# Forecasting Method of Power Consumption Information for Power Users Based on Cloud Computing

Chen Dai[1](✉), Yukun Xu[1], Chao Jiang[1], Jingrui Yan[2], and Xiaowei Dong[2]

[1] State Grid Shanghai Municipal Electric Power Company, Shanghai 200120, China
a2555421451@126.com
[2] LongShine Technology Group Co., Ltd., Shanghai 201600, China

**Abstract.** In order to realize the real-time balance of power demand and effectively avoid the waste of power, it is necessary to forecast the power consumption. Under this background, a forecasting method of power consumption information for power users based on cloud computing is designed. The prediction model framework is designed based on cloud computing technology. Carry out abnormal data processing, missing data filling and normalization for power consumption data. Calculate the correlation degree and select the influencing factors of power consumption of power users. Combined with multiple regression analysis, the forecasting model of electricity consumption information of power users is constructed. The results show that the mean absolute percentage error (MAPS), the root mean square error (RMSE) and the equalization coefficient (EC) of the method are the minimum and the maximum, which proves the accuracy of the method.

**Keywords:** Cloud Computing · Power Consumption Information of Power Users · Multiple Regression Analysis · Prediction Method

## 1 Introduction

Compared with other energy sources, electric energy has the characteristics of balance of generation, transmission and distribution and difficulty in large-scale storage. In order to achieve the real-time balance of power demand and effectively avoid the waste of electric energy, short-term forecasting of power consumption is required. Short term forecasting of power system energy can provide accurate data reference for market operation and market planning of power system. However, the short-term power consumption forecasting of power system is a complex problem, which is mainly reflected in: influenced by the type of power users, the fluctuation of power consumption is large and shows obvious periodicity; Electricity consumption is affected by weather, holidays, economic conditions and other factors, and its volatility is large. Both overestimation and underestimation of electricity demand will affect the reasonable allocation and planning of electric energy. Therefore, from the perspective of the power market, the development of


B. Wang et al. (Eds.): ICMTEL 2023, LNICST 534, pp. 304–317, 2024.
https://doi.org/10.1007/978-3-031-50577-5_22

a highly accurate power consumption forecasting method can not only provide accurate load demand data for power operators and power market participants, but also provide reasonable power consumption arrangements for power users and reduce power consumption to a large extent. With the continuous promotion of electric power reform, it is a new direction for the development of the current electric power industry to break the monopoly position of the power sales side and introduce a market-oriented competition mechanism. In order to adapt to the development of the trend, it is necessary to master the accurate power demand of power users in the process of power market transactions, which puts forward higher requirements for the accuracy of short-term power consumption prediction. Many studies have shown that the power consumption of a region is closely related to the development of its local economy, which is called a "barometer". As the most important secondary energy, electricity accounts for a high proportion of terminal energy consumption. Considering the non storage nature of power resources, oversupply will result in huge investment waste and energy consumption, which will have a negative impact on economic development. Therefore, reliable power consumption prediction is crucial for formulating correct energy development planning. Because the power consumption is affected by various uncertain factors, it is difficult to predict accurately. In order to solve this problem, many methods have been used to build prediction models. These methods can be divided into three types: time series method [1], statistical model [2]and machine learning method [3]. Time series model is one of the most basic and widely used methods in power forecasting. It predicts the future trend based on historical series. When the number of observations is large enough, time series methods can obtain higher prediction accuracy, but they are too dependent on past data and lack of interpretability. Statistical models have advantages over other models in terms of interpretability, simplicity and ease of deployment. Machine learning methods are more advanced methods to deal with electricity problems, and they are applicable to more complex calculations. Its main disadvantage is that its internal operating mechanism is unknown. Only in the case of big data can we obtain satisfactory prediction accuracy, which requires great efforts in data compilation. Under this background, a forecasting method of power consumption information for power users based on cloud computing is proposed.

