root mean square error. In addition, the researchers found that they have good correlation coefficients between their measured values and predicted values.

Ramos et al. [6] proposed a combined prediction of gray prediction method and neural network back propagation model, called gray neural network and input-output combination prediction model. The validity of the proposed model is validated using a real energy prediction case. Finally, the model tests show that the proposed model has higher simulation and prediction accuracy in terms of energy consumption than the gray model and other combined methods.

Combining the above learning methods, we can combine the energy consumption situation of Beijing to get corresponding methods to make relevant predictions. A simple theoretical description of the corresponding method will be given in Chapter 3.

## 3 Research Methods and Research Objectives

After analyzing the energy consumption data of Beijing, the data selection time is set as the statistical data from 1980 to 2017. In the establishment of the prediction model, the regression algorithm, gray prediction and neural network combined model in deep learning are used to model the data. Finally, the obtained model is verified and the data is compared, and the prediction accuracy of various models is obtained. The verification algorithm uses MSPE, and finally a total model suitable for Beijing energy consumption prediction is obtained. The historical data predicts the model of energy consumption data for several years in the future, and gives corresponding suggestions on energy consumption through data and some previous data.

### 3.1 Multiple Linear Regression

Regression techniques are used to predict response variables that contain numerical outputs, which may be numeric or categorical. However, what these algorithms have in common is that the output is a numerical value. Linear regression is a particularly old way of predicting statistics, dating back to the work of Carl Friedrich Gauss in 1795. The goal of linear regression is to fit a linear model be-tween the corresponding variable and the independent variable and use it to predict the output of a given independent observed variable. Regression analysis is a method that is not required for sample data and can process and predict mid- and long-term data with high accuracy [7]. The principle is simple, the calculation is simple, and it can be applied in many scopes, but its variables have correlation, and there is a perfect test method for the inspection at the later stage, which ensures the credibility of the prediction data.

### 3.2 Grey Prediction

The grey prediction model is an established model based on certain information and the predicted relationship information is uncertain. The model mainly uses the accumulation method to initialize the data to generate a sequence of data with strong regularity.

### 3.3 Artificial Neural Networks

The artificial neural network becomes a mathematical model that simulates the biological neural network to process information. After conducting physiological research on the brain, they rely on the results of the research as the basis, and the ultimate goal is to stimulate the results of some specific functions by simulating some mechanisms in the brain.

### 3.4 BP Neural Network

BP neural network is actually an algorithm with the function of back propagation algorithm. The characteristic of the algorithm is to directly estimate the error of the output layer error on the output layer by using the error after output, and use the error to estimate the error of the previous layer. The error estimate for each layer can form the process of passing the error presented by the output layer to the input layer of the network in the reverse of the input transmission. The propagation process of BP neural network is divided into two directions: forward propagation process and backward propagation process. It consists of three parts: the input layer, the hidden layer, and the output layer. However, the hidden layer can be a multi-layer or a single layer. The forward propagation process is that the known information is the input layer, the hidden layer is the limit value law, and the output layer is obtained [8]. The backward propagation process mainly uses the gradient descent method for each neuron, and the error is calculated by adjusting the weight, and the error signal is reduced. Small, making fewer error signals.

### 3.5 LSTM Long-Term and Short-Term Memory Neural Network

LSTM is a special kind of cyclic neural network. In order to solve some problems encountered in training the length sequence, such as gradient disappearance and gradient explosion, it can be said that this network can be in long sequence compared with the general cyclic neural network. There is a good performance. The LSTM network adds a state unit $c\_t$ to the hidden layer based on the RNN and controls it with three control switches: the forgetting gate $f\_t$, the input gate $i\_t$, and the output gate $o\_t$. These doors, like the storage in our computer, can perform a series of functions such as storage, writing, and reading [9]. The LSTM modules use these to control the information they flow. Each module has switches to determine storage, read and write, and then use the sigmoid function to determine the output. Here we should note that the output is a standardized number between 0-1 [10].

