Calculation method of distribution network limit line loss rate based on fuzzy clustering

X Y Yuan\textsuperscript{1,2} and Y L Tao\textsuperscript{1,2}

\textsuperscript{1}State Grid Gansu Electric Power Group Zhangye Power Supply Company, Zhangye 734099, Gansu Province, China

E-mail: yuanxy@gs.sgcc.com.cn/tyl@gs.sgcc.com.cn.

Abstract. In view of the characteristics of 10kV distribution network lines, in order to analyze the line loss rate and determine the loss reduction space, this paper proposes a calculation method of the limit line loss rate (LLLR) based on a series of current LLLR indicators. Firstly, according to the difference between different line characteristics and operation parameters of distribution network, the index which has a high impact on the LLLR is screened out, then the clustering index parameters are extracted, and clustering analysis is carried out by using the fuzzy C-means algorithm. According to the clustering results of feeders, the reference feeders are selected for each type of feeders, and the equivalent resistance method is used to calculate the LLLR under the limit conditions. For intra-class non-reference feeders, the adjustment coefficient method is used to calculate the LLLR, and the LLLR is corrected by checking whether the line loss rate obeys the normal distribution. Taking the distribution network data of a city as an example, the LLLR is calculated, which proves the validity of the method.

1. Introduction
Line loss (LL) and line loss rate (LLR) comprehensively reflect the economic operation level of power grid and the management level of power supply enterprises. LL and LLR are closely related to energy saving and emission reduction. How to reduce line loss has economic value and energy saving significance. At present, the mature calculation methods of line loss include root mean square (RMS) current method, average current method, maximum current method, loss factor method, voltage loss method, equivalent resistance method and power flow algorithm \cite{1-4}. The root mean square current method, equivalent resistance method and power flow algorithm are widely used. New line loss calculation methods include genetic algorithm, interval algorithm, clustering algorithm and artificial neural network method \cite{5-8}. Compared with traditional line loss calculation methods, these algorithms have some improvements, and are gradually applied to line loss calculation.

At present, there are few studies on the clustering of lines and the calculation of limit line loss. There are few studies on the application of line clustering to the calculation of LL. In reference \cite{8}, a new method is proposed to apply line clustering to the calculation of line loss by neural network, but the RBFNN model is relatively complex, and the LLR is calculated by power flow algorithm, so it is not easy to be popularized in practical engineering. In reference \cite{9}, the average LLR, standard deviation of LLR and gradient of LLR are used as clustering indicators to cluster feeders. This method only considers LLR and its digital characteristics, and does not calculate LLLR, and does not consider the relevant indicators affecting line loss, so it is not practical. In reference \cite{10}, a series of LLLR and loss reduction rate indicators are defined. However, for large-scale distribution network, the
calculation of limit line loss one by one will cause huge workload and increase the intensity of calculation of line loss.

In summary, a new method for calculating the LLLR of medium voltage distribution network is proposed in this paper. The present feeder is taken as the research object. Considering the differences between the basic parameters of the line and the operation data, the parameters with high impact on line loss are selected from the easily accessible data types. Based on the principle of compactness and separation, the feeder is clustered by fuzzy similarity. After categorizing the feeders, the reference feeders within the class are selected, and the LLLRs of them are calculated by the equivalent resistance method considering the limit conditions. For intra-class non-reference feeders, the LLLR of other feeders is calculated by adjusting the parameters. The LLLR of all feeders in the class is corrected by normal calibration. The LLLR of the distribution network is calculated according to the limit line loss rate of the feeder-like line and the total power of the feeder-like line. Taking a city distribution network as an example, the validity of the algorithm is verified.

2. Limit line loss rate and fuzzy clustering

2.1. Limit line loss rate
For a structure-determined distribution network, under the current load characteristics and load distribution, the line loss under the limit condition, that is, the line loss after adopting all available loss reduction measures, is called the LLLR, and the line loss rate at this time is called the LLLR.

