Health Assessment and Fault Diagnosis of Substation Equipment Based on Digital Twin

Lei Tang¹, Xingde Huang², Chong Zhang³*, Xing He³, Tao Zhu¹, Li Gu¹, Yilun Wan¹

¹ State Grid Shanghai Electric Power Company Pudong power supply company, Shanghai, 200120, China
² State Grid Shanghai Municipal Electric Power Company, Shanghai, 200120, China
³ School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 201100, China

*Corresponding author’s e-mail: zc2640@sjtu.edu.cn

Abstract. The equipment safety of substation is related to the normal operation of substation and even distribution network. In this paper, the condition monitoring and fault diagnosis of key equipment in substation are realized through the introduction of artificial intelligence algorithm and random matrix theory. A large number of PRPD spectrums are obtained by point tracing simulation of the measured PRPD spectra of switchgear. The characteristics of the spectra are designed and extracted to train the PRPD pattern recognition modelling and the accuracy of the test set of the model is 97%. Based on the random matrix theory, whether the oil chromatogram data generated by transformer operation is abnormal is monitored on-line. Through big data analysis, the equipment status is accurately evaluated and the type, position and severity of defects are accurately evaluated. The health assessment and fault diagnosis of substation equipment based on digital twin improves the intelligence and efficiency of substation equipment monitoring.

1. Introduction
As the most basic, key and huge component of energy Internet, substation is one of the solid material foundations of energy Internet. The maintenance and management of substation and its equipment are the daily and core task of substation operation and maintenance. With the development of modern communication and information technology, great changes have taken place in the mode of operation and maintenance and the means of condition monitoring. By installing condition monitoring system in substation equipment, using advanced intelligent monitoring equipment and collecting state information for equipment operation status assessment, hidden dangers of equipment operation can be found and dealt with in time.

Online monitoring can improve the utilization rate of electrical equipment, help to complete the transformation of substation operation and maintenance mode from periodic and preventive maintenance to condition-based maintenance, improve asset management and equipment life assessment, strengthen fault cause analysis and improve fault diagnosis accuracy. In reference [2], a set of station control layer equipment monitoring system is designed to monitor the hardware, software and core business process of the station control layer equipment in real time. In reference [3], a
simulation system based on augmented reality technology is introduced for substation operation and maintenance personnel from both visual and tactile aspects.

Digital twin technology was initially applied to the health maintenance and guarantee of aerospace vehicles. By analogy, the digital twin technology can be applied to the field of substation equipment operation and maintenance, using intelligent sensors, big data analysis and other technologies to build "twin" individuals in the virtual space that are mapped in real time with the real equipment in the substation, make full use of the collected data to guide the equipment operation and maintenance management. Based on big data science and technology such as high-dimensional statistical analysis and artificial intelligence, the potential value of measurement big data is comprehensively and efficiently excavated and the substation fault diagnosis and analysis are assisted.

2. Background of condition-based maintenance of main equipment in Substation

At present, transformer, GIS, switchgear cubicle and other important power equipment are the main objects of live detection test in substation.

2.1. Condition assessment of switchgear cabinet

The existing criterions of switchgear cabinet mainly rely on the measurement of ground wave, ultrasonic wave and gas monitoring quantity. As for the health status, the planned maintenance is still the main method, while the fault diagnosis index depends on a single measurement and a single moment. It fails to consider the impact drill, auxiliary air conditioning, mobile phone signal, base station, high-speed rail, fluorescent lamp and sensor error, so it is difficult to strike a balance between sensitivity and reliability. Mature cases based on high-dimensional data analysis and digital twin have not been reported yet.

2.2. Fault diagnosis of transformer

The insulating components of transformer oil mainly include mineral insulating oil and solid organic insulating material contained in petroleum. During the normal use of transformer oil, the insulating oil and other insulating materials contained in the transformer oil will gradually deteriorate and age, accompanied by a small amount of gases such as hydrogen, ethane, methane, acetylene, ethylene, carbon monoxide and carbon dioxide. Once the internal failure of the transformer, the release of such gases will increase rapidly. At present, the transformer oil chromatogram data is mainly monitored by threshold method, but since there may be abnormal noise in the data collected by the equipment, the threshold method is easy to cause the false alarm. High dimensional statistical index is used to describe the operation state of substation system by probability. Linear eigenvalue statistic (LES) is an important statistical index. Compared with the classical indicators, statistical indicators have a series of advantages:

- compatible with multi time, space and cross functional data, can make full use of data resources;
- pure mathematical steps, not limited to the sampling process and mechanism model, can fuse heterogeneous data by splicing, and connect the power grid by natural decoupling.