## 2 Prediction Model of Power Consumption Information for Power Users

### 2.1 Cloud Computing Technology Design Prediction Model Framework

Cloud computing is an effective way to process massive data proposed by the industry in recent years [4]. Cloud computing mainly includes three levels: Infrastructure as a Service, Platform as a Service and Software as a Service. Generally, cloud computing should have the following characteristics: rapidly deploy resources or obtain services based on virtualization technology; Realize dynamic and scalable expansion; Provide resources according to demand and pay according to usage; It is provided through the Internet for massive information processing; Users can easily participate; Flexible shape, free gathering and dispersing; Reduce the processing burden of user terminals; It reduces users' dependence on IT expertise. The core of cloud computing is Map/Reduce technology. The two steps of Map and Reduce are mainly used to process large-scale datasets in parallel. First, the Map function will first perform the specified operation on each element of the original data composed of many independent elements, and the original data will not be changed. In this step, multiple new lists (Key Value pairs) will be created to save the processing results of the Map, so the Map operation is highly parallel. After the map work is completed, the system will then shuffle and sort the newly generated multiple lists, and then reduce these newly created lists, and merge the elements in a list appropriately according to the Key value. Since Google put forward the concept of cloud computing and Map/Reduce, many companies have implemented it. The most famous and widely used one is Hadoop, the open-source version of Map/Reduce developed by Yahoo team led by Doug Cutting, the father of Lucene, and managed by Apache.

In the application of smart grid, the grid company has collected a large amount of power consumption information from users and stored it in the historical database as the historical data of power load forecasting. The power consumption information of a user forms a time series. In the comprehensive consideration of weather, politics and other factors, the user's power load data of the next day or hour can be predicted based on the time series or other technologies, and these data can be pushed to the user. The user can then adjust the forecast data according to his actual situation, and return the adjusted data to the grid company, So as to ultimately improve the accuracy of load forecasting of grid companies and improve the efficiency of power generation. However, short-term power load forecasting for tens of millions of users and massive data is a very challenging task, which requires a data processing center with strong computing power of power grid companies. Such a data center is expensive. Cloud computing can combine some low-cost devices to meet the rapid processing requirements of large amounts of data. This paper proposes a short-term load forecasting technology based on Map/Reduce. According to the implementation process of Map/Reduce, this paper divides the power consumption prediction based on Map/Reduce into the following steps.

1) Preprocessing: preprocess the original data to improve the quality of sampling data.
2) Map: Select an appropriate mapping function, such as remainder operation, to process data in parallel and map data with the same user ID to the same node.
3) Shuffle: Each node sorts its data by time, and the sorting is performed by month, day, and hour.
4) Reduce: The sorted data on each node forms a time series, which uses a mature power load forecasting technology to forecast the power load data of the next period.
5) Output result: get the power load forecast data of each user in the next time period, and push it to the user through the broadband network, so as to ensure that the user can obtain these data in a timely manner.

## 2.2 Power Consumption Data Preprocessing

With the promotion of smart grid construction, the traditional manual recording of electricity has been replaced by smart meters. The accuracy of power consumption recording has been greatly improved through smart meters. However, due to certain risks in the operation of power grid equipment, special circumstances or emergencies may cause the loss or mutation of electricity data. Therefore, when forecasting power consumption, it is first necessary to preprocess the collected original data. The purpose of preprocessing is to improve the quality of the sampled data, reasonably predict and fill the missing data, so as to discover the real power consumption law and improve the prediction accuracy [5]. At the same time, in the process of power consumption prediction, due to many factors affecting power consumption, there may be some correlation and redundant information between each factor, which will lead to the increase of the network size of the prediction model and the increase of time in the calculation process, thus reducing the prediction accuracy of the model. Therefore, the accuracy of the prediction model can be improved by extracting the features of correlation and redundant information of related factors based on intelligent algorithms.

