

# Monthly Electricity Consumption Forecast Based on Multi-Target Regression

Haiming Li, Ping Chen

School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China

Email: lhm@shiep.edu.cn, 754786991@qq.com

**How to cite this paper:** Li, H.M. and Chen, P. (2019) Monthly Electricity Consumption Forecast Based on Multi-Target Regression. *Journal of Computer and Communications*, 7, 231-242.

<https://doi.org/10.4236/jcc.2019.77019>

**Received:** September 19, 2018

**Accepted:** July 21, 2019

**Published:** July 24, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Urban grid power forecasting is one of the important tasks of power system operators, which helps to analyze the development trend of the city. As the demand for electricity in various industries is affected by many factors, the data of relevant influencing factors are scarce, resulting in great deviations in the accuracy of prediction results. In order to improve the prediction results, this paper proposes a model based on Multi-Target Tree Regression to predict the monthly electricity consumption of different industrial structures. Due to few data characteristics of actual electricity consumption in Shanghai from 2013 to the first half of 2017. Thus, we collect data on GDP growth, weather conditions, and tourism season distribution in various industries in Shanghai, model and train the electricity consumption data of different industries in different months. The multi-target tree regression model was tested with actual values to verify the reliability of the model and predict the monthly electricity consumption of each industry in the second half of 2017. The experimental results show that the model can accurately predict the monthly electricity consumption of various industries.

## Keywords

Forecasting, Multi-Target Tree Regression, Electricity, Monthly Electricity Consumption, Predict

## 1. Introduction

In the electricity market environment, monthly electricity forecasting of urban power grids helps to better operate and maintain generators. It can also help power system operators adjust the progress of the grid [1]. More importantly, accurate electricity consumption forecasting is an important issue in power grid planning and development. Since electricity has a great impact on society, it is necessary to find a fast, feasible and accurate method for predicting power con-

sumption. Accurately predicting power consumption is influenced by a range of factors, such as population density, economic growth, power facilities, and climate factors, making forecasting a challenging and complex task.

In the actual prediction process of electricity consumption, there are the following challenges: 1) the monthly data on electricity consumption has the non-linear characteristics of the two trends of volatility and trend. It is difficult for a single prediction model to describe this nonlinear change process, which is difficult to achieve the demand for forecasting accuracy [2]; 2) the monthly data on electricity consumption has its own variation law, and it has certain uncertainty due to internal and external multi-factors. It is difficult to introduce relevant variables to reflect these two in the mathematical modeling process. 3) Because the change of the electricity quantity data has a certain randomness [3] [4], the prediction model with high fitting precision to the historical data does not necessarily have high prediction accuracy for future power. In order to solve these thorny problems, many scholars have proposed many prediction techniques. Each method has its own advantages and disadvantages, and there is no prediction method or all predictions of the model.

Most of the previous electricity forecasting work only predicts the total electricity consumption of the whole society or the electricity consumption of a single industry. There is very little work to predict the electricity consumption of each industrial structure at the same time. In order to solve the problems faced in the actual forecasting work, this paper proposes a method to predict the monthly electricity consumption of various industries. This paper fully considers the external factors other than electricity consumption and the correlation between electricity consumption in various industries. Mathematical models are constructed with different dimensions to collect data which contains correlation factors. The data will be combined with actual power consumption data. A method, combining variance and covariance, is proposed. Our proposed multi-target tree regression model can more accurately predict the total monthly electricity consumption of the whole society.

## 2. Related Work

In the past few decades, researchers have done a lot of research on electricity consumption forecast. Commonly used electricity consumption forecast methods can be divided into two categories: one is based on the time series method. The other is based on algorithms in machine learning, including multiple linear regression [5], support vector machines [6], random forests, GBDT, XGBoost, BP algorithms and LSTM algorithms in neural network algorithms [7] [8]. However, monthly power demand forecasting through multivariate time series analysis is a complex task. Most previous studies assume that the input and output variables are fixed and the statistical model is applied [9] [10]. However, internal factors of actual monthly electricity demand and external variables that may affect power demand have been found to be non-stationary. When one or

more external variables are non-stationary, it is necessary to consider complex models that can describe the series of nonlinear inputs and outputs [11] [12].

