Hui WANG, Chengjun HUANG, Linpeng YAO, Yong QIAN, Xiuchen JIANG

Shanghai Jiao Tong University

Application of Gustafson-Kessel clustering algorithm in the pattern recognition for GIS

Abstract. *This paper simulates four typical defects in GIS for PD detection, and uses the pulse, amplitude, phase and number of PD to form the three-dimensional PQN matrix. Based on the PQN, three two-dimensional distributions of Hqmax~Phi, Hqmean~Phi and Hn~Phi can be achieved. Then the new G-K clustering method is introduced to separate the four different defects according to the parameters of Sk, Ku, Pe, Q and CC. At last the comparison with different clustering methods is achieved based on the evaluation criterions which are PC, CE, PI and SI.*

Streszczenie. *W artykule zasymulowano cztery typowe defekty w łącznikach gazowych wynikające z wyładowań niezupełnych oraz wykorzystano amplitudę, fazę i liczbę impulsów do sformowania trzywymiarowej macierzy PQN. Następnie wykorzystano nową metodę klastrową do rozdzielenia różnych defektów na podstawie parametrów Sk, Ku, Pe,Q i CC. Nas™epnie przeprowadzono porównanie różnych metod. (**Zastosowanie algorytmu klastrów Gustafsona-Kessela do rozpoznawania wzoru przy analizie rozdzielnic gazowych)*.

Keywords: gas insulated switchgear (GIS), partial discharge (PD), Gustafson-Kessel (G-K) method, pattern recognition
Słowa kluczowe: rozdzielnice gazowe, wyładowanie niezupełne, rozpoznawanie znaków.

Introduction

Measurement of partial discharge (PD) for gas insulated switchgear (GIS) is an important tool for evaluating the insulating system constructions. Because of the high reliability, low maintenance, and compact size have made the GIS an attractive option in many substations. Although the reason above, it is well known that GIS breakdown is preceded by PD activity inside the GIS chamber, caused by defects such as: (i) protrusion on the high voltage conductor or enclosure; (ii) free moving particles; (iii) floating electrode; (iv) void in the resin spacer and etcetera[1-3].

Thus, it is highly desirable not only to detect the PD but also to identify the pattern of the harmful defects before GIS failure. The pattern recognition of PD can be divided to three partitions: model construction, feature extraction and pattern classification. The pattern can be represented by the three-dimensional distribution which is PQN matrix (discharge amplitude – discharge magnitude – discharge number) and the three two-dimensional functions of the phase angle which are the maximum discharge amount phase distribution $H_{qmax}(\varphi)$, the mean discharge amount phase distribution $H_{qmean}(\varphi)$ and the pulse count phase distribution $H_n(\varphi)$. Then, according to the statistical parameters Skewness (Sk), Kurtosis (Ku), the number of peaks Pe, the corss-correlation factor CC and the discharge factor Q, to feature the fingerprint from the two-dimensional distributions[4-6].

This paper prepares four typical defects in GIS, and uses the IEC 270 method and the ultra high frequency (UHF) sensor to acquire the PD signals emitted from the defects. With calculation of the parameters on both positive and negative half phases from the three two-dimensional distributions, we can achieve 24 statistical fingerprints, and then use the classification algorithm of Gustafson-Kessel (G-K) to separate and cluster the different PD pulses from the four defects[7-10,16].

Defect model and measurement device

Experiments are conducted by using an experimental PD detection system shown in Figure 1, which is consist of transformer, protecting resistor, coupling capacitance, measuring impedance, GIS chamber, UHF sensor, oscilloscope and expert system. The applied 50Hz~220V AC voltage is manually increased to 35kV. And then the PD signals can be acquired by using the IEC 270 and UHF methods[11-12]. Four types of schematic diagram of defects in GIS are designed in Figure 2, and the physical models and the GIS chamber are showed in Figure 3.