### 2.2.1 Abnormal Data Processing

In fact, in the process of long-term stable operation of the power system, there will be no sudden change in the power consumption of the system, and the historical data of power consumption is also a continuous and stable time series, so the difference between the power consumption data at a specific time and the power consumption data at adjacent times in the historical data of power consumption is not large [6]. When the change value of electricity consumption at a certain time is too large, it needs to be corrected by horizontal processing method. First, use Formula 1 to judge the abnormal value.

$$\max[|S(T,t) - S(T,t-1), |S(T,t) - S(T,t+1)] > A(t) \tag{1}$$

Then use Formula 2 to correct the abnormal value.

$$S(T,t) = \frac{S(T,t-1) + S(T,t+1)}{2} \tag{2}$$

where, $S(T,t)$ is the threshold value; $t$ is the sampling point; $S(T,t+1)$ is the power consumption at time $t+1$ on day $T$; $S(T,t)$ is the electricity consumption at time $t$ on day $T$; $S(T,t-1)$ is the power consumption at $t-1$ on day $T$.

### 2.2.2 Defect Data Completion

In the actual big data of power consumption, data loss often occurs due to storage or human factors. The rapid detection of outliers in the previous section will also bring data loss points. The existence of these defects will make it difficult to model, increase the error of calculation results, and even cause the collapse of the analysis program. What needs to be discussed in this section is how to effectively complete the missing data of power consumption big data [7]. Here, KNN filling algorithm is selected to complete the missing data with big TV data.

The data filling algorithm based on KNN mainly includes three main stages: missing point search, missing point calculation and missing point recursion. First, search the number and location of missing points in the original data set, then use the KNN algorithm to find the K-nearest neighbor of each missing point in turn and calculate the distance between the missing point and the K-nearest neighbor, and calculate the filling value of the missing point by giving the weight that the K-nearest neighbor is inversely proportional to the distance. After each missing point is calculated, update the original data set for circular calculation until there is no missing value. Next, calculate the difference between the current filling value vector and the last round of recorded filling value vector. If the difference is less than the predefined allowable error, stop recursion, otherwise repeat the missing point calculation process until the conditions are met. However, KNN filling algorithm uses all the nearest neighbors of target data (data records with missing items) to participate in the filling of target data. It ignores the problem that the k-nearest neighbor of target data records may have noise [8]. Therefore, the accuracy of KNN filling algorithm in filling missing data depends to a large extent on the quality of the original data. In the real world, the noise in the data always exists. When the k nearest neighbor has noise, filling missing data with KNN algorithm will produce large deviation. To solve this problem, the ENN-KNN missing data filling algorithm proposed in this paper can effectively eliminate the impact of such noise on the filling results by comparing the true nearest neighbor degree between the nearest neighbors of the target data records, thus improving the accuracy of filling [9]. The ENN KNN missing data filling algorithm is identical to the KNN algorithm in obtaining the k-nearest neighbor of the target data.

By comparing the Euclidean distance between the target data and the perfect value data, k nearest neighbors of the target data are obtained. In this paper, $P(x_i)$ is used to represent the nearest neighbor set of the target data $x_i$. After getting the nearest neighbor of target data $x_i$, use the same method to find the nearest neighbor of each nearest neighbor of target data. In this paper, $PP(x_i) \quad i = 1, 2, ..., k$ represents the nearest neighbor set of the $i$ nearest neighbor in the nearest neighbor set $P(x_i)$. The specific process is as follows:

1) Data initialization and construction of complete value data matrix;
2) Calculate the Euclidean distance $d_i(x_i, X)$ between the target data and all data records in the complete value data matrix, as shown in Formula 3;

$$d_i(x_i, X) = \sqrt{(x_i - X)^T (x_i - X)} \tag{3}$$

where, $x_i$ represents target data; $X$ stands for complete value data matrix.

3) The $k$ data records with the minimum Euclidean distance are selected as the $k$ nearest neighbor of the target data, and their positions in the complete value data matrix are stored in the array $B$;
4) Select the $k$ data records with the minimum Euclidean distance to the nearest neighbor of each target data from the complete value data matrix, and store their positions in the complete value data matrix into the two-dimensional array $\hat{B}$;
5) Initialize the neighbor importance $Q_{KK(x_i)}$ of the nearest neighbor of each target data;
6) Calculate the importance of each data record in the complete value data matrix, such as Formula 4.

$$q_i = \sum_{i=1}^{k} \sum_{i \in KK(x_i)} \frac{Q_{KK(x_i)}}{k} \tag{4}$$

where, $q_i$ represents the importance of the $i$-th complete value data record in the data set.

