

PAPER • OPEN ACCESS

Research on Influencing Factors of Line Loss Based on Multi Model Analysis

To cite this article: Junling Liu *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **237** 032101

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

Research on Influencing Factors of Line Loss Based on Multi Model Analysis

Junling Liu¹, Xiuqing Cui², Yan Chen¹, Jiakui Zhao¹, Yuxi Liu¹, Yanghong Ou¹, Bao Yuan¹

¹Beijing China-power information technology Co., Ltd. Haidian District Beijing, China

²Marketing Department, Stage Grid Corporation of China, China

*Corresponding author e-mail: 13552462300@163.com

Abstract. Line loss is an important factor to measure the economy of power system. In order to explore the main factors affecting line loss, the line loss data of 60 stations in a city are screened out by using power information collection system. With the voltage level data as the index, the stations are clustered by the fuzzy clustering algorithm to form effective differential grouping. According to different types of stations, the curve similarity method and regional factor analysis method are used to analyze the weather conditions and the influence degree of the station factors on line loss. The main influencing factors of line loss are obtained. According to the analysis results of each type of stations, the gray correlation analysis algorithm is used to rank the significant factors according to the influence degree. Taking a certain area line loss data as an example, the influencing factors of line loss abnormality are studied. The results show that according to the characteristics of each type of platform area, the model can analyze the factors of each type of platform area and judge the influencing factors at different levels, so as to achieve the effect of targeted and hierarchical governance for different regions.

1. Introduction

In recent years, with the rapid development of social economy and the acceleration of industrialization, the power supply capacity of distribution lines has been increasing. At the same time, the problem of distribution line loss has become increasingly prominent. The level of line loss not only reflects the economic operation status of the power grid, but also is closely related to the economic benefits of power grid enterprises [1]. Inadequate management, technology and other reasons can lead to abnormal line loss rate, resulting in waste of energy. Therefore, line loss management must be carried out to effectively improve the level of line loss management and refined level of grid operation management. In power marketing, there are many related factors and multi-terminal transformations, including abnormal power loss caused by active power loss, such as abnormal power consumption, leakage, metering failure, etc., and abnormal line loss caused by passive power loss, such as circuitous power supply, line aging, equipment aging, which results in a great waste of energy. It is not conducive to the development of the power industry. Therefore, it is necessary to conduct a



comprehensive and effective analysis of line loss and accurately locate the causes of abnormal line loss in different areas.

The power loss of distribution network is related not only to the structure and load nature of power grid, but also to the management level of enterprises. Because management line loss has no rules to follow, and it is difficult to measure, it is usually called "unknown loss". The power loss caused by inadequate management occupies a large proportion of the actual line loss, and in some places, some lines are even quite serious. Through the analysis of line loss, this paper can deeply understand the causes, nature, proportion of each component and other factors of line loss, find out the main factors affecting the loss, and take corresponding measures to achieve greater loss reduction effect and economic benefits with less investment.

Traditional line loss management methods [2] generally use multi-system professional collaboration and data resources integration to divide components, and divide the distribution network into sections, monitor each section online, and then eliminate the factors one by one. However, most of these methods rely on the experience of business personnel, which are subjective and have little substitution. It takes a long time and lacks the necessary scientific analysis methods, resulting in a low accuracy of the factors.

Based on the electric power information acquisition system, 12 valid data such as line loss, transformer type, voltage grade, rated capacity, short circuit loss, no-load loss, rated current at low voltage side and number of households are screened. First, the fuzzy clustering algorithm is used to cluster the stations. Then, according to different types of stations, this paper use Spearman analysis method and factor analysis method to analyze the weather conditions and the influence degree of the station factors on the line loss, and get the main factors affecting the line loss. Finally, the grey correlation method is used to sort the main factors according to the significance. This method can accurately locate the significant factors affecting line loss in different areas, and judge the degree of impact by different levels, so as to achieve targeted and hierarchical treatment for different areas. It provides theoretical support for power grid enterprises to analyze distribution line loss factors, and it is of great significance for strengthening the safety of distribution network operation and promoting the sustainable development of distribution network.

2. Study on comprehensive method for influencing factors of line loss

The comprehensive analysis method of influencing factors of line loss is to make a significant judgment of weather factors and station factors from the perspective of station area. This method not only judges whether a factor index is the cause of abnormal line loss, but also considers the influence of other factors comprehensively, and ranks the significant factors one by one according to the degree of influence. In order to find out the main and secondary factors that lead to abnormal line loss in different regions, and achieve the effect of targeted treatment in different stations, and the analysis results are more scientific and effective [3-4].