Multi-objective prediction, also known as multivariate prediction or multi-output regression, refers to the task of predicting multiple continuous variables using a common set of input variables. The final task is to predict multiple real-valued target variables. Although multi-target prediction is a relatively young topic, it has been applied in many fields, such as: predicting the impact of airflow noise on vehicle engines, forest monitoring, and ecological model establishment. There are also many energy-related forecasts, such as, wind and solar production forecasts. The multi-objective regression method also provides an effective modeling method, which not only considers the basic relationship between the feature and the corresponding target, but also considers the relationship between the targets, thus ensuring better representation and interpretability of true values. Another advantage of multi-objective methods is that they can produce simpler models with better computational efficiency. The multi-objective prediction method in this paper is based on the single-target prediction algorithm of AdaBoost regression tree. On this basis, consider the relevance and order between the targets, and then consider the dependence relationship between multi-objective variables, and predict the multi-objective variables by expanding the space of the input variables.

### 3. Method

#### 3.1. Data Set and Preprocessing

The data used in this paper is from the Shanghai Municipal Government Data Service Network, and the monthly electricity consumption data for Shanghai from 2014 to 2017. The economic factor data comes from data released by the Shanghai Municipal Bureau of Statistics, including GDP growth dynamics, urbanization rate and GDP of the three major industrial structures. The weather factor data comes from the Shanghai Meteorological Network, which can check the weather conditions of the day and the next few days, as well as the historical weather conditions. Meteorological conditions mainly include time, weather conditions, temperature, wind speed, humidity and air pressure.

#### 3.2. Data Description

The economic influencing factors include economic aggregate, economic aggregate of each industry, per capita GDP, industrial structure, population growth, consumer price index, energy consumption, meteorological environment, etc. This paper adopts the Pearson correlation coefficient identification method, such as formula (1) [13]:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$

where  $X$  and  $Y$  represent different sequences and  $N$  represents the number of variables. According to the correlation coefficient  $r$ , the matching index related to the electric quantity is obtained, and is extracted as a modelable index.

There are multiple records in the acquired meteorological data for one hour, and there is no case of meteorological records. As shown in **Table 1**, there are three meteorological records at six o'clock because we process meteorological data on an hourly basis. In this case, the weather is treated in the order of snow, rain, fog, and fine weather. That is, when there are multiple weather records in an hour, if one of them is snowing, then the snow is pressed for that hour. If there is a rain, press the rain. If there is a fog, it will be treated with fog. If it is all fine, follow the weather. The temperature, wind speed humidity and wind speed take the average of multiple values. For the hour when there is an empty hour, the weather condition of the last hour is taken as the weather condition of the hour. If there are multiple hourly null values, the most recent weather condition is taken as the weather condition for that hour. In order to obtain the daily weather conditions, the weather with the most occurrences is taken as the weather of the day, and the average value of all temperatures, wind speeds, and humidity of the day is taken as the value of the day.

### 3.3. Multi-Target Prediction

Multi-target prediction is also called multivariate prediction or multi-output regression [14] [15]. The final task is to predict multiple real-valued target variables instead of binary class labels. The multi-target prediction method is based on the research of single-target prediction [16] [17]. Considering the dependence relationship between multi-target variables, predicting multi-target variables by expanding the space of input variables.

### 3.4. Single-Target Prediction.

The establishment of the single-objective prediction model is based on the Ada-Boost iteration method: transforming the weak regression model into a strong regression learning model. First of all, we should consider the problem of weak regression model selection. In single-target prediction, this paper chooses Classification and Regression Tree (CART) to establish a weak regression learning model. The main steps are as follows:

1) Building tree: First, find the best feature to be sliced; if it can no longer be split, save the node as a leaf node and return [18] [19]. Then, according to the best segmentation feature, the data set is divided into left and right subtrees. If

**Table 1.** Daily weather conditions.

Time	Weather condition	Temperature	Wind speed humidity	Wind speed's average
06:20	Light Snow	-2°C	20 > 33 km/h	74%
06:36	Light Snow	-3°C	26 > 35 km/h	80%
06:51	Snow	-2°C	15 > 35 km/h	74%

the feature value is greater than a given value, it belongs to the left subtree, and vice versa. Finally, build the left and right subtrees separately.