Limit Conditions: Make the parameters affecting line loss reach the extreme value under allowable conditions, and take measures to reduce line loss as far as possible. 1) The maximum allowable for the selection of conductor cross section. 2) Reactive power compensation to the limit. 3) Optimize transformer to limit type (e.g. three-phase 13 three-dimensional winding transformer (S13) or Three-phase 15 three-dimensional winding transformer (S15)). 4) Balancing three-phase load.

2.2. Explanation of limit line loss rate
The LLLR is that under the limit conditions. The LLLR in this paper is the LLLR under the given load characteristics and load distribution conditions.

LL is different under different load curves, and the LLLR is also different. Generally, the load curve of typical day is chosen for calculating LLLR. Generally, the amount of calculation for a line is not large, but for the whole distribution network, the amount of calculation will become huge. Therefore, the clustering algorithm is considered when calculating the LLLR of distribution network, so it is not necessary to calculate the LLLR of each line.

2.3. Fuzzy C-mean clustering (FCM) algorithm
In reference [11], clustering is to discover potential similarity patterns in data without any prior information and classify the data, so that the data classification meets the principle of maximizing similarity within the class and maximizing the difference between different classes. FCM is a clustering algorithm based on data partition. Its goal is to maximize the similarity between objects classified into the same class and minimize the similarity between different classes. The partition of data by FCM is a kind of fuzzy and flexible partition. The characteristics of FCM algorithm are used to cluster feeders in distribution network.

Let the parameter of feeder index be $p$. For a data set with $n$ feeder samples, $X=\{X_1, X_2, \ldots X_n\}$, whose clustering number is $K$.

The objective functions expressed as follow.

$$J_m(U,V) = \sum_{i\in C} \sum_{x_{ij}\in l_i} \mu_{ij}^m d^2 (x_j, v_i)$$  \hspace{1cm} (1)

The constraints of membership degree $u_{ij}$ are as follows.
\[
0 \leq \mu_{ij} \leq 1, 1 \leq i \leq K, 1 \leq j \leq n
\]  
(2)

\[
\sum_{i=1}^{K} \mu_{ij} = 1, 1 \leq j \leq n
\]  
(3)

\[
0 < \sum_{j=1}^{n} u_{ij} < n, i = 1, 2, \ldots, K
\]  
(4)

Where: \( U \) is the membership degree matrix and \( V \) is the cluster center matrix of \( p \times K \). \( u_{ij} \) refers to the membership degree of the \( j \) sample feeder \( x_j \) belonging to the class \( i \); \( K \) refers to the clustering dimension, \( m \) refers to the ambiguity parameter, and \( d^2 (x_j, v_i) \) refers to the Euclidean distance from the sample point \( x_j \) to the center \( v_i \).

After determining the optimal number of clusters, firstly, the FCM algorithm initializes the clustering center, and then updates the membership matrix and clustering center matrix through iteration calculation until the difference between the two iterations of the objective function \( J_m (U, V) \) is within a given range, then stops the iteration calculation and outputs the clustering results.

The initial clustering center expression is as follows.

\[
u_i = \frac{1}{n} \sum_{x_j \in C_i} x_j
\]  
(5)

The updating formulas of membership \( u_{ij} \) and distance center \( v_i \) are respectively as follows.

\[
\mu_{ij} = \frac{1}{\sum_{r=1}^{K} \left[ d(x_j, v_i)/d(x_j, v_r) \right]^{2/(m-1)}}
\]  
(6)

\[
v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{i=1}^{n} \mu_{ij}^m}
\]  
(7)

At the same time, considering the compactness and separation of data partition, the optimal clustering number is determined, and the clustering validity index function is selected, so that the partitioning results meet the principle of intra-class compactness and inter-class separation.

\[
XB(-) = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n} \mu_{ij}^m d(x_j, v_i)}{n \times \min_{i,j} d(v_i v_j)}
\]  
(8)

After clustering calculation, all kinds of reference feeders can be selected according to the principle of maximum membership degree. The feeders corresponding to the highest degree of membership in the final membership matrix are used as reference feeders.

3. Data collection and clustering index selection

The number of distribution network lines is large, and the number of branch lines and components of a single line are also large. Considering the huge amount of data and the existence of large redundancy, firstly, the data types that can be collected are pre-processed to reduce data redundancy, select the indicators that have a high impact on the line loss, and use the principle of correlation to reduce the dimensionality of data, so as to reduce the amount of clustering calculation.

