

## Power System Load Forecasting Based on Dislocation Samples and Cloud Model

Wang Huizhon<sup>1</sup>, \*Liu Ke<sup>2</sup>, Yang Shiliang<sup>3</sup> and Zhu Hongyi<sup>4</sup>

<sup>1, 2, 3</sup>Lanzhou University of Technology, Lanzhou 730050, China)

<sup>4</sup> State Grid Gansu Electric Power Research Institute, Lanzhou 730050, China)

<sup>1</sup>38878809@qq.com, <sup>2</sup>415797893@qq.com, <sup>3</sup>375434860@qq.com,

### Abstract

*Against short-term load forecasting, this paper established a comprehensive sample system, to evaluate the factors for load forecasting by analyzing the impact of each factor and load data, and select the most appropriate factors. This paper presents the concept of dislocation sample and proposes a new model to combine the cloud model and LSSVM. Finally, use the actual data to simulation comparison by establish Cloud-LSSVM model. Experimental results show that the forecasting accuracy of the new model was significantly higher than the general forecasting methods. Test shows that the prediction method is feasible and effective.*

**Keywords:** Correlation analysis; Dislocation sample; Cloud model

### 1. Introduction

With the construction of the smart grid and the marketization of the smart grid further deepen, Not only the higher request for the safe and stable operation of power grid, but also the speed and precision of the short-term power load forecasting put forward higher requirements [1]. Constantly improving the method of load forecasting and improving the speed and precision of prediction is important significance for the market main body. For the electric power production, short-term load forecasting is the key of the electric power enterprise power generation planning and the power supply and demand balance. For the electricity market, as the operation modes of electric power from the monopoly to market competition, short-term load forecasting is becoming more and more important, it became the power plant according of the quotation, it is also the important premise of guarantee the safe and stable operation of power grid, its prediction accuracy directly affects the economic benefits of power enterprises. In general, accurate short-term power load forecasting can ensure power grid operation security and stability [2].

Traditional load forecasting methods are improved on the algorithm and ignore the importance of influencing factors [3]. In this paper, combining with the research results of scholars at home and abroad about short-term load forecasting, proposing the concept of dislocation sample. Select the appropriate sample for load forecasting. Least squares support vector machine (LSSVM) is widely used on the load forecasting of power system, but this method has many shortcomings in dealing with uncertainty problems. In order to improve the accuracy of selecting the parameters of the kernel function, to deal with uncertainty factors and to improve the accuracy of short-term load forecasting, this paper proposes a new model which is combined by the cloud model and LSSVM. Finally according to the real load data and weather data of China's southern city, using the Cloud-LSSVM model for load forecasting.

## 2. The Factors Influencing Power Load

The main factors influencing power load are natural and social factors [4]. The paper includes as much as possible historical load data and historical weather data, classify these data and then doing correlation analysis and comparison, choose the biggest impact factors for load.

Meteorological factors influencing the load most, summer air conditioning load can lead to a summer peak load, the winter heating load will make the load increased dramatically. Other weather conditions also directly or indirectly affected the load. For example, rainfall will directly affect the agricultural load, humidity and temperature will increase the change of power load. In recent years, taking into account the combined effects produced by multiple meteorological factors, integrated meteorological factors emerged, and the effect of these factors are not separate, they tend to together to influence the power load. In this paper, we consider the comprehensive meteorological factors including effective temperature, temperature humidity index and human comfort and so on. Reference formula is as follows:

1) Effective temperature (TE): Also known as the effective temperature, the temperature felt in the stationary saturation of atmospheric conditions [5]. The formula is:

$$TE = 37 - \frac{37 - T_a}{[0.68 - 0.14R_h + 1 / (1.76 + 1.4v^{0.75})]} - 0.29T_a(1 - R_h) \quad (1)$$

In which,  $TE$  is effective temperature,  $T_a$  is the temperature, °C;  $R_h$  is relative humidity, %;  $v$  is wind speed, m/s.

2) Temperature humidity index (THI): It expressed comfort level awareness of environmental temperature and humidity. The formula is:

$$THI = 32 + T_a \times 1.8 - 0.55(1 - R_h) \times (T_a \times 1.8 - 26) \quad (2)$$

In which, THI is the temperature humidity index;  $T_a$  is the temperature, °C ;  $R_h$  is relative humidity, %.

3) Human comfort (SD): It refers to the combined effects of humidity, temperature and wind speed on the human body, the human comfort level in the atmosphere [6]. The formula is:

$$SD = 1.8T_a - 0.55(1 - R_h) - 3.2\sqrt{v} + 32 \quad (3)$$

In which, SD is human comfort ;  $T_a$  is the temperature, °C ;  $R_h$  is relative humidity, % ;  $v$  is wind speed, m/s.