2.1 Data screening

In this paper, weather factors and area factors are selected as data indicators for analyzing line loss. Eleven complete and effective factor data and the line loss rate of area are screened out through the electricity information collection system. Among them, the factors of the station area mainly include: transformer type, voltage grade, rated capacity, short circuit loss, no-load loss, rated current of low-voltage side, total household number, transformer use nature, power consumption category, operation capacity and supply voltage.

2.2 Platform Division

Based on the electricity information acquisition system, the data indicators of many stations are screened out. Because the effects of different regional factors are different, and the characteristics of the stations are different, but the decomposition boundaries between the data indicators are not clear, so the fuzzy clustering algorithm is used to cluster the stations.

(1) establish data matrix.

Set the universe $U = \{u_1, u_2, \dots, u_{60}\}$ is a categorized object, and there are Omega indices for each area object.

$$x_i = \{x_{i1}, x_{i2}, \dots, x_{i\omega}\}, i = 1, 2, \dots, 60 \quad (1)$$

The original data matrix $X = (x_{ij})_{60 \times \omega}$ is obtained.

(2) data normalization

$$x'_{ij} = \frac{x_{ij}}{\max_{1 \leq k \leq n} \{x_{kj}\}}, i = 1, 2, \dots, 60; j = 1, 2, \dots, \omega \quad (2)$$

(3) establishing fuzzy similarity matrix

$$\delta_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m (x_{ik} \vee x_{jk})} \quad (3)$$

(4) clustering and drawing dynamic clustering diagram

The transfer closure matrix $t(R)$ of the fuzzy similarity matrix R is obtained by using the transfer closure method. The dynamic clustering graph is drawn by clustering according to the size of lambda.

2.3 Algorithm for similarity of curves

Curve similarity measurement algorithm is used to solve the problem of similarity measurement between curves. Pearson correlation coefficient is often used to describe the similarity of two curves [5-6], the formula is:

$$r = \frac{\sum_{i=1}^n (w_i - \bar{w})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (w_i - \bar{w})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

Among them, n is the sample size, \bar{w} and \bar{y} are the observed values and mean values of weather variables and line loss rate respectively. R describes the degree of linear correlation between two variables. The value of R is between -1 and $+1$, if $r > 0$, it indicates that two variables are positively correlated, that is, the larger the value of one variable, the larger the value of another variable; if $r < 0$, it means that two variables are negatively correlated, that is, the larger the value of one variable, the smaller the value of another variable. The greater the absolute value of R , the stronger the correlation. It is important to note that there is no causal relationship here. If $r = 0$, it indicates that the two variables are not linearly correlated, but may be correlated in other ways (such as curve).

Using the sample correlation coefficient to infer whether the weather factors are related to line loss, the original hypothesis that the overall correlation coefficient is 0 can be tested by t statistics. If t -test is significant, the original hypothesis is rejected, that is, weather factors are correlated with line loss; if t -test is not significant, the original hypothesis can not be rejected, that is, weather factors are not linearly correlated with line loss.

2.4 regional single factor analysis model

Considering the factors affecting line loss y , A has k different levels of A_1, A_2, \dots, A_k , at the same time, at each level of A_r , the total X_1, X_2, \dots, X_k is independent of each other and satisfies $X_r \sim N(\mu_r, \sigma^2)$, ($r = 1, 2, \dots, k$). At each level of A_r , there are n_r ($n_r \geq 2$) line loss data samples $X_{1r}, X_{2r}, \dots, X_{nr}$ ($r = 1, 2, \dots, k$). For fixed R , they come from the same whole, that is, $X_{ij} \sim N(\mu_j, \sigma^2)$, ($i = 1, 2, \dots, n_j, j = 1, 2, \dots, r$).

Order $\varepsilon_{ij} = X_{ij} - \mu_j$, ($i = 1, 2, \dots, n_j, j = 1, 2, \dots, r$), the ε_{ij} of epsilon A_r is the error of the I factor under the horizontal level, and it is the unobservable random error. X_{ij} can be expressed as:

$$\begin{cases} X_{ij} = \mu_j + \varepsilon_{ij}, & (i = 1, 2, \dots, n_j, j = 1, 2, \dots, r) \\ \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases} \quad (5)$$

Among them, μ_j and σ^2 are unknown parameters, and each ε_{ij} is independent. Formula (5) is called Regional univariate analysis model [7].