For the specific power system, the threshold of LES can be calculated by combining with historical data, that is, when LES is not within the threshold range, the system will lose stability with a certain probability.

3. Spatiotemporal big data and data-driven

Big data mining system includes the basic theories, mathematical tools and processing algorithms involved in the modeling and analysis of high-dimensional data. The difficulty of the implementation is that high dimension, not large data volume, is also the main feature of big data. The high dimension (i.e. multi measurement points) increases the spatial dimensions of the data set, so that the correlation between the variables can be calculated through the high-dimensional statistical analysis, that is, the
high-dimensional statistical information can be obtained. The fusion of high dimension and high density (i.e. high sampling rate) constitutes a high-dimensional spatiotemporal data structure. [6]

### 3.1. Random Matrix Theory (RMT)

RMT takes the matrix as a unit and the spectrum as the main research object in high-dimensional space. Through data modeling and analysis, RMT studies the spatial-temporal correlation of data. The data with measurement error, small perturbation and white noise will show a sort of statistical randomness. However, when there is a signal anomaly source in the system, the operating mechanism of the system will be changed, and the statistical randomness of the data of the system will be broken.

For the matrix with $N$ rows and $T$ columns, $\tilde{X} \in \mathbb{C}^{N \times T}$. The transition matrix, $\tilde{X} \in \mathbb{C}^{N \times T}$ can be obtained by the basic transformation formula (1).

$$\tilde{x}_{i,j} = (\tilde{x}_{i,j} - \mu(\tilde{x}_i)) \times (\sigma(\tilde{x}_i)/\sigma(\tilde{x}_j)) + \mu(\tilde{x}_i) \quad (1)$$

Where $\mu(\tilde{x}_i) = 0$, $\sigma^2(\tilde{x}_i) = 1, i = 1, \ldots, N$. The singular value equivalent matrix of the transition matrix, $X_u = \mathbb{C}^{N \times N}$ can be obtained by the basic transformation formula (2).

$$X_u = \sqrt{\tilde{X}^H \tilde{X}} U \quad (2)$$

Where $U$ is Haar unitary matrix, $X_uX_u^H = \tilde{X}\tilde{X}^H$, the superscript $H$ stands for conjugate transpose. $\tilde{x}_u$ can be obtained according to the common formula (3), and $\sigma^2(\tilde{x}_u) = 1/N$.

$$\tilde{x}_u = x_u / \left[ \sqrt{N} \sigma(x_u) \right], i = 1, \ldots, N \quad (3)$$

The definition of LES is shown in formula (4):

$$\mathcal{N}_N[\varphi] = \sum_{i=1}^{N} \varphi(\lambda_i) \quad (4)$$

Where $\varphi$ is the continuous test function, $\lambda$ is the eigenvalue of $\tilde{X}_u$, $N$ is the number of eigenvalues $\lambda$.

### 3.2. Random forest algorithm

Random forest (RF) is a classical bagging method in ensemble learning. [7] The basic learner of random forest is decision tree, which is a simple algorithm, which has strong interpretability and is in line with human intuitive thinking. This is a supervised learning algorithm based on if then else rule. Random forest is composed of many decision trees, and there is no association between different decision trees. When we carry out the task of classification, each decision tree in the forest will be trained by samples separately. Each decision tree will get its own classification result. Which one of the classification results of the decision tree is the most, then the random forest will take this result as the final result.

### 4. Case studies

#### 4.1. Pattern recognition model of partial discharge in switchgear cabinets

In this paper, four typical partial discharges are studied. The typical defects include suspension discharge, tip discharge, internal insulation discharge and particle discharge. [8] The measured PRPD patterns of various partial discharge types and the PRPD spectrums generated by point tracing simulation are shown in the figure 1, figure 2, figure 3 and figure 4.
Figure 1. The measured and the simulated PRPD spectrum of tip discharge.