4) Cold and humidity index (CHI): To measure the degree of cold in the winter, in addition to considering temperature, and wind speed, humidity also should be considered. The formula is:

$$CHI = (33 - T_a) (3.3\sqrt{v} - v / 3 + 20) e^{0.005|R_h - 40|} \quad (4)$$

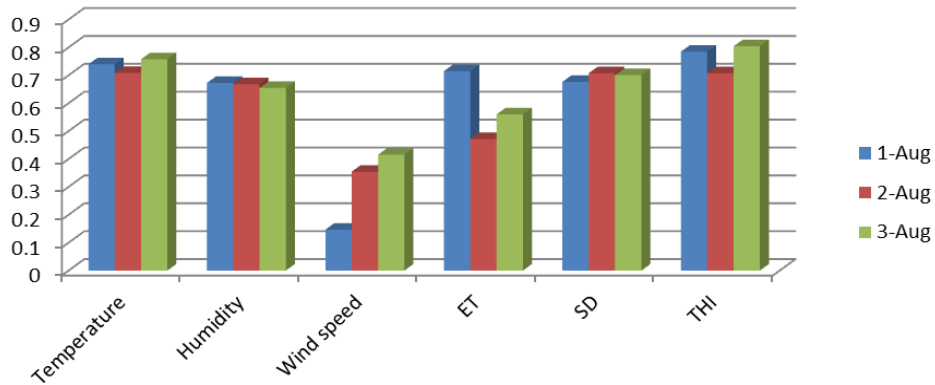
In which, CHI is cold and humidity index ;  $T_a$  is the temperature ;  $R_h$  is relative humidity ;  $v$  is wind velocity.

In this paper, the main single meteorological factors are: temperature, relative humidity, rainfall, wind speed, *etc.* Due to seasonal variations of regional climate, so not all meteorological factors is suitable as the train sample of forecast the load [7]. In order to select suitable train samples, using the correlation coefficient to analysis the single meteorological factors and the integrated meteorological factors relationally, and then choose the most favorable load forecasting meteorological factors. Formula is as follows:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \sqrt{DY}} \quad (5)$$

In which, X and Y is binary random variable,  $\rho$  is the digital features to measure the linear correlation close degree between variables X and Y.

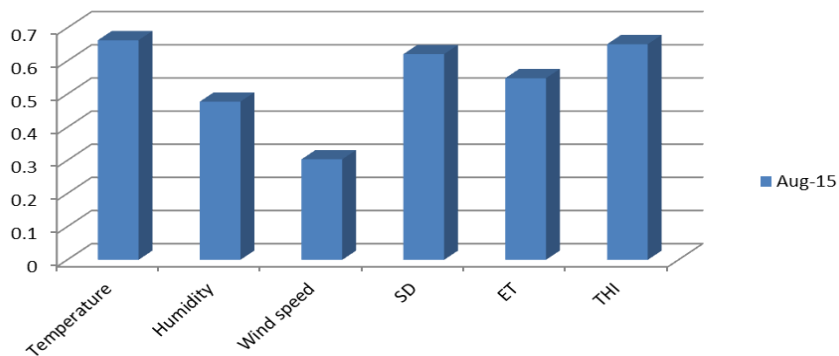
Doing correlation analysis based on a variety of meteorological data and practical value in power load of one city in south in August 2015. Because of using the historical load data are all summer and fall, so does not consider the THI. Figure 1 shows the correlation coefficient of meteorological factors and load data of daily 24 o'clock of one southern city between August 1 to 3, 2015.



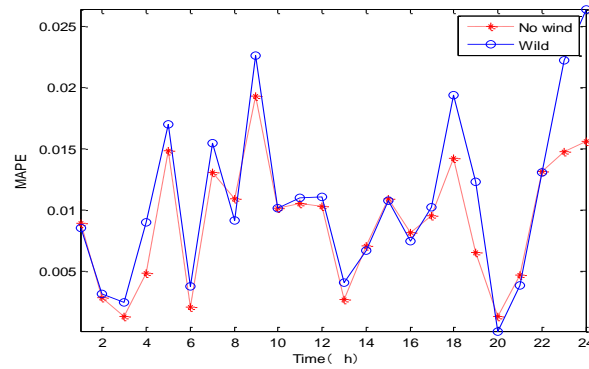
**Figure 1. The Correlation Coefficient of Meteorological Factors and Load**

From Figure1, the correlation of daily temperature, humidity, temperature humidity index is largest and the most stable, the correlation of wind speed and power load is smallest. And the correlation of effective temperature and human comfort is not stable. With only a few days of rain in August, so rainfall correlation can be ignored.