2) Selection of the best to be sliced features: First, the eigenvalues of the feature are traversed, and the error of the data set segmentation according to the feature value is calculated. Then, select the feature with the smallest error and its corresponding value as the best segmentation feature and return.

3) Regression tree based prediction: First, it is judged whether the current regression tree is a leaf node, and if so, a prediction is made, and if not, the next step is performed. Then, the feature value on the data set feature is compared with the current regression tree. If the feature value of the data set is large, it is determined whether the left subtree of the current regression tree is a leaf node. If it is not a leaf node, a regression tree based prediction is performed. If it is a leaf node, it is predicted. Accordingly, the right subtree also makes such a judgment.

In regression prediction, AdaBoost's algorithm idea is: For the same sample points, continuously update its weights to train multiple basic classifiers (weak regression models). Then combine the different weight weak regression models to form a final strong classifier (strong regression model). The specific process of the AdaBoost algorithm is divided into the following steps:

1) Initialize the training set, each training sample is given the same weight of  $1/N$  at the beginning.  $D_1 = (w_{11}, w_{12}, \dots, w_{1i}, \dots, w_{1N})$ ,  $w_{1i} = \frac{1}{N}$ ,  $i = 1, 2, \dots, N$

2) Perform multiple iterations, where  $m$  represents the number of the iteration,  $m = 1, 2, \dots, M$  ( $m \leq$  maximum number of iterations).  $M = 100$ , indicating 100 iterations, resulting in 100 weak classifiers.

3) Combine each weak classifier to get the final classification predictor.

#### Algorithm 1. MTS-Tree.

**Input:** Training set  $D = \{(x^1, y^1), \dots, (x^n, y^n)\}$

**Output:** Regression chain model  $h_j = X \rightarrow R$

1. generate  $D_1$
2.  $D_1^* = \{(x^1, y_1^1), \dots, (x^n, y_1^n)\}$
3. For  $j = 1$  to  $m$
4.  $h_j = D_j^* \rightarrow R$
5. If  $j < m$
6. Generating data sets  $D_{j+1}^*$
7.  $h_j^i = D_j^* \setminus D_j^i \rightarrow R$
8. for  $x_j^i \in D_j^i$
9.  $x_{j+1}^i = x_j^i$
10.  $y_j^i = h_j^i(x_j^i)$
11.  $x_{j+1}^i = [x_j^i, y_j^i]$
12.  $D_{j+1}^* = D_{j+1}^* \cup (x_{j+1}^i, y_{j+1}^i)$

