

A Study of China's Electricity Supply Based on Correlation and Time Series Modeling

Liangyu Qin, Huihui Wang, Qihang Yao*

School of Electronic and Information Engineering, Beihang University, Beijing, China, 100191

*Corresponding author: yaoqh@buaa.edu.cn

Abstract. In this study, the relationship between electricity supply and various indicators is deeply explored through multi-factor correlation analysis, quantitative analysis and prediction modeling. First, the Pearson correlation coefficient analysis was used, and it was revealed that electricity supply is significantly and positively correlated with gross domestic product (GDP), per capita GDP, urban population ratio, and consumption level of residents. Subsequently, linear regression methods were used for quantitative analysis, and forecasting models for electricity supply were established, including ARIMA model and Holt-Winters seasonal exponential smoothing model. Finally, the forecast value of China's electricity supply for the next 37 years was derived by synthesizing the forecasting models. The results of this study provide an important reference for optimizing the structure of electric power supply and improving the supply efficiency, and are of guiding significance for the formulation of relevant policies and planning.

Keywords: Electricity supply, Correlation analysis, Forecasting models.

1. Introduction

With the continuous development of the economy and society, electricity supply plays a crucial role as an important infrastructure to support the functioning of modernized societies. However, in the process of modernization, the challenges and problems faced by power supply have become increasingly prominent. The purpose of this paper is to explore the relationship between power supply and various indicators, and to solve the related problems by establishing a model, with a view to providing theoretical support and practical guidance for optimizing the power supply structure and improving the supply efficiency^[1-2].

First of all, the rapid growth of China's economy poses a great challenge to the demand for electricity. With the acceleration of industrialization and urbanization, the demand for electricity is growing, and how to effectively meet this huge demand has become an important issue in front of us.

Secondly, the supply of electricity is closely related to the development of the national economy, and the question of whether the existing electricity supply structure is reasonable and whether it can effectively support economic growth is also an urgent issue to be resolved.

In addition, with the increasing prominence of environmental issues, how to ensure the stability of power supply under the premise of realizing the sustainable use of clean energy is also one of the current challenges that need to be resolved.

In view of this, this paper will start from multi-factor correlation analysis, quantitative analysis, prediction modeling, etc., aiming to study the status quo of power supply, problems and their solutions, and to provide theoretical and practical guidance for the sustainable development of China's power industry. Websites, where we collect data, are listed in Table 1.

Table 1: Notations used in this paper

Database Names	Database Websites
(China) National Bureau of Statistics (NBS)	https://www.stats.gov.cn/
Chinese national data	https://data.stats.gov.cn/
Resource Discipline Innovation Platform	https://www.data.ac.cn/
baidu	https://www.baidu.com/
National Weather Science Data Center	https://data.cma.cn/

2. Modeling and Solving

2.1. Model 1: Multi-factor Correlation Analysis

2.1.1 Qualitative Analysis

Since all data variables are quantitative variables, the correlation can be analyzed using the Pearson correlation coefficient^[3]. The Pearson correlation coefficient is widely used to measure the degree of linear correlation between two variables, and its value range is between -1 and 1, where positive values indicate a positive correlation, negative values indicate a negative correlation, and values closed to 0 indicate there is almost no linear correlation between two variables, and its calculation formula is:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Table 2: Pearson Correlation Analysis

	GDP	GDP per Capita	Urban Population Ratio	Resident Consumption Level	Electricity Consumption
Pearson Correlation	0.984	0.986	0.995	0.987	1.000
Significance (Two-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
Number of Cases	25				

It can be seen from the Table 2 that the correlation between all indicators and electricity supply is very close to 1, and the significance level is less than 0.001. The null hypothesis is rejected, indicating that there is a significant positive correlation between electricity supply and GDP, GDP per capita, urban population ratio, consumption level of residents and electricity consumption, which is very strong.