For model (5), test hypothesis $H_0: \mu_1 = \mu_2 = \dots = \mu_r, H_1: \mu_1, \mu_2, \dots, \mu_r$. The $\mu_1, \mu_2, \dots, \mu_r$ is not all equal. The form is shown in Table 1:

Table 1. One way ANOVA

Data sources	Sum of squares	Freedom f	Mean square	F value	Saliency
factor A	S_A	$r - 1$	$\bar{S}_A = \frac{S_A}{r - 1}$	$\frac{\bar{S}_A}{\bar{S}_e}$	
error	S_e	$n - r$	$\bar{S}_e = \frac{S_e}{n - r}$		
total	S_T	$n - 1$			

In Table 1, S_T is referred to as the sum of total deviations, or the sum of total squares for short. It represents the sum of the differences between all factor data X_{ij} and the total average \bar{X} and reflects the fluctuation degree of all factor data X_{ij} . S_e is called the sum of squares or the sum of squares of errors, which reflects the degree of random fluctuation in factor data. S_A is called the sum of squares or the sum of squares of errors between groups. The sum of squares of effects of factor A reflects the degree of difference between the mean μ_j of factor data. In the saliency column of Table 1, if $F \geq F_{\alpha}(r - 1, n - r)$, the indicator A is the significant factor affecting line loss; if $F < F_{\alpha}(r - 1, n - r)$, the indicator A is not the significant factor affecting line loss.

2.5 regional multivariate analysis model

(1) establish a multivariate analysis model.

Considering that in the analysis of the influence of a station factor on line loss, due to the intersection of factors and effects, other factors will also have some immeasurable impact on line loss, then these other variables are defined as covariates x_1, x_2, \dots, x_m , line loss loss is defined as dependent variable y , the factor considered is A, and k level [8].

There is a linear relationship between x_1, x_2, \dots, x_m and dependent variable y , and the relationship is consistent in each group, that is, the covariates of each group are parallel to the regression linear model established by dependent variable. The dependent variable y and covariate x_1, x_2, \dots, x_m . The least square method was used to establish the model. The linear regression model of x_1, x_2, \dots, x_m is used to test whether the slope of the linear regression model is consistent at different levels of K group.

In order to test whether there is a significant difference in the mean of the dependent variable y at different levels from the perspective of regression analysis, a dummy variable is introduced.

$$G_r = \begin{cases} 1, & y \text{ is observed at the } A_r \text{ level} \\ 0, & Y \text{ is observed at non } A_r \text{ level.} \end{cases}, \quad r = 1, 2, \dots, k$$

With $G_1 + G_2 + \dots + G_k = 1$, the regional multivariate model can be expressed as:

$$Y = \beta'_0 + \beta_1 X_1 + \dots + \beta_m X_m + \gamma_1 G_1 + \gamma_2 G_2 + \dots + \gamma_{k-1} G_{k-1} \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n) \quad (6)$$

(2) parameter estimation of multivariate analysis model

The least square method is used to estimate the coefficients beta and gamma in the model. According to the principle and conclusion of the least square method, the estimators beta and gamma satisfy the normal equation.

$$(X, G)'(X, G) \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (X, G)'Y$$

Among, $X = (X_1, X_2, \dots, X_m)$, $G = (I, G_1, G_2, \dots, G_{k-1})$, $\gamma = (\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_{k-1})'$, $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$, $Y = (y_1, y_2, \dots, y_k)'$

At this point, the sum of squares of residuals in model (6) is

$$S_e(x_1, \dots, x_m, G_1, \dots, G_{k-1}) = (Y - X\hat{\beta} - G\hat{\gamma})'(Y - X\hat{\beta} - G\hat{\gamma})$$

For a single factor covariance model at a single level, the model (6) can be reduced to:

$$Y = \beta'_0 + \beta_1 X_1 + \dots + \beta_m X_m + \gamma_{k-1} G_{k-1} + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n) \quad (7)$$

Among, $\beta'_0 + \gamma_{k-1} G_{k-1} = \beta_0$, The model (7) is simplified to:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m \quad (8)$$

The coefficients in the multivariate analysis model can be transformed into the coefficients of multivariate linear regression. If n_r line loss data samples exist at each level, the model (8) can be transformed into:

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, n_r \end{cases} \quad (9)$$

Matrix expressions are often used for convenience. Remember:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n_r1} & x_{n_r2} & \dots & x_{n_rm} \end{bmatrix},$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{n_r} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_m \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_{n_r} \end{pmatrix}$$

According to the least square method, the coefficient estimated in the regional multivariate analysis model is: $\hat{\beta} = (X'X)^{-1}X'Y$.

