

Research on Autoregressive Distribution Lag Modeling Method for Power Load Forecasting

Linghua Tong¹, Lihua Luo¹ and Liangzhong Fan²

¹State Grid Zhejiang Cixi Power Supply Co., Ltd., Ningbo, China

²Ningbo Institute of Technology, Zhejiang University, Ningbo, China

Abstract—Based on the autoregressive distribution lag model, this paper studies the short-term power load forecasting problem in a certain area of Ningbo in summer. Firstly, the characteristics of time periodicity and temperature sensitivity of power load are analyzed. Secondly, on the basis of this analysis, the autoregressive distribution lag model is proposed, and the stepwise regression algorithm is used to overcome the multicollinearity problem of variables. Finally, the autoregressive distribution lag model method is applied to the practice of Ningbo summer power load forecasting, and the accurate forecasting model is obtained. The average relative error of power load forecasting is controlled within 4%, which verifies the validity and accuracy of the model.

Keywords—temperature; autoregressive distribution lag Model; load forecasting; stepwise regression method

I. INTRODUCTION

With the rapid development of industry, power production also has a great space for development. With the increasing demand for electric power in various industries, China has broken the traditional idea of monopoly management and marketized the electric power industry. In order to reduce the production cost of electric power and improve the competitiveness among enterprises, as an effective way of safe operation and cost reduction of power grid, the importance of power load forecasting is gradually recognized by people.

With the maturity of power market, more and more countries in the world pay attention to the forecasting of power load. Power load forecasting technology is also constantly developing, from the previous offline analysis to the present online application, and has moved towards the road of automation and intellectualization. For a long time, many scholars and research institutes have devoted themselves to the study of new methods for forecasting power load, and have made great achievements so far. Among them, Du Zibing proposed to apply EMD-SVM algorithm to short-term load forecasting of power system[1]. In [2], a new power load forecasting method based on improved regression method was proposed. The singular values were eliminated by ridge regression, and then the main factors affecting the load were extracted by principal component regression, and the analytical form of the model was obtained. Du Li[3] proposed a kind of power load forecasting method based on principal component analysis (PCA) and BP neural network. The authors extracted

the influencing factors of power load by PCA, and trained the new variables by BP neural network. In [4], a hybrid power load forecasting model combining fruit fly optimization algorithm and generalized regression neural network was proposed, where the fly optimization algorithm was used to automatically select the appropriate spread parameter value for the generalized regression neural network power load forecasting model.

The autoregressive distributed lag(ADL) model is the major workhorse in dynamic single-equation regressions. This method is simple, easy to use and does not require much data. By making use of the autoregressive distribution lag model, this paper will study the short-term power load forecasting in Ningbo based on the temperature factors.

II. ANALYSIS OF POWER LOAD

Power load is one of the most important indicators of power system planning, design and operation management. Studying the characteristics of load and its changing rules is the first condition for power departments to carry out load forecasting and scientific power supply allocation.

Power load series have obvious periodicity from data characteristics, including annual periodicity and 24-hour periodicity[5]. The periodicity of data is an important auxiliary feature for sequence prediction. The periodicity of power load sequence should be fully utilized in power load forecasting and modelling. As shown in Figure I below, the graph shows the daily power load curves of a certain area of Ningbo on July 1, July 15, August 1 and August 15, 2019. Sample points of power load are collected every five minutes. So there are 12 power load sample points per hour and 288 power load sample points per day. The daily variation of power load in the graph can be followed regularly. Due to the periodicity of power load, the historical load data of the previous days can be used to predict future load data.