### 3.5. Multi-Target Prediction Algorithm Based on Target Stacking (MTS)

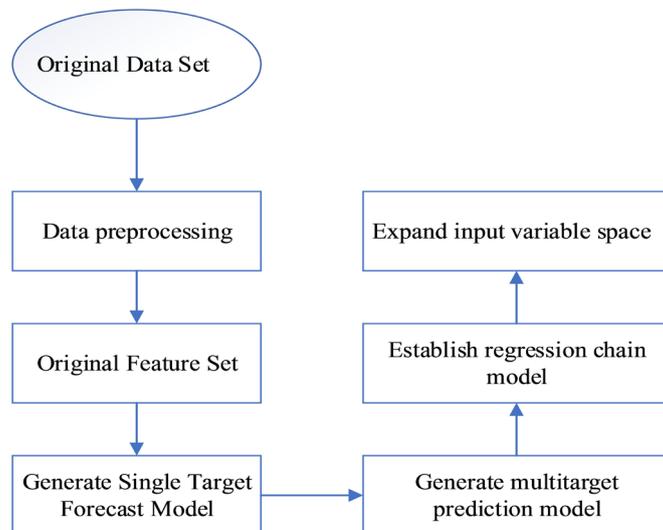
As shown in Algorithm 1, definition  $X$  and  $Y$  are 2 random vectors, where  $X$  contains  $d$  input variables  $X_1, \dots, X_d$ ,  $Y$  contains  $m$  output variables  $Y_1, \dots, Y_m$ .  $\alpha = R^d$  and  $\beta = R^m$  respectively represent the domain space of  $X$  and  $Y$ , and are often used as input space and output space. In a sample point  $(x, y)$ ,  $x = [x_1, \dots, x_d]$  is the input vector,  $y = [y_1, \dots, y_m]$  is the output vector. Given a data set  $D = \{(x^1, y^1), \dots, (x^n, y^n)\}$ , with  $n$  training samples, multi-target regression prediction aims to learn a model.  $h: \alpha \rightarrow \beta$ , this model can make an input vector  $x^q$ , can predict its output vector  $\bar{y}^q = h(x^q)$ , this predicted output value is as similar as possible to the actual true output value  $y^q$ . Algorithm flow chart is **Figure 1**.

The specific steps can be divided into the following:

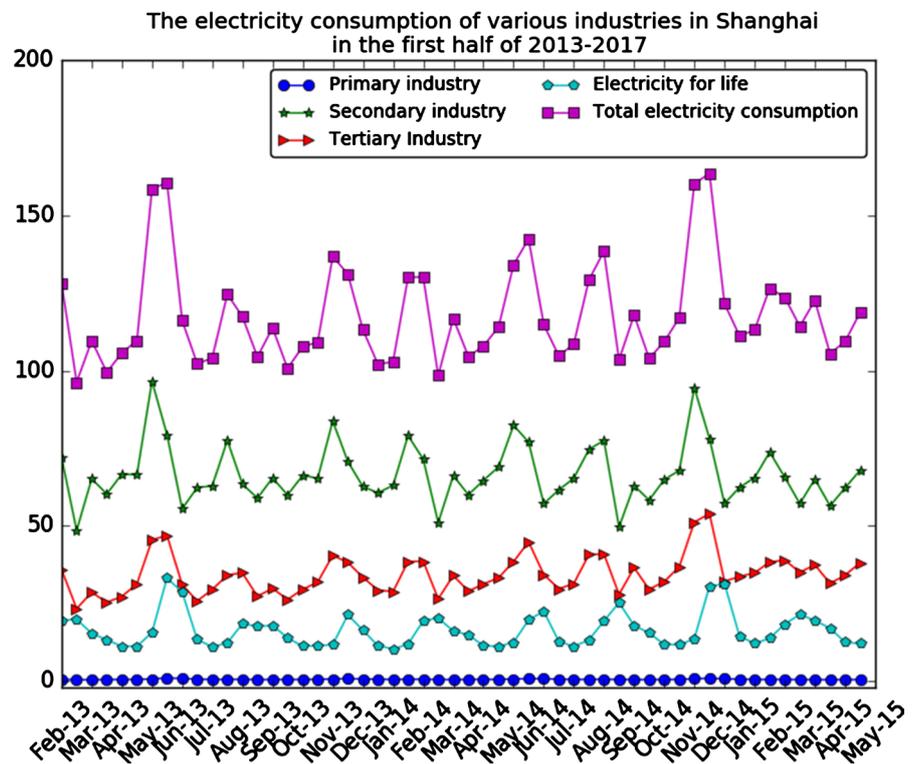
- 1) Single target prediction model generation: let  $h(x) = Y_M(X)$ , each target variable is calculated based on the algorithm of the regression tree AdaBoost iteration, on the regression tree model, each target variable is iterated  $M$  times by AdaBoost,  $m$  target variables, which is equivalent to the final iteration of  $M \times m$  times, and finally forms  $m$  independent single-target prediction models  $h_j = X \rightarrow R$ .
- 2) Generation of multi-target prediction models: added additional training phase to the regression model for single-target prediction, for each target  $Y_j$ , learn  $m$  models  $h_j^*: X \times R^{m-1} \rightarrow R$ . In the learning of each model  $h_j^*$ , the training set is constantly changing  $D_j^* = \{(x^{*1}, y_i^1), \dots, (x^{*n}, y_i^n)\}$ .

