

PAPER • OPEN ACCESS

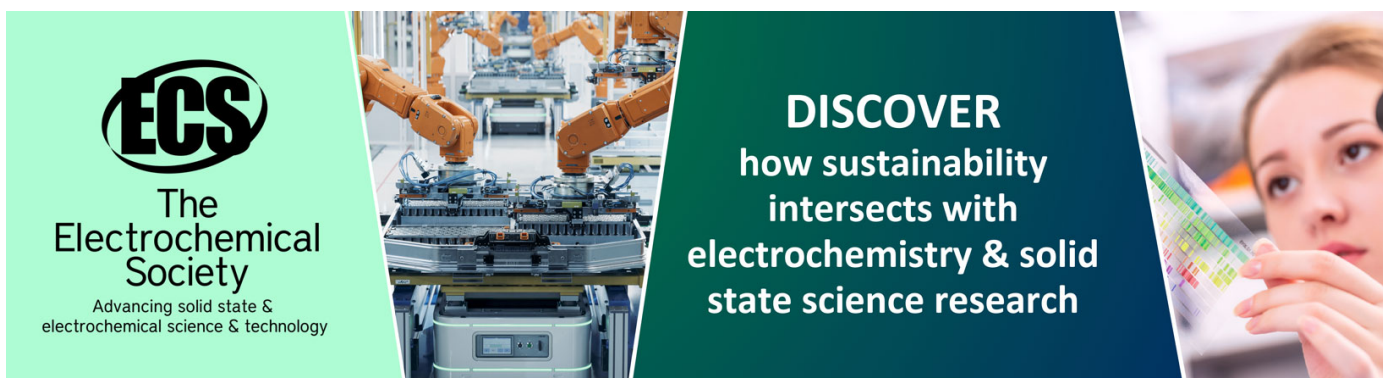
Research on the Application of Power Grid Big Data in Census Work

To cite this article: Siyao Wang *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **693** 012014

View the [article online](#) for updates and enhancements.

You may also like

- [CEWQO Topical Issue](#)
Mirjana Bozic and Margarita Man'ko
- [The Earth radiation balance as driver of the global hydrological cycle](#)
Martin Wild and Beate Liepert
- [HST/WFPC2 and VLT/ISAAC Observations of Proplyds in the Giant H II Region NGC 3603](#)
Wolfgang Brandner, Eva K. Grebel, You-Hua Chu et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Research on the Application of Power Grid Big Data in Census Work

Siyao Wang^{*}, He Hao, Yuan Liu, Yuanyuan Zhao, Yimeng Wang and Ye Du

Beijing Guodiantong Network Technology Co., Ltd., Block A, Dongxu International Center, Fengtai District, Beijing.

Email: 895692408@qq.com

Abstract. Aiming at the problem of "No one at home" in China's census work, a solution based on big data of power grid to the census work is proposed. With the support of big data technology, models for identifying vacant houses and predicting households at home are established by analyzing the electricity consumption and trends of each house. Besides, a data analysis platform based on power grid big data was developed to provide census staff with data support, such as the vacancy status of residential houses and the probability of residents at home. It has been verified that the research program is effective, which can reduce the workload and improve the efficiency of the census work.

1. Introduction

Census work is an important measure to ensure the balanced and full development of China. The country's population, population structure, distribution, etc. will be investigated to achieve the purpose of adjusting and improving population policies. In the entire census process, household census is the most important step. However, during the household census, the following problems will be encountered:

- The real vacancy situation of the house is different from the recorded data in the community.
- Seasonal migrants are absent from their place of residence regularly and for a long time.
- People with special schedules may not go home due to work or other factors.

These problems lead to "No one at home" during the household census, which increases the cost of investigation time and reduces the efficiency of the census. Solving the problem of "no one at home" is an important guarantee for the smooth development of the census.

2. Power Grid Big Data and Technical Advantages

2.1. Power Grid Big Data

In recent years, with the country's strong support for digital information technology, smart grid and information construction are developing rapidly. At the same time, a large amount of power industry data has been accumulated. These data covers all links of the power grid, including power generation, transmission, transformation, distribution, power consumption, and dispatch. They meet the 5V characteristics of big data: volume, velocity, variety, Value, Veracity. They are gradually becoming a new type of asset for power companies.