When the factor index A keeps parallel at different levels, the regression slope of the single factor covariance analysis method is the same. It is considered that the single factor covariance analysis method is effective.

If the test results are covariates x_1, x_2, \dots, x_m has no effect on the dependent variable y, and the variance of y can be directly analyzed by one-way ANOVA

(3) significance test of factor A

For $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_{k-1} = 0$, the test can be conducted with the following F statistics:

$$F = \left(\frac{S_e(x_1, \dots, x_m)}{S_e(y) - \hat{\beta}'R_0} - 1 \right) \cdot \frac{n-m-k}{k-1} \quad (10)$$

At a given level of significance, the negative field of H_0 is:

$$W = \{F \geq F_\alpha(k-1, n-m-k)\}$$

Among them, $S_e(x_1, \dots, x_m)$ represents the dependent variable y established for all line loss sample values on the dummy variable x_1, \dots, x_m . If F is greater than the critical value $F_\alpha(k-1, n-m-k)$, then the original hypothesis is rejected. That is to say, factor index A has significant difference in line loss y at different levels of each group, i.e. the factor indicator is considered A is a significant factor affecting line loss.

2.6 significant sorting of regional factors

For each station area, several main factors affecting line loss are extracted by using curve similarity algorithm and factor analysis model. According to the different influencing factors of each station area, the main factors are sorted successively according to the significance of line loss, and the main sum of each station area is extracted by using grey relational degree algorithm [9] Secondary factors.

In the power system, it is difficult to find the main contradiction and find the main characteristics. Because the relationship between the factors in the station area is grey and the relationship between the factors is not clear. The grey relational degree algorithm is a method to analyze the relational degree of each grey factor in the system. In calculating the relational degree, it is necessary to calculate the relational coefficient first. Taking the line loss value of area as a reference sequence, the main factor value is taken as the comparison sequence [10].

Selected reference sequences and comparison sequences:

Reference sequence: $Y_0 = \{y_0(1), y_0(2), y_0(3), \dots, y_0(n)\}$,

Comparison sequence: $Y_1 = \{y_i(1), y_i(2), y_i(3), \dots, y_i(n)\}$,

Dimensionless processing of parameter sequences and comparison sequences is carried out using initialization:

$$\hat{X}_0(k) = \frac{y_0(k)}{y_0(1)} = \{\hat{X}_0(1), \hat{X}_0(2), \dots, \hat{X}_0(n)\} \quad (11)$$

$$X_i(k) = \frac{y_i}{y_i(1)} = \{X_i(1), X_i(2), \dots, X_i(n)\} \quad (12)$$

Calculate correlation coefficient:

$$\eta_{0i}(k) = \frac{\min\min|X_0(k) - X_i(k)| + \rho \max\max|X_0(k) - X_i(k)|}{|\hat{X}_0(k) - X_i(k)| + \rho \max\max|\hat{X}_0(k) - X_i(k)|} \quad (13)$$

In the form:

(1) The $|\hat{X}_0(k) - X_i(k)|$ is the absolute error of K point X_i and \hat{X}_0 .

(2) $\min\min|\hat{X}_0(k) - X_i(k)|$ is the two level minimum difference. $\min|\hat{X}_0(k) - X_i(k)|$ is the first minimum difference, which means finding the minimum difference between each point and $X_i(k)$ on the $\hat{X}_0(k)$ sequence; $\min\min|\hat{X}_0(k) - X_i(k)|$ is the second minimum difference, which means finding the minimum difference in all sequences on the basis of finding the minimum difference on each sequence.

(3) $\max\max|\hat{X}_0(k) - X_i(k)|$ is the second largest difference, and its meaning is similar to the minimum difference.

(4) ρ is called resolution, $0 < \rho < 1$, and generally takes $\rho = 0.5$.

(5) For sequences with different units and initial values, initialization should be performed before calculating the correlation coefficient, i.e. dividing all the data of the sequence by the first data.

