

# Long-term load combination forecasting method considering the periodicity and trend of data

YANG Jingying<sup>1,\*</sup>, JIANG Minglei<sup>1</sup>, LI Xin<sup>1</sup>, SHE Xin<sup>1</sup>, XIN Haokuo<sup>1</sup>, ZHOU Lin<sup>1</sup>, and LIU Chang<sup>2</sup>

<sup>1</sup>Economic and Technical Research Institute of State Grid Jilin Electric Power Company, Lvyuan District, Changchun 130062, China

<sup>2</sup>State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Changping District, Beijing 102206, China;

**Abstract.** In order to solve the problems of insufficient accuracy of long-term power load forecasting and poor applicability of the model, this paper considers the coupling of a number of macro indicators, such as regional economic development and social development indicators, with the time series data of regional power load. BP neural network and Autoregressive integrated moving average model (ARIMA) are used to integrate and improve the forecasting model, so as to improve the trend forecasting ability of annual load forecasting model. The non parametric function method is used to forecast the periodic load data in the monthly load data, the annual load forecast is combined with the monthly load forecast to improve the overall forecasting accuracy of the model. Finally, through the comparison of grey prediction and other models and the verification of MAPE error analysis method, the prediction accuracy of the model method considering the combination of data periodicity and trend is significantly improved, which is suitable for the long-term prediction of regional power load.

## 1 Introduction

The accurate prediction of long-term power load has important strategic significance for power network planning and power infrastructure construction. In recent years, long-term power load prediction is mainly based on the overall trend prediction of annual load value<sup>[1]</sup>, and less consideration is given to the inertia growth of data, periodic changes and the cumulative effect of numerical values in the prediction, thus affecting the accuracy of load prediction. Therefore, the model modeling needs to consider that the annual load growth should be based on the monthly load accumulation, so as to improve the segmenting accuracy of the model interval prediction and ensure the accuracy of the prediction. Based on this work, the combined method of data trend and periodic prediction will be studied.

Power load forecasting methods include parameterized forecasting methods and nonparametric forecasting methods. The analysis algorithms based on data parameterization characteristics mainly include mechanical learning algorithm and time series algorithm. Mechanical learning, such as neural network learning<sup>[2]</sup>, SVM support vector machine<sup>[3]</sup>, decision tree. Time series ARIMA and other methods realize the analysis and prediction of the development trend of data itself through manual parameter setting. In order to improve the prediction accuracy of time series data, a variety of machine learning algorithms can be combined with ARIMA method to realize the coupling correction of data<sup>[4]</sup>. Nonparametric method is a kind of response

analysis method obtained directly or indirectly from the experimental analysis of the actual system, such as the impulse response or step response of the system obtained from the experimental records<sup>[5]</sup>. This kind of algorithm can fully mine the periodic characteristics of the data itself, and predict the trend of the data in the interval unit through the dimensionality reduction of the data based on the periodicity of the data. In order to further improve the accuracy of long-term power load trend forecasting, the annual and monthly power load data are decomposed, Arima<sup>[6]</sup> method, which is used to study the law of data development, is integrated into BP neural network algorithm, and an improved bp-arima load trend forecasting model is proposed to realize the annual power load forecasting function under the comprehensive influence of multiple factors; the function nonparametric method is introduced to analyze the monthly data of the past years, and the periodical prediction of the future data is made by the function time series prediction model, and the component fusion of trend prediction and periodical prediction is carried out to get the new combined prediction model, so as to improve the accuracy of the long-term power load prediction.

## 2 Model construction

The BP-ARIMA forecasting model based on the functional nonparametric method decomposes the time series data, uses the BP-ARIMA model to forecast the trend of the annual load data of the time series data, and

\* Corresponding author: [fumengdi97@163.com](mailto:fumengdi97@163.com)

uses the functional nonparametric method to estimate the monthly load data. Through the component combination of the two types of data, the stability and accuracy of the long-term prediction of the model are improved. The process is shown in Figure 1.