2.2. Technical Advantages

After the big data technology was introduced into the power grid system, unified management of power data is realized. By integrating and analyzing different types of data resources, reference data



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

can be provided for the formulation of corporate policies and production plans, which is conducive to the stable development of companies. The operation of the system can be monitored remotely by real-time monitoring of power equipment system data. And faults can be identified in a very short time, which can guarantee the stable and reliable operation of power equipment. In the meanwhile, this technology can analyze and warn the risks in the power production process, which can remind the staff to take preventive measures in time to ensure the safety and reliability of power production.

3. The Application of Power Grid Big Data in Census Work

Power grid big data can predict the vacancy and the occupancy status by analyzing the electricity consumption and trends of the house. This kind of forecast data can provide information on residential usage and household living rules before the census household survey. Such applications of power grid big data can reduce the personnel cost and time cost of the census work, which can improve work efficiency and reduce the workload of the census work.

Our research team developed a data analysis platform based on power grid big data for the census. To realize the functions of identifying vacant houses and the prediction of household probability, we called the housing power data from big data center. And use data analysis and mining technology to analyze those big data. The realization of the functions is mainly based on the following models.

3.1. Housing Vacancy Recognition Model

3.1.1. Weekly vacant recognition model. The vacancy of houses is identified on the weekly dimension. For various reasons, the electricity consumption of vacant houses will be non-zero, together with slight numerical fluctuations. For this reason, the concepts of weekly vacancy threshold (mark as $TH1$) and weekly coefficient of variation (mark as c_{TH1}) are introduced. The vacancy threshold is the limit value of the average weekly electricity consumption of a vacant house. When the weekly electricity consumption of the house is less than the weekly vacancy threshold ($TH1$), the house is deemed vacant. The weekly variation coefficient is used to reflect the fluctuation of weekly electricity consumption, and its calculation formula is:

$$c_v = \frac{\sigma}{\mu} = (\bar{x})^{-1} \left(\frac{\sum_{i=1}^7 (x_i - \bar{x})^2}{7} \right)^{1/2} \quad (1)$$

Among them, σ is the standard deviation of electricity consumption. μ is the average electricity consumption. x_i is the daily electricity consumption on the i day of the week. $\bar{x}$ is the average daily electricity consumption of the house during the week. The vacant week variation coefficient threshold is the limit value of the vacant house week variation coefficient. When the house week variation coefficient is less than the vacancy variation coefficient threshold (c_1), the house is deemed vacant.

It can be seen that the selection of the threshold is particularly important. The calculation method is as follows:

Step 1: Data cleaning. And exclude values less than 0 or greater than 2500 in the sample.

Step 2: Calculate the average weekly electricity quantity (mark as A_k) and weekly variation coefficient (mark as c_k) for each house, where k represents the k th house.

Step 3: Take P-quantile for A_k and c_k of all samples. Mark the average weekly electricity quantity and weekly variation coefficient corresponding to the P-quantile as A_p and c_p . Continuously adjust the value of P. In the meanwhile, calculate the proportion of housing with $A_k < A_p$ and $c_k < c_p$. When the proportion is 5%, at this time, A_p is the vacancy threshold, and c_p is the vacancy coefficient of variation threshold, which is $A_p = TH1$, $c_p = c_{TH1}$.

When using the above model to calculate vacant houses, some vacant houses used at very low frequency are excluded. For this reason, we need to conduct a second inspection on the “non-vacant houses” output by the above model. According to actual observations, when the weekly electricity consumption of a house is less than another weekly vacancy threshold (mark as TH_2), and the weekly coefficient of variation is also less than another vacant coefficient of variation threshold (mark as σ_{TH_2}), the house is a vacant house used with very low frequency.

The calculation process of the two threshold parameters of the secondary test is similar to the previous one. Just in step 3, Take Q-quantile for A'_k and take 2Q-quantile for c'_k of all samples. Mark the average weekly electricity quantity and weekly variation coefficient corresponding to the Q-quantile as A_Q and c_{2Q} . Continuously adjust the value of P. In the meanwhile, calculate the proportion of housing with $A'_k < A_Q$ and $c'_k < c_{2Q}$. When the proportion is 1%, at this time, A_Q is the vacancy threshold, and is the vacancy coefficient of variation threshold, which is $A_Q = TH_2$, $c_{2Q} = c_{TH_2}$.

In the actual calculation of the threshold, the daily electricity consumption information we use are from three districts A, B, and C in Heilongjiang Province. A total of 6,141,869 electricity consumption information for each house from January 1, 2018 to June 31, 2020. The calculation results are as follows:

Table 1. The result of calculating the threshold of the vacant housing model.

