Load Forecasting of Distributed System Host Based on Linear Time Series Model

Fawazen Almulihi*

Jawaharlal Nehru University, India

*corresponding author

Abstract: With the development of communication network, wireless data transmission technology is becoming more and more important in people's daily life. As a new structure, distributed system has strong applicability and adaptability. This paper mainly studies the load forecasting of urban cloud computing platform based on time series model. Firstly, the load forecasting method and principle are introduced. Secondly, according to the actual situation, the appropriate range of random disturbance parameters is determined and a stable and reliable index system is established. Finally, through simulation experiments, the load rate curves of different time resolution and different types of distributed systems under uniform load distribution are verified, and the stability and accuracy are improved by adjusting the load degree. From the test results, it can be seen that the host load prediction results for linear time series have small errors, which indicates that the prediction model has high accuracy.

1. Introduction

With the rapid development of China's economy, energy and environmental problems are increasingly serious, and the investment in renewable energy power generation systems is also increasing [1-2]. As a kind of clean, efficient, environmental protection and low cost, distributed power generation has many advantages. In the case of insufficient power supply, it can effectively relieve the tension of power consumption in the peak load of the power grid. Meanwhile, the distributed power supply enterprise adopts the decentralized management mode for centralized regulation and control, which makes it have certain flexibility and adaptability [3-4].

For the research of distributed system, foreign countries started earlier and developed rapidly. At present, many scholars have done a lot of theoretical analysis on it. The United States calculates and describes the relationship between time series based on discrete points and continuous function equations, and there is a gap between the predicted value and the actual data. It establishes the change law of distributed network load rate in the period based on time series, and proposes the idea and method of using Bayesian algorithm to optimize it [5-6]. Domestic scholars started their

research on load forecasting late, but also made some achievements. Some scholars have established mathematical models and analyzed them with examples. They believe that with the correlation and regular change trend (i.e., linear) between time series and random data, the development of non-linear relationship and the improvement and in-depth study of relevant theories all show that the load forecasting based on time series model can be realized quickly and accurately. Some scholars put forward a new method after analyzing the application status of distributed energy generation technology in China's power industry [7-8]. Therefore, based on the linear time series model, this paper studies the load prediction of distributed system host.

Aiming at the traditional load forecasting of buildings based on time series model, combined with some relevant index data, a distributed system load calculation method is established. And combine it with randomness and non-stationary characteristics. This paper uses a new processing scheme. The algorithm uses linear regression analysis and discrete variance transformation to solve the nonlinear problem. At the same time, it uses the linear criterion method to combine the distributed theoretical knowledge and the advantages of traditional mathematical modeling technology to propose a time series model method to study the load forecasting of buildings, and finally obtains satisfactory and reliable results.

2. Research on Load Forecasting of Distributed System Host Based on Linear Time Series Model

2.1. Host Load Prediction

Host load prediction is based on the data of time series analysis method. When estimating the actual output power, the relative change law is determined by calculating and comparing the values of various parameters. First, compare the actual input, output flow and relevant historical data in each time period with the results measured in the original model [9-10]. Then the corresponding values are obtained by different algorithms, and then the computer software is used to predict and analyze these data. Finally, the load situation in each time interval is judged according to the predicted trend curve. In the time series model method and the traditional system load calculation, the traditional historical data is used as the basis for analysis. By sorting, statistics and processing the original time series data, and then using computer technology to establish a complete, reliable, reasonable, economic and high-precision system that can be applied to various fields. According to the actual situation, select appropriate methods to further predict the parameters. The research work is based on the consideration of the load characteristics of the host. In the actual problems, it is necessary to fully combine various influencing factors to better achieve and accurately solve the objectives [11-12]. Fig. 1 is a load prediction model.
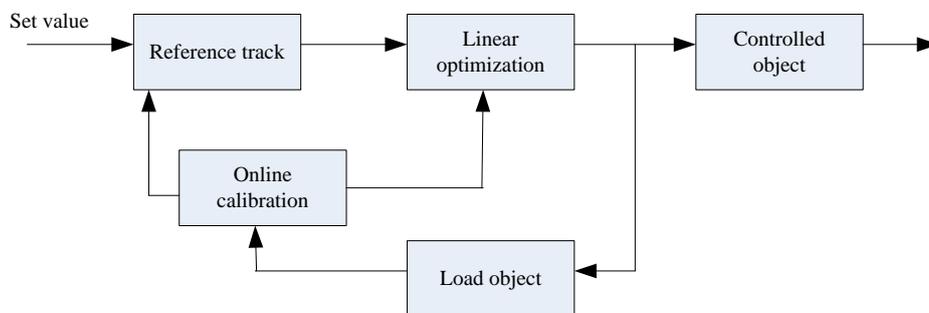

