Sensitivity Analysis and Forecast of Power Load Characteristics Based on Meteorological Feature Information

Hu Li\textsuperscript{1,}\textsuperscript{*}, Jian Tan\textsuperscript{1}, Jun Han\textsuperscript{1}, Yi Ge\textsuperscript{1} and Dehua Guo\textsuperscript{2}

\textsuperscript{1}State Grid Jiangsu Electric Power Co., Ltd. Economic and Technological Research Institute, Jiangsu, China
\textsuperscript{2}Beijing Qingruan Innovation Technology Co., Ltd, Beijing, China

\textsuperscript{*}Corresponding author e-mail: 2453305270@qq.com

Abstract. The rapid development of big data and artificial intelligence technology provides a good way to analyze the factors of power demand changes. In this paper, the meteorological factors that affect power demand are studied in depth, and a meteorological factor index system for power demand change is established. Based on the identification method of dominant meteorological factors, the coupling relationship between meteorological factors and power loads is quantitatively evaluated; The sensitivity of load changes in commercial, residential and industrial industries under typical scenarios is analyzed, and the relationship between dominant meteorological factors and quantification affecting the load changes in summer and winter is studied. Finally, the validity of this model is verified by the sensitivity analysis and prediction of power load characteristics to meteorological information based on the data of Nanjing power network.

1. Introduction
Power load is one of the important indicators of power system planning, design and operation management. Studying the characteristics of load and its changing rules is the primary condition to achieve power network security, stability, high-quality and economic operation. Because the factors that affect the power load are various, especially a considerable number of factors cannot be given accurately and quantitatively, the changing characteristics of power load are manifested as time variability, Randomness, complexity, and diversity. The load changes of power grids in different regions are also different. In recent years, power shortage has become a concern of the whole society. With the continuous improvement of people's living standards and the gradual adjustment of industrial structure, the proportion of residential and tertiary industry electricity consumption has been increasing. Both types of electricity consumption are inseparable from meteorological conditions, and their proportion in total power demand has further increased, which makes the relationship between power load changes and meteorological conditions closer [1-3]. However, due to the different climatic conditions of different power grids, the economic structure and development level of different regions are also varied, and the relationship between the load of each power grids and meteorological factors is also different. Therefore, it is necessary to further study the relationship between power load and meteorological factors to provide a basis for power load predicting, forecasting and operation scheduling [4-5].
At present, domestic scholars have partially carried out research on the relationship between regional or urban power load and meteorological factors. Through the study of the relationship between Central China Power Grid and temperature, Chen Zhenghong et al. found that the daily electricity consumption in summer has a significant positive correlation with the average daily temperature, but the correlation is not significant in winter [6]. Unlike Central China, Zhang Lixiang et al. analyzed the electricity consumption and meteorological factors in Shenyang from 1988 to 1998. The results showed that the seasons significantly affected by meteorological conditions are transitional winter and summer, while the autumn and peak winter seasons are less affected [7]. In addition, Cai Xinling et al. [8-10], Duan Hailai et al. [11], Chen Li et al. [12], Wu Xiangyang et al. [13] have done similar research work on power load and meteorological factors in other cities and regions respectively. It can be seen that the relationship between power load and meteorological factors also has different response relationships in different regions. Among many meteorological factors, the response of power load to temperature is the most direct [14-16], but the influence of other meteorological factors besides air temperature on power load cannot be ignored. Therefore, there are also some studies to analyze the relative humidity, precipitation, wind speed and other factors to explore the relationship between changes in power load and meteorological factors [18, 19].

In this paper, the random forest algorithm is used to summarize the meteorological factors that will affect the power demand, establish the index system of meteorological factors, study the identification method of dominant meteorological factors, and model the quantitative evaluation of the relationship between meteorology and power system; Collect key customer information through the negative control system, extract the load characteristics of different industries, residents, and major industries in the typical city of Nanjing, and study the leading meteorological factors and quantitative relationships that influence the load changes in summer and winter.