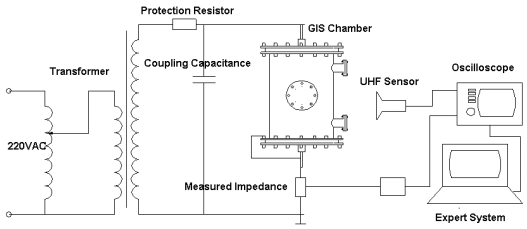

Fig.1. The measurement device in the laboratory

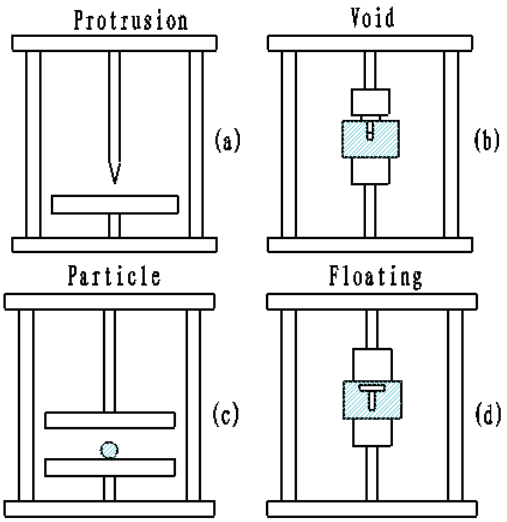

Fig.2. The schematic diagram of four defects. (a) is the protrusion discharge, (b) is the void discharge, (c) is the free particle, (d) is the floating discharge.

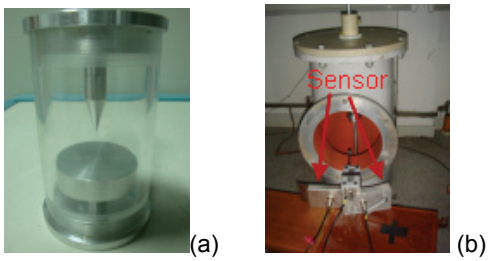

Fig.3. (a) is the physical model of protrusion, (b) is the GIS chamber with UHF sensors, the top is connected with high voltage, the bottom is connected with ground.

Feature extraction and statistical recognition

It is known that a strong relationship exists between the shape of phase-resolved PD (PRPD) patterns and the originating discharge source (type of defects). Each discharge source with its own geometry, location, dielectric properties and applied field, is characterized by a specific sequence of PD. Analysis of these sequence is thus a good means of discrimination between different discharge sources.

To improve the efficiency and accuracy of the PD classification, it is crucial to extract the most effective features from the pre-selected signals to form a three-dimensional PQN matrix which is also feature space. Based on the PQN matrix, we acquire three two-dimensional distributions, which are shown in Figure 4. As shown in this example, there are significant differences between the positive and the negative half of the voltage cycle.

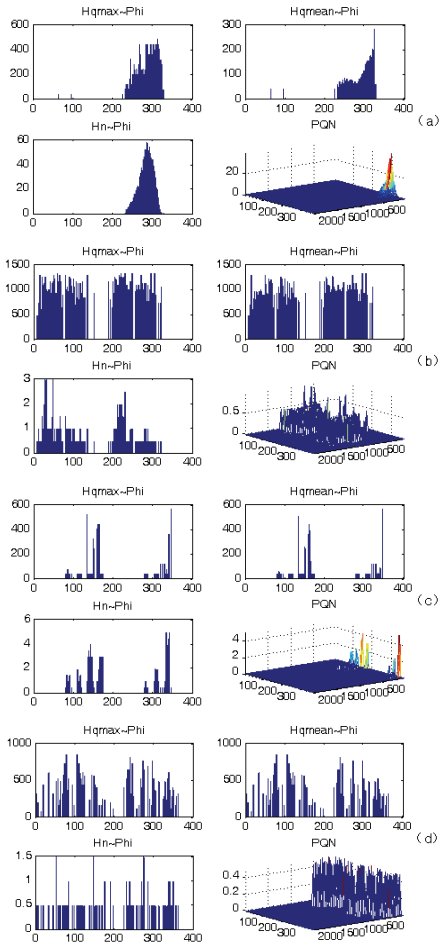

Fig.4. Phase-resolved distributions from the four typical defects. (a) is the protrusion, (b) is the floating electrode, (c) is the void, (d) is the free particle.

As shown in Figure 4, the phase-resolved PD distributions, e.g.: $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$ and $H_n(\varphi)$ of the four defects can be characteristic for the type of a defect. For this reason each distribution is analyzed by means of statistical parameters. To evaluate the shape of $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$ and $H_n(\varphi)$ the following statistical parameters are used to describe the properties of the measurement[8,13].