2.1.2 Quantitative Analysis

Based on the results of the qualitative analysis, in order to better describe the relationship between power supply and various indicators^[4-5], the linear regression method is used for quantitative analysis, and the structure of the linear regression model based on the least squares method is given as follows:

$$y = \beta_0 + \beta_1 x_i + \epsilon \quad (2)$$

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (4)$$

Using SPSS to solve, the results are as Table 3:

Table 3: Linear Regression Fitting

	GDP	GDP per Capita	Urban Population Ratio	Resident Consumption Level	Electricity Consumption
R-squared	0.969	0.973	0.990	0.975	0.999
Regression Coefficient β_1	9281.8	7868.4	-81688.2	-3710.3	-385.3
Regression Coefficient β_0	0.069	0.989	250030.1	38.498	1.015
Significance	<0.001	<0.001	<0.001	<0.001	<0.001

3. Model 2: Forecast of Power Supply

First, ARIMA and Holt-Winters seasonless exponential smoothing models are established respectively, and then the standard deviation is calculated based on the prediction error of the test set to weight the prediction results^[6-7].

3.1. Establishment of ARIMA Model

The essence of the ARIMA model is a combination of the difference operation and the ARMA model. Among them, the autoregressive part captures the persistence in the time series, the moving average part captures the influence of the error term in the time series, and the difference part is about comparing the sequence with its past values. Comparisons are made to make the series stationary. Its model structure is:

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t \quad (5)$$

In the formula, X_t is time series data, that is, electricity supply. B is a lag operator (backshift operator), such that $B^k X_t = X_{t-k}$; d is the degree of difference, used to make the data stationary. $(1 - B)^d$ is a difference operator. $\phi(B)$ is the polynomial of the non-seasonal autoregressive part. $\theta(B)$ is the polynomial of the non-seasonal moving average part. Z_t is the white noise error term. The algorithm steps of ARIMA are shown in the figure 1:

The power supply data of the past years were divided into training and testing sets. After traversing different (p, d, q) parameter combinations, calculating AIC and BIC respectively and taking the parameter combination corresponding to the smallest value as the final parameter of the model, the result of which came out to be $(0, 2, 2)$. Then a model was established based on the optimal parameters. The model construction results are shown in table 4:

Table 4: ARIMA model building results

	value	standarderror	Tstatistic	Pvalue
Constant	186.82	449.38	0.41573	0.67761
MA1	-0.16312	0.54586	-0.29883	0.76507
MA2	-0.83688	0.65339	-1.2808	0.20025
Variance	2.7884e6	1.8872e6	1.4775	0.13953

Absolute error analysis was performed on the test set. The results are as table 5:

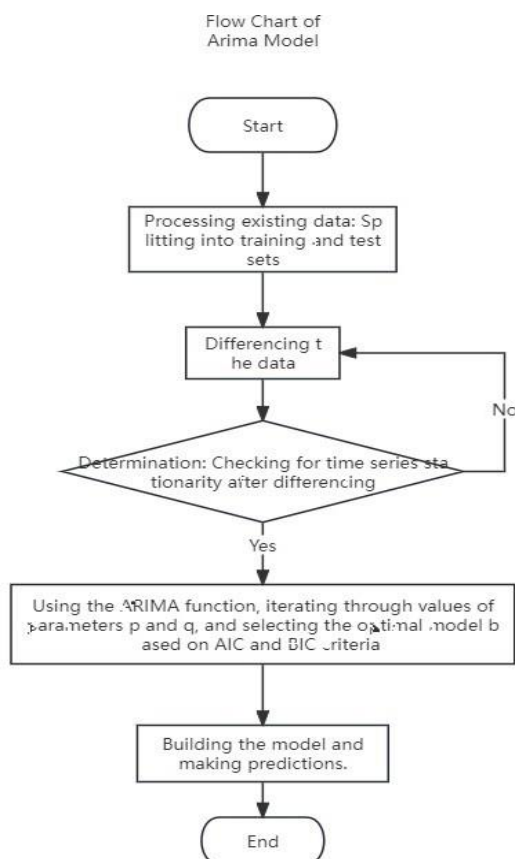


Figure 1: Flow Chart of ARIMA Model

Table 5: ARIMA test set error

years	2018	2019	2020	2021	2022
absolute error for each item	1454.2	212.3	1833.0	730.5	1300.2

According to figure 2, the average relative error is 0.014, and it passes the autocorrelation and partial autocorrelation tests, indicating that the model fitting effect is very good.

