Comprehensive Analysis of Power Grid Energy Saving and Loss Reduction Based on Power Big Data Platform

Haiqing Lu¹, Hongyang Huang¹, Jun Wu¹, Feng Chen¹, Chengyu Lu¹, Weiwei Xu¹, Qi Qian*, Pengcheng Wang²

¹State Grid Zhejiang Electric Power Co., LTD. Research Institute, Hangzhou 310014, China
²Nanjing Soft Core Technology Co. LTD, Nanjing 310007, China

*Corresponding author: wangyan@chinasoftcore.com

Abstract. Whether the original power data is accurate and complete is a key factor affecting the accuracy of line loss calculations. This research is based on a big data platform to obtain the distribution network model data, topology data, and operating data required for theoretical line loss calculation and technical high loss analysis. The power system analysis algorithm is combined with the big data analysis algorithm to automatically check the distribution network. To locate the high-loss links of the power grid to assist grid planners in their energy-saving and loss-reducing planning and transformation work.

Keywords: Power big data platform, power grid, energy saving and loss reduction.

1. Introduction
In the era of big data, the importance of power big data is self-evident. The development of economy and people's lives are inseparable from electric power enterprises. Electricity big data is not only related to public data about electricity, but can also be associated with different types of data. The collection and sorting of electric power big data is costly, so it is a necessary choice to fully excavate and rationally use the potential value of electric big data. Big data mining can improve the management, power services and user experience of power companies. With the construction of smart grids, geographic information systems, smart substations, and smart meters have been quickly promoted and applied. At the same time, the 4 types of data center platforms built by the Grid Corporation have accumulated a large amount of relevant data resources, which can lay a foundation for enterprises to efficiently conduct big data analysis. basis. In the context of energy saving and emission reduction, the downward pressure on the growth rate of power sales by power grid companies has increased [1]. Power grid companies must reduce costs and increase efficiency and strengthen the management of line losses. The theoretical line loss calculation of the distribution network is to calculate the power loss generated by each element in the process of electric energy transmission and distribution and the proportion of various losses, and determine the change law of the line loss of the distribution network. Energy saving and loss reduction of the distribution network is a comprehensive economic and technical management work of the power sector. Due to the complexity of the distribution network structure, the diversity of parameters and the imperfect data, as well as the
lack of real-time monitoring equipment, the previous energy saving and loss reduction work relied on manual experience methods to make loss reduction decisions.

In the past technology to reduce line loss work, line loss management personnel generally according to the planning or management experience to summarize suitable for the local power grid to reduce loss method, this method is high accuracy, but also in line with the actual needs, intelligent degree is not good, at the same time, the efficiency is very low, the new Internet technology used in this study, can use machine learning to replace the human brain to do analysis, has the advantages It has a scientific guiding scheme to effectively reduce line loss.

In order to solve this problem, many scientific researchers continue to study the methods of energy saving and loss reduction of distribution network from theory to practice, hoping to develop a new method more suitable for the theoretical line loss calculation of distribution network, and pass the massive distribution network more quickly and accurately. Grid data is used to analyse the key links that cause high loss in the distribution network to meet the analysis and management needs of the power sector's distribution network line loss.

This research is based on a big data platform to obtain the distribution network model data, topology data, and operating data required for theoretical line loss calculation and technical high loss analysis. The power system analysis algorithm is combined with the big data analysis algorithm to automatically check the distribution network. To locate the high-loss links of the power grid to assist grid planners in their energy-saving and loss-reducing planning and transformation work.

2. Data Architecture
Based on the energy management system and PMS production management system, obtain the equipment parameters and topology of the 10 kV piezoelectric network, including the line and equipment model, length, user and transformer side electricity, load, power, voltage, current and other data. Based on the data channel of the electricity consumption information acquisition system, the grid model parameters and real-time operating parameters required for the theoretical line loss calculation are obtained, and the real-time theoretical line loss calculation is performed [2]. As shown in Figure 1, it is a data management system for power grid energy saving and loss reduction.

Figure 1. Data management system for power grid energy saving and loss reduction
2.1. Distribution transformer file data
(1) The basic data of distribution transformer mainly comes from measurement automation system, marketing system and GIS system. Data obtained in the metering automation system: substation number, line number, transformer number, station area number, household number, metering point number. Data obtained in the marketing system: substation number, line number, transformer number, station area number, account number, metering point number. Obtain the transformer ID, transformer name, substation ID, transformer model, and transformer capacity in the GIS system. (2) The operating data of the distribution transformer mainly comes from the metering automation system. Data obtained in the metering automation system: ID, transformer name, time, active power at the first end, reactive power at the first end, and current at the first end.