## 4. Experiments

Forecasting the monthly electricity consumption of various industries is the focus of grid companies. As shown in **Figure 2**, this experiment uses the data set training model of Shanghai city in the first half of 2013-2017 to predict the electricity consumption of various industries in the second half of 2017, and make



**Figure 1.** Algorithm flow chart.



**Figure 2.** Monthly trend of three major industries' electricity consumption, domestic electricity consumption and total electricity consumption.

predictions and compare them with real values.

#### 4.1. Datasets

Electricity consumption data for various industries in the first half of 2013-2017: We collected data on electricity consumption in various industries in recent years from the Shanghai Municipal Government Data Service Network, including the primary industry, the secondary industry, the tertiary industry, household electricity, and total electricity consumption. The basic unit is billion kWh, and the chart shows the monthly electricity consumption and total monthly electricity consumption for each industry.

GDP growth of various industries in the first half of 2013-2017: In order to accurately predict the monthly electricity consumption of various industries, we consider the monthly GDP growth of various industries. We collected the monthly GDP output and growth rate of various industries in recent years in the official website of the Shanghai Municipal Bureau of Statistics.

Monthly weather conditions for the first half of 2013-2017: We have not only considered the factors of GDP growth, but also considered the weather factor. We collected monthly specific weather conditions from the Shanghai Meteorological Network for the past few years, including the number of days of cloudy, sunny, cloudy, rainy, and snowy days (in [Figure 2](#)). In addition, the monthly average maximum temperature and average minimum temperature are collected in [Figure 3](#). [Figure 4](#) shows the monthly weather conditions in recent years.

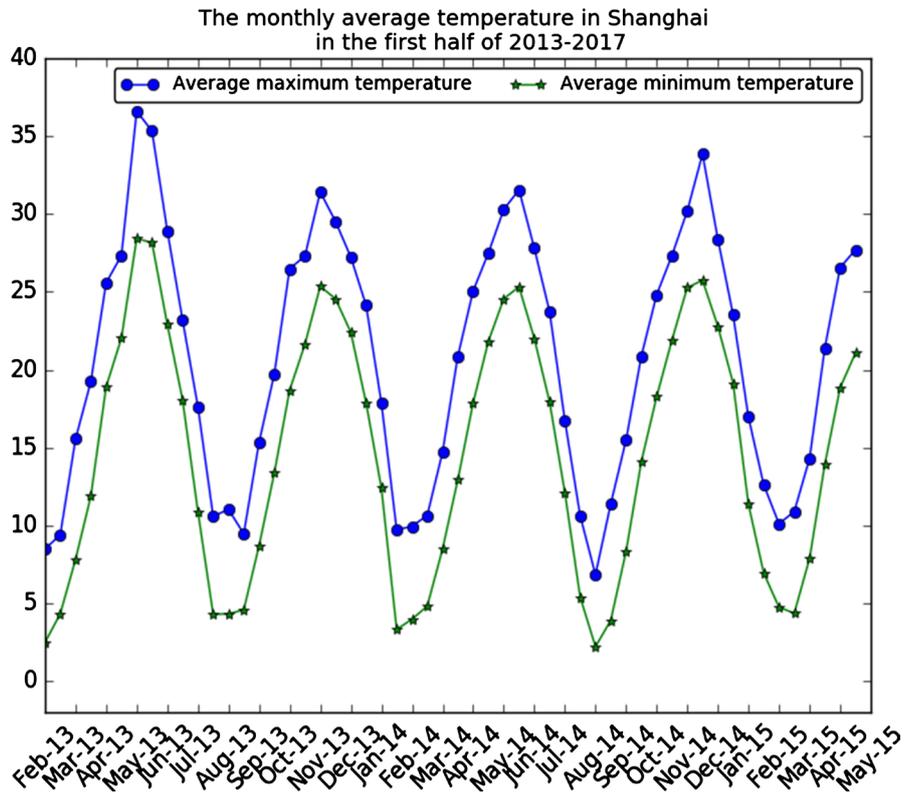


Figure 3. Monthly average minimum and maximum temperature trends.