1. Sk (Skewness), which describes the asymmetry of the $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$ and $H_n(\varphi)$ distributions of each positive and negative cycle.

$$Sk = \frac{\sum (x_i - \mu)^3 f_i}{\sigma^3} \tag{1}$$

2. Ku (Kurtosis), which describes sharpness of the $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$ and $H_n(\varphi)$ distributions of each positive and negative cycle.

$$Ku = \frac{\sum (x_i - \mu)^4 f_i}{\sigma^4} - 3 \tag{2}$$

3. The number of peaks Pe, in order to distinguish between $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$ and $H_n(\varphi)$ distributions of each positive and negative cycle with single top and distributions with several tops.

4. CC (Cross-Correlation), which describes the difference in the distributions of the positive and negative half cycle.

$$CC = \left[\sum x_i^+ x_i^- - \sum x_i^+ \sum x_i^- / \left(N^+ + N^-\right)\right] / \sqrt{\left[\sum \left(x_i^+\right)^2 - \left(\sum x_i^+\right)^2 / N^+\right]\left[\sum \left(x_i^-\right)^2 - \left(\sum x_i^-\right)^2 / N^-\right]} \tag{3}$$

where x_i is a discrete value, μ is the mean value of the distribution and σ is the standard deviation of the distribution. *Sk* is zero for a symmetric distribution, positive when the median is smaller than the mean of the distribution and negative when the mean is less than the median. *Ku* is zero for a normal distribution, positive for a sharper than normal distribution and negative for a smoother than normal distribution.

These statistical parameters are significant for the shape of the distribution, and the discharge patterns of the four typical defects can now be described by the values of the 24 mentioned statistical parameters which are shown in table 1. Because of the recognition process strongly depends on the test conditions, so, we acquire 100 samples and calculate the mean value in order to decrease errors for each defect.

Table.1. Parameters of the two-dimensional distributions of four typical defects

	Parameters	Protrusion	Floating	Void	Particle
Hqmax ~ Phi	Sk+	0	0.33	1.07	0.33
	Sk-	-0.01	0.42	1.15	0.42
	Ku+	-0.02	1.03	1.75	-1.03
	Ku-	-0.01	1.95	3.31s	-0.96
	Pe+	0	0.23	0.2	0.23
	Pe-	0.01	0.22	0.17	0.21
	Q	0.3	1	1.02	1
	CC	0.01	0.35	0.51	0.35
Hqmean ~ Phi	Sk+	0	0.29	1.03	0.3
	Sk-	-0.01	0.36	1.05	0.34
	Ku+	-0.02	1.06	1	-1.07
	Ku-	-0.01	1.1	2.9	-1.02
	Pe+	0	0.33	0.2	0.23
	Pe-	0.01	0.25	0.17	0.22
	Q	0.1	1.01	1	1.01
	CC	0	0.32	0.51	0.32
Hn ~ Phi	Sk+	0	0.66	1.47	0.67
	Sk-	0	0.89	1.64	0.89
	Ku+	0	1.64	2.15	-0.65
	Ku-	-0.01	1,25	2.56	-0.26
	Pe+	0	0.18	0.16	0.18
	Pe-	0.01	0.13	0.11	0.13
	Q	2.28	0.9	1.02	0.9
	CC	0	0.34	0.22	0.34

Gaustafson-Kessel (G-K) clustering algorithm

Gaustafson and Kessel extended the standard fuzzy c-means algorithm by employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set [9, 14]. Each cluster has its own norm-inducing matrix A_i, which yields the following inner-product norm:

$$(4) \qquad D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \ 1 \leqslant i \leqslant c, \ 1 \leqslant k \leqslant N$$

The matrices A_i are used as optimization variables in the c-means functional, thus allowing each cluster to adapt the distance norm to the local topological structure of the data. Let A denote a c-tuple of the norm-inducing matrices: $A = (A_1, A_2 \ldots A_c)$. The objective functional of the G-K algorithm is defined by:

$$(5) \qquad J(X;U,V,A) = \sum_{i=1}^{c}\sum_{i=1}^{N}(\mu_{ik} D_{ikAk}^2)$$

Using the Lagrange multiplier method, the following expression for A_i is obtained:

$$(6) \qquad A_i = [\rho_i \det(F_i)]^{1/n} F_i^{-1}$$

Where F_i is the fuzzy covariance matrix of the i-th cluster by:

$$(7) \qquad F_i = \sum_{k=1}^{N}(\mu_{i,k})^m (x_k - v_i)(x^k - v^i)^T / \sum_{k=1}^{N}(\mu^{i,k})^m$$

