Research on Dynamic Relationship between Urban Rail Transit and Power Consumption

Liu Zhe¹, Du Fengqing¹ and Liu Xu²*

¹State Grid Shanghai Electric Power Research Institute, Shanghai, Shanghai, China
²North China Electric Power University, Baoding, Hebei, China

527532129@qq.com

Abstract. Aiming at the problem of insufficient feature quantity and lack of correlation between feature quantity and load, based on time series analysis method, ARIMNA model is used to fit the load forecasting problem. The study found that the urban orbit data has a good autocorrelation. Compared to the Linear Regression method, when the ARIMA model is used, the algorithm can predict the trend of urban rail transit load data better.

1. Introduction
Forecasting is the premise of decision making and provides guidance for planning and scheduling [1]. In the field of power systems, one of the important tasks of the power sector and the dispatching department is to make accurate predictions of the load. In general, accurate prediction of medium and long-term load [2] is conducive to long-term planning of power systems, and fuel purchase plans can be formulated in advance. Similarly, short-term load forecasting can be a guide for short-term power generation planning.

Load forecasting can be generally classified into long-term forecasts, medium-term forecast, short-term forecast for forecasting daily load, and ultra-short-term forecast. At present, the development of the power industry tends to be stable. Foreign research’s hotspots mainly focus on short-term load forecasting [3-4]. In China, industrial adjustment has a great impact on power demand [5]. Therefore, the study of load forecasting is a combination of long-term and short-term.

Since the development of load forecasting, many scholars at home and abroad have proposed many prediction methods based on their own research [6]. In literature [7], the support vector machine method is used to perform regression analysis on the load data and the grid search method is used to find the optimal parameters to predict the load data. Another way to predict the load is to combine multiple prediction methods. In [8], Markov chain was used to select two prediction models that satisfy the accuracy from several models. Using this method to filter the model can improve the prediction accuracy. Since the development of artificial intelligence, neural networks have also begun to be used in load forecasting. In [9], the ridge regression-BP neural network was used to predict the load on the premise of using the ridge map to screen the indicators and achieve high precision. With the development of big data, the research hotspot of load forecasting is transformed into parallel processing of data using related technologies of big data [10]. The study found that when the amount of data increases, the prediction accuracy is also improved [11-12].

The above research has laid the foundation for the application of data science in power systems, as well as the relationship between load data and factors such as region and climate. However, the above
research does not reveal the trend of the load in time. When faced with the problem of insufficient feature quantity and insufficient correlation between feature quantity and load, the prediction is not very accurate. The study found that the seasonal characteristics of the load data are obvious and have a linear relationship with time. Based on this, the time series analysis method is used to analyze the load. The energy consumption data of Shanghai Line 6 was selected as the research object, and the time series model was established. The study found that the energy consumption data has a high degree of autocorrelation, and the fitted curve is high when performing load forecasting.

2. Sequence Prediction Model

The sequence method divides the continuous load into a quantifiable sequence, and then uses the sequence research method to predict the value of the next segment by fitting the previous sequences. The load forecasting abstract expression of the sequence model is shown in equation (1):

\[ y = f(S, X, t) \]  

(1)

In the load model, \( X \) represents the \( m \) factors that affect the load, \( X \) recorded as \( X = [x_1, x_2, \ldots, x_n]^T; t \) represents the time serial number; \( S \) is a coefficient matrix, sort by the order of \( t \) from small to large, when the load model is a first-order equation \( y = a + bt, S = [a, b]^T \).

The source of data in load forecasting has two aspects, one is the influencing factor data obtained from the economic and meteorological departments, and the other is the load data obtained from the power sector. The collected data is divided into a training data set and a test data set. The composition of the training sample is the value of the influencing factor \( y_t \) and the magnitude of the load obtained over the known time period \( t(n + 1 \leq t \leq N) \). The composition of the predicted sample is the value of the influencing factors \( X_t = [x_{t+1}, x_{t+2}, \ldots, x_{n}]^T \) obtained in the future time period \( t(n + 1 \leq t \leq N) \). The coefficient matrix \( S \) is estimated by the above data and the value of \( y_t(n + 1 \leq t \leq N) \) is determined.