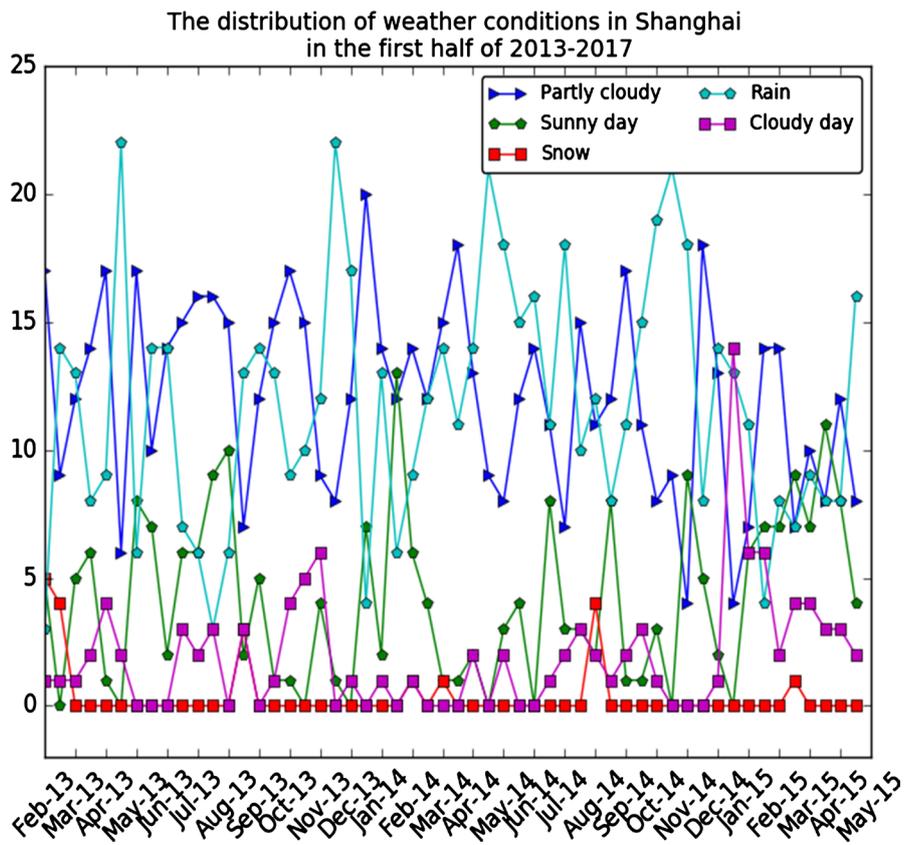


Figure 4. Monthly statistics for cloudy, sunny, rainy, cloudy, and snowy days.

Tourism in the first half of 2013-2017 and other factors: In addition, we collected statistics on the number of monthly tourists and the frequency of travel seasons from the official website of the Shanghai Tourism Bureau. We also divide the December, quarterly, holiday frequency, comfort, etc. as a reference factor.

## 4.2. Results and Analysis

This study considers the rationality of the data and the proportion of the invalid values, and the fitting interval to be used for the monthly data. Using the data from January 2013 to June 2017 to train the MTS model, predict the electricity consumption of each industry in July 2017, and calculate the error rate of the total and actual value of the predicted electricity consumption of each industry in July. The 2017 July training set sample was constructed using the July 2017 forecast and the characteristic attributes of the relevant model to update the MTS model and predict the electricity consumption of each industry in August 2017. Calculate the error rate of the total of the predicted electricity consumption values of each industry in August and the true value. Cycle through the previous steps until you forecast the electricity consumption of each industry in December 2017. Calculate the error rate of the total of the December electricity consumption forecast values for each industry and the total value. The electricity consumption of various industries is affected by many factors such as business conditions, industry environment, macro policies, and seasons. The main features of the current construction model are electricity history data and other data, which cannot accurately predict the long-term behavior of electricity in various industries.

