

# Selection and Evaluation of influencing Parameters for Heat Load Forecasting Model

Yuli Wu, Rui Li\*

School of Environment and Energy Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

**Abstract.** This paper analyses the factors affecting the heating consumption of a heating substation. The input parameters of neural network prediction model are analysed and selected. The average absolute error, average absolute percentage error, and mean square error are used to evaluate the effect of the prediction model. The results show that when the model input parameters are the maximum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load in the previous three days, the heating load in the previous two days, the heating load in the previous day and when the model input parameters are the maximum outdoor temperature, the minimum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load of the previous three days, the heating load of the previous two days, the heating load of the previous day, the effects are better.

## 1 Introduction

Building energy consumption accounts for a large proportion of various energy consumption in China, building energy saving is of great significance to energy saving. Among the total energy consumption of buildings in China, urban heating energy consumption accounts for the largest proportion. Statistics in 2017 show that the national heating energy consumption for the whole year is about 223 million tons of standard coal, accounting for 21% of the total energy consumption of the whole society [1]. An important reason is that heating in the north consumes a lot of energy in winter. If heating on demand can be achieved, energy can be effectively saved. The premise of realizing on-demand heating is to accurately predict the heat load. The accurate prediction of the heat load can provide powerful data support for the operation and management of the heating system, thereby reasonably guiding the operation and management.

Since there are many influencing factors of heat load, including meteorological factors, system factors, architectural factors. Werner used statistical analysis methods to analyze the influence of outdoor temperature, solar radiation, wind speed and other factors on the heating load. Finally, he concluded that the outdoor temperature has the greatest influence on the heat load of the system, followed by solar radiation. The impact of wind speed is minimal.[2] System factors mainly refer to the thermal inertias of the building. Wang Su Yu adopted a data mining method for heating operation. This method takes into account the thermal inertia of the building and introduces the concept of equivalent outdoor air temperature, and finds out the relationship between the heat load and the

outdoor air temperature in previous days.[3] Architectural factors include building age, building type, building function and building envelope structure. Matteo Caldera et al used the sensitivity analysis method to analyze the relationship between the geometric and thermophysical parameters of the building and the heating energy consumption of the building. In addition, they also pointed out that the construction age of the building will also affect the heating energy consumption of the building. The later the building was built, the lower the heating energy consumption was.[4]and the heat load is nonlinearly distributed, it is difficult to establish an accurate mathematical model. Because artificial neural network has strong nonlinearity and adaptability, it can be used to predict the thermal load of a thermal station. When using neural network to forecast heating load, it is very important to determine and select the input parameters, because different input parameters will affect the accuracy of the prediction results. Input parameters of neural network model that can be selected include outdoor temperature, indoor temperature, wind speed, relative humidity, atmospheric pressure, primary water supply temperature of domestic hot water, primary return water temperature of domestic hot water, and other parameters on the day. In this paper, the factors that affect the accuracy of heat load forecasting model are researched, and the prediction results of different input parameters are compared and analysed to obtain the input parameters to ensure the accuracy of load forecasting.

\*Corresponding author's e-mail: [lirui@bucea.edu.cn](mailto:lirui@bucea.edu.cn)

## 2 Parameter selection and evaluation method

### 2.1 Model Evaluation Criteria

In this paper, the average absolute error (MAE), average absolute percentage error (MAPE), and mean square error (MSE) are used as the evaluation indicators of the model test effect.

$$MAE = \frac{1}{n} \sum_{i=1}^n (X'_i - X_i) \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{X'_i - X_i}{X_i} \times 100\% \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (X'_i - X_i)^2 \quad (3)$$

Where:  $X'_i$  is the predicted value;  $X_i$  is the actual value;  $n$  is the number of samples. Among them, the smaller the  $MAE$ ,  $MAPE$ , and  $MSE$ , and the closer  $R^2$  is to 1, the higher the prediction accuracy.

The average absolute error can more intuitively see the prediction effect of the model on the heat load data, and characterize the reliability of the model; the average absolute percentage error can compare the reliability of the prediction results of the heat load data by different models;

**Table 1.** Correlation analysis related methods

Methods	Scope of application	Effect
Pearson	Quantitative data, data meets normality	Normal graph can see normality, scatter graph shows data relationship
Spearman	Quantitative data, data does not meet normality	Normal graph can see normality, scatter graph shows data relationship
Kendall	Quantitative data, consistency judgment	Used for scoring consistency level research

It is analysed whether the sample data of each influencing factor and the heat supply of the day are normally distributed, so as to determine which analysis method to use to confirm the correlation between each influencing factor and the heat supply of the day. The K-S

**Table 2.** Correlation analysis related methods

Classification	K-S Value
Minimum outdoor temperature	0.062
Average outdoor temperature	0.002
Average indoor temperature	0.200
Average wind direction	0.200
Average wind speed	0.000
Heating load for the previous two days	0.003
Heating load for the previous three days	0.011

From Table 2, it is seen that the minimum outdoor temperature, the average indoor temperature and the average wind direction obey the normal distribution, and the average outdoor temperature, average wind speed, the heat load of the previous three days and the heat load of the previous two days do not follow the normal distribution.

the average error is a representation of the prediction. The average deviation of the value relative to the actual value, and the larger individual deviation has a greater impact on this error, which can be used to evaluate the stability of the model[5].

### 2.2 Selection of input parameters of neural network prediction model

Li Rui et al pointed out in the article that there are many factors that affect the load of thermal power stations, and the load characteristics of each thermal power station are also different. Therefore, it is necessary to select suitable methods for thermal load prediction according to different thermal power stations.[6]

In order to select reasonable model input parameters, it is necessary to analyze the correlation between the influencing factors of heat load and the heat supply of the day. Correlation analysis is used to determine whether there is a relationship between variables and the closeness of the relationship. It is a very important method for the selection of model input parameters. The relationship between each influencing factor and the heat supply of the day can be specifically expressed with data.