Quantile (approximate)		Vacancy threshold	Value
TaInitial inspection	25%	Weekly average vacancy threshold 1	0.358571
		Weekly vacant coefficient of variation threshold 1	0.170590
Secondary inspection	30%	Weekly average vacancy threshold 2	0.798571
	60%	Weekly vacant coefficient of variation threshold 2	0.400590

3.1.2. Bi-weekly and monthly vacant housing model. Combining the above housing weekly vacancy model, longer-term housing vacancy rate model can be obtained. For example, if the house meets the weekly vacancy model for two consecutive weeks, you can get bi-weekly vacant model. If the housing meets the weekly vacancy model for four consecutive weeks, you can get the monthly vacant model. After the model is deployed in the power big data census system, the output effect is stable and good. The effect diagrams are shown in Figure 1, Figure 2. After actual verification, the housing vacancy rate series model can effectively cover houses in different vacant scenarios. The output results of the model are basically consistent with the actual verification results, which means they can provide strong support for the development of the census.



Figure 1. Operating output of the vacant model during the housing week.



Figure 2. Operating output of monthly vacancy model.

3.2. Prediction Model of Household Probability

This model is based on household historical electricity consumption data. We analyze the household electricity consumption behavior characteristics and monthly electricity consumption behavior characteristics. And build an ARIMA model using electricity consumption data for a continuous month. Finally, The forecast of household electricity consumption in the next week is realized. And according to this result of prediction, the probability of the household staying at home for each day of the next week is obtained.

3.2.1. Household electricity consumption prediction model. The electricity consumption information we use is from three districts A, B, and C of a certain city in Heilongjiang province. The electricity consumption information of each house in the past month is used as input data. The model is trained with a time series model which is suitable for processing non-stationary series. The output of the model is the predicted value of electricity consumption x'_i for each day in the next week.

3.2.2. Household Probability Prediction Model. For houses that have been vacant for two weeks or more, the probability of staying at home is regarded as zero. Excluding the influence of other external factors, the distribution of electricity consumption of other houses approximately obeys the normal distribution. For housing under other conditions, the distribution of electricity consumption of households approximately obeys a normal distribution when the influence of other external factors is excluded. Therefore, the household status can be obtained based on the difference between the predicted value of household electricity consumption x'_i and the monthly household electricity consumption statistics. The judgment criteria are shown in the following table:

Table 2. Correspondence table of household probability and household electricity consumption forecast.

Household probability	the numerical range of x'_i
100%	$x'_i \geq \mu + 2\sigma$
80%	$\mu + \sigma \geq x'_i > \mu + 2\sigma$
60%	$\mu \geq x'_i > \mu + \sigma$
40%	$\mu - \sigma \geq x'_i > \mu$
20%	$\mu - 2\sigma \geq x'_i > \mu - \sigma$
10%	$\mu - 2\sigma \geq x'_i$

Among them, μ is the average daily electricity consumption of the household in the previous month. σ is the standard deviation of electricity consumption of the household in the previous month.

The model is deployed in the census system based on power grid big data. The output effect is stable and good. The effect diagram is shown in Figure 3. It has been verified that the output of the model is almost the same as the accuracy, which means the algorithm and application can provide data support for the census work.



Figure 3. Operation output of household probability prediction model.

4. Conclusion

This article studies the application of power grid big data in China's census, and proposes a scheme. The prediction of housing vacancy and household occupancy information are realized by analyzing and mining of housing electricity information. A power big data analysis platform is designed for the census work. From the platform, it is found that the model output is stable, and the results are basically the same as the actual detection situation, which means this application can provide data support for the staff's of census work. It can help reduce the personnel and time costs and improve the efficiency of the census work. Power grid big data realizes the service to the government and the people.

5. References

- [1] Bessa R J . Future Trends for Big Data Application in Power Systems 2018 *J. Big Data Application in Power Systems* 223-242.
- [2] Angadi R V , Venkataramu P S , Daram S B . Role of Big Data Analytics in Power System Application 2020 *J. E3S Web of Conferences* 184(3):01017.
- [3] Mocanu, Elena. Big Data Application in Power Systems 2018 *J. Deep Learning for Power System Data Analysis* 125-158.
- [4] Yan S , Yang J , Li J . Design and Its Application of Power Time Synchronization Management System Based on TMU and IEC61850 Protocol 2019 *J. IOP Conference Series: Earth and Environmental Science* 252(3):032201 (10pp).
- [5] Zhang S , Zhang D , Zhang Y , et al. The research and application of the power big data 2017 *C. International Conference on Electronics & Information Engineering. International Society for Optics and Photonics*.