3.1. Data collection and index selection

According to the data collected from the distribution network and considering their influence on the limit line loss, the selected parameters of the distribution network line are as follows: power supply radius, cable rate, distribution capacity, average power, maximum current and maximum load rate.

- Power supply radius. The power supply radius is defined as the line length from the power point to the farthest load point. The length of power supply radius directly affects the equivalent resistance of feeder, and has a high impact on the limit line loss.
• Cable rate. Cable rate is the percentage of cable length to total line length. Because eddy current causes losses in cables, the losses of cables are larger than those of overhead cables. Therefore, the higher the cable rate, the greater the limit line loss.

• Total capacity. The total capacity is the sum of the rated capacity of all distribution transformers on the line. Generally, the larger the total capacity is, the more transformers there are. The equivalent resistance of different types of transformers with the same capacity is different. The economic operation range is slightly different for different iron losses.

• Average electricity. The average annual active power supply reflects the load of distribution network, and the load directly affects the limit line loss.

• Maximum current. The maximum current is the short-term maximum operating current of the distribution feeder without affecting the safety of the equipment. The maximum current is an effective index of the load curve, and the line loss is proportional to the quadratic of the current, so the maximum current has a high impact on the limit line loss.

• Maximum load rate. Transformer loss has copper loss and iron loss, the same capacity transformer, when the load rate is different, iron loss is the same; with the increase of load rate, copper loss increases. The maximum load rate reflects the limit loss to some extent.

3.2. Selection of clustering index parameters

The six indicators in Section 3.1 are closely related to line loss and can be used as line loss rate indicators, but whether they can be used as cluster index parameters needs further analysis.

The selection of clustering index parameters should be related to line loss, and be able to characterize line attributes, meanwhile minimizing data size and dimension. Clustering indices are selected by considering the linear correlation coefficient between line indices and the influence degree of line loss.

The following aspects should also be taken into account in the selection of cluster index parameters:

• Whether the index reflects the basic attributes of the line, so as to distinguish different lines;

• Whether the index can objectively reflect the normal load level of transmission line;

Through the above analysis, the influence degree of line loss is defined to ensure that the indicators involved in the clustering index system are clear and have a greater impact on the line loss rate.

The formula for calculating the correlation coefficient is as follows.

\[ \rho = \frac{E(xy) - E(x)E(y)}{\sqrt{D(x)D(y)}} \]  

(9)

Where \( \rho \) is the correlation coefficient between index parameter \( x \) and \( y \). \( E(x) \), \( E(y) \) and \( E(xy) \) are the mathematical expectations of the product of index \( x \), \( y \) and \( xy \) respectively; \( D(x) \), \( D(y) \) are the variances of index \( x \) and \( y \) respectively. When the correlation coefficient is greater than 0.7, it is considered that there is a high correlation between the two indicators [12].

4. Algorithm design

4.1. Fuzzy C-means clustering algorithm

Figure 1 shows the flow chart of calculating the LLLR by FCM algorithm.
Data collection and processing
Screening Line Loss Rate Indicators
calculate the optimal cluster number
Cluster computation

Yes

No

Output membership matrix
Determine the reference feeder
Calculation and Correction of Limit Line Loss Rate
Calculate the comprehensive limit line loss rate

End

Figure 1. FCM algorithm to calculate the LLLR flow chart.

4.2. Calculation of LLLR of intra-class reference
After clustering the feeder samples with FCM algorithm, all kinds of reference feeders are obtained, then the limit line losses of them are calculated. For the calculation of LLLR of the feeder, the line parameters of the reference feeder are calculated according to the definition of the limit line loss. Before calculating, we restrict the limit parameters in the following aspects:

- According to the definition of LLLR of section, all feeder sections of the feeder are replaced by the largest section type conductors specified by the power supply company.
- According to the definition of LLLR of optimal transformer, the transformer of feeder is replaced by the optimal type transformer specified by the power supply company.
- According to the definition of LLLR of three-phase equilibrium, the calculation condition of LLLR is assumed to be three-phase equilibrium.