The specific prediction results are shown in the table. In view of the short time interval of monthly data and the most obvious seasonal fluctuations, the multi-target regression model was used to train the electricity consumption data and give the fitting error and prediction. **Tables 2-6** show the prediction results of three industries, household electricity consumption and total electricity consumption. The multi-target regression model predicts results for four different targets. The model has a significant effect on the data fitting with more obvious seasonal fluctuations. Still converging to more accurate accuracy due to insufficient training samples or some poor quality data nodes.

**Table 2.** First industry forecast results.

Forecast year and month	Actual value/GWh	Forecast value/GWh	Relative error rate
July 2017	0.70	0.66	5.71%
August 2017	0.95	1.03	8.42%
September 2017	0.80	0.67	16.25%
October 2017	0.64	0.80	25%
November 2017	0.49	0.50	2.04%
December 2017	0.49	0.55	12.24%

**Table 3.** Secondary industry forecast results.

Forecast year and month	Actual value/GWh	Forecast value/GWh	Relative error rate
July 2017	103.87	110.88	6.75%
August 2017	75.55	81.32	7.64%
September 2017	57.20	55.36	3.22%
October 2017	59.72	70.06	17.31%
November 2017	64.56	61.09	5.37%
December 2017	77.90	80.33	3.12%

**Table 4.** Tertiary industry forecast results.

Forecast year and month	Actual value/GWh	Forecast value/GWh	Relative error rate
July 2017	60.73	55.96	7.85%
August 2017	54.13	60.09	11.01%
September 2017	34.24	40.12	17.17%
October 2017	34.74	40.65	17.01%
November 2017	35.88	39.83	11.01%
December 2017	42.75	44.31	3.65%

**Table 5.** Life electricity forecast results.

Forecast year and month	Actual value/GWh	Forecast value/GWh	Relative error rate
July 2017	13.24	14.36	8.46%
August 2017	39.75	40.67	2.31%
September 2017	33.45	32.86	1.76%
October 2017	14.86	14.09	5.18%
November 2017	12.34	13.73	11.26%
December 2017	13.99	15.07	7.72%

**Table 6.** Total electricity forecasting results.

Forecast year and month	Actual value/GWh	Forecast value/GWh	Relative error rate
July 2017	178.54	181.86	1.86%
August 2017	170.38	183.11	7.47%
September 2017	125.69	129.01	2.64%
October 2017	109.97	125.60	14.21%
November 2017	113.26	115.15	1.67%
December 2017	135.13	140.26	3.80%

## 5. Conclusion

In the research process, this paper combines Shanghai's social economy, weather conditions, distribution density of tourists, seasonal division and actual use of electricity in the whole society. In-depth analysis of the relationship between the