2.2. Feeder file data
(1) The basic data of the feeder mainly comes from the measurement automation system. Data obtained in the metering automation system: ID, line segment name, voltage level, substation to which it belongs, wire type, line length.
(2) The operating data of the feeder mainly comes from the metering automation system. Data obtained in the metering automation system: ID, line segment name, time, active power at the first end, reactive power at the first end, current at the first end, active end at the end, reactive end at the end, and end current. According to the real-time theoretical line loss calculation results, the high loss cause diagnosis analysis is performed. Find out the problems in power grid planning, marketing management, equipment performance, etc., so as to provide scientific loss reduction methods, so that the ability of analysis and auxiliary decision-making can be fundamentally improved, and the economic benefits of enterprises can be improved.

3. Design Ideas

3.1. Equipment operation monitoring
Based on the theoretical line loss calculation results, the operation of branch lines, line segments, and distribution transformers are monitored, and the power factor, load rate, line loss rate and other equipment operation indicators are monitored and early warning in accordance with management evaluation indicators. Figure 2 shows the line loss power flow calculation ideas.

![Figure 2. Line loss power flow calculation ideas](image)

3.2. Calculation of line loss contribution
Based on the power flow calculation algorithm, calculate the influence of line segments and distribution transformers on the line loss of the feeder, and deeply explore the contribution factors of
each component to the line loss of the entire feeder [3]. The iterative search method of line loss contribution degree based on power flow algorithm uses daily average load to calculate the contribution degree of branch line, line segment, and distribution line loss according to the grid space structure:

Firstly, the contribution degree algorithm of the branch line loss rate.

\[ a = \text{The calculated loss of the 10kV feeder when the branch is removed.} \]
\[ b = \text{The actual loss of the feeder before removal.} \]
Contribution of branch line loss = \(1 - \frac{a}{b}\).

Secondly, the line segment line loss rate contribution degree algorithm.

\[ a = \text{The loss calculated by simulating the line segment R, X, G, and B to 0 and the 10kV feeder.} \]
\[ b = \text{The actual loss of the feeder before removal.} \]
Contribution of line segment line loss = \(1 - \frac{a}{b}\).

Thirdly, the contribution degree algorithm of distribution transformer line loss rate.

\[ a = \text{Simulate the loss calculated by removing the distribution transformer and the 10kV feeder.} \]
\[ b = \text{The actual loss of the feeder before removal.} \]
Distribution line loss contribution = \(1 - \frac{a}{b}\).

3.3. Analysis of Loss Reduction Potential

Taking the large feeder as the analysis object, combined with the big data learning algorithm, analyse the energy saving and loss reduction space and the direction of the loss reduction transformation of the large feeder, and provide reference suggestions for the planning and transformation of the feeder [4]. Based on the big data regression analysis algorithm, the regression algorithm model is trained online, and the feeder’s power supply radius, trunk section, load rate and other indicators are used to predict the theoretical line loss rate of the feeder.

3.3.1. Line loss regression model. Random forest belongs to the Bagging algorithm. In the training phase, the random forest uses bootstrap sampling to collect multiple different sub-training data sets from the input training data set to train multiple different decision trees in turn; in the prediction phase, the random forest makes multiple internal decisions the predicted results of the tree are averaged to get the final result. The RFR implemented in this paper is a packaged combination of multiple binary decision trees [5]. Training RFR is to train multiple binary decision trees. When training a binary decision tree model, it is necessary to consider how to choose segmentation variables (features), segmentation points, and how to measure the quality of a segmentation variable and segmentation point. For the selection of segmentation variables and segmentation points, this implementation adopts an exhaustive method, which is to traverse each feature and all the values of each feature, and finally find the best segmentation variable and segmentation point; The quality of the segmentation variable and the segmentation point is generally measured by the impurity of the node after segmentation, that is, the weighted sum \(G(x, v_j)\) of the impurity of each child node. The calculation formula is as follows:

\[
G(x, v_j) = \frac{n_{left}}{N_x} H(X_{left}) + \frac{n_{right}}{N_x} H(X_{right})
\]

Among them, \(x\) is a segmentation variable, \(v_j\) is a segmentation value of the segmentation variable, \(n_{left}, n_{right}, N_x\) is the number of training samples of the left child node, the number of training samples of the right child node and all training samples of the current node after the segmentation The number. \(X_{left}, X_{right}\) is divided into the training sample set of the left and right child nodes, \(H(X)\) is the function to measure the impurity of the node, and the classification and regression tasks generally
use different impurity functions [6]. The training process of a node in the decision tree is mathematically equivalent to the following optimization problem:

\[
(x^*, v^*) = \arg\min_{x,v} G(x, v)
\]  