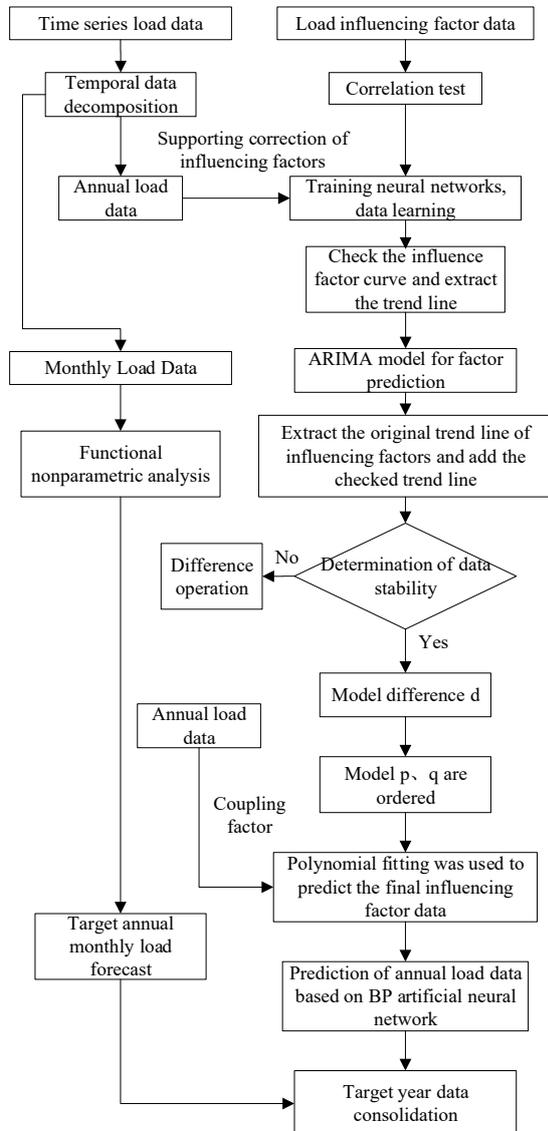


Fig. 1. Description of load forecasting model

### 2.1. Annual load forecasting based on BP-ARIMA

The model relates the regional power load with economic and social indicators such as regional GDP, industrial added value and disposable income<sup>[7]</sup>. Based on the coupling characteristics of each influencing factor, the fitting analysis of the data is adopted to clarify the accurate trend of the influencing factor data, and the BP neural network data training is carried out according to the trend of each influencing factor and the annual load data to obtain the annual power load trend forecast value. MATLAB data analysis software is used to compile the annual load prediction model of BP-ARIMA. The modeling steps are as follows:

(1)The degree of correlation between influencing factors and annual load value was preliminarily tested by correlation test method.

(2)The annual load value is used to train the neural network of each influencing factor value, Based on the checked factor prediction data, the trend data of influencing factors are extracted, and the trend line can accurately judge the trend and numerical range of the data.

(3)The ARIMA model is used for trend prediction of various influencing factors, and the checked trend line is used to replace the original trend line of influencing factors. In this process, the correlation test can be carried out by trend line prediction, and the influencing factors with weak correlation can be eliminated, and the predicted value of factors can be output by polynomial fitting method.

(4)The annual load trend forecasting is carried out by coupling the influence factor data after checking with the annual load data.

### 2.2 Related factors of long-term load forecasting

From the existing research, scholars generally use economic and social indicators to analyze the impact on power load. As regional GDP, regional financial investment and regional population growth play a positive role in promoting regional power load growth, there is an inevitable relationship between regional power load growth and power infrastructure construction and regional economic and social index growth<sup>[8]</sup>. Based on the existing research, the relevant data are collected, sorted and analyzed from the perspective of economic development, social development and industrial production. The relevant influencing factors are shown in Table 1.