Figure 2 shows the correlation analysis of meteorological factors and the load actual value of one southern city in August 2015.



**Figure 2. The Correlation Coefficient of Meteorological Factors and Load**



**Figure 3. The MAPE of Wind Speed on July 30**

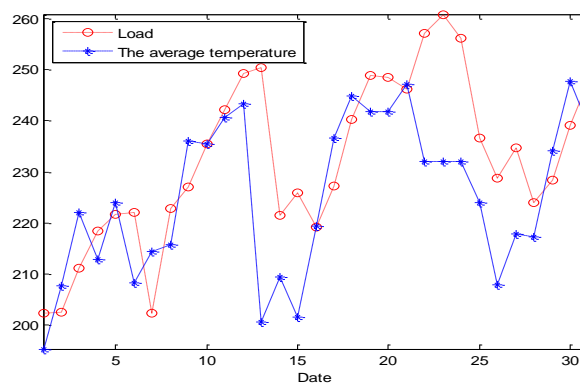
From Figure 2, can see, in meteorological factors of August, the correlation of the temperature, temperature humidity index and load is highest.

In order to verify the influence of the correlation to selected samples, forecast the load of August 30, respectively, in the condition of considering wind speed and wind speed is not considered. The prediction error of two samples as shown in Figure 3, it is obviously show that the error of only considering temperature and humidity below in the condition of joined the wind speed.

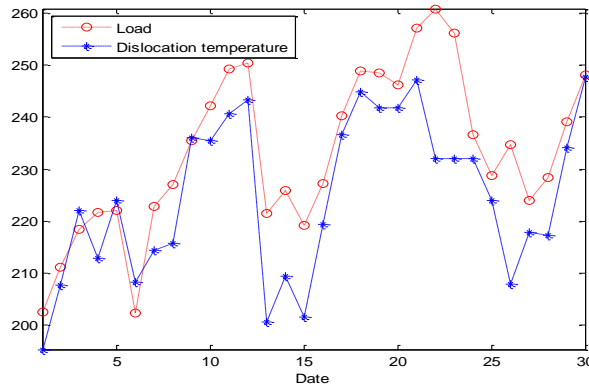
The above experiment shows that the train sample dimension is not the more the better, at the same time, verified the effectiveness of the correlation analysis. Sometimes the more sample dimension the lower accuracy, even the comprehensive meteorological index is not everything. Not every meteorological factors are applicable in every seasons. For the selection of training samples, the local climate characteristics and the seasonal habits must be considered, so as to improve the forecast accuracy when increase the number of samples dimension. This request power load forecasting must analyze the local climate and seasonal change, thereby selecting suitable meteorological factors as training samples.

### 3. Dislocation Samples

In the factors influencing the load, there have a special meteorological factors is the average temperature, low temperature and high temperature. It only has one data every day.



**Figure 4. Load and Average Temperature Comparison Chart**



**Figure 5. Load and Dislocation Average Temperature Comparison Chart**

Figure 4 is the south city daily load and average temperature contrast figure in August 2015. The temperature multiplied by eight in order to better contrast. The average temperature follow the trend of load. Dislocation samples concept has been proposed in this paper.

The traditional method, select the weather factors of n days before the forecast day as the training samples. The dislocation samples, select the weather factors of from n+1 days to two days before the forecast day as the training samples. Figure 5 shows daily total load and the dislocation samples of average temperature in August 2015. You can see the fitting degree of the dislocation temperature and load is higher. So, join the dislocation samples into the training sample is very necessary.

## 4. Cloud-LSSVM Model

### 4.1. Principles of New Improved Model

Cloud is the model of the uncertain transition between qualitative and quantitative concept, which is described by language, it completely collected fuzziness and randomness together. The cloud model can be reflected by its three characteristics: Expected value ( $Ex$ ), Entropy ( $En$ ) and Hyper-Entropy ( $He$ ) [8].