(6) After calculating the number of $\hat{X}_0(k)$ sequences and $X_i(k)$ sequences, the average values of all kinds of correlation coefficients are calculated.

$$\gamma_{0i} = \frac{1}{n} \sum_{k=1}^n \eta_{0i}(k) \quad (14)$$

This average γ_{0i} is called the correlation degree between the $\hat{X}_0(k)$ sequence and the $X_i(k)$ sequence. It is also the correlation degree between the reference sequence Y_0 and the comparison sequence Y_1 , that is, the correlation degree between the line loss of the station area and the main factors [11].

3. Empirical analysis

3.1 Differentiated area processing

Based on the factor indexes of 60 stations in a city selected in the electricity information acquisition system, the fuzzy clustering algorithm is used to cluster the stations according to the characteristics of the stations, which is the data basis for the targeted treatment of the line loss of the stations.

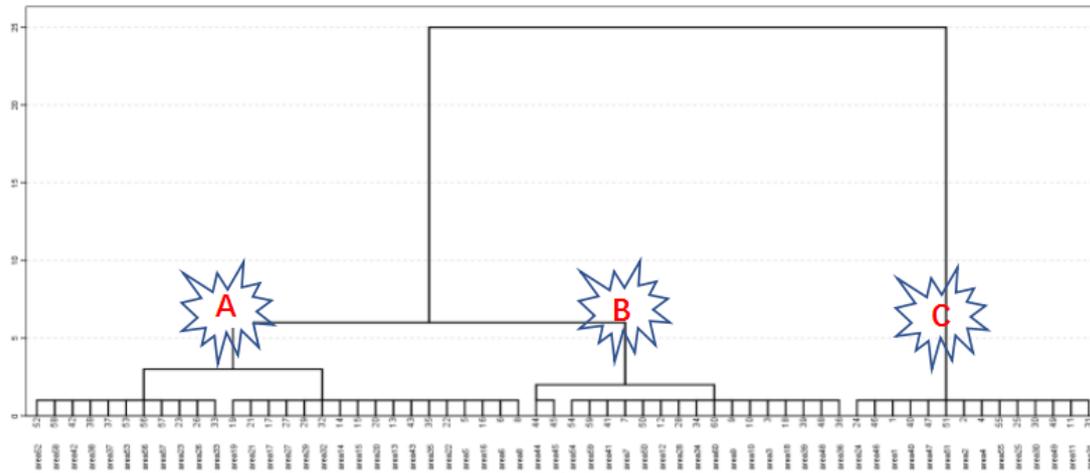


Figure 1. Cluster map

According to the cluster map, 60 stations can be divided into three categories: A, B and C according to the characteristic attributes of each station. The specific classification results are shown in Figure 2.

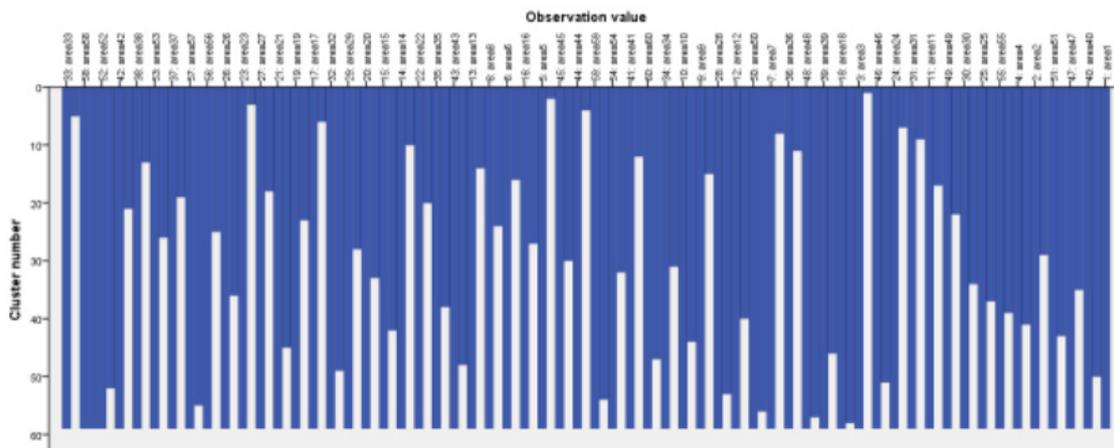


Figure 2. Regional icicle map

Through the regional icicle map, we can clearly see the classification results of each category of the proposed area. Based on the results of the ice column map, the causes of the line loss anomalies in each category of the area are analyzed.