**4 Energy Consumption Forecasting Model**

*4.1 Definition of the Concept of Total Energy Consumption*

The research object of this paper is the total energy consumption in Beijing. In many current literatures, total energy consumption refers to the sum of various energy sources consumed by a certain region (geographical or administrative region) in daily life and production during a certain period of time, including terminal energy consumption, The energy processing conversion loss, the energy transportation and the loss of management process are three parts. It can be seen that the research object of this paper refers to the sum of the terminal energy consumption, the energy processing conversion loss and the energy loss in the energy transportation and management process consumed by Beijing in a certain period of time [11].

*4.2 Energy Consumption Prediction Model Based on Multiple Linear Regression*

For the multiple linear regression model, there are some aspects that are not excellent. First, we analyze the advantages and disadvantages of the overall model. Regression analysis is simpler and more convenient when analyzing multi-factor models. Using the regression model, as long as the model and data are the same, the standard results can be used to calculate the unique results [12]. However, in the form of graphs and tables, the interpretation of the relationship between the data often varies from person to person, and different analysts draw. The fitting curve is probably also different. However, which factor is used in some regression analyses and what expression to use for this factor is only a speculation, which affects the diversity and unpredictability of certain factors, making regression analysis limited in some cases.

There are many factors that affect the total energy consumption, but the relationship is very complicated. The influencing factors will be the regional industrial structure, some local policy reasons, the economic level of the place, the pricing of energy, and the ability of science and technology. However, many of these factors are non-quantitative factors, so the selection of regression variables in the establishment of predictive models of Beijing energy consumption does not consider these non-quantitative factors. According to the data compiled and comprehensive considerations (the unit of all variables here is 10,000 tons of standard coal), the factors selected in the modeling are as follows:

The representative of the economic level is the GDP data of Beijing in recent years, which is calculated as x1.

The representative of the industrial structure selects the consumption of the tertiary industry, which accounts for a larger proportion of the total production value in recent years, as a data factor, which is calculated as x2.

Select the per capita income level as the influencing factor and record it as x3.

The proportion of urban population to the population can also be seen as a very influential factor, which is considered to be x4.

Y is the actual value of all energy consumption over the years.

Based on the above factors, I first established a Y-multiple linear regression model for Beijing's total energy consumption:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u \tag{1}$$

Data from the sum of Beijing's energy consumption and various explanatory variables from 1980 to 2017 were selected and introduced into STATA15. Using the least squares method to fit the multiple regression model, the following results were obtained.

| Source | SS | df | MS | | Number of obs | = | 38 |
|--------|-----|-----|-----|---|---------------|---|-----|
| | | | | | F(4, 33) | = | 1018.19 |
| Model | 110275589 | 4 | 27568897.2 | | Prob > F | = | 0.0000 |
| Residual | 893520.302 | 33 | 27076.3728 | | R-squared | = | 0.9920 |
| | | | | | Adj R-squared | = | 0.9910 |
| Total | 111169109 | 37 | 3004570.52 | | Root MSE | = | 164.55 |

| Y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|-------|------------|-----------|-------|-------|-----------|-----------|
| x1 | -.1270255 | .0484515 | -2.62 | 0.013 | -.2256007 | -.0284502 |
| x2 | 1.620507 | .1851076 | 8.75 | 0.000 | 1.243903 | 1.997112 |
| x3 | .0337739 | .0185693 | 1.82 | 0.078 | -.0040056 | .0715534 |
| x4 | 43.24889 | 7.158476 | 6.04 | 0.000 | 28.68486 | 57.81292 |
| _cons | -946.5521 | 405.1645 | -2.34 | 0.026 | -1770.866 | -122.2386 |

**Figure 1:** Multivariate linear regression fit graph using Stata15 software

After the multi-collinearity test and correction, heteroscedasticity and auto-correlation test and correction of the above data, we can finally determine the Beijing energy consumption regression model as (Due to space reasons, only the data for nearly ten years is listed here):