Given the data set X, choose the number of clusters 1<c<N, the weighting exponent m>1, the termination tolerance ε>0 and the norm-inducing matrix A. Follow the following four steps:

Repeat for l=1, 2…

Step 1 Calculate the cluster centres.

$$(8) \qquad v_i^{(l)} = \sum_{i=1}^{N}(\mu_{ik}^{(l-1)})^m x_k / \sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m, \quad 1 \leqslant i \leqslant c$$

Step 2 Compute the cluster covariance matrix.

$$(9) \qquad F_i = \sum_{k=1}^{N}(\mu_{ik})_m (x_k - v_i)(x_k - v_i)^T / \sum_{k=1}^{N}(\mu_{ik})^m$$

Step 3 Compute the distances.

$$(10) \qquad D_{ikA}^2 = (x_k - v_i)^T A (x_k - v_i), \quad 1 \leqslant i \leqslant c, \quad 1 \leqslant k \leqslant N$$

Step 4 Update the partition matrix.

$$(11) \ \mu_{ikA}^l = 1 / \left(\sum_{j=1}^{c}(D_{ikA} / D_{jkA})^{2/(m-1)}\right), \quad 1 \leqslant i \leqslant c, \quad 1 \leqslant k \leqslant N, \ until \left\|U^{(l)} - U^{(l-1)}\right\| < \varepsilon .$$

Validity Measures

Cluster validity refers to the problem whether a given fuzzy partition method fits to the parameters of the four typical defects at all. The clustering algorithm always tries to find the best fit for a fixed number of clusters and the parameterized cluster shapes. However, this does not mean that even the best fit is meaningful at all. Either the number of clusters might be wrong or the cluster shapes might not correspond to the groups in the data.

In this paper, we use the G-K algorithm and the fuzzy c-mean method to classify the parameters shows in table.1, and compare the two methods using the validity measures, which are described below【15,16】.

1. Partition Coefficient (PC): measures the amount of overlapping between clusters.

$$(12) \qquad PC(c) = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2$$

Where $\mu_{i,j}$ is the membership of data point k in cluster i.

2. Classification Entropy (CE): measures the fuzzyness of the cluster partition.

$$(13) \qquad CE(c) = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}\log(\mu_{ij})$$

3. Patition Index (PI): is the ratio of the sum of compactness and separation of the clusters.

$$(14) \qquad PI(c) = \sum_{i=1}^{c}\frac{\sum_{j=1}^{N}(\mu_{ij})^m \left\|x_j - v_i\right\|^2}{N_i \sum_{k=1}^{c}\left\|v_k - v_i\right\|^2}$$

4. Separation Index (SI): on the contrary of PI, the SI uses a minimum-distance separation for partition validity.

$$(15) \qquad SI(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2 \left\|x_j - v_i\right\|^2}{N \min_{i,k}\left\|v_k - v_i\right\|^2}$$

Clustering analysis

It has been shown in the previous sections that the PD pulse samples obtained from the four typical defects with the acquisition device are transformed into the statistical parameters. Each kind of defect is consist of 100 samples, after the calculation of PQN matrix and the three two-dimensional distributions (Hqmax ~ Phi, Hqmean ~ Phi, Hn ~ Phi and PQN matrix), the discharge pattern of four particular defects can now be described by the values of the 24 mentioned statistical operators, such as Sk+, Sk-, Ku+, Ku-, Pe+, Pe-, Q and CC.

Figure.5 shows the two defects of floating electrode and void which are clustered by the G-K and FCM method, respectively, using the parameters of Ku+ and Pe+ for the distribution of Hqmax~Phi.

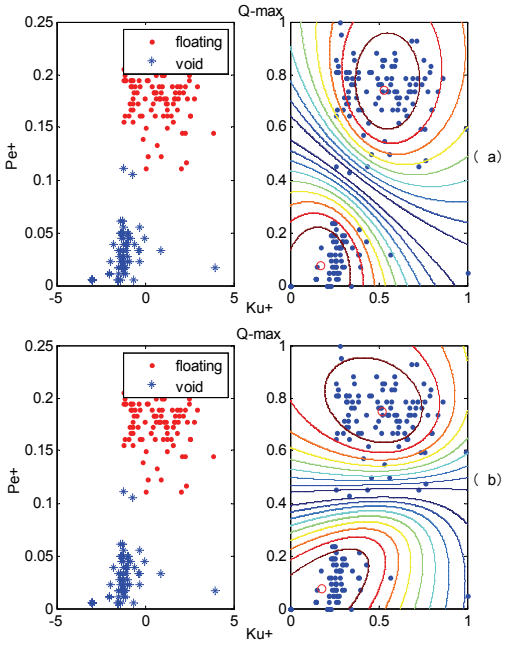

Fig.5. The partition result with the parameters of Ku+ and Pe+ in the Hqmax~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.6 shows the two defects of particle and void which are clustered by the G-K and FCM method, respectively, using the parameters of Sk+ and Pe+ for the distribution of Hqmax~Phi.