3.2 Regional weather condition analysis

Based on the weather data collected from three types of stations A, B and C through Internet channels, the curve similarity algorithm is used to study the curve similarity with the line loss rate of three types of stations respectively. As shown in Figure 3.

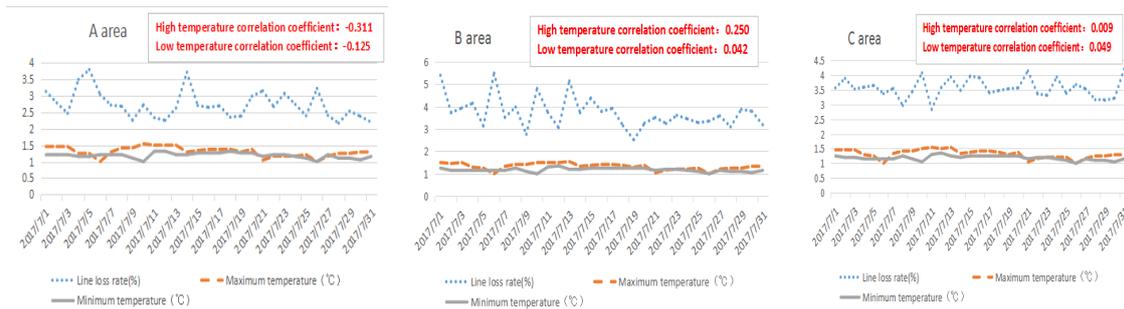


Figure 3. Analysis of three types of regional weather factors

According to the weather factor analysis tables of A, B and C, it can be concluded that the maximum temperature of A/B platform area has the most significant effect on line loss, and the correlation coefficients are -0.313 and 0.25, respectively. The maximum temperature of A platform area is negatively correlated with line loss, that is, the higher the maximum temperature of A platform area, the lower the line loss; It is a positive correlation, that is, the lower the maximum temperature of the B class area, the lower the line loss. The lowest temperature in C station has the most significant effect on line loss, and the correlation coefficient is 0.049, which is positive correlation. That is, the lower the lowest temperature in C station, the lower the line loss rate.

3.3 Multi model regional factor analysis

Based on the electric power information collection system and the influence business application system, all the electric power customer files are found in the same period, which is used as the data support for the analysis of the influence factors of line loss. For each type of station area, the number of households, supply voltage, operation capacity, type of power consumption, type of transformer, nature of transformer use, voltage grade of transformer, rated capacity of transformer, short-circuit loss of transformer, no-load loss of transformer, rated current of low-voltage side of transformer are selected as the factors affecting line loss.

(1) index selection. Taking A station as an example, the influence degree of transformer type on line loss is analyzed. The influencing factors of the station area collected by the electric power information acquisition system include the total number of households, the type of transformer, the rated capacity of transformer and the rated current of low voltage side of transformer.

(2) discriminant model. The scatter plot is used to analyze whether there is a linear relationship between the total number of households as covariates, the rated capacity of transformers, the rated current of low voltage side of transformers and the line loss rate in the region, and whether the regression coefficients are equal. The analysis results are shown in Fig. 4.

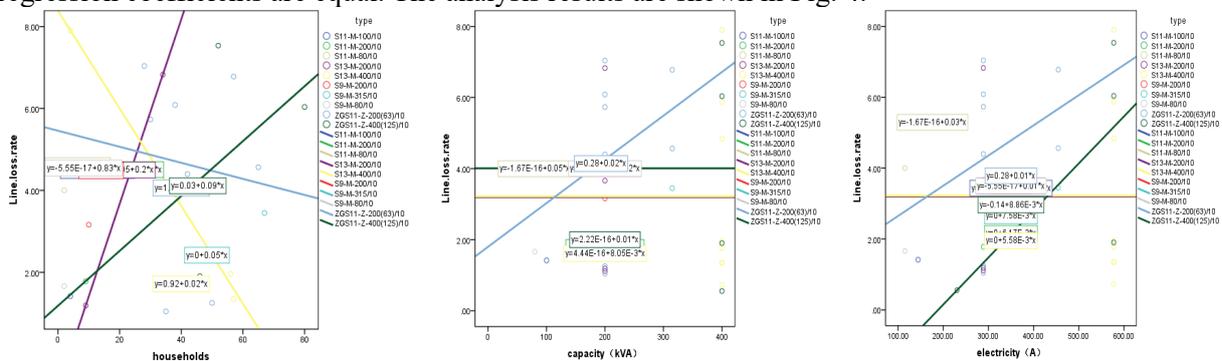


Figure 4. The scatter diagram of the line loss rate of the transformer in the form of other covariates with transformer type as fixed factor