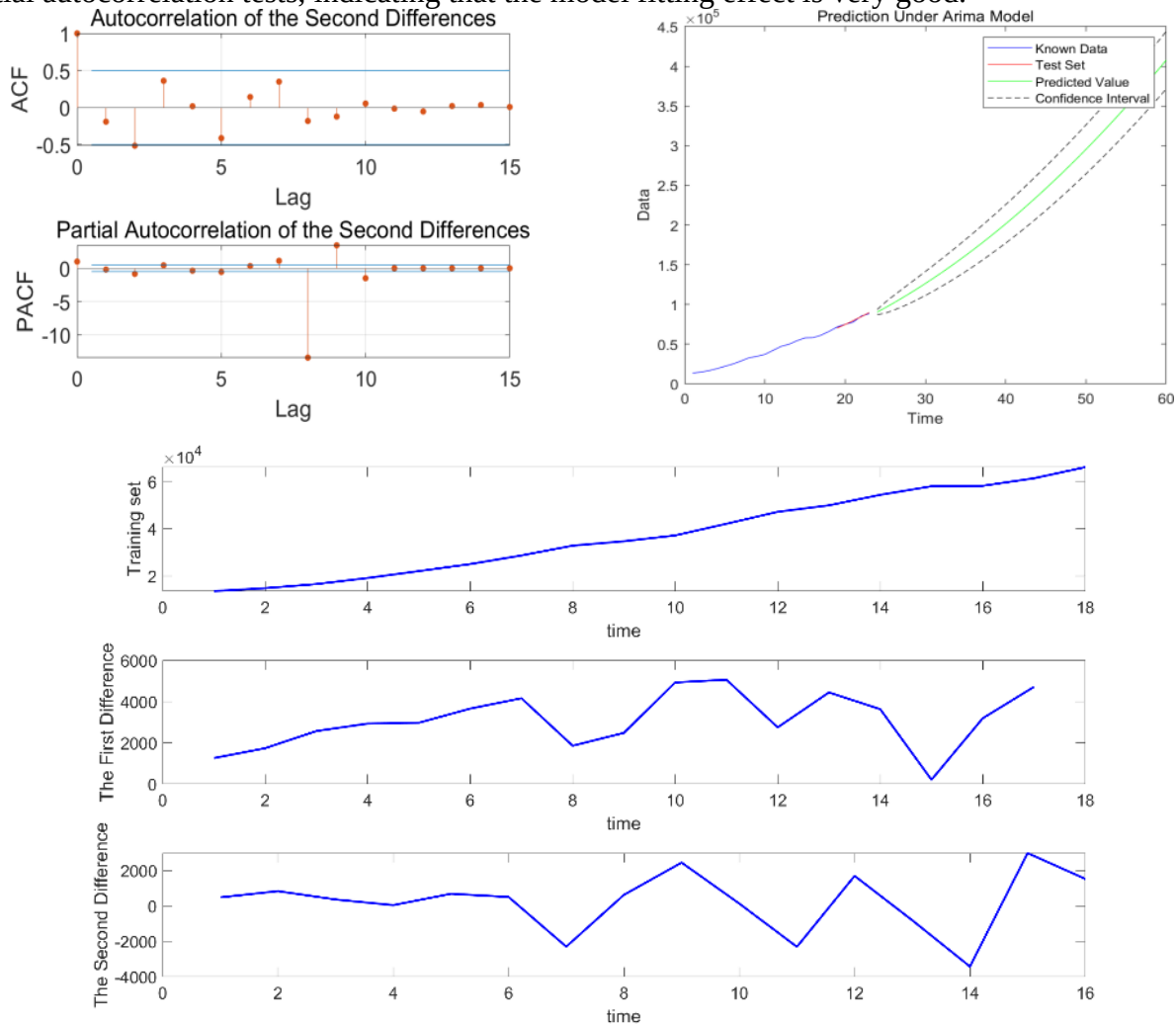


Figure 2: Code result visualization of ARIMA model

3.2. Holt-Winters Seasonless Exponential Smoothing Model

The Holt-Winters seasonality-free exponential smoothing model is a method for time series forecasting that is particularly suitable for data without obvious seasonal patterns. Although electricity supply does not appear to be seasonal, the trend indeed exists. So, this model can be used [8-9]. This model extends the simple exponential smoothing method to improve forecast results through trends. Its structure is:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}); b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (6)$$

In the formula, l_t is the level estimate of time point t ; α is the horizontal smoothing parameter, the value range is $0 < \alpha \leq 1$; y_t is the actual observation value at time point t ; b_{t-1} is the trend estimate at time point $t - 1$; b_t is the trend estimate at time point t ; β^* is the trend smoothing parameter, and the value range is $0 < \beta^* \leq 1$. And the algorithm steps of the Holt-Winters seasonality-free exponential smoothing model are shown in the figure 3:

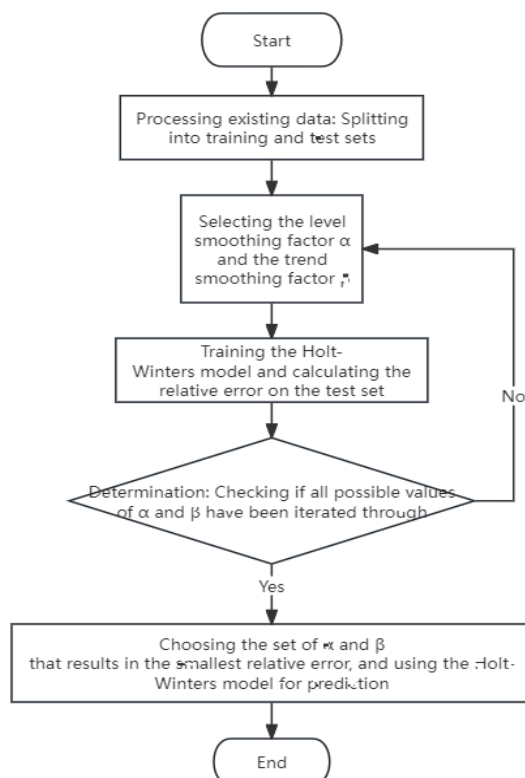


Figure 3: Flow Chart of Non-seasonal Holt-Winters method

The power supply data of the past years were divided into training and testing sets. After traversing different combinations of (α, β^*) parameters to calculate the average relative error, and taking the parameter combination corresponding to the smallest value as the final parameter of the model, the result is shown in table 6:

Table 6: Holt-Winters model errors and parameters

MSE	α	β^*
0.011	0.95	0.90

Then perform absolute error analysis on the test set, and the results are as table 7:

Table 7: Holt-Winters model test set error

years	2018	2019	2020	2021	2022
absolute error for each item	1209	66.21	1693.2	1343	28.07

The result of running the MATLAB code is as figure 4:

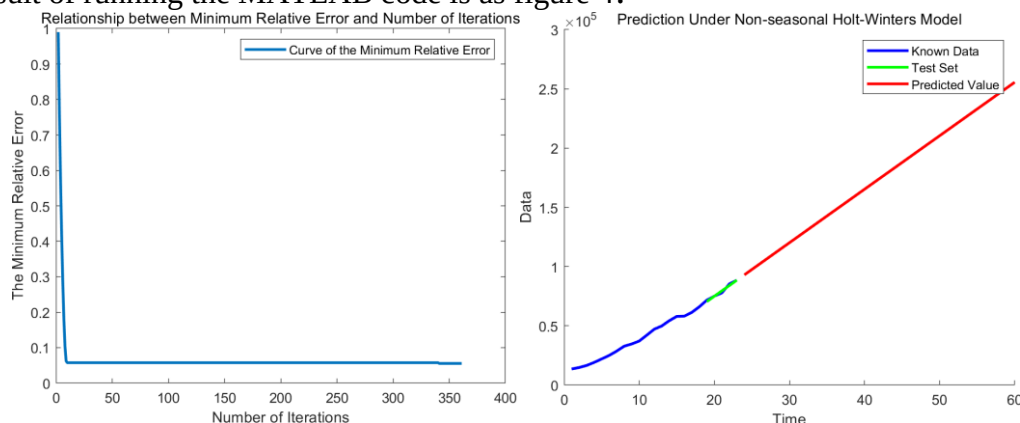


Figure 4: Visualization of code results for the Holt-Winters model

3.3. Combined Forecasting Model

In order to make full use of the effective information reflected in a single model, overcome the defects of a single model, reduce the randomness of prediction and improve the prediction accuracy, the establishment of a combined prediction model can be considered^[10], and the standard deviation method can be used to determine the combination weights. Assume that the forecast error standard deviations of ARIMA and Holt-Winters models are respectively σ_1 , σ_2 , Then the weight is determined as:

$$\omega_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}, \omega_2 = \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad (7)$$

Calculate the standard deviations of the two models respectively and substitute them into the formula to get: $\sigma_1=0.55$,