Figure 2. The measured and the simulated PRPD spectrum of suspended discharge.

Figure 3. The measured and the simulated PRPD spectrum of particle discharge.

Figure 4. The measured and the simulated PRPD spectrum of hole discharge.
Through the method of generating discharge pulse points by random function and adding noise points, 100 groups of PRPD data are generated for each type of partial discharge. By extracting the characteristics of each phase in the PRPD spectrum, including discharge frequency of each phase, discharge amplitude quartile of each phase (25%, 50%, 75%, 100%), namely a total of five features, the shape of $5 \times 360$ data matrix. The characteristics of every five phases are compressed, that is, the discharge frequency of every five phases are added, and the quartile value of the maximum discharge amplitude of the five phases is selected as the final characteristic value of the discharge amplitude quartile to form a shape of $5 \times 72$ characteristic matrix. Since the input of random forest algorithm is one-dimensional feature vector, so the characteristic matrix is reshaped into a shape of $1 \times 360$ dimensional vector as the input of partial discharge pattern recognition model based on random forest algorithm. The accuracy of the test set of the model is 97%.

4.2. Condition monitoring of transformer oil chromatogram data

A spatiotemporal data matrix with 8 rows and 1186 columns is formed from 1186 sampling values of 8 gases, the oil chromatogram data are shown in figure 5.

Figure 5. Oil chromatogram data
A moving window is used to acquire a shape of $8 \times 42$ data matrix. RMT is introduced to process the oil chromatography, and linear characteristic statistics (LES) is used as the evaluation index to avoid the influence of noise value and monitor the operation of transformer. The obtained LES index is as shown in figure 6. LES curve presents a U-shape in some periods and at the time of occurrence of the abnormal value of LES value, it is found that some gases appear abnormal rise phenomena, as shown in the place circled in red in Fig 4, which mean some potential threats of transformer fault.

![LES curve of oil chromatogram data](image)

Figure 6. LES curve of oil chromatogram data

5. Conclusion
In this paper, the condition monitoring and fault diagnosis of key equipment in substation are realized through the introduction of artificial intelligence algorithm and random matrix theory. Based on simulation technology and random forest algorithm, a partial discharge pattern recognition model of switchgear cabinet is constructed. Based on the random matrix theory, whether the oil chromatogram data generated by transformer operation is abnormal is monitored on-line. To realize the real-time state evaluation and defect dynamic early warning of substation equipment, more data mining models need to be further explored in the future.

Acknowledgments
This report is supported by the project "Health assessment and fault diagnosis of substation equipment based on digital twin" of State Grid Shanghai Municipal Electric Power Company.

References
[1] Du J. Analysis on on-line monitoring of substation electrical equipment [J]. China Equipment Engineering, 2018, 000 (006): 115-116
[2] Liu Z., Jian X., Shi B., et al. Operation and maintenance simulation of intelligent substation equipment based on Augmented Reality Technology [J]. China Southern Power Grid technology, 2019, 013 (011): 69-75
[3] Hou H., Sheng G., Sun Y., et al. Partial discharge location method for substation based on electromagnetic wave signal propagation attenuation model [J]. Journal of electrotechnics, 2014, 29 (6): 326-332
[4] Ling H., Liu J., Li Z., Transformer oil chromatogram analysis and fault diagnosis [J] hydraulic pneumatic and sealing, 2012 (5): 46.50
[5] Xu X., he X., AI Q., et al. Distribution network operation state correlation analysis method based on random matrix theory [J]. Power grid technology, 2016 (phase 3): 781-790
[6] He X., Qiu C., Al Q., et al. Research on spatio-temporal big data of distribution network based on random matrix theory [J]. Power supply and consumption, 2017, 34 (006): 14-19
[7] Wang T., Liu J., Zhu S., Wang Z., Qin H., Zhang Y., Zhang X.. Power system transient stability assessment and emergency control strategy based on random forest [J]. Power grid technology, 2020, 44 (12): 8
[8] Song H. GIS risk assessment method based on partial discharge depth learning [D]. Shanghai Jiaotong University, 2018