2. Model Establishment
Based on the random forest algorithm, analyze the influence of weather factors such as maximum temperature, minimum temperature, average temperature, wind speed, air pressure on load change. Then the dataset is normalized and divided into training set, test set and validation set, and the support vector machine model is used for prediction.

![Figure 1. Sensitivity analysis and prediction model of power load characteristics for meteorological feature information.](image)
2.1. Characteristic Contribution Analysis of Random Forests

RF generates a new subset of samples by random sampling from the initial sample set B in a Bootstrap Sampling method, and then generates a forest collection of K decision trees from each subset. It is essentially a combinatorial decision tree algorithm, which combines multiple decision trees to obtain a more general learning ability model. Generally, each tree in a forest has the same distribution, and the fitting error depends on the correlation between the learning abilities of each tree. During the sampling process, the remaining unselected samples are out-of-pocket samples and are defined as sets. Where C and $\overline{C}$ are subsets of B and $\overline{B}$ respectively. Assuming $\chi^{n-p}$ is an n-dimensional dataset with P features and y is an n-dimensional label vector, the RF algorithm calculates the importance of features by rearranging the fitting errors before and after the features. When T trees are created, T out-of-pocket sample sets are used as test sets. The characteristic importance index S is calculated as follows:

$$S(x_j) = \frac{1}{T} \sum_{i \in B_k} \frac{1}{|B_k|} \left( \sum_{i \in B_k} I(h_{k,i}^!(i) \neq y_i) - I(h_{k,i}^!(i) \neq y_i) \right)$$  \hspace{1cm} (1)$$

Where $y_i$ is the fitting attribute of the i-th out-of-pocket data, $I$ is the error representation function, $h_{k,i}^!(i)$ is the fitting attribute of the samples predicted by the dataset $B_k$, and $h_{k,i}^!(i)$ is the fitting attribute after replacing the feature $x_j$.

2.2. Support Vector Machine (SVM) Model

Support vector machine is a classification model proposed by Vapnik et al., which has been applied in the recognition of voltage sag disturbance [12]. The basic idea of using support vector machine to classify samples is to map the input features of linear non separable disturbance signals to the feature space with higher dimension, and establish an optimal hyper plane in the high dimension feature space, so that the distance between the hyper plane and the edge of the samples to be divided is maximized, and when the training set is small, the classification accuracy can also be guaranteed to the maximum extent.

The support vector machine is derived from the binary classification problem. For the binary classification sample set $(x_i, y_i), x_i \in \mathbb{R}^d, y \in \{-1,1\}$, where $x_i$ is the sample vector to be divided and $y_i$ is the sample label. Establish the hyper plane $w \cdot x + b = 0$, $w$ is the normal vector of hyper plane, $b$ is the real number, $x$ is the d-dimensional input sample size. The training process of SVM is a process of seeking the optimal hyper plane through iterative calculation. In other words, obtain the values of $w$ and $b$ when the classification edge is maximized. The training sample points on the closest hyper plane parallel to the classification plane are defined as support vectors.

The problem of solving the optimal hyper plane can be expressed as follows:

$$\min \phi(w) = \frac{1}{2} ||w||^2$$  \hspace{1cm} (2)$

s.t. $y_i[(w \cdot x_i) + b] - 1 \geq 0, i = 1,2,..., n$  \hspace{1cm} (3)$

Define the Lagrange function:

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^{n} \alpha_i \{y_i[(w \cdot x_i) + b] - 1\}$$  \hspace{1cm} (4)$

Where $\alpha_i$ is the Lagrange Coefficient ($x_i, \alpha_i > 0$), using the principle of duality, the optimization problem is solved under the constraint conditions, and then the optimal classification function is obtained as follows:

$$f(x) = \text{sign}\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\}$$  \hspace{1cm} (5)$$
Where \( b^* \) represents the threshold value of sample classification, and its value can be obtained by substituting any support vector into formula (6). In the case of linear indivisibility, the penalty factor \( C \) and relaxation variable \( \xi_i \) are introduced. Thus, the problem of solving the generalized optimal classification surface can be transformed into:

\[
\phi(w, \xi) = \frac{1}{2} ||w||^2 + C(\sum_{i=1}^{n} \xi_i) \tag{6}
\]