### 2.3 Monthly load forecasting based on functional nonparametric method

The regional monthly load data has the characteristics of periodic increasing. For this kind of time series data, the functional data is generated by taking the whole year (period  $T = 12$ ) as the unit, and the random variable  $f$  is generated. The observation values of  $F$  from  $t = a$  to  $t = a + NT$  constitute a continuous time series  $f(T)$ . Monthly load data can be regarded as repeated statistics by cycle  $t$ . The expression of functional data<sup>[9]</sup> is as follows:

$$\{F(t), t \in [a, a + nT]\} \quad (1)$$

$$f_i = \{f(a + (i - 1)T) + r, r \in [0, T)\} \quad (2)$$

$$a \in \{1, 2, \dots, 12\}$$

Where,  $f_i$  is the load function data of year  $i$ , According to equation (2), the continuous time series  $F(T)$  on  $[a, a + NT]$  of the observation interval can be transformed into  $\{f_1, f_2, \dots, f_n\}$ , according to  $\{f_1, f_2, \dots, f_n\}$  can forecast the monthly load data of  $f_{n+1}$  and other periods. The prediction model is based on

$\{f_1, f_2, \dots, f_n\}$  to construct functional data pairs. The formula of kernel function and semi metric parameter is as follows:

$$K(u) = 1.5(1 - u^2)I_{[0,1]}(u) \quad (3)$$

$$d(x_i, y_i) = \sqrt{\int [x_i^{(q)}(r) - x_j^{(q)}(r)]^2 dt} \quad (4)$$

The selection of semi metric parameters is based on the derivative of the periodic curve, and the q value in  $d(x_i, y_i)$  affects the fitting degree of the curve. In the functional nonparametric method, the fitting degree of the curve is affected by the window width ( $h$ ) in the kernel function, and the window width is automatically obtained by cross validation, That is:

$$m_{(-n)}(x) = \sum_{i=1, i \neq j}^n Y_i K(h^{-1}d(x, X_i)) / \sum_{i=1, i \neq j}^n K(h^{-1}d(x, X_i)) \quad (5)$$

$$CV(h) = \sum_{j=1}^n (Y_j - m_{(-j)}(x_j))^2 \quad (6)$$

$$h_{opt} = \arg \min CV(h) \quad (7)$$

Remove the last n-th term from the observed  $(X_i, Y_i)$  data pairs, and analyze the data by functional regression based on the first  $n - 1$  term. The minimum value of  $CV(h)$  is used to determine the value of h. The accuracy of the periodic function prediction is determined by comparing the predicted n-th data with the real load value.

The comprehensive forecasting model combines the annual trend load forecasting with the monthly periodic load forecasting, and can allocate the weight of the monthly load and annual load through the average absolute percentage error (MAPE) weight method<sup>[10]</sup>, so as to improve the reliability and accuracy of the model forecasting. The weight distribution function of load forecasting is as follows:

$$P_{load} = w_m p_m + w_y p_y \quad (8)$$

$$w_{m,y} = MAPE_{y,m} / MAPE_m + MAPE_y \quad (9)$$

$$MAPE = \sum_{i=1}^n |real_i - predict_i / real_i| \cdot 100 / n \quad (10)$$

Where,  $P_{load}$  is the combined predicted value of load components;  $p_m$  and  $p_y$  are monthly and annual load values respectively.  $w_m$  and  $w_y$  are weights of monthly and annual load values.

### 3 Analysis of calculation examples

#### 3.1. Instance data

The data are shown in Table 1. In order to improve the accuracy of load forecasting, the monthly load indexes from 2006 to 2018 are introduced, as shown in Table 2.