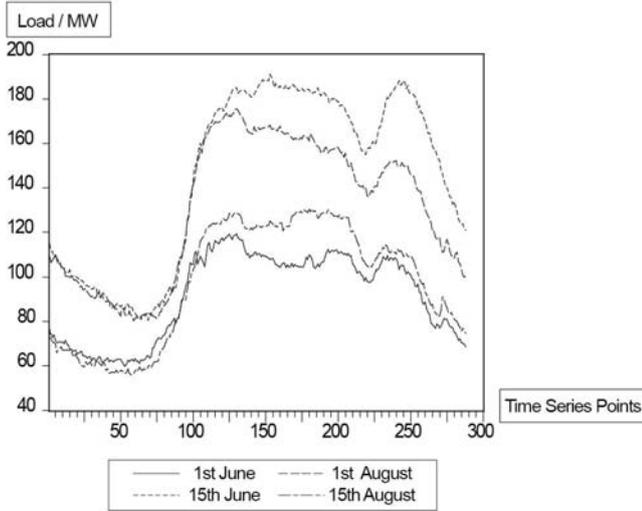


FIGURE I. THE DAILY POWER LOAD CURVES

Power load is affected by many factors, but meteorological factor is one of the important factors affecting the change of power load. Sensitivity analysis of load and temperature can clarify the impact of temperature factors on load at each time point. For two random variables ξ and η , the correlation coefficient is

$$r_{\xi\eta} = \frac{E(\xi - E\xi)(\eta - E\eta)}{\sqrt{\text{Var}\xi\text{Var}\eta}},$$

where

$$E\xi = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad E\eta = \frac{1}{n} \sum_{i=1}^n \eta_i$$

and

$$\text{Var}\xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - E\xi)^2, \quad \text{Var}\eta = \frac{1}{n} \sum_{i=1}^n (\eta_i - E\eta)^2.$$

The load and meteorological factors at each time point are regarded as random variables, load history data and weather history data in July and August as samples of random variables, and their correlation coefficients are calculated as the following Table I.

As can be seen from the table above, there is a strong correlation between temperature factors and power load. Especially for short-term power load forecasting, temperature should be a major factor.

TABLE I. CORRELATION COEFFICIENT BETWEEN LOAD AND TEMPERATURE

Time	Daily average temperature	Daily maximum temperature	Daily minimum temperature
11:00	0.597306223	0.605864866	0.434939759
11:05	0.590412782	0.612532962	0.415752575
11:10	0.594817632	0.615668725	0.418003774
11:15	0.609893355	0.622855377	0.443230056
11:20	0.61766479	0.609890182	0.464518906
11:25	0.616064186	0.603457212	0.468109515
11:30	0.611512854	0.616740583	0.447957095
11:35	0.606347376	0.624854634	0.433272413
11:40	0.608921133	0.633145012	0.429853801
11:45	0.621201185	0.651802162	0.458679074
11:50	0.613352869	0.656565912	0.430076353
11:55	0.623881894	0.666747274	0.441130181
12:00	0.624757168	0.673091456	0.437095169

III. AUTOREGRESSIVE DISTRIBUTION LAG MODEL

A. Establishment of the General Model

Autoregressive distributed lag(ADL) model is an important tool in dynamic one-equation regression, and is widely used in the economic model of time series. The general form of the autoregressive distribution lag model given in reference [6,7] is as follows:

$$X_t = \alpha_0 + \sum_{i=1}^m \alpha_i X_{t-i} + \sum_{j=1}^p \sum_{i=1}^n \beta_{ji} Z_{jt-i} + \mu_t$$

where

$$\mu_t \sim N(0, \sigma^2).$$

Among them, X_t is the explained variable, Z_{jt} ($j=1,2,\dots,p$) is the exogenous variable, μ_t is a white noise sequence and obeys normal distribution, p is the number of exogenous variables and, m and n are the maximum lag time of X_t and Z_{jt} , respectively. The autoregressive distribution lag model was recorded as $ADL(m, n, p)$.

In the autoregressive distributed lag model, the explained variable X_t is expressed as a function of its lag variable X_{t-i} , other exogenous variables' synchronous value Z_{jt} and lag variable Z_{jt-i} .

B. Stepwise Regression Algorithms to Select Variables

The choice of explanatory variables is very important when using regression analysis to build a model. The optimal linear regression model should be understood as follows: (1) the

model contains all explanatory variables that have a significant impact on interpreted variables; (2) the number of explanatory variables contained in the model is as small as possible.

If there is a correlation between two or more explanatory variables, there is a multicollinearity. Once the multicollinearity occurs, many problems will arise, such as the difficulty of parameter estimation and the unreasonable economic significance of parameter estimation. So in the regression model, we need to find a way to overcome multicollinearity.