That is to find the smallest segmentation variable and segmentation point of G.

\[
G(x, v) = \frac{1}{N_s} \sum_{i \in x} (y_i - \bar{y}_{ag})^2 + \sum_{j \in x} (y_j - \bar{y}_{ag})^2
\]

Considering N times of replacement and repeated sampling, the probability of each sample being selected for each sampling is 1/N, and for N times of sampling, the probability of being selected is

\[
1 - (1 - 1/N)^N 
\]

\[
\Rightarrow 1 - \lim_{N \to \infty} \left(1 - \frac{1}{N}\right) = \left(1 - \frac{1}{e}\right) \approx 0.633
\]

3.3.2. Line loss regression fitting. Enter the power supply radius, wire length, number of distribution transformers, distribution transformer capacity, load distribution and other indicators to predict the theoretical line loss rate.

3.3.3. Diagnosis of high damage causes. Based on the line loss regression model, with the goal of reducing loss, analyse the line loss influencing factors after the treatment of each parameter of the feeder, and locate the crux of high loss [7].

3.3.4. Analysis of loss reduction space. Based on the line loss regression model, simulate the target feeder line loss rate and energy saving and loss reduction space after feeder parameter treatment

4. Data case display
The author uses this processing method in the theoretical line loss calculation and analysis software jointly developed with the Electric Power Bureau, combined with the "network maximum flow method", and has achieved good results. As shown in Table 1 to Table 4 for the data case display.

| Name                              | 1 line | 2 line | 3 line | 4 line | 5 line | 6 line | 7 line |
|-----------------------------------|--------|--------|--------|--------|--------|--------|--------|
| Theoretical line loss rate (%)    | 2.15   | 2.26   | 1.74   | 2.33   | 9.96   | 2.07   | 4.61   |
| Line load rate (%)                | 14.07  | 10.18  | 11.71  | 10.31  | 19.59  | 17.98  | 14.64  |
| Distribution transformer load rate (%) | 10.57  | 7.57   | 13.2   | 20.33  | 38.39  | 0      | 337.69 |
| Length (km)                       | 28479.1 | 15583.1 | 15803.1 | 10354.6 | 8212.7 | 4205   | 10476.5 |
| Number of distribution transformers (sets) | 67     | 48     | 57     | 35     | 17     | 19     | 31     |
| Trunk section (mm2)               | 150    | 150    | 120    | 150    | 150    | 240    | 240    |
| Power supply radius (km)          | 6560.28 | 8633.05 | 7024.5 | 5831.96 | 6594.4 | 3595.5 | 8789.14 |
| Load distribution factor          | 0.55   | 0.45   | 0.58   | 0.51   | 0.59   | 0.66   | 0.79   |
| Target line loss rate (%)         | 1.47   | 1.47   | 1.47   | 1.22   | 1.47   | 1.86   | 0.87   |
### Table 2. Contribution of transformer line loss

| Transformer name | Copper loss (kWh) | Iron loss (kWh) | Copper to iron ratio | Loss rate (%) | Percentage of bus loss (%) | Line loss contribution (%) |
|------------------|------------------|----------------|----------------------|--------------|---------------------------|----------------------------|
| Change 1#        | 22.12            | 19.99          | 1.11                 | 1.55         | 4.43                      | 9.89                       |
| Change 2#        | 22.06            | 19.99          | 1.10                 | 1.08         | 5.37                      | 16.27                      |
| Change 3#        | 22.01            | 19.99          | 1.10                 | 1.46         | 4.98                      | 13.19                      |
| Change 4#        | 5.81             | 5.29           | 1.10                 | 1.05         | 0.91                      | 0.91                       |
| Change 5#        | 21.95            | 19.99          | 1.10                 | 1.19         | 15.05                     | 9.47                       |
| Change 6#        | 21.92            | 19.99          | 1.10                 | 1.20         | 4.28                      | 40.21                      |
| Change 7#        | 2.89             | 2.65           | 1.09                 | 0.35         | 1.46                      | 25.20                      |
| Change 8#        | 11.48            | 10.58          | 1.08                 | 0.68         | 1.62                      | 27.70                      |
| Change 9#        | 2.85             | 2.65           | 1.08                 | 1.00         | 24.97                     | 15.90                      |