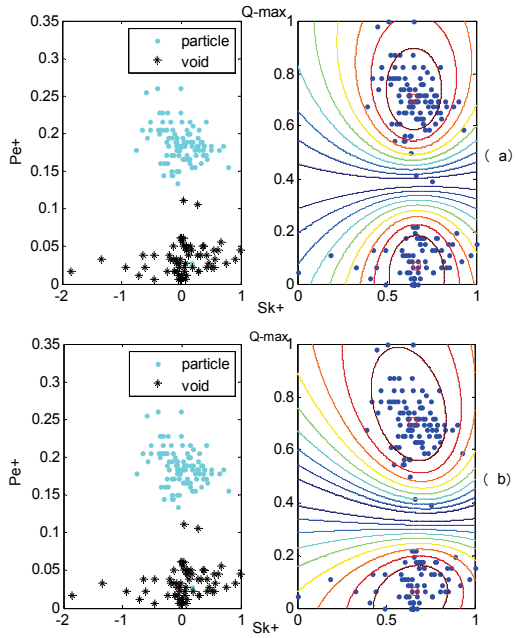

Fig.6. The partition result with the parameters of Sk+ and Pe+ in the Hqmax~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.7 shows the two defects of particle and void which are clustered by the G-K and FCM method, respectively, using the parameters of Ku+ and Pe+ for the distribution of Hqmean~Phi.

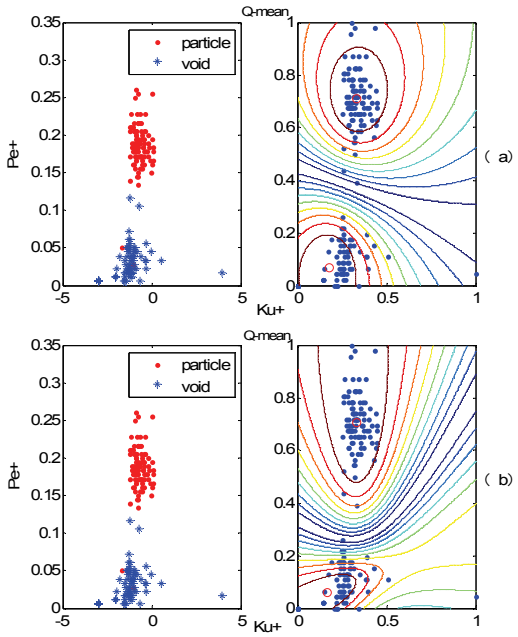

Fig.7. The partition result with the parameters of Ku+ and Pe+ in Hqmean~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.8 shows the two defects of particle and protrusion which are clustered by the G-K and FCM method, respectively, using the parameters of Sk+ and Pe+ for the distribution of Hqmean~Phi.

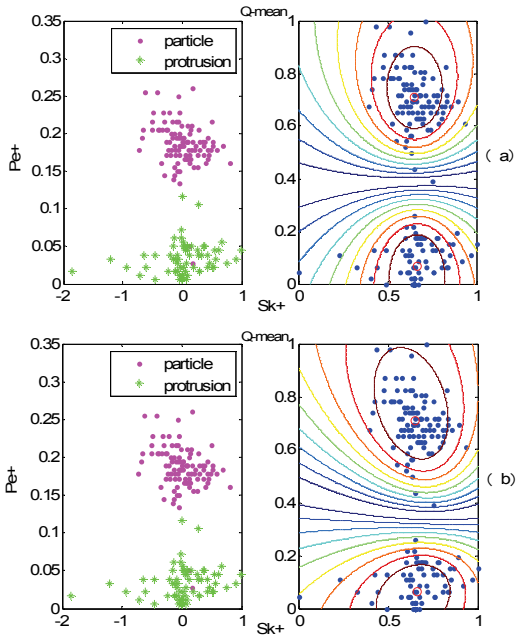

Fig.8. The partition result with the parameters of Sk+ and Pe+ in Hqmean~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.9 shows the two defects of floating and particle which are clustered by the G-K and FCM method, respectively, using the parameters of Sk+ and CC for the distribution of Hn~Phi.