Support vector machine maps the indivisible sample vectors in the low-dimensional space to the high-dimensional feature space through the nonlinear mapping operation of kernel function, and converts the point product operation \( (x_i \cdot x) \) in the optimal classification surface into kernel function \( K(x_i, x) \). Then, the discriminant function is obtained:

\[
f(x) = \text{sign}\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x) + b^*\} \tag{7}
\]

Kernel function is essential for the construction of support vector machine. Common kernel functions include linear kernel function, polynomial kernel function, radial basis kernel function and Sigmoid kernel function. In dealing with nonlinear problems, the radial basis function has excellent performance, so this kernel function is used in this paper, and its expression is \( K(x_i, x) = \exp (-\gamma ||x_i - x||^2) \).

The parameter setting of SVM is the key to determine its performance. Considering the limitations in practical application, GWO is used to optimize the parameters of SVM.

3. Example Analysis

The calculation example uses three types of loads in Nanjing: industrial load, commercial load and residential load. In this paper, root mean square error (RMSE) and mean absolute percentage error (MAPE) are selected as the evaluation indexes of the model prediction effect:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{8}
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{9}
\]

Where \( y_i \) represents the real power load value at time \( i \); \( \hat{y}_i \) represents the predicted load value under the model adopted at time \( i \). The smaller the values of RMSE and MAPE are, the more accurate the prediction value is, which proves that the model used is more effective.

3.1. Analysis of the Influence of Weather on the Power Load in Each Season

Random forest algorithm is used to calculate the correlation degree between residential, commercial and industrial loads and the weather information in winter and summer. It can be seen from Figure 2 and Figure 3 that the load change has the greatest correlation with the temperature information, and there is also a certain correlation with relative humidity. The air pressure and wind speed have little impact on various loads. In addition, a more in-depth analysis of the weather impact characteristics shows that the residential load is more sensitive to weather changes, and the maximum temperature in summer and the minimum temperature in winter have a greater impact on the residential electricity consumption. Due to the strong regularity of commercial and industrial load, compared with the residential load, it is relatively less affected by the weather than residential loads. Industrial load is
more susceptible to temperature, and its production schedule has a certain correlation with the average temperature.

**Figure 2.** Degree of weather influence on various types of loads in summer.

**Figure 3.** Degree of weather influence on various types of loads in winter.
Figure 4. Corresponding relationship between residential load and temperature.

Figure 4 further analyzes the relationship between residential load and temperature. There is a linear relationship between the temperature and the power load. Generally, the temperature in this area is relatively comfortable, and there are some periods of lower temperature. When the temperature is low, the use of air conditioning or other heating equipment increases the power load demand. The analysis of the relationship between them proves the accuracy of the correlation analysis of meteorological characteristics in the random forest algorithm.

3.2. Forecast Effect Analysis

Based on the analysis of the weather characteristics in Section 3.1, the load is predicted separately. It can be seen from Figure 5 that the daily load within one month is selected for the test data range analysis. It can also be seen that the prediction effect of using RF-SVM model in this paper is better than the neural network model. This is because after analyzing the correlation of meteorological features through RF, it is possible to select features that have a greater impact on the prediction of future trends, which makes the model have good generalization ability and prediction effect.

Figure 5. Analysis of load forecasting results.

4. Conclusion

In this paper, the random forest algorithm is used to collect the meteorological factors that will affect the power demand, establish the meteorological factor index system, study the dominant
meteorological factor identification method, and model the quantitative evaluation of the relationship between meteorology and power system; The load characteristics of different industries, residents and major industries are extracted by collecting the information of key customers through the negative control system, and the leading meteorological factors and quantitative relationship that affect the load changes in summer and winter are studied. Finally, an example is used to verify that the prediction model after meteorological feature selection has good prediction ability and generalization performance.

Acknowledgments
This work was financially supported by Theoretical Research on Comprehensive Analysis of Power Economy Based on Big Data Theory, No: J2015039.