4.3. Calculation of LLLR by equivalent resistance method
Considering the characteristics of medium voltage distribution lines and the above definition of limit line loss, the equivalent resistance method is used to calculate the limit line loss. Steps are as follows:

- The network topology is obtained by calculating the equivalent circuit of the limit parameters of the reference feeder, and the typical daily load curve is obtained.
- The loss of reference feeder is calculated by equivalent resistance method. When calculating, the following assumptions are made to limit the load: the load is proportional to the rated capacity of the transformer, that is, the load rate is the same. Load point power factor is the
same, node voltage is the same, voltage drop is not considered.

\[ R_{eq} = \sum S_N_i^2 (R_i + r_0 l_i) \]

\[ \delta_i = \frac{3 I_j f R_{eq} t}{E_l} \times 100\% \]

Where \( R_{eq} \) is the equivalent resistance of feeder; Assuming that the line is divided into \( m \) segments, where \( S_N_i \) and \( S_N \Sigma \) are respectively the rated transformer capacity of the branch line of section \( i \) and total distribution of the line; \( R_i \) is the equivalent resistance of transformer on the branch line of section \( i \); \( r_0 \) is the AC resistance per unit length of the feeder line of section \( i \); \( l_i \) refers to the length of feeder line in section \( i \); \( \delta_i \) is the LLLR, \( I_j f \) is RMS current; \( t \) is the running time; \( E_l \) is the feeder power supply.

- Calculate the LLLR of class-i feeder

For the calculation of other feeder loss in the same cluster, the difference between the index parameters of the same cluster and the reference feeder is used to calculate the LLLR, thus omitting the steps of (1), (2) and simplifying the calculation.

The adjustment coefficients between defined intra-class non-reference feeders and reference feeders can be expressed as

\[ P_{ij} = \sum_{i=1}^{q} \frac{x_{ij} - x_{i0}}{x_{i0}} \]

Where \( q \) is the total number of selected indicators in section 3.1. \( x_{ij} \) is the index parameter value of item \( i \) of the \( j \) feeder in a certain class; \( x_{i0} \) is the index parameter value of item \( i \) of the reference feeder.

The LLLR of a feeder is calculated as follows:

\[ \delta_j = (1 + P_{ij}) \delta_0 \]

\( \delta_j \) is the LLLR of the \( j \) in a cluster feeder, \( \delta_0 \) is the LLLR of reference feeder.

- Correction of the LLLR of a Type of Feeder

In order to reduce the deviation, the LLLR of reference feeder is revised.

According to the statistical theory in reference [13], if some data obey normal distribution, the mean value should be used to describe its overall characteristics; if the data do not obey normal distribution, the mid-value can be used to describe its overall characteristics. Whether the LLLR of all feeders in all kinds of feeders obeys normal distribution is checked. If the distribution of the LLLR of the inner feeder follows the normal distribution, the average value is calculated as the revised LLLR; otherwise, the median value is selected as the revised LLLR.

- Calculate the LLLR of Medium Voltage Distribution Network

The modified LLLR of feeder-like line and power supply quantity are used to calculate the LLLR of medium-voltage distribution network.

\[ \delta = \frac{\sum_{s=1}^{f} \delta_{ls} \cdot E_{ls}}{\sum_{s=1}^{f} E_{ls}} \]

In the formula, \( \delta \) is the LLLR of medium voltage distribution network and \( \delta_{ls} \) is the revised LLLR of \( s \)-type feeder. \( E_{ls} \) is the sum of the power supply of the class \( s \) feeder, \( f \) is the total number of feeder classifications.

5. Example analysis

According to the actual data of 87 10kV lines in a city’s distribution network, 87 lines are the data that have been eliminated lossless lines and total loss lines (power lines) do not meet the requirements. According to the characteristics of the collected data, the LLLR calculated in this paper is the average
5.1. Selection of clustering index parameters

The collected parameters are as follows: power supply radius ($X_1$), average electric quantity ($X_2$), maximum current ($X_3$), cable rate ($X_4$), total capacity ($X_5$) and maximum load rate ($X_6$). The six received line parameters are analyzed and synthesized in this section to get the final clustering index parameters.