The stepwise regression algorithm[8] is a combination of forward regression and backward regression, which can be used to deal with multicollinearity. The main idea of the stepwise regression algorithm is as follows: Take X_i as the explained variable. The explanatory variables are introduced one by one to form a regression model and to estimate the model. According to the change of goodness of fit, it is decided whether the newly introduced variables can be replaced by the linear combination of other variables rather than as independent explanatory variables. If the goodness-of-fit varies significantly, the newly introduced variable is an independent explanatory variable. If the goodness of fit does not change significantly, it shows that the newly introduced variable is not an independent explanatory variable. In this case, the newly introduced variable can be replaced by a linear combination of other variables, that is to say, it has a multicollinear relationship with other variables, which should be removed from the model.

IV. EVENING PEAK LOAD FORECASTING

In this section, the autoregressive distributed lag model will be applied to forecast the evening peak load at 20:00 in summer. The explained variable L_t denotes the load at time t . Because this paper studies the short-term power load forecasting, it does not consider the level of GDP, population growth rate, changes in industrial and agricultural modes and other factors. Temperature is the main factor to be considered. The temperatures factor mainly include daily maximum temperature, daily minimum temperature, daily average temperature and 8 o'clock temperature, which can be denoted by $MAXT_t, MINT_t, AVGT_t, EIGT_t$.

A. Establishment of Evening Peak Load Forecasting Model

In this paper, 45 sets of power load data for July and August 2019 are selected as samples from a certain area of Ningbo. The autoregressive distributed lag model assume that $m = n = 1$, and there is a linear relationship between the explained variable L_t and the following variables:

$$L_t = 4.694 + 0.961L_{t-1} + 0.488\left(\frac{MINT_t - MINT_{t-1}}{MINT_{t-1}}\right)L_{t-1} + 1.125\left(\frac{AVGT_t - AVGT_{t-1}}{AVGT_{t-1}}\right)L_{t-1} + \left(\frac{MAXT_t - MAXT_{t-1}}{MAXT_{t-1}}\right)L_{t-1} + \left(\frac{EIGT_t - EIGT_{t-1}}{EIGT_{t-1}}\right)L_{t-1}$$

The stepwise regression method[8] is applied to overcome the multicollinearity and the independent exogenous variable can be selected. By using EViews statistical software, the forecasting model of the evening peak load (20:00) is obtained

$$L_t = 4.694 + 0.961L_{t-1} + 0.488\left(\frac{MINT_t - MINT_{t-1}}{MINT_{t-1}}\right)L_{t-1} + 1.125\left(\frac{AVGT_t - AVGT_{t-1}}{AVGT_{t-1}}\right)L_{t-1} \quad (1)$$

From the above model, we can see that the change rates of the average temperature and the minimum temperature appear in the model as influencing factors, but the change of the maximum temperature does not appear in the model. We think that this is because there is a great correlation between the change rate of average temperature and the change rate of maximum temperature, so that the influence of the change rate of average temperature on goodness of fit can replace the influence of the change rate of maximum temperature on goodness of fit.

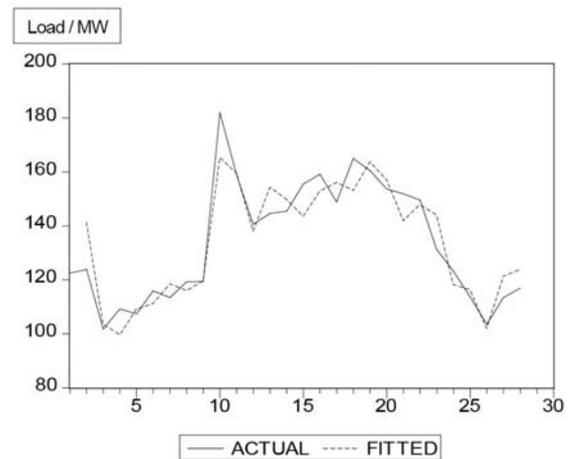


FIGURE II. THE FITTING GRAPH OF THE EVENING PEAK LOAD

Figure II is a fitting diagram of the evening peak load based on the autoregressive distribution lag model. As can be seen from Figure II, the difference between the predicted value and the actual value is small, which shows that the fitting effect of the model is better. In fact, the average relative error is 3.97%.