### Table 3. Contribution of branch line loss

| Branch name | Load factor | Copper to iron ratio | Power factor | Active loss | Branch line loss rate | Percentage of feeder loss | Line loss contribution |
|-------------|-------------|----------------------|--------------|-------------|-----------------------|--------------------------|------------------------|
| Spur 1      | 2.93        | 0.17                 | 0.91         | 3.24        | 0.15                  | 0.02                     | 3.12                   |
| Spur 2      | 3.26        | 0.17                 | 0.92         | 4.12        | 0.17                  | 0.01                     | 1.30                   |
| Spur 3      | 8.51        | 0.17                 | 0.96         | 10.78       | 0.12                  | 0.03                     | 2.55                   |
| Spur 4      | 6.72        | 0.14                 | 0.82         | 10.92       | 0.16                  | 0.03                     | 37.66                  |
| Spur 5      | 0.70        | 0.13                 | 0.93         | 1.32        | 0.13                  | 0.00                     | 93.11                  |
| Spur 6      | 0.24        | 0.13                 | 0.97         | 0.97        | 0.38                  | 0.00                     | 42.42                  |
| Spur 7      | 12.05       | 0.13                 | 0.97         | 7.86        | 0.15                  | 0.04                     | 32.01                  |
| Spur 8      | 0.80        | 0.12                 | 0.94         | 0.74        | 0.12                  | 0.01                     | 9.55                   |
| Spur 9      | 2.18        | 0.12                 | 0.98         | 13.06       | 0.74                  | 0.06                     | 3.52                   |

### Table 4. Contribution of line loss

| Line segment name | Line loss | Power factor | Line loss rate | Percentage of bus loss | Line loss contribution |
|-------------------|-----------|--------------|----------------|------------------------|------------------------|
| Line segment 1    | 0.00631818 | 0.99603257 | 0.0004429     | 0.0004429              | 0.000138266            |
| Line segment 1    | 0.03060458 | 0.89306508 | 0.0009584     | 0.000676016             | 0.000130236            |
| Line segment 1    | 0.38087993 | 0.99993478 | 0.0064547     | 0.003793198             | 0.00311274             |
| Line segment 1    | 0.10167101 | 0.99893102 | 0.0025008     | 0.00250083              | 0.00125177             |
| Line segment 1    | 0.12093653 | 0.98254101 | 0.0031754     | 0.003174874             | 0.001857277            |
| Line segment 1    | 0.14362079 | 0.98260280 | 0.0037707     | 0.00377039              | 0.002205505            |
| Line segment 1    | 0.14363022 | 0.98266449 | 0.0037706     | 0.003770638             | 0.002205505            |
| Line segment 1    | 0.02691847 | 0.99488912 | 0.0015754     | 0.00058485              | 0.000145064            |
| Line segment 1    | 0.01162998 | 0.99271984 | 0.0003831     | 0.000252681             | 0.00006267             |
5. Conclusions
This paper proposes a new idea to improve the accuracy of power grid line loss calculation. It makes full use of the available grid data, rearranges it, and calculates the line loss according to the continuous load curve to obtain more accurate and reliable results. Through comparison and analysis with conventional line loss algorithms, the effectiveness and practicability of the proposed algorithm are demonstrated. The algorithm has been tested on actual lines and the effect is good.

References
[1] Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., Stevens, S., Zhou, Q., & Prasanna, V. Cloud-based software platform for big data analytics in smart grids. Computing in Science & Engineering, 15(4) (2013) 38-47.
[2] Al-Ali, A. R., Zualkernan, I. A., Rashid, M., Gupta, R., & Alikarar, M. A smart home energy management system using IoT and big data analytics approach. IEEE Transactions on Consumer Electronics, 63(4) (2017) 426-434.
[3] Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., & Chiroma, H. The role of big data in smart city. International Journal of Information Management, 36(5) (2016) 748-758.
[4] Wu, J., Guo, S., Li, J., & Zeng, D. Big data meet green challenges: Big data toward green applications. IEEE Systems Journal, 10(3) (2016) 888-900.
[5] Jin, X., Wah, B. W., Cheng, X., & Wang, Y. Significance and challenges of big data research. Big Data Research, 2(2) (2015) 59-64.
[6] Baccarelli, E., Cordeschi, N., Mei, A., Panella, M., Shojaifar, M., & Stefa, J. Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study. IEEE Network, 30(2) (2016) 54-61.
[7] Anastasi, G., Conti, M., & Di Francesco, M. A comprehensive analysis of the MAC unreliability problem in IEEE 802.15. 4 wireless sensor networks. IEEE Transactions on Industrial Informatics, 7(1) (2010) 52-65.