Figure 1. Load prediction model

2.2. Problems in Host Load Prediction

In actual work, the load of the host is a random change state rather than a continuous variable, so it is impossible to determine that there is no connection between the load curve and the power spectrum. Moreover, due to energy consumption, output fluctuation and other factors, the node voltage and current are unstable or even serious problems, resulting in an increase in the probability of system collapse or abnormal operation of the host shutdown, thus affecting the stability and reliability of the entire network. Since the time series model method is a random simulation based on probability and statistical characteristics, it does not take into account the impact of time factors on the actual output power change, so this method is subjective to a certain extent [13-14]. But doing so will also cause errors, because the algorithm requires a large amount of data in the calculation process. And these data often have the characteristics of high frequency, large amplitude fluctuation and instability. It may lead to large differences between different times, and this result will also affect the prediction accuracy. For example, it is difficult to deal with these problems with the simplest and most effective linear function when the data changes or is unstable. In order to avoid these negative effects as much as possible, it is necessary to continuously optimize the parameter selection and adjustment methods in practical applications to ensure that the final output results are more accurate and reliable. For example, after determining the main data set, how to carry out appropriate load calculation according to the specific situation has also become an important factor. Therefore, in actual work, it is often necessary to establish a variety of data information indicators such as different types and the same type of use, similar performance, and can accurately reflect the change trend of the total demand of the system under various conditions to determine the relationship and relationship between various parameters [15-16].

2.3. Linear Time Series Model

Linear time series model is a common random data processing tool, which can solve noisy and nonstationary indexes such as system parameters and power spectral density [17-18]. In practical applications, it is usually unnecessary to consider the signal strength and frequency response characteristics. Information can be directly obtained from the sample space and converted into discrete vectors to establish a correlation matrix or vector model. The original data sequence and its corresponding wavelet coefficients and corresponding time domain variables, such as step function and periodic spectral density, can also be obtained by calculation. Given the parameters (such as model coefficients, autocorrelation probabilities, etc.) and expectation functions required by the time series model, the calculation steps are given as follows. First, a undetermined physical quantity is established as the input variable to analyze the prediction problem, then the correlation in the Gaussian white noise log data processing method is used, and finally, the appropriate modeling algorithm is selected according to the actual needs. As for a random event, there are uncertain factors that affect the occurrence and duration of the event. Secondly, due to the various characteristics of the target, there may be errors or omissions in the process of processing. These factors may lead to inaccurate prediction results or even failure of the model. Figure 2 is a flowchart of a linear time series model.
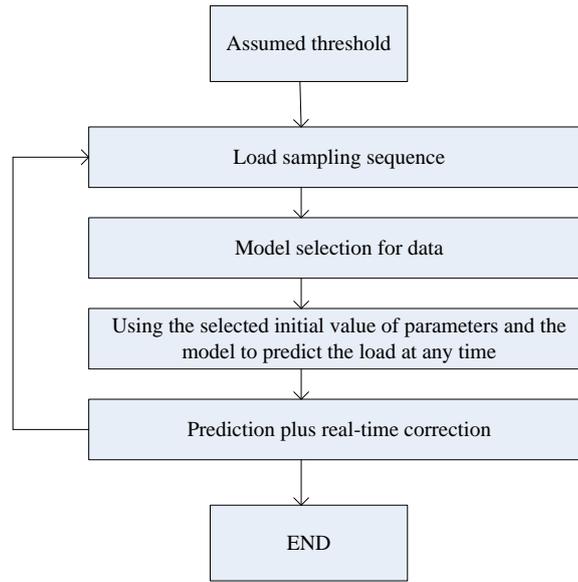
Figure 2. Linear time-series model

Linear time series algorithm is a statistical method, which combines data processing and analysis. In practical application, due to the influence of external random factors such as weather and temperature, and some unpredictable factors such as earthquake activity, etc. Therefore, we need to estimate these random variables to a certain extent to establish their theoretical models or estimate parameters so as to obtain an approximate value or more accurate results as the research object. At the same time, we can also calculate the output error after data processing through the linear time series algorithm, and then use this method to analyze and evaluate the problems in the data.

$$E_{PA} = \frac{1}{N} \sum_{i=1}^{N} [P - o] \tag{1}$$

$$E_R = \frac{\sum_{i=1}^{N} [P_i - P]}{(N-1)S} \tag{2}$$

Equation 1 reflects the correlation between the deviation of the predicted value near its mean value and the deviation of the observed value near its mean value. The value is between + 1 and - 1. EPA = 1 indicates that the predicted curve and the observed value curve are the same after normalizing their mean and standard deviation respectively.