After obtaining the parameter estimation value \( \hat{S} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k]^T \), the load model can be fitted or predicted. Substituting \( \hat{S} \) into equation (1) can be obtained:

\[ \hat{y}_t = f(\hat{S}, X_t, t), t = 1, 2, \ldots, n \]  

(2)

The fitted residual is:

\[ u_t = y_t - \hat{y}_t, t = 1, 2, \ldots, n \]  

(3)

The goal of model optimization is to minimize the value of \( u_t \) by adjusting the value of \( \hat{S} \). There are many methods for parameter optimization. The most classic methods are least squares method and intelligent algorithms represented by ant colony algorithm and particle swarm algorithm.

3. Time Series

A time series is a type of sequence that refers to a sequence of numbers arranged in chronological order. Due to the existence of fluctuations and error terms, the changes are widely present in the time series. Time series is especially good at dealing with changes in trends and periodic composition. Time series analysis is the process of regression analysis and prediction of the obtained data. Commonly used methods are autoregressive (AR), autoregressive moving average (ARMA), and vector autoregressive model (VAR).

From the idea of time series, the predicted value \( y_t \) can be written as a linear multiplication of the constant coefficient with the known value and added to the offset. Define the p-order AR model as shown in equation (4):

\[ y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \alpha_t \]  

(4)
Where \( p \) represents a \( p \)-order function, \( \varphi_i, \varphi_i = 1, 2, \cdots, p \) represents a constant coefficient, and the offset \( \theta_i \) is a value of white noise at a predicted time \( t \).

The MA model is to study the relationship between predicted values and noise sequences. The \( q \)-order MA model is defined as shown in equation (5):

\[
y_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\]

Where \( q \) is a \( q \)-order function, \( \theta_i, \varphi_i = 1, 2, \cdots, p \) represents a constant coefficient, offset \( a_{t-i}, i = 0, 1, \cdots, p \) is the white noise at the prediction time \( t \) value.

Combining the AR model with the MA model, a more general model ARMA model is obtained, as shown in equation (6):

\[
y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \cdots - \varphi_p y_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\]

The ARMA model is applied to a smooth time series. In the load forecasting, the load data shows the non-stationary period characteristics. In order to meet the conditions of the ARMA model, a differential operation can be added before the model fitting to remove the periodicity and non-stationarity.

The steps of time series analysis are shown in Figure 1.

As can be seen from Figure 1. Firstly, the energy consumption data of the subway operation is acquired, and the missing value and the error value are modified. Then, the obtained data is subjected to stationarity detection. If the data is not stable, a difference operation is performed, and the data for performing the difference operation is again subjected to stationarity detection, and this step is repeated until the data is smooth. White noise detection is then performed. Model fitting is performed on non-white noise data, and the commonly used fitting model in time series analysis is ARIMA. Finally, data prediction and evaluation of the model are performed.
4. Examples

As an important livelihood project of urban infrastructure construction, urban rail transit reflects the strategic positioning and competitiveness of urban development. Taking Shanghai Metro as the research object, analyzing the electricity consumption model of urban rail transit and forecasting demand, it can provide methods and models for rail transit power consumption forecast of new lines in Shanghai. The Shanghai Metro was built in 1990. By December 2018, there were 16 subway lines in Shanghai, with a total of 415 stations with a total mileage of 705 kilometers.

This section combines passenger flow and annual operating mileage data to analyze the changes in energy consumption of Line 6. The specific data is shown in Table 1:

| Years | Passenger Flow (10,000 people) | Annual Operating mileage (10,000 kilometers) | Electricity Consumption (10,000 degrees) |
|-------|-------------------------------|---------------------------------------------|----------------------------------------|
| 2011  | 8711.613                      | 1383.411                                    | 6853.1                                 |
| 2012  | 9568.014                      | 1431.229                                    | 7145                                   |
| 2013  | 10622.05                      | 1603.415                                    | 8080.8                                 |
| 2014  | 11598.74                      | 1934.383                                    | 11441.2                                |
| 2015  | 12970.33                      | 2020.484                                    | 12008                                  |
| 2016  | 14126.9                       | 2045.305                                    | 14370                                  |
| 2017  | 14683.95                      | 2033.448                                    | 14040                                  |

In order to make the data image smoother, the First-order difference operation is performed on the original data. The calculated data is shown in Figure 2.