**Table 1.** Statistics of electricity consumption and related factors in 1995-2018

particular year	Regional GDP (100 million yuan)	Investment in fixed assets (100 million yuan)	Per capita disposable income of urban residents (yuan)	Industrial added value (100 million yuan)	Regional resident population (10000)	Electricity consumption of secondary industry (100 million kwh)	Regional electricity consumption (100 million kwh)
1995	1 811.7	532.5	3 795.3	562.4	6 013.8	143.2	288.9
1996	2 093.5	613.3	4 512.7	634.2	6 040.3	159.3	308.1
.....							
2016	24 408.7	26 758.1	29 156.3	9 912.4	6 195.5	1 218.5	1 795.1
2017	27 519.6	29 186.4	31 640.5	10 804.5	6 254.8	1 292.5	1 921.4
2018	30 006.8	34 440.0	34 393.4	11 809.4	7 082.9	1 406.3	2 135.1

**Table 2.** Statistics of monthly electricity consumption in 2006-2018

Year	January	February	March	April	May	June	July	August	September	October	November	December
2006	44	45.2	47.9	51.9	53.4	59.1	58.9	61.3	61	57.2	62.2	59.6
2007	49.7	52.6	54.2	63.5	65.4	68.1	69.9	73.3	72.2	68.3	67.8	63.4
.....												
2016	133.8	128.9	136.8	152.4	152.1	161.3	166.3	173.1	165.2	140	144.2	140.3
2017	123.6	119.6	139.4	163.3	164.9	179.3	186.6	190.3	172.8	159.2	162.2	159.7
2018	145.6	147.3	142.3	168.6	176.3	185.6	205.6	210.5	200.4	187	180.1	185.3

#### 3.2. The data analysis

The relevant general parameters of BP-ARIMA model and the difference parameters of improved ARIMA

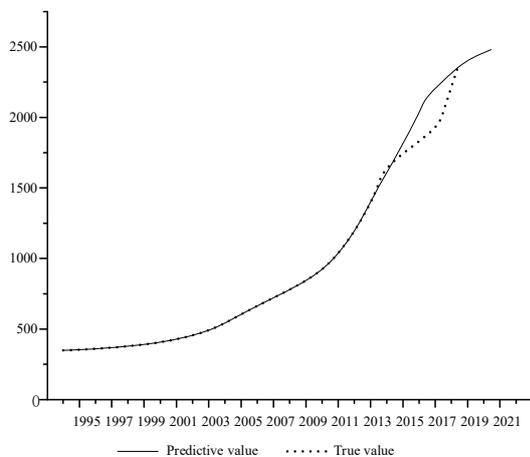
algorithm are shown in Table 3, where n is the highest power of polynomial in polynomial fitting method:

**Table 3.** General and differential parameters of BP-ARIMA mode

Branch model	The training function	Number of hidden layer variables	Training times	Training requires precision	Minimum gradient requirement	Learning rate	-
BPnet	Traingdx	21	3 000	0.1	0.00001	0.01	-
Branch model	parameter	Regional GDP	Investment in fixed assets	Per capita disposable income	Industrial output	Permanent population	Electricity consumption of secondary industry
ARIMA	Partial autocorrelation coefficient (P)	2	2	2	3	3	3
	Autocorrelation coefficient (q)	2	2	2	3	3	3
	Difference times (d)	1	2	2	2	3	1
	Highest power (n)	2	2	2	2	2	3

The correlation test was carried out between the influencing factors and the annual load value to screen out the influencing factors with strong correlation; The annual load value is used to train the neural network of each influencing factor value to check the trend line of each influencing factor; The verified ARIMA model can predict the other influential factors with strong correlation in the target year based on the characteristics of the data structure.

Through BP-ARIMA prediction model, the screened and checked influencing factors and annual power load are combined for multi factor prediction. The model is a kind of prediction based on the trend development of influencing factors and load series data. The annual regional power consumption prediction data is shown in Figure 2. In the figure, the predicted value of power load model is from 2019 to 2021, and the power load will continue to grow in the next three years, But the growth rate will be slowed down.



**Fig.2** Predicted annual power load value of BP-ARIMA prediction model

The BP-ARIMA model and functional nonparametric method are combined to forecast, and the trend forecast value and periodic forecast value from 2016 to 2018 are combined with MAPE weight method. The MAPE value

of the functional nonparametric prediction method is 1.17%, which is less than 1.93% of BP-ARIMA. After the component fusion of the two methods, the MAPE value of the new combined model is 1.59%. The error result is relatively ideal, and the load forecasting data is shown in Table 4.