References
[1] Zhenghong Chen, Bin Hong. Evaluation of the relationship between daily electricity consumption and air temperature in four provinces of Central China Power Grid [J]. Acta Geographica Sinica, 2000, 55 (Supplement): 34-38.
[2] Haidong Zhang, Zhaobo Sun, Yan Zheng, et al. The influence of temperature change on urban power load in Nanjing [J]. Journal of Atmospheric Sciences, 2009 (04). Reference to a chapter in an edited book:
[3] Xiaoling Zhang, Yingchun Wang. Forecast and Analysis of the Relationship between Electricity Depletion and Meteorological Conditions during Summer in Beijing City [J]. Meteorological Monthly, 2002 (02): 17-21.
[4] Douglas M, Le Cornte, Henry E W. Modeling the impact of summer temperatures on national electricity consumption [J]. Appl Meteorol, 1981, 20 (2): 1415-1419.
[5] Jiacheng Zhang, Suhua Gao, Yaru Pan. Investigation on the Relationship between Temperature Change and Winter Heating in China [J]. Quarterly Journal of Applied Meteorology, 1992, 3 (1): 70-75.
[6] Jianglin Hu, Zhenghong Chen, Bin Hong, et al. A Relationship between Daily Electric Loads and Meteorological Elements in Central China [J]. Meteorological Monthly, 2002 (03).
[7] Lixiang Zhang, Liqiang Chen, Minghua Wang, et al. The Relationship of Electricity Supply and Weather Condition in a City [J]. Meteorological Monthly, 2000, 26 (7): 27-31.
[8] Xinling Cai, Hong Xu, Qiuwen Qiao. A relationship between meteorological electric quantity and meteorological elements in Shaanxi Province [J]. Journal of Northwest Hydroelectric power, 2003, 19 (4): 40-43.
[9] Xinling Cai, Hao He, Hongyan Gao. Relationship between summer power load and meteorological conditions and forecast, 2004.
[10] Xinling Cai, Hao He, Jianke Li, et al. Meteorological forecast model of daily electricity consumption and maximum load in Shaanxi Province [J]. Shaanxi Meteorology, 2004 (3): 40-42.
[11] Hailai Duan, Huaisui Qian. Responses of the Electric Power Consumption to Climate Change in Guangzhou City [J]. Journal of Applied Meteorological Science, 2009, 20 (1): 80-87.
[12] Li Chen, Shuai Li, Xiuqi Fang, et al. Influence Factor Analysis of Urban Residential Cooling Energy Consumption by Air Conditioners in Beijing During 1996-2007 [J]. Advances in Climate Change Research, 2009, 5 (4): 231-236.
[13] Xiangyang Wu, Haidong Zhang. Econometric Analysis on Beijing Temperature Influence upon Electricity Load [J]. Journal of Applied Meteorological Science, 2008, 18 (5): 531-538.
[14] Haidong Zhang, Zhaobo Sun, Yan Zheng, et al. Impact of Temperature Change on Urban Electric Power Load in Nanjing [J]. Transactions of Atmospheric Sciences, 2009, 32 (4): 536-542.
[15] Yi Quan. Study on the relationship between power grid load and temperature [D]. Huazhong University of science and technology, 2006.
[16] Lihua Zhong, Shaoyi Zhou, Yong Li, et al. Guangxi Electrical Network Power Load Change Characteristic and Temperature Relations and Its Forecast [J]. Journal of Meteorological Research and Application, 2007 (1): 56-59.

[17] Xiaoyun Zhang, Yuekun Liu, Jian Xiao, et al. Relationship between electrical energy supply and air temperature from June to September in Tianjin, China [J]. Journal of Meteorology and Environment, 2009, 25 (3): 62-65.

[18] Mingming Xiong, Mingcai Li, Yu Ren, et al. Characteristics of Electrical Load and Its Relationship Meteorological Factors in Tianjin [J]. Meteorological Science and Technology, 2013, 41 (3): 577-582.

[19] Lihua Zhong, Yong Li, Dianxiu Ye, et al. The Influence of Comprehensive Meteorological Factors on Guangxi Electric Power Load [C]. Urban meteorology and urban sustainable development branch of annual meeting of China Meteorological Society, 2008: 31-37.