There are many methods of correlation analysis, each of which has different characteristics and scope of application, as shown in the Table 1.

value of each influencing factor are shown in Table 2. The basis for judging whether the data obeys the normal distribution is: if the K-S value is greater than 0.05, the data obeys the normal distribution, otherwise, the data does not obey the normal distribution.

Therefore, the correlation analysis between the minimum outdoor temperature, the average indoor temperature, the average wind direction and the heat supply of the day uses the Pearson correlation analysis method. The correlation analysis between the heat load of the previous three days, the heat load of the previous two days, and the heat load of the previous day and the heat

supply of the day adopts the Spearman correlation analysis method.

The basis for judging whether the influencing factor is significantly related to the day's heat supply is: if Pearson's significance value is less than 0.01, then the influencing factor is significantly related to the day's heat supply at the level of 0.01, otherwise, it is not significantly correlated at the level of 0.01. If the significance value of Spearman is less than 0.01, it means that the influencing factor is significantly correlated with the day's heat supply at the 0.01 level, otherwise, it is not significantly correlated at the 0.01 level.

Through the correlation analysis, it is indicated that minimum outdoor temperature, the average indoor temperature, the average outdoor temperature, heating load of the previous three days and heating load of the previous two days are significantly related to the heating value of the day. They are important factors affecting the heating load, and are selected as the input parameters of prediction model.

### 3 Results and discussion

The input of parameters is divided into four types, the first is the maximum outdoor temperature and the average outdoor temperature; the second is the minimum outdoor temperature and the average outdoor temperature; the third is the maximum outdoor temperature and the minimum outdoor temperature; The fourth type is the maximum outdoor temperature, the minimum outdoor temperature, and the average outdoor temperature. Using them as input temperature parameters to build a model, four different prediction models can be obtained, and the

average absolute error (MAE), average absolute percentage error (MAPE), and mean square error (MSE) are used as the evaluation indicators of the model test effect.

The input parameters of the A model are the maximum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load of the previous three days, the heating load of the previous two days, the heating load of the previous day. The input parameters of the B model are the minimum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load of the previous three days, the heating load of the previous two days, and the heating load of the previous day. The input parameters of the C model are the maximum outdoor temperature, the minimum outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load of the previous three days, the heating load of the previous two days, the heating load of the previous day. The input parameters of the D model are the maximum outdoor temperature, the minimum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load for the previous three days, heating load for the previous two days, the heating load for the previous day.

The neural network model is used for load forecasting by input the above four groups of parameters respectively. The prediction effects of the four situations of input parameters for prediction model are shown in Table 3.

**Table 3.** The prediction effect of four situation of input parameters for prediction models

Prediction models	MAE	MAPE	MSE
A	-0.2406	0.1350	64.0679
B	-0.2641	-0.0125	80.9637
C	1.2137	0.7681	70.9609
D	-0.3594	0.0023	61.3146

It can be seen from Table 3 that the absolute value of the MAE of the four different input models is ranked as A < B < D < C; the absolute value of the MAPE of the four different input models is ranked as D < B < A < C; The absolute value of the MSE of the different input models are ranked as D < A < C < B; the comprehensive prediction effect of the A model and the D model are better.

### 4 Conclusions

This paper analyses the factors affecting the heating consumption of a heating substation. The input parameters of neural network prediction model are analysed and selected. The average absolute error (MAE), average absolute percentage error (MAPE), and mean square error (MSE) are used to evaluate the effect of the prediction model. The conclusions are as follows:

- It is very important to determine and select the input parameters, because different input

parameters will affect the accuracy of the prediction results.

- Minimum outdoor temperature, average indoor temperature, average outdoor temperature, heating load for the previous three days, heating load for the previous two days, the factors affecting the heating load of the previous day are significantly related to the heating load of the day.
- when the model input parameters are the maximum outdoor temperature, the average outdoor temperature, the average temperature difference between the primary supply and return of domestic hot water, the heating load in the previous three days, the heating load in the previous two days, the heating load in the previous day and when the model input parameters are the maximum outdoor temperature, the minimum outdoor temperature, the average outdoor temperature, the average

temperature difference between the primary supply and return of domestic hot water, the heating load of the previous three days, the heating load of the previous two days, the heating load of the previous day, prediction effects of heating load are better.

## Acknowledgments

This paper was supported by ‘Beijing Advanced Innovation Center For Future Urban Design, Beijing University Of Civil Engineering And Architecture’ (UDC2019011121).

## References

1. Chi Y L. (2008) Research on China's Wind Power Industry Development Based on the Perspective of Low-Carbon Economy. Jilin University.
2. Werner S. (1984) The Heat Load in District Heating Systems. Stockholm: Chalmers University of Technology.
3. Wang S Y, Jiang Y C, Fang X M, Liu L T, Zheng R Y. (2011) Appropriate heating technology based on heating data mining and load forecasting[J]. HVAC, 41(07): 1-5.
4. Matteo C, Stefano P C, Marco F. (2008) Energy demand for space heating through a statistical approach: application to residential buildings. Energy and Buildings, 40(10):1972-1983.
5. Liu Y. (2020) Research on heating load forecasting method based on wavelet neural network and support vector machine. Northeast Dianli University.
6. Li R, Hao X J, Fu B, Wang L X, Sun Z Y, Dong Y. (2018) Analysis and prediction methods of influencing factors of thermal load of urban thermal power stations. District Heating, (04): 25-29.