![First-order difference map](image)

**Figure 2.** First-order difference map

White noise detection is performed on the differential data. The results are shown in Table 2.

| Significant Level | P Value |
|-------------------|---------|
| 1.82327468        | 0.17692398 |

Observed in Table 2, the $P$ value is significantly smaller than the significance level, and the data is non-white noise.

The above model is used to predict the data and predict the data from 2015 to 2017. The results are shown in Figure 3.
Figure 3. Energy forecasting chart

The blue line represents the raw data and the red line represents the predicted value. Errors are introduced to analyze the prediction results. The values are shown in Table 3.

Table 3. Real and predicted values

| Year | Real value | Linear Regression Model Predictive Value | ARIMA Model Predictive Value |
|------|------------|-----------------------------------------|-----------------------------|
| 2015 | 12008      | 11390                                    | 11358                       |
| 2016 | 14370      | 12816                                    | 13508                       |
| 2017 | 14040      | 14242                                    | 14732                       |

It can be seen from the calculation that the error value of the load prediction by the linear regression model is 2837644, and the error value of the load prediction by the ARIMA Model is 1644408. The accuracy of load forecasting with time series is higher.

5. Conclusion

This paper takes Shanghai Metro data as the research object and time series as the analysis model to predict the energy consumption data. The study found that the urban orbit data has a good autocorrelation. Compared to the Linear Regression method, when the ARIMA model is used, the algorithm can predict the trend of urban rail transit load data more better.

Acknowledgments

Project Supported by State Grid Shanghai Electric Power Research Institute.

References

[1] Kang Chongqing, Xia Qing, Zhang Boming. Review of Power System Load Forecasting Research and Discussion on Its Development Direction. Automation of Electric Power Systems, 2004, (17):1-11.

[2] Zhang Zhi, Du Yanling, Cui Huijun, Wang Yang, He Zhe, Lai Xiaowen. Intelligent mid-long electricity load forecast method considering associated factors. Power System Protection and Control, 2019, 47(02): 24-30.

[3] Barakat E H, Al-rashed S A. Long Range Peak Demand Forecasting Under Conditions of High Growth. IEEE Trans on Power Systems. 1992, 7(4):1483-1486.

[4] Kandil M S, El-Debeiky S M, Hasamien N E. Overview and Comparison of Long-term Forecasting Techniques for a Fast Developing Utility: Part1. Electric Power Systems Research, 2001, 58(1):11-17.

[5] Yang Fangyuan, Shi Yuchao, Hou Yucheng. Correlation Analysis of Electricity Consumption Index and Load Forecasting Considering Adjustment of Industrial Structures. Electrical
[6] Zhou Chao, Xing Wenyang, Li Yulong. A Survey of Power System Load Forecasting Methods. Journal of Power Supply, 2012(06):32-39.

[7] Wang Ning, Xie Min, Deng Jialiang, Liu Mingbo, Li Jialong, Wang Yi, Liu Sijie. Mid-long term temperature-lowering load forecasting based on combination of support vector machine and multiple regression. Power System Protection and Control, 2016, 44(03):92-97.

[8] Zhang Dongliang, Yan Jian, Li Xiaobo, Ren Xiaoda, Zhang Jinzhong, Zhang Fulai. Mid-long term load forecasting based on Markov chain screening combination forecasting models. Power System Protection and Control, 2016, 44(12):63-67.

[9] Wen Ruiying, Wang Hongyong. A Forecasting Method of Controller’s Workload Based on Ridge Regression—BP Neural Network. Journal of Transportation Systems Engineering and Information Technology, 2015, 15(01):123-129.

[10] Wang Dewen, Sun Zhiwei. Big Data Analysis and Parallel Load Forecasting of Electric Power User Side. Proceedings of the CSEE, 2015, 35(03):527-537.

[11] Lei Zhengxin, Han Bei, Nie Meng, Zhen Ying, Wang Keyou, Li Guojie. Research and application of multi-node load forecasting algorithm under the environment of distribution network’s big data. Power System Protection and Control, 2016, 44(23):68-78.

[12] Xu Yuan, Cheng Qianshan, Li Yang, Zhang Hao, Yu Wei, He Bing. Mid-long Term Load Forecasting of Power System Based on Big Data Clustering. Proceedings of the CSU - EPSA, 2017, 29(08):43-48.