3. Experimental Process of Load Forecasting of Distributed System Host Based on Linear Time Series Model

3.1. Distributed System Host Load Prediction Model

In this paper, the method based on time series model is used to process the prediction data used by traditional methods. The algorithm mainly uses linear regression analysis and grey system theory. It can take the actual parameters in the original historical data that are similar or identical to the future with high attribute values or large load coefficients as reference objects to establish

differential equations to calculate the output power and power of each node of the distributed system and their change regularity (i.e. load) at each coordinate point, and predict the time series model method. For the time series model method, it is mainly aimed at linear prediction, but in practical application, it is often adjusted according to different conditions (as shown in Figure 3). Therefore, data processing, parameter estimation and state space should be reasonably divided. Firstly, a simple continuous tree is established to describe the characteristics of things or processes. Secondly, discrete points are decomposed into finite independent nodes to form a sub layer by using continuous functions, and the two adjacent independent elements on each node are connected to form an overall network structure. Then, the minimum distance method is used to calculate the load value of each local link.
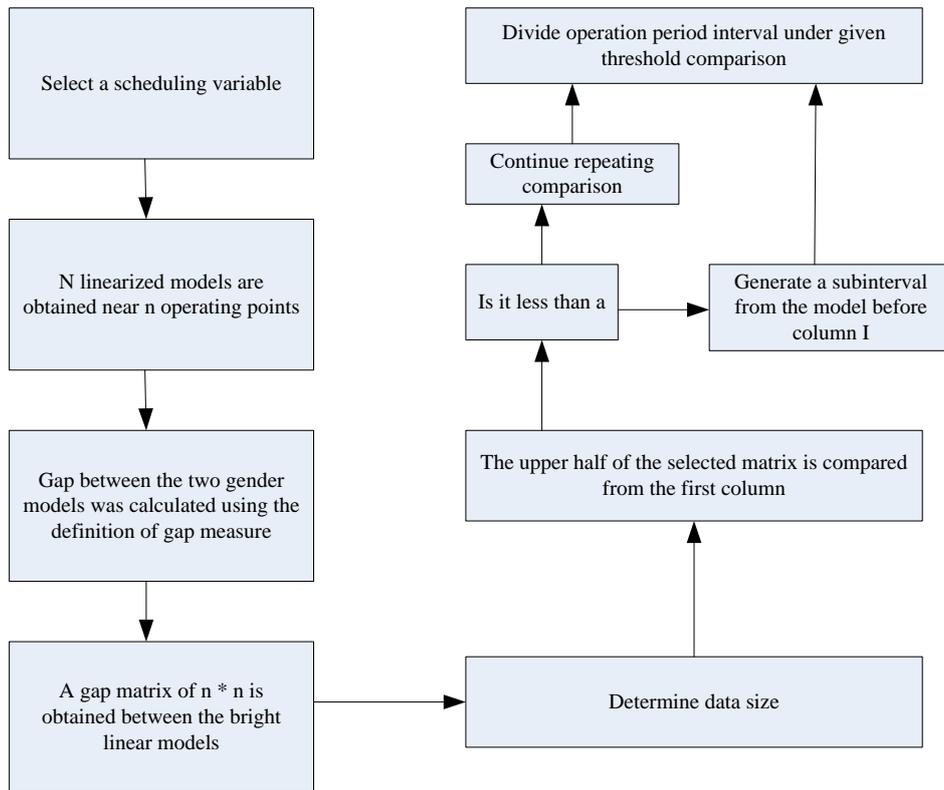
Figure 3. Distributed system host load prediction model

3.2. Load Forecasting Performance Test of Distributed System Host Based on Linear Time Series Model

First, the model is analyzed, and then the indicators are determined. A corresponding relationship shall be established between the initial parameters and the calculation results before the preliminary design is completed. If the selected method is set based on the user's needs, the relevant data should be recorded in advance and saved to provide reference for future use. Secondly, the appropriate sample size, sample size, time series and other information should be selected as the test standard according to the actual situation. Finally, whether the expected effect is achieved can be judged through experiments, so as to determine that the model is suitable for the load forecasting scheme in this paper.