7) Calculate the true nearest neighbor degree of k nearest neighbors of target data $x_i$, as shown in Formula 5.

$$\hat{q}_i = \sum_{i \in KK(x_i)} q_i \tag{5}$$

where, $\hat{q}_i$ represents the true nearest neighbor degree of the $i$-th nearest neighbor of the target data.

8) Remove the noise nearest neighbor of the target data. The ENN KNN missing data filling algorithm determines the nearest neighbor of noise based on the true nearest neighbor degree of the nearest neighbor [10]. The judgment standard is shown in Formula 6:

$$Z_i = \begin{cases} 1, \ \hat{q}_i < \alpha \overline{q} \\ 0, \ \text{else} \end{cases} \tag{6}$$

In the formula, $\overline{q}$ represents the average of the true nearest neighbor degree. $Z_i$ represents the noise judgment result of the $i$-nearest neighbor of the target data record. When the result is 1, it means that this nearest neighbor is the noise nearest neighbor of the target data. When the result is 0, this nearest neighbor is the non noise nearest neighbor of the target data. $\alpha$ is the elimination coefficient, and the value range is (0, 1). When $\alpha \to 0$, ENN-KNN filling algorithm is equivalent to KNN filling algorithm. When $\alpha \to 1$, the ENN-KNN filling algorithm is equivalent to filling the target data with the nearest neighbor with the greatest degree of real nearest neighbor.

9) According to $Z_i$, calculate the nearest neighbor weight $w_i$ of the target data;
10) Estimate the value of missing data and fill it, as shown in Formula 7;

$$\hat{x}_i = \sum_{i=1}^{k} w_i c_i \tag{7}$$

In the formula, $c_i$ represents the value of the corresponding position of the nearest neighbor of the target data $x_i$;

11) Repeat steps 2 to 10 until there is no missing data in the dataset.

### 2.2.3 Normalization of Power Consumption Data

In the process of processing the historical data of electricity consumption, because the factors that affect electricity consumption also need to be used in the prediction process, and the units of electricity consumption and various influencing factors are different, the relevant variables need to be normalized before the prediction to convert into data without dimensions [11], thus having little impact on the prediction of the model. Therefore, in the process of power consumption data processing, all variables should be normalized to ensure that each influencing factor is within the same numerical range. Normalization can be used to process each variable. The normalized data processing process of power consumption data is Formula 8:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

where, $x_i$ is the power consumption data to be normalized; $\hat{x}_i$ represents the data obtained after normalization; Intact data of previous and subsequent days; $x_{\max}$ is the maximum of the sample data used; $x_{\min}$ is the minimum of the sample data used.

After the above steps, the user power data is cleaned and normalized to the maximum value, and its data is transformed into a complete and comparable time series without outliers.

### 2.3 Selection of Influencing Factors for Power Consumption of Power Users

With the continuous improvement of living standards, people's preferences for electricity are also changing. Through reading the literature, the previous studies involved the following three factors affecting power consumption: economy, temperature and industrial structure. As far as residential power consumption is concerned, it has certain particularity. For example, as far as industrial structure is concerned, it represents the production status of the primary, secondary and tertiary industries, and more reflects production power consumption than residential power consumption; For example, household electricity consumption will be more sensitive to the reform of electricity price policy. Through summary and induction, the influencing factors of corresponding indicators are given, and then the influencing factors and power consumption are analyzed by grey correlation to obtain several influencing factors with the largest correlation with power consumption, providing basis for subsequent modeling [12]. Grey correlation analysis is an important part of the grey system theory, which is widely used in education, economy, agriculture and other fields. The degree of correlation between factors is judged by the degree of similarity of the development trend among the research objects, that is, the "grey correlation degree". The calculation formula of correlation degree is shown in Formula 9:

$$\psi_{0i} = \frac{\sum_{k=1}^{m} \lambda_i(k)}{m}, \quad k = 1, 2, \ldots, m \tag{9}$$

Where, $\lambda_i(k)$ is each index in the comparison sequence and the reference sequence, and its absolute difference is calculated; $m$ is the number of influencing factors.