city's electricity development and external factors, scientifically establishes a multi-target tree regression model of electricity forecasting model. The results show that the model can feedback the industry's future power growth trend and the accuracy of power forecasting to a certain extent. It is expected that the proposed model can help to more accurately predict short-term electricity demand in various industries. The ability to accurately predict short-term electricity demand can help power system operators ensure sustainable power planning decisions and ensure supply of electricity to consumers.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Li, Y. and Lei, J. (2011) The Application and Research of Electric Power Load Forecasting Technology Based on the Time Series Model. *Science Technology and Engineering*, **11**, 860-864.
- [2] Huang, D. and Fang, P. (2017) Power Consumption Forecasting Application Based on XGBoost Algorithm. *Modern Information Technology*, **1**, 10-12.
- [3] Zhang, J. (2017) Multi-Target Prediction Algorithm Based on AdaBoost Regression Tree. *Computer and Modernization*, No. 9, 89-95.
- [4] Jiang, Z. (2010) Research on Power Load Forecasting Base on Support Vector Machines. *Computer Simulation*, **27**, 282-285.
- [5] Wang, J.Q., Wang, F.L., Dong, Z.G., *et al.* (2017) Electrical Load Forecasting Based on Improved BP Neural Network. *Mathematics in Practice & Theory*, **47**, 276-284.
- [6] Ma, L. and Li, Y. (2015) A Characteristic Extraction and Forecast Method for Interval Power Loads. *Control Engineering of China*, **22**, 645-648.
- [7] Vilar, J., Aneiros, G. and Raña, P. (2018) Prediction Intervals for Electricity Demand and Price Using Functional Data. *International Journal of Electrical Power & Energy Systems*, **96**, 457-472. <https://doi.org/10.1016/j.ijepes.2017.10.010>
- [8] Borchani, H., Varando, G. and Bielza, C. (2015) A Survey on Multi-Output Regression. *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery*, **5**, 216-233. <https://doi.org/10.1002/widm.1157>
- [9] Ding, S., Hipel, K.W. and Dang, Y. (2018) Forecasting China's Electricity Consumption Using a New Grey Prediction Model. *Energy*, **149**, 314-328. <https://doi.org/10.1016/j.energy.2018.01.169>
- [10] Melki, G., Cano, A., Kecman, V., *et al.* (2017) Multi-Target Support Vector Regression via Correlation Regressor Chains. *Information Sciences*, **415-416**, 53-69. <https://doi.org/10.1016/j.ins.2017.06.017>
- [11] Spyromitrosxioufis, E., Groves, W., Tsoumakas, G., *et al.* (2012) Multi-Label Classification Methods for Multi-Target Regression. *Computer Science*, **104**, 55-98. <https://doi.org/10.1007/s10994-016-5546-z>
- [12] Li, Y., Zheng, Y., Zhang, H., *et al.* (2015) Traffic Prediction in a Bike-Sharing System. In: *SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, New York, 1-10. <https://doi.org/10.1145/2820783.2820837>
- [13] Xu, N., Dang, Y.G. and Gong, Y.G. (2017) Novel Grey Prediction Model with Non-

- linear Optimized Time Response Method for Forecasting of Electricity Consumption in China. *Energy*, **118**, 473-480. <https://doi.org/10.1016/j.energy.2016.10.003>
- [14] Al-Hamadi, H.M. and Soliman, S.A. (2004) Short-Term Electric Load Forecasting Based on Kalman Filtering Algorithm with Moving Window Weather and Load Model. *Electric Power Systems Research*, **68**, 47-59. [https://doi.org/10.1016/S0378-7796\(03\)00150-0](https://doi.org/10.1016/S0378-7796(03)00150-0)
- [15] Groves, W. and Gini, M. (2011) A Regression Model for Predicting Optimal Purchase Timing for Airline Tickets. Technical Report, University of Minnesota, Minneapolis.
- [16] Friedman, J.H. (1999) Stochastic Gradient Boosting. Technical Report, Stanford University, Stanford.
- [17] Madjarov, G., Kocev, D., Gjorgjevikj, D. and Dzeroski, S. (2012) An Extensive Experimental Comparison of Methods for Multi-Label Learning. *Pattern Recognition*, **45**, 3084-3104. <https://doi.org/10.1016/j.patcog.2012.03.004>
- [18] Zhang, M.-L. and Zhou, Z.-H. (2006) Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization. *IEEE Transactions on Knowledge and Data Engineering*, **18**, 1338-1351. <https://doi.org/10.1109/TKDE.2006.162>
- [19] Aho, T., Zenko, B. and Dzeroski, S. (2009) Rule Ensembles for Multi-Target Regression. *9th IEEE International Conference on Data Mining*, Miami, 6-9 December 2009, 21-30. <https://doi.org/10.1109/ICDM.2009.16>