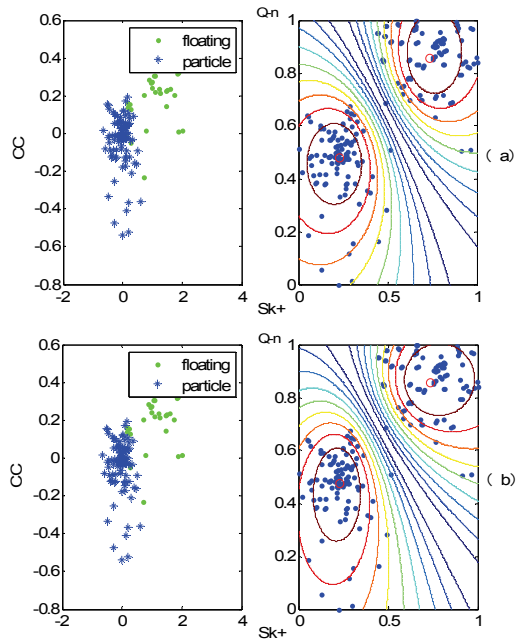

Fig.9. The partition result with the parameters of Sk+ and CC in Hn~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.10 shows the two defects of floating and particle which are clustered by the G-K and FCM method, respectively, using the parameters of Sk+ and Pe+ for the distribution of Hn~Phi.

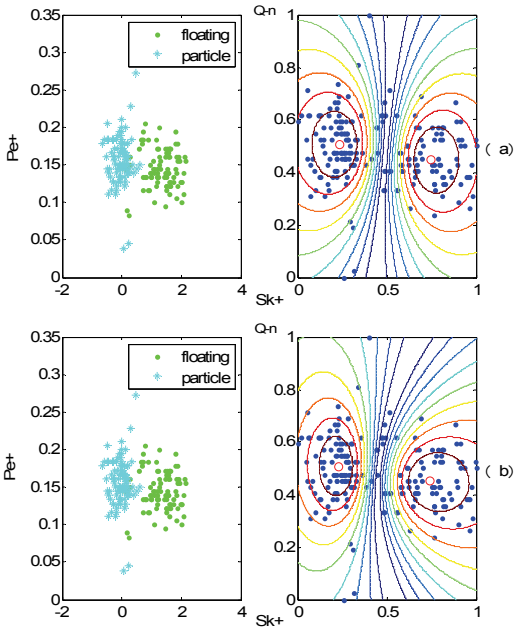

Fig.10. The partition result with the parameters of Sk+ and Pe+ in Hn~Phi distribution. (a) is the result of FCM, (b) is the result of G-K

Figure.5 to Figure.10 are the partition result for the statistical parameters of the three kinds of two-dimensional distributions, which use the FCM and G-K clustering algorithm, respectively, and two groups of different parameters are chosen to verify the goodness for each of the distributions. The classification results show us that both the FCM and the G-k algorithms have wonderful partition effects on the different parameters from the four defect models, and according to the validity measurements, four parameters are calculated which are shown in Table.2.

Comparison with the three parameters CE, PI and SI in Table.2, we find that the result of the G-K algorithm is smaller than the FCM. And on the contrary of parameter PC, the result of the G-K is greater than the FCM, which all confirms the more precision can be achieved by the G-K algorithm to separate the different kinds of PD pulses with different parameters [16].

Table.2. Evaluation criterion of FCM and G-K algorithm

Parameter		Algorithm	PC	CE	PI	SI
Hqmax~Phi	Ku+,Pe+	FCM	0.898	0.192	0.685	0.0034
		G-K	0.928	0.139	0.587	0.0033
	Sk+,Pe+	FCM	0.922	0.151	0.662	0.0032
		G-K	0.938	0.127	0.644	0.0031
Hqmean~Phi	Ku+,Pe+	FCM	0.926	0.149	0.636	0.0035
		G-K	0.934	0.128	0.513	0.0032
	Sk+,Pe+	FCM	0.921	0.155	0.669	0.0033
		G-K	0.936	0.132	0.651	0.0033
Hn ~ Phi	Sk+,C	FCM	0.891	0.202	0.854	0.0043
		G-K	0.895	0.196	0.831	0.0042
	Sk+,P	FCM	0.852	0.257	1.269	0.0063
		G-K	0.857	0.246	1.252	0.0061

Conclusion

Based on the four typical defects in GIS, we use the IEC 270 and the UHF methods to detect and extract the PD pulses. With the statistics of the positive and negative half phase for the maximum discharge amount phase distribution Hqmax(φ), the mean discharge amount phase distribution Hqmean(φ) and the pulse count phase distribution Hn(φ), 24 statistical parameters are acquired which are Sk+, Sk-, Ku+, Ku-, Pe+, Pe-,Q and CC.