The residential electricity consumption is mainly affected by climate, economy and industrial characteristics. According to the above correlation analysis, seven influencing factors are selected to form an indicator system (see Table 1), and relevant data are collected and sorted out.

Table 1 residential electricity consumption index system specifically includes climate, population, living area, economy, total energy consumption, electricity price, consumption habits, etc. More detailed analysis will be carried out below.

(1) Influence of climate on electricity consumption

   As far as climate is concerned, some studies show that the annual average temperature has little impact on power consumption, which is probably because the fluctuation of the national average annual temperature is relatively small within the studied time interval, so the impact of season and geographical climate on the national power consumption cannot be shown, but this does not negate the impact

**Table 1.** Residential electricity consumption index system

| Dependent variable | Independent variable |
|---|---|
| Customer power consumption (10000 DWh) | Average temperature (°C) |
| | Total population (10000 persons) |
| | Permanent population (10000 persons) |
| | Residential area (m2) |
| | Per capita disposable income (yuan) |
| | GDP (100 million yuan) |
| | Total energy consumption (100 million tons of coal) |

of season and temperature on power consumption. As the research in this paper is a medium and long-term power load forecast on a monthly basis, the temperature will have a greater impact on power consumption, such as cooling in summer and heating in winter. Especially in Chengdu, where there is no heating, electric heaters are mainly used for heating in winter; At the same time, the high temperature will urge families to turn on air conditioners, electric fans and more refrigerators, leading to an increase in power consumption. Therefore, the monthly average temperature in Chengdu is taken as the first influencing factor index and recorded as R1.

(2) Impact of economy on electricity consumption

Population, living area, GDP, etc. directly or indirectly reflect the economic development level of a region. According to literature research, economy also has a great impact on power consumption. To be specific, the impact of population on electricity consumption can be imagined. The more regional population, the more electricity consumption will naturally bring. It should have a strong positive correlation with residential electricity consumption. The total population and permanent population in Chengdu are taken as the influencing factor indicators and recorded as R2 and R3.

In fact, the living area indirectly reflects the population. From a common sense point of view, residents' electricity use will be more sensitive to electricity prices. The rise in electricity prices will improve people's awareness of saving electricity in daily life, but it often has little impact on industrial production. Of course, whether it is as expected in this paper depends on the results of grey correlation analysis. Take the per capita residential building area as an indicator of influencing factors and record it as R4.

The level of regional economic development greatly affects the power consumption of urban residents. The per capita disposable income of urban residents and the gross regional product reflect the level of economic development. Here, the per capita disposable income and annual gross domestic product of urban residents in Chengdu are included in the influencing factor indicators, which are recorded as R5 and R6 respectively.

(3) Impact of industrial characteristics on power consumption

For the energy industry of electricity, from the total amount of macro energy consumption to the micro electricity price, as well as the energy consumption habits of the end consumers themselves, will affect the electricity consumption behavior of the end residents.

The impact of total energy consumption on electricity consumption should be positive correlation, as it is common sense that the increase of total energy consumption will drive residents' use of electricity. The total energy consumption in Chengdu is taken as the influencing factor indicator and recorded as R7.

### 2.4 Prediction Model Construction

For the long-term development of a region, power system planning, design and dispatching are all factors affecting its development, so it is necessary to forecast the medium and long-term power consumption. With the great changes of power demand, some new characteristics have emerged in load characteristics. How to effectively improve the accuracy of medium and long-term power consumption forecast, first, fully consider the relevant factors, and second, accurately grasp the load characteristics.