According to the scatter plots of total households, transformer rated capacity and transformer low-voltage side rated current in Figure 4, it can be seen that the total households, transformer rated capacity and transformer low-voltage side rated current have linear relationship with the average line

loss in the region respectively, but the slope of each plot is obviously different and does not conform to the region. The condition of multi-factor analysis model, so it is invalid to use transformer type, transformer rated capacity and transformer low-voltage side rated current as covariable, transformer type as fixed factor and regional line loss rate as regional multi-factor analysis model. That is to say, the regional single factor analysis model should be used to determine whether the type of transformer in class A area is the cause of abnormal line loss.

(3) model checking. Using the regional single factor analysis model, we need to satisfy the normal population with the same variance in each group. For the index data of A area, the test results are shown in Table 2.

Table 2. Homogeneity test of variance

F	$Df1$	$Df2$	Saliency
0.646	3	20	0.595

According to Table 2, we can see that the probability value is 0.595, which is greater than 0.05. So we accept the original hypothesis that the variance of each group has no significant difference at the level of 0.05, which satisfies the homogeneity test of variance.

(4) model results.

Table 3. Regional single factor analysis table

Index	Sum of squares	Df	Mean square	F	Saliency
Between indicators	57.484	4	14.371	6.297	0.018
Within the index	45.640	20	2.282		
Total	103.124	24			

From Table 3, we can see that the probability value is 0.018, less than 0.05, so we reject the original assumption that there is a significant difference between transformer models at the level of 0.05, that is to say, the transformer model is a significant factor affecting line loss.

(5) factor selection. According to the results of regional single factor analysis model analysis, transformer type is one of the factors leading to abnormal line loss rate. However, there are many types of transformer in class A area. It is necessary to distinguish which type of transformer has the highest line loss rate through multiple linear comparison.

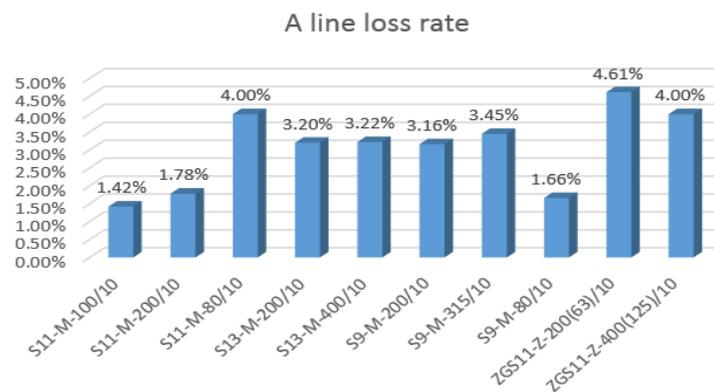


Figure 5. Line loss rate under different transformer types

According to Fig. 5, when the transformer model is ZGS11-Z-200(63)/10, the maximum line loss is 4.61%, and when the transformer model is S11-M-100/10, the minimum line loss is 1.42%.

(6) discrimination of regional factors.

Similar to the discriminant method of factor analysis in A platform area, the comprehensive analysis of factor in B/C platform area is carried out, and the following results are obtained:

Table 4. Comparison table of significant factors in different stations

Region	Significant factors	Factor level of maximum line loss	Factor level of minimum line loss
A	Transformer type	ZGS11-Z-200(63)/10	S11-M-100/10
B	Transformer type	SBH15-M-160/10 SBH15-M-200/10 SBH15-M-315/10 SBH15-M-400/10	S11-M-30/10
	Transformer no load loss	149w	128w
	Rated current at low voltage side of transformer	630A	43A
C	Transformer short-circuit loss	4165w	2500w
	Transformer no load loss	465w	480w

According to the analysis of the influencing factors of A, B and C stations, the following conclusions can be drawn: For the A area factors, the factors affecting the line loss are: transformer type; For the factor of B station area, the factors affecting line loss are transformer type, no-load loss of transformer and rated current of low voltage side of transformer; Among the factors of C station area, the factors affecting line loss are: short circuit loss of transformer and no-load loss of transformer.

3.4 Significant sorting of regional factors

Based on the analysis of the multi model area factors, This paper can draw a conclusion that the factors affecting the line loss are transformer type, transformer no load loss, transformer low voltage side rated current and transformer short circuit loss.