**Table 1:** Multivariate regression model for analysis of accuracy

| Year | Actual value | Fitted value | Relative error |
|------|--------------|--------------|----------------|
| 2007 | 5747.7 | 5731.107 | 0.29% |
| 2008 | 5786.2 | 5979.616 | 3.34% |
| 2009 | 6008.6 | 6134.267 | 2.09% |
| 2010 | 6359.5 | 6278.113 | 1.28% |
| 2011 | 6397.3 | 6435.025 | 0.59% |
| 2012 | 6564.1 | 6573.508 | 0.14% |
| 2013 | 6723.9 | 6698.192 | 0.38% |
| 2014 | 6831.2 | 6817.743 | 0.20% |
| 2015 | 6852.6 | 6856.771 | 0.06% |
| 2016 | 6961.7 | 7149.786 | 2.70% |
| 2017 | 7132.8 | 6962.615 | 2.39% |

In the error table of this linear regression, we can see that the error is very large several times, but it was in the 1880s and 1990s. With the passage of time, after 2010, the linear regression model is also very good. The possible error is several years because of the mutation of the explanatory variable or some other reason, but the model can still be approved.

### 4.3 Energy Consumption Prediction Model Based on Grey Prediction and Neural Network

For the statistical model of Beijing's energy consumption from 1980 to 2017, it is divided into the first, second and third industries to calculate the total amount and the gray neural network model for its data. The energy consumption data is as follows: the energy consumption is y, the energy consumption per 10,000 yuan of GDP is recorded as x1 in the table, and the per capita income level is selected as the influencing factor, which is calculated as x2, and the representative of the industrial structure chooses the proportion in recent years. The consumption of the tertiary industry, which accounts for a large proportion of GDP, is counted as x3 as a data factor. Another factor is the choice of the total population of the city [13].

The grey BP model combines the grey prediction model with the BP neural network [14]. We can obtain a new prediction model that contains information about the two prediction models, which can relatively avoid the limitations in the pre-diction of a single model. The specific steps are as follows:

First use the 38-year energy consumption data from Beijing from 1980 to 2017 as the original sequence data. Here we first use the gray prediction method to normalize the energy consumption data. In a neural network, all we have to do is initialize the relative weights of the network and its thresholds, and determine some parameters like the learning rate. The original sequence data is divided into training data and test data (test data does not participate in neural network training), wherein the training data selects the data of 1980-2015, and the test data selects the data of 2016-2017. Get the forecast data.

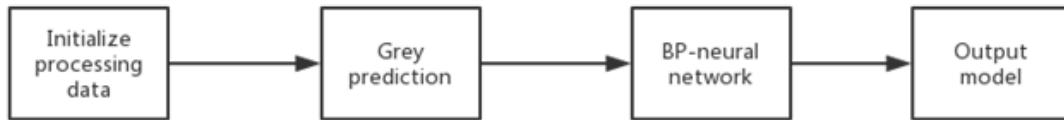Construct a grey BP neural network prediction model, the specific flow chart is as follows:



**Figure 2:** Grey BP neural network model flow

After understanding the flow of the grey BP neural network model, we will begin to construct the model: First, gray prediction for each data. For the number of neurons, five were selected, and the BP neural network prediction model was trained.

Firstly, we standardized the zero-mean data and brought it into the 3-layer neural network prediction model established by Beijing Energy Consumption Forecast (5 layers in the input layer, 10 hidden nodes, and 1 layer in the output layer) to obtain energy consumption in 2019. The value of the quantity is 80,679,300 tons of standard coal.