B. Forecast of Evening Peak Load (20:00)

On the basis of validating the validity of the model, the model (1) is used to predict the evening peak power load in the next 7 days, and the results are compared with the actual values, as shown in Table II and Figure III. The average relative error of the evening peak load forecasting model is 3.89%.

TABLE II. THE RELATIVE ERROR OF THE EVENING PEAK LOAD (20:00) IN THE NEXT 7 DAYS

Sequence point	1	2	3	4	5	6	7
Actual load /MW	110.7	110.5	127.9	124.6	130.6	146.5	171.9
Forecasting Value /MW	114.4	109.0	120.5	127.8	124.6	140.5	162.4
Absolute value of error /MW	3.7	1.5	7.4	3.2	6.0	6.0	9.5
relative error/%	3.34	1.35	5.78	2.56	4.59	4.09	5.52
Average relative error				3.89%			

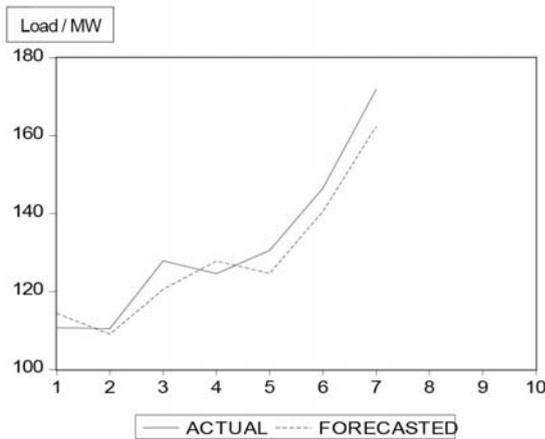


FIGURE III. THE FORECASTING GRAPH OF EVENING PEAK LOAD IN THE NEXT 7 DAYS

In this paper, only the load value of the evening peak (20:00) is predicted. But from the process of modeling and analysis, the method can be applied to the prediction of the corresponding load value at other times.

V. CONCLUSION

In this paper, the short-term power load forecasting problem in summer is studied by using the autoregressive distributed lag model. In this study, not only the time periodicity of late power load is considered, but also the important influence of meteorological factors on power load is taken into account. Therefore, the forecasting model obtained in this paper has good forecasting accuracy. In order to further improve the prediction accuracy of the model, the lag time can be enlarged. In addition, more meteorological factors, such as humidity and wind speed, can be added to the model as exogenous variables to improve the accuracy of the model.

ACKNOWLEDGMENT

This work was supported by State Grid Zhejiang Electric Power Co., Ltd. Science and Technology Project of Collective Enterprises in 2019: Power Emergency Assessment System for Bay Area Based on Vulnerability.

REFERENCES

- [1] Du Zibing. Research on Short-term Power Load Forecasting Method[D]. North China Electric Power University, 2013.
- [2] Li Po, Li Min, Liu Dichen. Power load prediction based on improved regression method[J]. Grid technology, 2006(1):99-104.
- [3] Du Li, Zhang Jianjun. Application of neural network in power load prediction[J]. Computer simulation, 2011, 28(10):297-300.
- [4] Li Hongze, Guo Sen, Li Chunjie. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm[J]. Knowledge-Based Systems, 2013, 37:378-387.
- [5] Lin Xiaohong, Xialihua et al. Preliminary study on the forecast model of summer power meteorological grade in Fuzhou[J]. Meteorological Science and Technology, 2006, 34(6):775-776.
- [6] An Heping. Study on Autoregressive Distribution Lag Model of Population Forecast in China[J]. Statistics and Decision-making, August 2005 (2):4-5.
- [7] Hassler U, Jürgen Wolters. Autoregressive distributed lag models and cointegration[J]. Allgemeines Statistisches Archiv, 2006, 90(1):59-74.
- [8] Fan Lishu, Econometrics[M]. Zhejiang: Zhejiang University Press, 2016.