In this paper, grey correlation algorithm is used to analyze the correlation degree between line loss and regional factors, and the factors are ranked significantly.

Taking the regional line loss rate as the reference sequence. The comparison series are transformer type, no-load loss of transformer, rated current of low-voltage side of transformer and short-circuit loss of transformer. The influence degree of the above factors is sorted by grey correlation algorithm. The result of correlation is shown in Table 5.

Table 5. Grey relational grade calculation results

Influence factor	Transformer type	Transformer no load loss	Rated current at low voltage side of transformer	Transformer short-circuit loss
Correlation degree	0.820	0.620	0.685	0.641

The results of grey correlation algorithm show that the correlation degree between the four major factors and the regional line loss rate is respectively: $\gamma_{01} = 0.820$, $\gamma_{02} = 0.620$, $\gamma_{03} = 0.685$, $\gamma_{04} = 0.641$. Because $\gamma_{01} > \gamma_{03} > \gamma_{04} > \gamma_{02}$. Therefore, the four major factors affecting the line loss rate from large to small are: transformer type, transformer low-voltage side rated current, transformer short-circuit loss, transformer no-load loss.

According to the results of regional factor ranking, this paper can draw the following experience: if there is an abnormal line loss in a certain station, according to the results of multi-regional factor analysis model, electric worker can first check the operation status and operation status of transformer models. It provides theoretical basis and targeted guidance for regional line loss management.

4. Conclusion

In this paper, factors indicators of 60 stations in a city are selected as the research object, and the weather factors leading to abnormal lines and station factors are analyzed by subregional analysis. According to the comprehensive study of regional multi-model factor analysis, it is found that the four factors leading to abnormal line loss are the transformer type, no-load loss of transformer, rated current of low-voltage side of transformer and short-circuit loss of transformer, and the influence degree decrease in turn. Combined with the method of multi-model factor analysis, the significant factors in each area can be identified. Based on the analysis results of this paper, for three types of A/B/C stations, line losses can be effectively reduced by adjusting the type of transformer. At the same time, the rated current of low-voltage side of transformer, short-circuit loss of transformer and transformer voltage can also be referred to. The indicator number of no-load loss is used to judge the line loss condition at that time. In the future development of power grid, this model can achieve the effect of targeted control of line loss in each region, more comprehensively reflect the line loss situation, and make the loss reduction effect reach the best state.

Acknowledgements.

This work gets the research and application of big data application project based on supporting cloud computing business.

References

- [1] Yongxian Huang. (2012) Study on energy saving and loss reduction of distribution network .theoretical research on Urban Construction. J., 29: 2-14.
- [2] Hongliang You, Jiang Xin. (2008) Study on the relationship between load variance characteristics and power loss in distribution lines. DSM.J., 10(2): 13-15.
- [3] Bing Ran, Xiaohui Song.(2009) Analysis of influencing factors of distribution network line loss . China Electric Power Academy.J., 10: 6-8.
- [4] Xingxing Lu, Qidong Huang, Feng Xue, Kun Yan. (2017) Analysis of the influencing factors of line loss in Taiwan . Songjiang Power Supply Company, Shanghai Power Company, Guodian, 5: 4-10.
- [5] Xiaojuan Shen. (2012) Symmetry Recurrence in protein sequence and structure with Pearson's correlation coefficients. In: BHI.Bei Jing .832-835.
- [6] Leo Egghe, Loet Leydesdorff. (2009) The relation between Pearson's correlation coefficient r and Salton's cosine measure. JASIST, 60(5): 1027-1036.
- [7] Taishu Cai, Yunyu Sun. (2001) Applied mathematical statistics. Wuhan: Wuhan University.42.
- [8] Minqiong Chen. (2014) A regression algorithm for single factor covariance analysis. Zhongshan University, 10:1-3.
- [9] Guoxiang Xu. (2000) Statistical prediction and decision. Shanghai: Shanghai University of Finance and Economics press, 2000:9-11.
- [10] Jingqin Liu. (2011) Analysis and control of influencing factors of line loss rate fluctuation. Xiamen Electric Power Bureau, Fujian Electric Power Co., Ltd., 2:2-9.
- [11] Dexiang Li, Chi Lu, Weibing Chen. (2016) Research on the improvement method of distribution line loss based on multi-factor analysis. Nanning Power Supply Bureau, Guangxi Power Grid Co., Ltd.,4: 12-13.