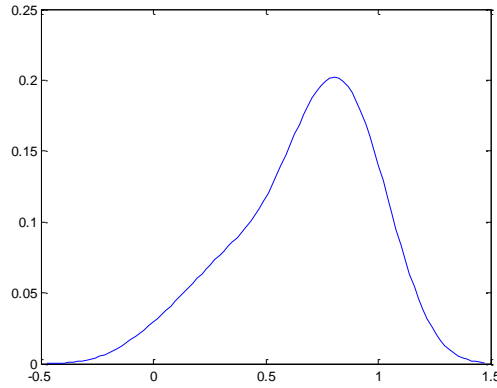
The kernel functions parameter  $\sigma$  and penalty parameter  $C$  of LSSVM had a great influence on its learning and generalization ability. The parameter optimization effect of Traditional method is not good in dealing with uncertainty, but Cloud model optimization LSSVM (Cloud - LSSVM) solved this problem.

This paper, data distribution curve and the normal cloud are obtained by the influencing factors cloud transform, then obtained digital characteristics of cloud, according to cloud generator algorithm. Combining cloud model and LSSVM, using cloud model  $Ex$  instead of penalty parameter  $C$ ,  $En$  instead of kernel function parameter  $\sigma$ . The new model can not only realize the unification of the randomness and fuzziness, but also can make use of LSSVM strong learning ability and nonlinear mapping ability.

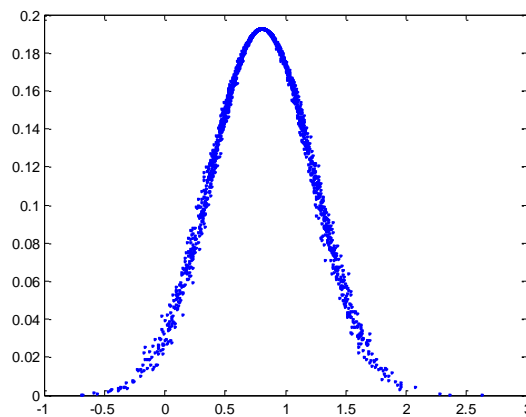
### 4.2. Cloud Transformation

Cloud Transformation means for converting any irregular data distribution, according to certain principles of mathematical transform, to several superimposed clouds. That is  $f(x) \approx \sum c_j * C_j(x)$ , in which,  $f(x)$  is data distribution function,  $c_j$  is coefficient,  $C_j(x)$  is probability density function of expectations based cloud.

In the southern city in August 2015, the temperature data as an example. First, the data normalization between [0-1], obtain the data distribution curve of temperature. Then cloud transform to the temperature factor, can be obtained one normal cloud, respectively, as shown in Figure 6 and Figure 7.



**Figure 6. Temperature Data Distribution**



**Figure 7. The Cloud of Temperature Data**

### 4.3. Cloud Generator

Normal cloud model is the most basic cloud model, its expectation curve is a normal characteristic curve, and the cloud distribution curve is:

$$y = e^{-\frac{(x - Ex)^2}{2(En)^2}} \quad (6)$$

The software or hardware generated cloud is called cloud generator [9]. Forward cloud generator based on three characteristics of cloud model, that is expectation, entropy, hyper-entropy, generate cloud droplets, backward cloud generator efficiently convert a certain number of precise numerical  $(x_i, y_i)$  to the appropriate qualitative language values  $(Ex, En, He)$ . N-dimensional backward cloud generator algorithm [10] is described as follows:

**Step1** Based on expectation  $(Ex_1, Ex_2, \dots, Ex_n)$  and variance  $(En_1, En_2, \dots, En_n)$  of each dimension of a sample, per dimension generate normal random number  $x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$ ,  $i \in [1, n]$ ;

**Step2** Based on variance  $(En_1, En_2, \dots, En_n)$  and hyper-entropy  $(He_1, He_2, \dots, He_n)$  of each dimension of a sample, per dimension generate k-dimensional normal random number  $x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$ ,  $i \in [1, n]$ ;

$$\mu_i = \exp \left[ -\frac{1}{2} \sum_{i=j}^n \frac{(x_{ij} - Ex_i)^2}{y_i^2} \right], i \in [1, n], \text{ let } (x_i, \mu_i) \text{ for cloud droplets;}$$

**Step3** Calculate

**Step4** Repeat the above three steps until the required number of cloud droplets generated.

## 5. Example Analysis

### 5.1. The Choice of the Training Sample

To predict the load in the southern city on August 29 to 31, 2015. Collected a variety of meteorological data and load data of August, select samples by correlation analysis. Take August 29, for example, the steps are as follows:

- 1) Weather and load of August 29 correlation analysis (Correlation on day);
- 2) Weather and load of 30 days before August 29 days correlation analysis (Correlation on month);
- 3) Get the highest correlation factors, selection of the same factors as training sample
- 4) Choose the highest correlation of dislocation samples as training samples.