Based on the main factors that affect the power demand and load characteristics analyzed above, this paper constructs a single mathematical model to predict the medium and long-term power consumption of the city in accordance with the four main factors that affect the power demand, namely, macroeconomic trend, industrial structure adjustment, development of key power use industries and climate conditions. The time span here is five to ten years. The power consumption regression model prediction technology is to use the regression analysis method in mathematics to count the historical data of power consumption and analyze the relationship between variables to obtain a definite curve, establish a mathematical model of power consumption and related variables, and extend the curve to a certain time to obtain the forecast value at that time, so as to achieve the purpose of prediction. In general, it is to use the mathematical model to predict future power consumption. The factors affecting electricity consumption in electricity consumption prediction are random, so take them as independent variables and electricity consumption as dependent variables [13]. The dependent variable changes with the independent variable, the independent variable is the cause, and the dependent variable is the result, and this relationship cannot be reversed. The power consumption forecasting problem just belongs to this kind of problem, that is, the fitting curve obtained can only be the fitting of random variables, and can not be used to describe the random changing power consumption in turn. The fitting curve can be established by either straight line fitting or curve fitting. The former is called linear regression and the latter is called nonlinear regression. Since there are many influencing factors (independent variables) of user electricity consumption, it is necessary to establish a multiple regression model.

Multiple linear regression is to study whether two or more independent variables and a dependent variable have a linear relationship of interdependence. This relationship can usually be expressed in terms of multiple regression equation, that is, the multiple regression equation depicts a linear regression model with two or more independent variables in the relationship equation between a dependent variable and multiple independent variables. In this model, the dependent variable is a linear function of multiple independent variables $R_1, R_2, .., R_M$ and error terms, such as Formula 10.

$$L = g_0 + g_1 R_1 + g_2 R_2 + ... + g_M R_M + E \tag{10}$$

where, $Y$ is the dependent variable (electricity consumption); $M$ is the number of independent variables (influencing factors); $g_0, g_1, ..., g_M$ is the regression coefficient; $E$ is the random error term.

In practical application, if $n$ group observation data $R_{j1}, R_{j2}, ..., R_{jM}$ $(j = 1, 2, ..., n)$ is obtained, as shown in Formula 11:

$$L_j = g_0 + g_1 R_{j1} + g_2 R_{j2} + ... + g_M R_{jM} + E_j \tag{11}$$

The corresponding matrix expression is Formula 12:

$$L = Rg + E \tag{12}$$

where, $L$ is the observed value vector of the dependent variable (the observed value of electricity consumption); $R$ is the observed value matrix of the independent variable (power consumption influencing factor); $g$ is the population regression parameter vector.

To facilitate parameter estimation of the model, the following basic assumptions are made for the equation:

(1) The independent variable $R_1, R_2, .., R_M$ is a deterministic variable and $R$ is a full rank matrix;
(2) The random error term has zero mean and equal variance, that is, it satisfies the Gauss Markov condition;
(3) The random error term obeys normal distribution;

The regression analysis steps for multiple variables are as follows:

Step 1: Draw a scatter chart and observe its distribution characteristics;
Step 2: Carry out the corresponding variable conversion according to the selected function;
Step 3: establish a linear regression model for the transformed data;
Step 4: Use the least square method to solve the equation of the urban road network capacity prediction model, and obtain the estimated values of various multivariate linear regression coefficients;
Step 5: Test the parameters of the linear regression model. The inspection includes $R$ inspection, t inspection and F inspection.

According to the regression analysis equation obtained, the value of the independent variable is directly brought into the value of the dependent variable, that is, the value of the influencing factor is brought into the value of the power consumption.

# 3 Method Validation

## 3.1 Power Consumption Data

The data used in this paper is from the daily hourly electricity consumption data of California from February 6, 2017 to April 16, 2017. This group of data has 24 observations every day, and each group of observations covers 2GB data. There are 480GB observations in 10 weeks. In this paper, the data of 9 weeks from February 6, 2017 to April 9, 2017 are used to forecast the average hourly power consumption from April 10, 2017 to April 16, 2017. The fluctuation of the original power consumption data is shown in Fig. 1.