**Table 4.** Prediction results of research models

Forecast method	2016	2017	2018	2019	2020	2021	MAPE index (%)
Functional nonparametric	1822	143.3	238.3	2290	2441	2611	1.17
BP-ARIMA	1854	1957	2150	2280	2374	2448	1.93
New combination method	1843	1952	2145	2287	2402	2566	1.59
True value	1795	1921	2135				

In order to reflect the prediction advantages of the new combination model, the grey prediction model GM (1,1), GM (1, N), BP neural network model and ARIMA model are respectively used to compare with the combination model. The error results are shown in Table 5. The new combination prediction model has the highest prediction accuracy, and the MAPE value is only 1.59%, which has obvious prediction advantages.

**Table 5.** General and differential parameters of BP-ARIMA model

prediction model	GM (1,1)	GM (1,N)	BP-Net	ARIMA	BP-ARIMA	Combination method
MAPE index (%)	4.72	3.94	2.49	3.26	1.93	1.59

### 4 conclusion

This paper proposes a new combined forecasting model, which combines the trend of annual load with the periodicity of monthly load. The model not only considers the trend and periodicity of data structure, but also increases the rationality of data trend forecasting through the coupling of influencing factors, and greatly improves the accuracy of data forecasting. BP-ARIMA load forecasting model combines the nonlinear

processing ability of BP with the linear forecasting ability of ARIMA, and improves the forecasting accuracy and data stability of influencing factors through trend line checking. The model can also screen out the influencing factors with poor correlation, and improve the stability of annual power load trend forecasting by using the coupling characteristics between factors. The functional nonparametric method can achieve the periodic forecasting of data by dimension reduction. The method can achieve good monthly load forecasting accuracy by selecting appropriate kernel function. The combined forecasting model can provide a more objective load forecasting method for long-term power load forecasting.

and Technology(Natural Science), 2018, 39(2): 87-93.

## References

1. Zhang Guanying, Xian Yiming, Ge Leijiao, et al. Medium and long-term load forecasting model based on Verhulst-SVM under new normal economy [J]. *Electrical Measurement & Instrumentation*, 2019, 56(1): 102-107.
2. Xie Wei, Zhao Qi, Guo Naiwang, et al. Application of the improved parallel fuzzy kernel C-means clustering algorithm in power load forecasting [J]. *Electrical Measurement & Instrumentation*, 2019, 56(11): 49-54.
3. Li Wei, Yan Ning, Zhang Zhengang. Study on long-term load forecasting of MIX-SVM based on rough set theory[J]. *Power System Protection and Control*, 2010, 38(13): 31-34.
4. Cui Herui, Mu Yupei, Peng Xu. Medium-term Power Load Forecasting Based on SARIMA Model with HP Filter[J]. *Journal of North China Electric Power University*, 2016, 43(4): 79-86.
5. Dai Xiujun, Shu Zhibiao. Implied Volatility Forecast Based on Nonparametric Regression Model[J]. *Journal of Fuzhou University*, 2018, 46(2): 156-162.
6. Zhao Binbin, Wang Ying, Wang Bin, et al. Photovoltaic power prediction in distribution network based on ARIMA model time series[J]. *Renewable Energy*, 2019, 37(6): 820-823.
7. Yu Xiaoxiao. Study on the relationship between power consumption and economic growth and short-term forecast of electricity demand in Anhui Province[D]. Hefei: Graduate School of Anhui University, 2017.
8. Zhang Zhi, Du Yanling, Cui Huijun, et al. Intelligent mid-long electricity load forecast method considering associated factors[J]. *Power System Protection and Control*, 2019, 47(2): 24-30.
9. Xu Liang, Sun Tao, Xu Jian, et al. Mid- and long-term daily load curve forecasting based on functional nonparametric regression model[J]. *Electric Power Automation Equipment*, 2015, 35(7): 89-94.
10. Li Pei, Peng Sijun. Application of New Combination Weight to Combination Forecasting Model[J]. *Journal of Henan University of Science*