The following images are a comparison of predicted and actual values predicted by BP neural networks.
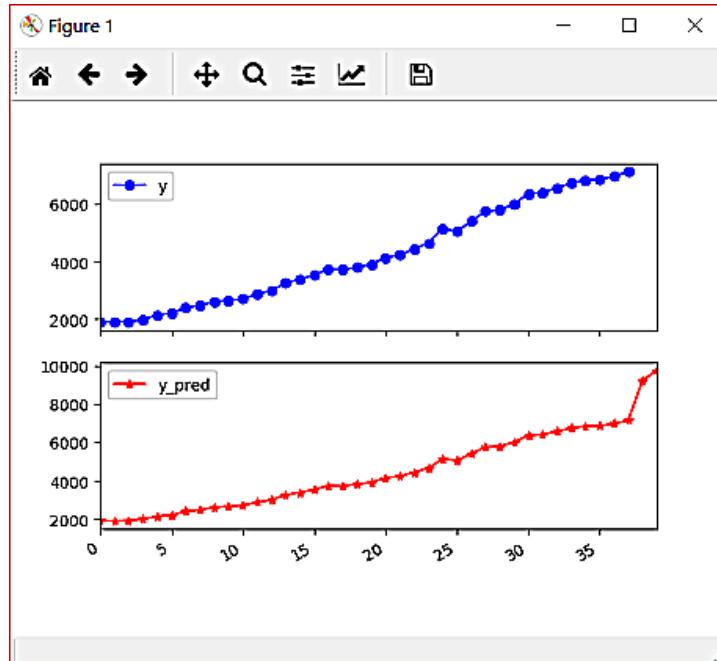


**Figure 3:** Comparison of predicted and actual value curves of BP neural network

After conducting the gray prediction to predict and predict the various characteristic data in 2018 and 2019, the generated data is read again and the BP neural network model is used for training prediction to predict the energy consumption. In the following two years, a comparison table (in 10,000 tons of standard coal) was predicted. Due to space reasons, only the data for nearly ten years is listed here.

**Table 2:** Comparison of actual predicted values after gray-BP neural network

| Year | Actual value | Predictive value | Relative error |
|------|-------------|------------------|----------------|
| 2007 | 5747.7 | 5740.491 | -0.13% |
| 2008 | 5786.2 | 5788.354 | 0.04% |
| 2009 | 6008.6 | 6011.184 | 0.04% |
| 2010 | 6359.5 | 6359.849 | 0.01% |
| 2011 | 6397.3 | 6403.969 | 0.10% |
| 2012 | 6564.1 | 6566.37 | 0.03% |
| 2013 | 6723.9 | 6722.984 | -0.01% |
| 2014 | 6831.2 | 6823.558 | -0.11% |
| 2015 | 6852.6 | 6838.107 | -0.21% |
| 2016 | 6961.7 | 6967.962 | 0.09% |
| 2017 | 7132.8 | 7138.217 | 0.08% |
| 2018 | | 7759.764 | |
| 2019 | | 8069.493 | |

For the prediction error of this model, the maximum and maximum error rate is a relatively long year in 1983. For the prediction of the past five to ten years, the average error rate is 0.1%, with the year. As the trend moves backwards, the performance of this model is getting better and better, and the error rate is getting lower and lower. It can be said that the combination of gray and BP neural net-work models is the best model in these research models.

### 4.4 LSTM-Based Energy Prediction Model

LSTM is an excellent variant of RNN that inherits the features of most RNN models and solves the problem of Vanishing Gradient due to the gradual reduction of the gradient back-propagation process. Specific to language processing tasks, LSTM is very suitable for dealing with problems that are highly correlated with time series. Although on the classification problem, it seems that the feed-forward network represented by CNN still has performance advantages, but LSTM is more in the long run. The potential for complex tasks is not comparable to CNN [15]. It more realistically characterizes or simulates the cognitive processes of human behavior, logical development, and neural organization. Especially since 2014, LSTM has become a very hot research model in RNN and even deep learning framework [16], and has received a lot of attention and research.