According to the statistical parameters, the Gustafson-Kessel (G-K) clustering algorithm is introduced to the partition of the different kinds of PD defect, and with the help of the validity measures which are Partition Coefficient, Classification Entropy, Partition Index and Separation Index, the better classification accuracy is confirmed while using the G-K method than the FCM method.

REFERENCES

[1] Pearson J.S., Farish O., Hampton B.F., Partial Discharge Diagnostics for Gas Insulated Substations, *IEEE Trans. On Dielectrics and Electrical Insulation*, (2) 1995, No. 5, 893-905

[2] Meijer S., Gulski E., Smit J.J. Pattern Analysis of Partila Discharge in SF6 GIS, *IEEE Trans. On Dielectrics and Electrical Insulation*, (5) 1998, No. 6, 830-842

[3] Judd M.D., Farish O., Hampton B.F. The Excitation of UHF Signals by Partial Discharge in GIS, *IEEE Trans. On Dielectrics and Electrical Insulation*, (3) 1996, No. 2, 213-228

[4] Gulski E., Digital Analysis of Partial Discharges, *IEEE Trans. On Dielectrics and Electrical Insulation*, (2) 1995, No. 5, 822-837

[5] Kreger F.H., Gulski E., Krivda A. Classification of Partial Discharge, *IEEE Trans. On Dielectrics and Electrical Insulation*, (28) 1995, No. 6, 917-931

[6] Cavalini A., Montanari G.C. A New Methodology for the Identification of PD in Electrical Apparatus: Properties and Applications. *IEEE Trans. On Dielectrics and Electrical Insulation*, (12) 1998, No. 2, 203-215

[7] Hoshino T., Koyama H., Maruyama S. Comparison of Sensitivity Between UHF Method and IEC 60270 for On-site Calibration in Various GIS, *IEEE Trans. On Power Delivery*, (21) 2006, No. 4, 1948-1953

[8] Judd M., Reid A.,Stewart B. Comparing IEC60270 and RF Partial Discharge Patterns. *International Conference On Condition Monitoring and Diagnosi*, 2008, 89-92

[9] Gustafson D.E., Kessel W.C. Fuzzy Clustering with a Fuzzy Covariance Matrix, *IEEE Conference. On Adaptive Processes*, (17) 1978, No. 1, 761-766

[10] Gath I., Geva A.B. Unsupervised Optimal Fuzzy Clustering, *IEEE Trans. On Pattern Analysis and Machine Inteligence*, (11) 1989, No. 7, 773-780

[11] Cleary G.P., Judd M.D. UHF and Current Pulse Measurements of Partial Discharge Activity in Mineral Oil, *IEE Proceedings-Science, Measurement and Technology*, (153) 2006, No. 2, 47-54

[12] Meijer S., Smit J.J. On-site Sensitivity Verification for UHF PD Dection in GIS. *Proceedings fo the International Conference On Electrical Engineering and Infromatics*, 2007, 30-33

[13] Krivda A. Automated Recognition of Partial Discharge. *IEEE Trans. On Dielectrics and Electrical Insulation*, (2) 1995, No. 5, 796-821

[14] Babuska R., Veen P.J. Improved covariance estimation for Gustafson-Kessel clustering. *IEEE Conference. On Fuzzy Systems*, (2)2002, 1081-1085

[15] Bensaid A.M., Hall L.O., Bezdek J.C. Validity-guided (re)clustering with applications to image segmentation, *IEEE Trans. On Fuzzy object*, (2) 1996, No. 2, 112-123

[16] Bezde J.C. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1981

***Authors**: Dr Hui Wang, Department of Electrical Engineering, Shanghai Jiao Tong University, No. 800 of Dongchuan Rd, Minhang District, Shanghai, CHINA, frankhery@hotmail.com*