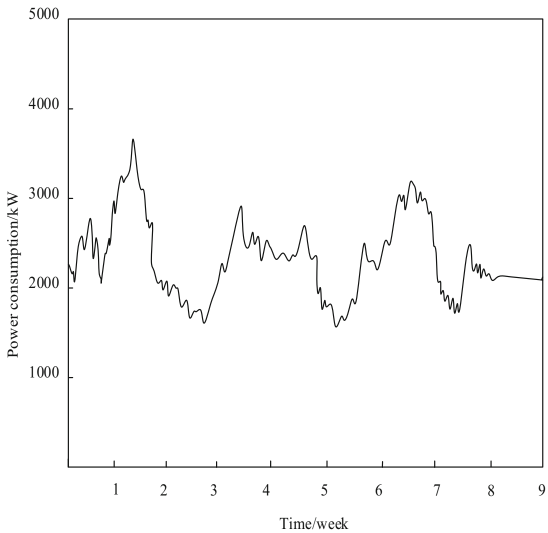


**Fig. 1.** Trend Chart of Data Power Consumption in 9 Weeks

It can be seen from Fig. 1 that the original series is non-stationary and nonlinear, and the seasonal effect of the series is very obvious.

### 3.2 Model Evaluation Criteria

In order to judge the prediction performance of each model, the indicators used to evaluate the quality of prediction models in this paper are: average absolute percentage error (MAPE), root mean square error (RMSE), and equalization coefficient (EC). In general, a small value of MAPE and RMSE indicates that the prediction model has a better effect. EC indicates the consistency between the real measured value and the predicted value. The larger the value, the better the prediction effect of the model. When EC $>$ 0.9, the prediction is meaningful. The indicators are defined as follows:

(1) Mean absolute percentage error (MAPE), as shown in Formula 13:

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{\beta_i - \hat{\beta}_i}{\beta_i} \right|}{n} \tag{13}$$

(2) Root mean square error (RMSE), as shown in Formula 14:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\beta_i - \hat{\beta}_i\right)^2}{n}} \tag{14}$$

(3) Equalization coefficient (EC), such as Formula 15:

$$EC = 1 - \frac{\sqrt{\sum_{i=1}^{n}\left(\beta_i - \hat{\beta}_i\right)^2}}{\sqrt{\sum_{i=1}^{n}(\beta_i)^2} + \sqrt{\sum_{i=1}^{n}\left(\hat{\beta}_i\right)^2}} \tag{15}$$

where, $\beta_i$ is the observed value of power consumption data, $\hat{\beta}_i$ is the predicted value of power consumption data, and $n$ is the number of load prediction results.

The above three evaluation indexes are used to evaluate the prediction performance of the model.

### 3.3 Prediction Results

The established linear regression model is used to predict the average hourly power consumption in the 10th week, and the results are shown in Fig. 2.

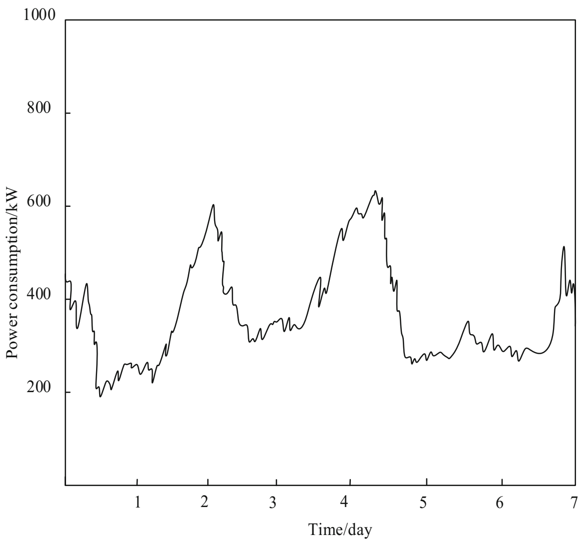


**Fig. 2.** Power Consumption Forecast Results

### 3.4 Prediction Accuracy Comparison

Under the same data, the proposed method, reference [1] time series method, reference [3] statistical model and reference [3] machine learning method are used to predict the average hourly power consumption in the 10th week, and then calculate the average absolute percentage error (MAPS), root mean square error (RMSE) and average coefficient (EC) between the prediction results and the actual results, as shown in Table 2.