First, the energy consumption data is entered into the csv file by year and the file is saved. The first step in the code is to obtain the pre-processed data. The pre-processing of the data will be divided into three steps: grouping the loaded data, normalizing (normalizing) the grouped data, and performing the data set [17]. Division. The LSTM network model is carried out to initially determine the hyperparameters of the network, send the preprocessed data to the network for training, and calculate the loss function and gradient during the training process. The specific process is as follows:
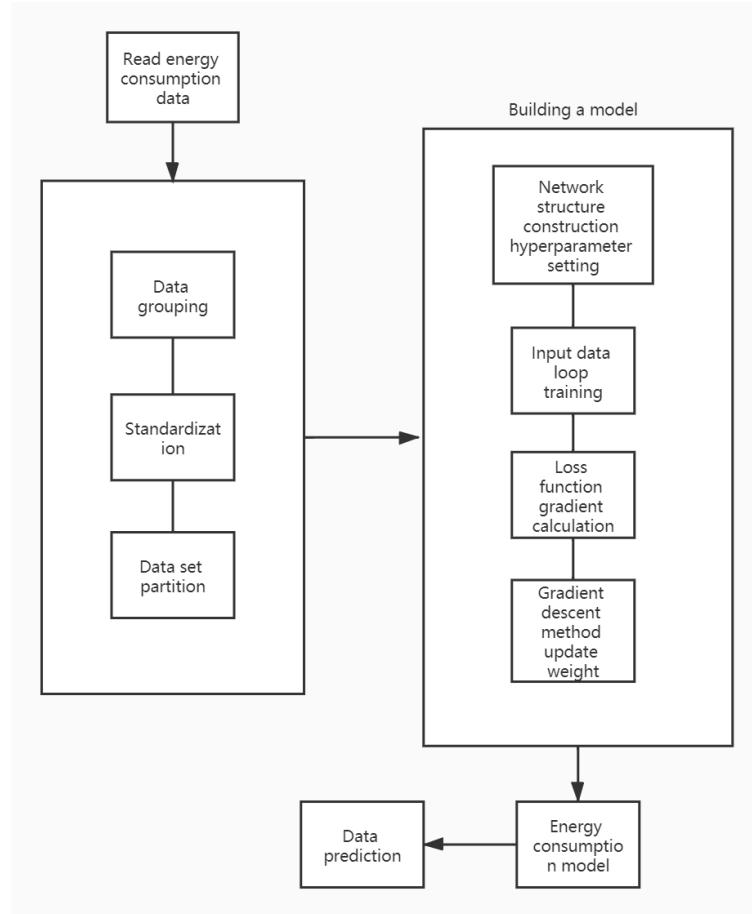
**Figure 4:** LSTM predictive energy consumption prediction model