4. Experimental Analysis of Load Forecasting of Distributed System Host Based on Linear Time Series Model

4.1. Load Prediction Performance Verification of Distributed System Host

Table 1 shows the host load prediction performance test data.

Table 1. Estimated performance

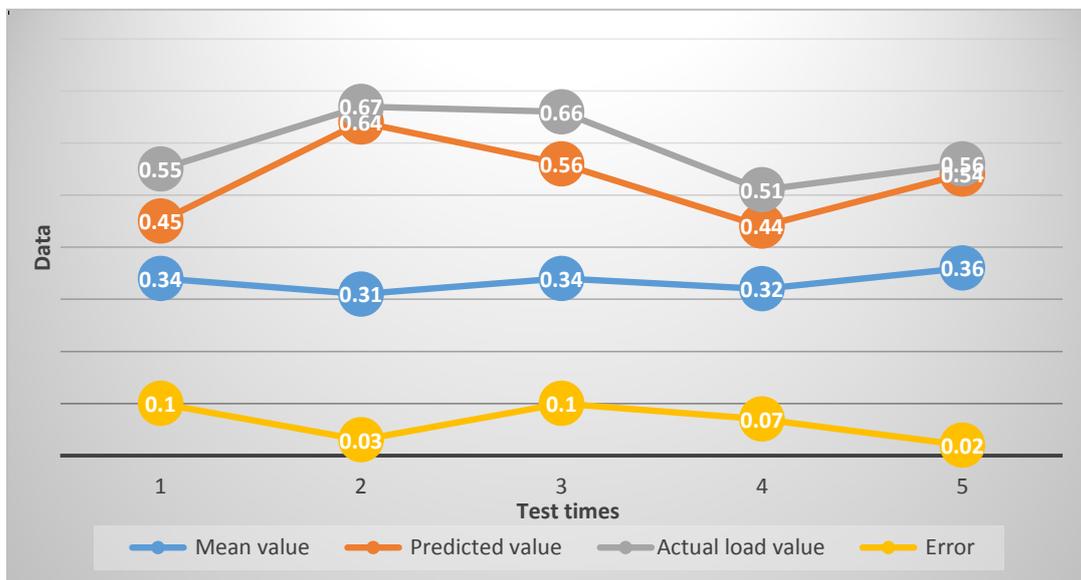| Test times	Mean value	Predicted value	Actual load value	Error
1	0.34	0.45	0.55	0.1
2	0.31	0.64	0.67	0.03
3	0.34	0.56	0.66	0.1
4	0.32	0.44	0.51	0.07
5	0.36	0.54	0.56	0.02

Figure 4. Predicted value

This paper mainly introduces the performance test and analysis of distributed system. Firstly, data collection, including cloud computing, large capacity server, network transmission, etc. Secondly, the virtualization model is established by MATLAB software and the simulation experiment is conducted to verify whether the load prediction results are consistent with the real values. Finally, the load change trend and fluctuation range as well as the prediction accuracy and accuracy are judged according to different types of load curves. It can be seen from Figure 4 that the error of the host load prediction result oriented to the linear time series is small, which indicates that the prediction model has high accuracy.

5. Conclusion

With the rapid development of network, distributed system plays an important role in enterprise management. It plays an irreplaceable role in realizing production process automation, improving

labor efficiency and reducing costs. This paper takes linear time series as the research object to carry out load forecasting. Firstly, the theory of forecasting method and relevant basic knowledge are introduced. Secondly, load modeling is carried out based on the model learning method. Finally, an example is given to verify that the algorithm can effectively deal with the randomness problem, and the corresponding improvement scheme is given to improve the load balance rate of the system, so as to maximize the economic benefits and achieve the expected objectives and requirements.

References

[1] Addison W. Bohannon, Vernon J. Lawhern, Nicholas R. Waytowich, Radu V. Balan:The Autoregressive Linear Mixture Model: A Time-Series Model for an Instantaneous Mixture of Network Processes. IEEE Trans. Signal Process. 68: 4481-4496 (2020).

[2] Epaminondas Markos Valsamis, Henry Husband, Gareth Ka-Wai Chan:Segmented Linear Regression Modelling of Time-Series of Binary Variables in Healthcare. Comput. Math. Methods Medicine 2019: 3478598:1-3478598:7 (2019).