According to the calculation results of correlation coefficient in table 1, we can see that the line loss indexes with correlation coefficient greater than 0.7 are $X_2$ and $X_3$, $X_2$ and $X_6$, $X_3$ and $X_6$, that is, the correlation between the indexes is strong, so we should select one of them as the clustering index parameter.

![Table 1. Correlation coefficients of index parameters.](image)

5.2. Fuzzy clustering computing

In this study, the fuzzy clustering calculation is carried out by using MATLAB2016. Firstly, $\text{XB} (-)$ is used to calculate the optimal number of clusters, and the optimal number of clusters is determined to be 6. Then the FCM algorithm is used to cluster the validity of the fuzzy clustering. The maximum number of iterations is 100, the clustering number $k = 6$, the ambiguity $m= 2$, and the change of the objective function value of the iteration termination condition is less than $1\times e^{-5}$. The clustering results are as follows:

In the figure, the abscissa is the power supply radius and the ordinate is the average power. Each point represents a 10 kV line. It can be seen from the graph that the classification results meet the criteria of separation between classes and compactness within classes. The clustering data are shown in table 2.

![Table 2. Statistics of each cluster data.](image)

The classification is clear and the results are of practical significance. The power supply radius and average power consumption of A-type feeder data are minimal, while those of E-type feeder are maximal. Moreover, the influence of power supply radius and average electric quantity on line loss rate is great, so the line loss rate of all kinds of feeders is relatively concentrated.

In the clustering result graph (figure 2), the clustering center usually does not fall at a certain data point, most of the cases are near a certain data point. According to the 6*87 membership matrix in the process, the feeder corresponding to the maximum membership degree is selected as the reference.
feeder of all kinds of feeders, and the LLLR is calculated. The membership degree of all kinds of reference feeders is above 0.94. The specific data are shown in Table 3.

From Table 3, we can see that the membership degree of the E-type reference feeder is 0.9451, and that of the other feeders is above 0.98. Although all kinds of clustering centers do not fall on the feeder data points, the benchmark feeders are very close to the clustering centers. These feeders can be selected as all kinds of benchmark feeders.

5.3. Calculation and correction of LLLR

After selecting the reference feeder, the network topology is drawn according to the actual data, and the line parameters are replaced to the limit parameters. Taking F28 of B-type as an example, the 3×240 cables out of the original line are replaced by 3×400 cables. JKLVJ-3×240 overhead cables are used for the main line, JKLVJ-3×150 overhead cables for the branch line and S15 for the distribution transformer. The topology diagram after replacement is shown in Figure 3. In Figure 3, the transformer capacity units are kVA.

The equivalent resistance method is used to calculate the LLLR of the reference line, and the LLLR is 1.275%. When calculating the line loss rate of other feeders in the class, considering the influence of other parameters on the line loss rate, the LLLR of each line can be obtained by calculating the difference coefficient of the line loss rate parameters such as the average power quantity, power supply radius of the lines in the class as shown in Table 4.

The Lillieforms function is used to test whether the above data satisfy the normal distribution and the loss rate of B-type feeder meets the normal distribution. The average value is taken as the limit loss rate of class B feeder after correction.
Figure 3. B-type reference feeder limit line loss topology.

Table 4. LLLR of each type B feeder.

| line number | LLLR % | line number | LLLR % |
|-------------|--------|-------------|--------|
| F15         | 0.534  | F34         | 1.133  |
| F16         | 0.804  | F35         | 0.738  |
| F20         | 0.933  | F36         | 1.932  |
| F25         | 0.776  | F39         | 1.519  |
| F26         | 0.92   | F40         | 1.529  |
| F27         | 1.268  | F42         | 1.893  |
| F29         | 1.363  | F49         | 1.996  |
| F30         | 0.948  | F56         | 2.447  |
| F31         | 1.275  |             |        |

Similarly, after calculation, the LLLRs of all kinds of reference feeders are shown in table 5. Lilliefors function is used to test whether the LLLR of all kinds of feeders conforms to normal distribution. After testing and analysis, the LLLR of each cluster feeder conforms to the normal distribution. The average value of the LLLR of each cluster feeder is calculated as the revised LLLR of the class feeder, as shown in table 6.