For the establishment of the LSTM energy consumption prediction model, five parameters should be considered, the number of sample training, the number of layers in the input layer, the selection of the number of layers in the hidden layer, the setting of the dimension of the hidden layer, and the selection of the dimension of the output variable. The training step size setting of the sample which is the sample length which is the setting of seq_len, is to consider whether the model can be trained because of these sample lengths. If this value is set too small, the model will not be trained to the corresponding law. If this value is selected too much, it will cause more irrelevant sequence generation, resulting in inaccurate prediction of the model. This model will set different sample lengths and then finalize a value. Since the input variables in this paper are energy consumption, we choose 1 for the input dimension. The number of layers of multiple hidden layers can theoretically fit any function. However, due to the fear of over-fitting, the appropriate amount is selected according to the resources and suggestions on the Internet.
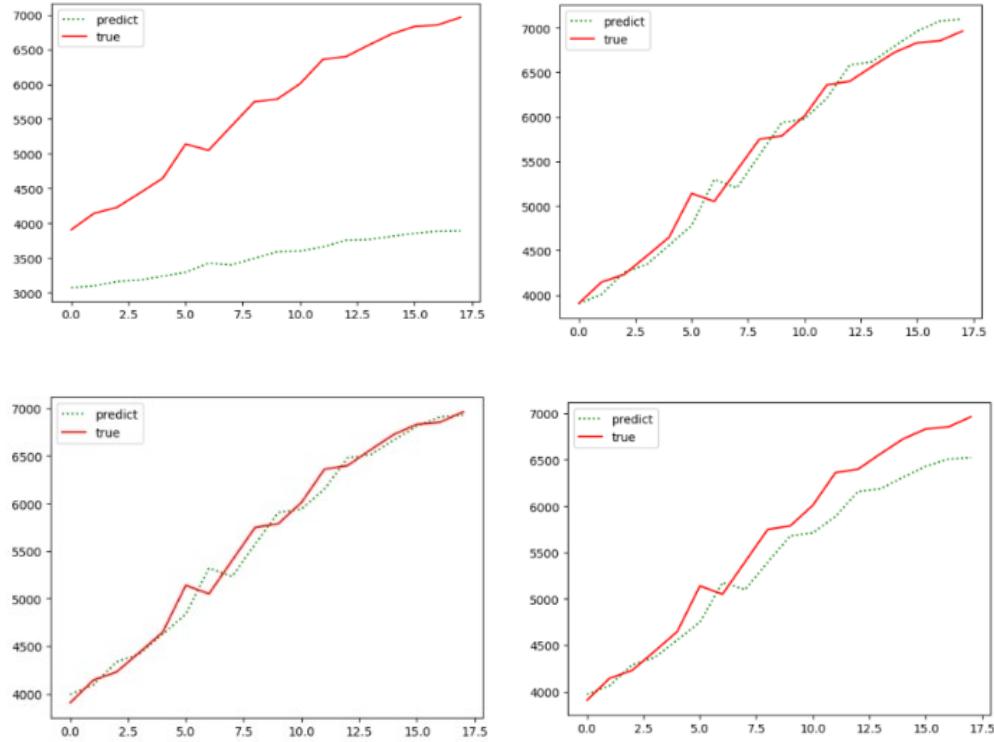
**Figure 5:** Comparison of results after adjusting parameters and epoch values multiple times

It can be seen from the above curves that after the epoch selected 30 times, 300 times, 600 times and 900 times, we can get about 600 times. It can be seen from the figure that the prediction effect is the best, and the error is the smallest. The model with epoch 600 times was chosen as the final energy consumption model.

**Table 3:** LSTM model error rate table

| Year | Actual value | LSTM | Relative error |
|------|-------------|------|----------------|
| 2007 | 5747.7 | 5574.966782 | –3.01% |
| 2008 | 5786.2 | 5909.290628 | 2.13% |
| 2009 | 6008.6 | 5945.960165 | –1.04% |
| 2010 | 6359.5 | 6156.691148 | –3.19% |
| 2011 | 6397.3 | 6485.308516 | 1.38% |
| 2012 | 6564.1 | 6520.420267 | –0.67% |
| 2013 | 6723.9 | 6674.677227 | –0.73% |
| 2014 | 6831.2 | 6821.410552 | –0.14% |
| 2015 | 6852.6 | 6919.354349 | 0.97% |
| 2016 | 6961.7 | 6938.831504 | –0.33% |
| 2017 | 7132.8 | 7098.360225 | –0.48% |

It can be seen from the above error table that under the prediction of the LSTM model, the error rate exceeds 5% only in 2014 and 2015, and the smallest error rate in the past five years is only -0.14% in 2014, which can be said to be The model performed better once. The average error of this model is -0.44%. I think it can represent an excellent prediction model.

## 5 Conclusion

The analysis results obtained from the above models can be obtained: the pre-diction value of the gray prediction model is the worst compared to the single prediction model, and the best is the linear regression model, the overall fitting effect and the accuracy of the prediction are required. Better than LSTM and gray models. But for the overall model, the grey prediction and BP neural network prediction combined model is the optimal model. In today's social development and steady economic growth, energy consumption will increase in the future. From Tab. 2 (units of 10,000 tons of standard coal), it can be known from the gray BP model that if Beijing's energy consumption and some other factors do not change significantly, Beijing's energy in 2019 The total consumption will be around 80,694,913 tons of standard coal. Judging from the current energy supply and demand situation in Beijing, the demand for future energy consumption supply will be very severe. We should suggest that the relevant units need to do the corresponding defense and supply work.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1]  J. H. Bu and M. Liu, "Analysis of Beijing energy conservation and environmental protection industry development and policy," *Green Building Materials*, vol. 1, pp. 44–47, 2019.