$\sigma_2=0.45$, so that a combined prediction model can be established:

$$P = P_1 \times 0.55 + P_2 \times 0.45 \quad (8)$$

In the formula, P is the combined prediction value; P_1 is the ARIMA prediction value; P_2 is the Holt-Winters prediction value. According to the combined forecast model, the final forecast value of China's power supply in the next 37 years can be obtained as table 8:

Table 8: Electricity supply forecast value from 2024 to 2060

*Total power generation unit: $\times 10^9$ KW.

years	Total power generation	years	Total power generation	years	Total power generation	years	Total power generation
2024	91571.1	2034	146977.1	2044	212601.1	2054	288442.9
2025	96651.9	2035	153079.7	2045	219725.4	2055	296589.0
2026	101834.8	2036	159284.5	2046	226952.0	2056	304837.4
2027	107120.0	2037	165591.4	2047	234280.7	2057	313187.9
2028	112507.3	2038	172000.6	2048	241711.6	2058	321640.6
2029	117996.8	2039	178511.9	2049	249244.7	2059	330195.5
2030	123588.6	2040	1.851254	2050	256880.0	2060	338852.5
2031	129282.4	2041	1.918410	2051	264617.5		
2032	135078.5	2042	1.986589	2052	272457.1		
2033	140976.7	2043	2.055789	2053	280398.9		

4. Conclusions

In this study, the relationship between electricity supply and various indicators is deeply explored through multi-factor correlation analysis and quantitative analysis, and a corresponding prediction model is established. Through Pearson correlation coefficient and linear regression analysis, it is found that electricity supply shows a significant positive correlation with gross domestic product (GDP), per capita GDP, urban population ratio, and consumption level of residents. The prediction results based on ARIMA and Holt-Winters models show that the trend of China's future electricity supply will continue to grow stably.

Combined with the comprehensive forecast model, we obtained the forecast value of China's power supply in the next 37 years. These results have important reference significance for formulating development strategies for the power industry, improving power supply efficiency, and promoting economic growth.

But we also need to see that the power supply field also faces challenges such as supply and demand balance, environmental sustainability, and clean energy utilization. Therefore, further in-depth research is needed in the future to formulate more comprehensive and effective policies and measures to promote the healthy and sustainable development of the power industry.

References

- [1] Cai Zhihao, Shi Hongyan. Sparrow search algorithm optimization for RF-BILSTM short-term power load forecasting [J]. Journal of Heilongjiang Engineering Institute, 2024, 38 (01): 15-20.
- [2] Qiang Wang, Hongwei Liu, Zifan Nie. Two-stage short-term load forecasting based on VMDT-POA-DELM-GPR [J]. Foreign Electronic Measurement Technology, 2024, 43 (01): 101-109.
- [3] Ji Yang. Research on short-term power load forecasting based on VMD-LSTM model [J]. Mechatronics Information, 2024, (01): 15-17+21.
- [4] Zuo J. En. Research on load forecasting and optimal scheduling method of power system based on big data analysis [J]. Appliance Maintenance, 2024, (01): 52-54.
- [5] SHEN Hongtao, LI Fei, SHI Lun et al. Short-term power load forecasting based on weather data dimensionality reduction and hybrid deep learning [J]. Power Construction, 2024, 45 (01): 13-21.
- [6] Lianbing Li, Guoqiang Gao, Weiguang Chen et al. Ultra-short-term load power prediction considering feature reorganization and BiGRU-Attention-XGBoost model [J/OL]. Modern Power, 1-11[2024-02-29].
- [7] Gao Hongyu, Xu Ning. Short-term power load forecasting based on Gaussian process with mixed distribution estimation [J]. Electric Power Science and Engineering, 2023, 39 (12): 51-59.
- [8] Xie W.L., Zhang L., Wang S.B. et al. Short-term load forecasting of long-short-term memory networks based on variational modal decomposition-aligned entropy-improved pelican optimization algorithm [J]. Hunan Electric Power, 2023, 43 (06): 82-92.
- [9] Gao Zhino. Research on electric load forecasting method based on deep neural network [J]. Electrical Technology and Economy, 2023, (10): 29-31.
- [10] QIN Yaokai, TAO Tao, CHEN Xingyan et al. Electricity prediction model of panel customized furniture workshop based on WOA-BP neural network [J]. Forest Industry, 2023, 60 (12): 50-56+68.