[3] Leonardo Di Gangi, Matteo Lapucci, Fabio Schoen, Alessio Sortino:An efficient optimization approach for best subset selection in linear regression, with application to model selection and fitting in autoregressive time-series. Comput. Optim. Appl. 74(3): 919-948 (2019).

[4]Thach Le Nguyen, Severin Gsponer, Iulia Ilie, Martin O'Reilly, Georgiana Ifrim:Interpretable time series classification using linear models and multi-resolution multi-domain symbolic representations. Data Min. Knowl. Discov. 33(4): 1183-1222 (2019).

[5]Masoumeh Heidari Kapourchali, Bonny Banerjee:Unsupervised Feature Learning From Time-Series Data Using Linear Models. IEEE Internet Things J. 5(5): 3918-3926 (2018).

[6]Luca Faes, Alberto Porta, Michal Javorka, Giandomenico Nollo:Efficient Computation of Multiscale Entropy over Short Biomedical Time Series Based on Linear State-Space Models. Complex. 2017: 1768264:1-1768264:13 (2017).

[7]Abhimanyu Patra, Soumya Das, Sarojananda Mishra, Manas Ranjan Senapati:An adaptive local linear optimized radial basis functional neural network model for financial time series prediction. Neural Comput. Appl. 28(1): 101-110 (2017).

[8] Enielma Cunha da Silva, Ozy D. Melgar-Dominguez, Rubén Romero:Simultaneous Distributed Generation and Electric Vehicles Hosting Capacity Assessment in Electric Distribution Systems. IEEE Access 9: 110927-110939 (2021).

[9]Beibei Li, Rongxing Lu, Wei Wang, Kim-Kwang Raymond Choo:Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system. J. Parallel Distributed Comput. 103: 32-41 (2017).

[10]Naqash Ahmad, Yazeed Ghadi, Muhammad Adnan, Mansoor Ali: Load Forecasting Techniques for Power System: Research Challenges and Survey. IEEE Access 10: 71054-71090 (2022).

[11]Negin Alemazkoor, Mazdak Tootkaboni, Roshanak Nateghi, Arghavan Loughghalam:Smart-Meter Big Data for Load Forecasting: An Alternative Approach to Clustering. IEEE Access 10: 8377-8387 (2022).

[12]Thamer Alquthami, Muhammad Zulfiqar, Muhammad Kamran, Ahmad H. Milyani, Muhammad Babar Rasheed:A Performance Comparison of Machine Learning Algorithms for Load Forecasting in Smart Grid. IEEE Access 10: 48419-48433 (2022).

[13]George Rouwhorst, Edgar Mauricio Salazar Duque, Phuong H. Nguyen, Han Slootweg:Improving Clustering-Based Forecasting of Aggregated Distribution Transformer

Loadings With Gradient Boosting and Feature Selection. IEEE Access 10: 443-455 (2022).

[14]*Khalid Ijaz, Zawar Hussain, Jameel Ahmad, Syed Farooq Ali, Muhammad Adnan, Ikramullah Khosa:A Novel Temporal Feature Selection Based LSTM Model for Electrical Short-Term Load Forecasting. IEEE Access 10: 82596-82613 (2022).*

[15]*Azfar Inteha, Nahid-Al-Masood, Farhan Hussain, Ibrahim Ahmed Khan:A Data Driven Approach for Day Ahead Short Term Load Forecasting. IEEE Access 10: 84227-84243 (2022).*

[16]*Mithun Madhukumar, Albino Sebastian, Xiaodong Liang, Mohsin Jamil, Md. Nasmus Sakib Khan Shabbir:Regression Model-Based Short-Term Load Forecasting for University Campus Load. IEEE Access 10: 8891-8905 (2022).*

[17]*Arash Moradzadeh, Behnam Mohammadi-Ivatloo, Mehdi Abapour, Amjad Anvari-Moghaddam, Sanjiban Sekhar Roy:Heating and Cooling Loads Forecasting for Residential Buildings Based on Hybrid Machine Learning Applications: A Comprehensive Review and Comparative Analysis. IEEE Access 10: 2196-2215 (2022).*

[18]*Alfredo Nespoli, Sonia Leva, Marco Mussetta, Emanuele Giovanni Carlo Ogliari:A Selective Ensemble Approach for Accuracy Improvement and Computational Load Reduction in ANN-Based PV Power Forecasting. IEEE Access 10: 32900-32911 (2022).*