[2]  J. Y. Zhang and Y. J. Zhang, "Optimization of energy consumption in metallurgical enterprises based on genetic algorithm," *Ferroalloys*, vol. 3, pp. 22–26, 2017.

[3]  F. H. Ding and P. Liu, "Short-term prediction model of building energy consumption based on genetic optimization decision tree," *Computer Engineering*, vol. 5, pp. 1–11, 2019.

[4]  Z. G. Wang and D. L. Yao, "Effect analysis of multiple linear regression model in the prediction of total output value of construction industry in Jiangsu Province," *Journal of Taiyuan Urban Vocational and Technical College*, vol. 5, no. 202, pp. 58–60, 2018.

[5]  P. Sutthichaimethee and D. Ariyasajjakorn, "Forecast of carbon dioxide emissions from energy consumption in industry sectors in Thailand," *Environmental and Climate Technologies*, vol. 1, no. 22, 2018.

[6]  C. Ramos, A. S. García, B. Moreno and G. Díaz, "Small-scale renewable power technologies are an alternative to reach a sustainable economic growth Evidence from Spain," *Energy*, vol. 2019, no. 167, 2019.

[7]  L. L. Dong, F. Cheng, Z. Xiang and C. Cao, "Coal mine water inrush prediction based on LSTM neural network," *Coal Geology & Exploration*, vol. 47, no. 2, pp. 137–143, 2019.

[8]  J. Y. Qu, Y. Meng and Q. Xing, "A airport delay prediction model based on regional residuals and LSTM network," *Journal on Communications*, vol. 40, no. 2, pp. 149–159, 2019.

[9]  C. Yu, "Research and implementation of base station energy consumption data prediction method based on data mining," M.S. dissertation, Beijing University of Posts and Telecommunications, 2016.

[10] Y. Chen, "Prediction algorithm of PM2.5 mass concentration based on adaptive BP neural network," *Computing*, vol. 100, no. 8, 2018.

[11] C. Anitescu, E. Atroshchenko, N. Alajlan and T. Rabczuk, "Artificial neural network methods for the solution of second order boundary value problems," *Computers, Materials & Continua*, vol. 59, no. 1, pp. 345–359, 2019.

[12] H. K. Shin, Y. H. Ahn, S. H. Lee and H. Y. Kim, "Digital vision based concrete compressive strength evaluating model using deep convolutional neural network," *Computers, Materials & Continua*, vol. 61, no. 3, pp. 911–928, 2019.

[13] Y. Shen, Y. Li, J. Sun, W. Ding, X. Shi *et al.*, "Hashtag recommendation using LSTM networks with self-attention," *Computers, Materials & Continua*, vol. 61, no. 3, pp.1261–1269, 2019.

[14] B. Wang, P. Liu, Z. Chao, J. Wang, W. Chen *et al.*, "Research on hybrid model of garlic short-term price forecasting based on big data", *Computers, Materials & Continua*, vol. 57, no. 2, pp. 283–296, 2018.

[15] N. Wang, M. He, J. Sun, H. Wang, L. Zhou *et al.*, "ia-PNCC: noise processing method for underwater target recognition convolutional neural network," *Computers, Materials & Continua*, vol. 58, no. 1, pp. 169–181, 2019.

[16] Z. Gao, S. Zhang, L. Hao and N. Cao, "Modeling and analysis the effects of EMP on the Balise system," *Computers, Materials & Continua*, vol. 58, no. 3, pp. 859–878, 2019.

[17] X. Zhuang and S. Zhou, "The prediction of self-healing capacity of bacteria-based concrete using machine learning approaches," *Computers, Materials & Continua*, vol. 59, no. 1, pp. 57–77, 2019.