Table 5. LLLR of various types of reference feeders.

| Cluster number | LLLR of each datum Feeder % |
|----------------|----------------------------|
| A              | 0.949                      |
| B              | 1.575                      |
| C              | 1.435                      |
| D              | 2.477                      |
| E              | 1.585                      |
| F              | 1.847                      |

Table 6. LLLR after various feeders are corrected.

| Cluster number | Corrected LLLR % |
|----------------|------------------|
| A              | 1.127            |
| B              | 1.446            |
| C              | 1.343            |
| D              | 2.49             |
| E              | 1.487            |
| F              | 1.958            |
According to formula (14), the LLLR of the distribution network is 1.859%. Under the limit condition, the ideal assumption is adopted when the equivalent resistance method is used to calculate the loss rate of the reference feeder, and the calculation of the limit loss rate is a kind of calculation. According to the clustering results, the correction calculation of the LLLR is carried out. In the process of clustering by fuzzy similarity, there will be some errors in the process, but under the effect of adjusting coefficient correction calculation, the influence parameters of line loss rate should be considered as comprehensively as possible to reduce the existing errors. Because the LLLR is a way of taking the limit of line loss rate, its value can only approximate the actual limit value in the hypothesis as far as possible. Comparing the LLLR with the existing line loss rate data, it meets the requirement of comprehensive line loss rate (deducting lossless line) less than 5% for urban lines of 10 kV and below, which is lower than the current statistical line loss rate of 5.5% and the theoretical line loss rate of 3.2%. It is considered that the LLLR has reached an ideal LLLR to some extent and can be used as the LLLR of the current distribution network.

6. Conclusion
In this paper, the fuzzy C-means algorithm is applied to the calculation of the LLLR of distribution network, and the parameters of power supply radius and average electric quantity clustering index are selected. The parameters that affect the LLLR are fully considered in the calculation of adjustment coefficient. Reducing the calculation amount of the LLLR, calculating the distribution network LLLR under the limit conditions, comparing with some current line loss rate indicators, can reflect the current line loss level and determine the loss reduction space.

References
[1] Wang K and Liu L P 2002 Error analysis of theoretical calculation method of 10 kV line loss J. Elec. Power (J.) 17 251-3
[2] Tang X Y 2014 Research on the calculation method of line loss in distribution network (Hunan, China: Hunan University)
[3] Zhang G Q 2010 Line loss calculation of distribution network (Nanjing, China: Nanjing University of Technology)
[4] Tian J Y and Zhu Y 2010 A method for online calculation of line loss in distribution networks based on CIM CICED 2010 Proceedings pp 1-4
[5] Yuan H M and Guo X Q 2002 A new method for calculating line loss in medium voltage distribution network J. Power Sys. Automat. 26 50-3
[6] Ou Y S and Feng T R 2016 Models for estimating line loss rate of medium voltage distribution network considering feeder clustering characteristics J. Power Automat. Eq. 36 33-9
[7] Cheng S and Liu S 2002 Theoretical calculation of distribution network line loss based on interval algorithm J. Power Sys. Automat. 2 22-7
[8] An M 2006 A new method for calculating distribution network line loss based on dynamic adaptive clustering algorithm RBF network (Tianjin, China: Tianjin University)
[9] Liu T and Yang J F 2013 Cluster-based line loss analysis J. East China Power 41 1452-5
[10] Liu J and Duan J Q 2013 Distribution network limit line loss analysis and loss reduction measures optimization J. Power Sys. Prot & Control 41 27-35
[11] Zhou K L and Yang S L 2014 A review of cluster validity J. Sys. Eng. Theory & Pract. 34 2417-31
[12] Qu F H and Cui G C 2011 Fuzzy clustering algorithm and its application [M] 63-4(Beijing: China National Defence Industry Press)
[13] Fan J C and Mei C L 2010 Data Analysis [M] 1-7 (Beijing Science Press)