**Table 2.** Comparison of Forecast Accuracy

| Method | MAPE | RMSE | EC |
|---|---|---|---|
| Research methods | 1.3622 | 21.3255 | 0.9987 |
| Time series method | 2.6247 | 84.2521 | 0.9421 |
| Statistical model | 2.0124 | 63.1425 | 0.9632 |
| Machine learning method | 2.8742 | 72.2725 | 0.9425 |

It can be seen from Table 2 that compared with time series method, statistical model and machine learning method, the average absolute percentage error (MAPS), root mean square error (RMSE) of the studied method are smaller, and the equalization coefficient (EC) is higher, which proves the prediction accuracy of this method.

## 4 Conclusion

In order to improve the applicability of power user power consumption information prediction, this paper proposes a cloud computing based power user power consumption information prediction method. In this paper, based on cloud computing technology, a power consumption information prediction model for power users is constructed. On this basis, the influencing factors of power consumption are analyzed. The experimental results show that the proposed method has good prediction accuracy.

## References

1. Saoud, L.S., Ghorbani, R.: Metacognitive octonion-valued neural networks as they relate to time series analysis. IEEE Trans. Neural Netw. Learning Syst. **31**(2), 539–548 (2020)
2. Sathishkumar, V.E., Shin, C., Cho, Y.: Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city. Build. Res. Inform. **49**(1), 127–143 (2021)
3. Geetha, R., Ramyadevi, K., Balasubramanian, M.: Prediction of domestic power peak demand and consumption using supervised machine learning with smart meter dataset. Multimed. Tools Appl. **80**(13), 19675–19693 (2021)
4. Praveenchandar, J., Tamilarasi, A.: Dynamic resource allocation with optimized task scheduling and improved power management in cloud computing. J. Ambient. Intell. Humaniz. Comput. **12**(3), 4147–4159 (2021)
5. Liang, Y., Hu, Z., Li, K.: Power consumption model based on feature selection and deep learning in cloud computing scenarios. IET Commun. **14**(10), 1610–1618 (2020)
6. Ding, S., Li, R., Wu, S., et al.: Application of a novel structure-adaptative grey model with adjustable time power item for nuclear energy consumption forecasting. Appl. Energy **298**(3), 117114 (2021)
7. López, M., Valero, S., Sans, C., et al.: Use of available daylight to improve short-term load forecasting accuracy. Energies **14**(1), 95 (2020)
8. Kim, Y., Kim, S.: Forecasting charging demand of electric vehicles using time-series models. Energies **14**(5), 1487 (2021)

9. Wu, X., Dou, C., Yue, D.: Electricity load forecast considering search engine indices. Electr. Power Syst. Res. **199**(11), 107398 (2021)
10. Ren, C., Xu, Y., Dai, Bijian, Zhang, Rui: An integrated transfer learning method for power system dynamic security assessment of unlearned faults with missing data. IEEE Trans. Power Syst. **36**(5), 4856–4859 (2021)
11. Narimani, M.R., Molzahn, D.K., Crow, M.L.: Tightening QC relaxations of AC optimal power flow problems via complex per unit normalization. IEEE Trans. Power Syst. **36**(1), 281–291 (2021)
12. Ahn, H.K., Park, N.: Deep RNN-based photovoltaic power short-term forecast using power IoT sensors. Energies **14**(2), 436 (2021)
13. Wu, W., Song, Y., Wei, S.: Research on power consumption forecasting based on improved prophet mode. Comput. Simul. **38**(11), 473–478 (2021)