**Table 1. The Correlation of Influence Factor on August 29**

Correlation on day	Correlation on month	Dislocation samples
THI 0.832	THI 0.663	The average temperature 0.88
Temperature 0.742	Temperature 0.645	The highest temperature 0.754
ET 0.702	SD 0.6	The lowest temperature 0.458

Table 1 shows the correlation of influence factor on August 29. So choose temperature, temperature humidity index (THI), dislocation samples of average temperature as the training samples.

### 5.2. The Simulation

In order to verify the dislocation samples and cloud model influence on prediction model, this paper also carried out two experiments were compared. The first set without dislocation samples. The second set without dislocation samples and Cloud model.

**Table 2. Load Value Compared with the Actual Value**

Date	Actual Value (MW)	NEW Method		No Dislocation Samples		General Method	
		Predictive Value	MAPE	Predictive Value	MAPE	Predictive Value	MAPE
29	966.5	972.1	0.58%	975.5	0.93%	985.5	1.97%
30	971.4	968.6	0.60%	959.7	1.20%	955.3	1.66%
31	907.8	901.1	0.74%	919.7	1.31%	923.9	1.77%
Average MAPE		0.64%		1.15%		1.8%	

Table 2 shows three kinds of load forecast method on July 29 to 31, 12 am. Can obviously see the MAPE of this method is only 0.64%, the MAPE of the method without dislocation sample is 1.15%. The MAPE of general methods is 1.8%. In this paper, the method makes the prediction error reduced more than one percent.

## 6. Conclusion

This paper analyzes the various factors influencing the load, through the correlation analysis of each factor and the load data to evaluate the effects of various factors on load forecasting, select the most appropriate factors as the training sample. Put forward the concept of dislocation samples. Using the Cloud-LSSVM model can not only realize the unification of the randomness and fuzziness, but also can make use of LSSVM strong learning ability and nonlinear mapping ability. Experimental results show that the accuracy of power system load forecasting based on dislocation sample was significantly higher than the general forecasting methods. Test show that this short-term load forecasting method is effective and feasible and has a broad development prospects.

## References

- [1] Z. Ming, C.-Q. Lü and K. Tian, "Least Squares-support Vector Machine Load Forecasting Approach Optimized by Bacterial Colony Chemotaxis Method", Proceedings of the CSEE, (2011).
- [2] C.-L. Tang, J.-P. Lu and Y.-Z. Xie, "Improved Data Stream On-Line Segmentation Based Ultra Short-Term Load Forecasting", Power System Technology, vol. 7, no. 38, (2014), pp. 2015-2020.
- [3] W.-R. Mo, B.-M. Zhang, H.-B. Sun and Z.-H. Hu, "Application of Extended Short-term Load Forecasting", Proceedings of the CSEE, (2003).
- [4] B.G. Koo and M. S. Kim, "Short-Term Electric Load Forecasting Using Data Mining Technique", Intelligent Systems and Control (ISCO), (2013); Coimbatore, India.
- [5] L.-H. Zhong and Y. Li, "The Impact of Composite Meteorological Factors on the Electrical Load in Guangxi", Meteorological Monthly, vol. 34, no. 5, (2008), pp. 31-37.
- [6] Z. Wei, "A distribution short-term load forecasting based on human comfort index", Power System Protection and Control, vol. 9, no. 41, (2013), pp. 74-79.
- [7] R. Li and C.-Y. Guo, "Relevance Analysis of Meteorological Index and Peak Load in Tianjin-Beijing-Tangshan Power Grid", Power System Technology, vol. 32, no. 6, (2008), pp. 87-92.
- [8] Y.-Y. Zhang, R.-J. Liao and L.-J. Yang, "An Assessment Method for Insulation Condition of Power Transformer Based upon Cloud Model", Transactions of china electro technical society, vol. 27, no. 5, (2012), pp. 13-19.
- [9] X.-P. Mi and X.-M. Li, "Study on fault diagnosis of gear pump based on normal cloud neutral network", Computer Science, vol. 40, no. 8, (2013), pp. 266-267.
- [10] G.-W. Zhang, D.-Y. Li and Z. Q. Luo, "A new kind of subjective trust mode", Wuhan University Journal of Natural Sciences, vol. 11, no. 6, (2006), pp. 1457-1461.



## Authors



**Wang Huizhong**, Lanzhou University of Technology, China.  
38878809@qq.com.



**Liu Ke**, Lanzhou University of Technology, China. The research  
direction is electric power system dispatching automation.  
415797893@qq.com.

