USE OF MACHINE LEARNING FOR PARTIAL DISCHARGE DISCRIMINATION

L. Hao P. L. Lewin and S. G. Swingler

University of Southampton, UK

ABSTRACT

Partial discharge (PD) measurements are an important tool for assessing the condition of power equipment. Different sources of PD have different effects on the insulation performance of power apparatus. Therefore, discrimination between PD sources is of great interest to both system utilities and equipment manufacturers. This paper investigates the use of a wide bandwidth PD on-line measurement system to facilitate automatic PD source identification. Three artificial PD models were used to simulate typical PD sources which may exist within power systems. Wavelet analysis was applied to pre-process the obtained measurement data. This data was then processed using correlation analysis to cluster the discharges into different groups. A machine learning technique, namely the support vector machine (SVM) was then used to identify between the different PD sources. The SVM is trained to differentiate between the inherent features of each discharge source signal. Laboratory experiments indicate that this approach is applicable for use with field measurement data.

INTRODUCTION

PD on-line monitoring reveals advantages over conventional PD measurement in many aspects, particularly in terms of monitoring the condition of equipment in service [1]. In practical power systems, more than one discharge source may exist within power apparatus and can be active at the same time. These PD sources can be different discharge types, of the same type but at different locations and of different sizes. Therefore, identification of multiple PD sources is of great importance for health assessment of in-service power assets. Characteristics that represent PD events and sources can be categorized in general into time and frequency domain components. Time domain based methods are suitable for representing the characteristics of a single PD source and type. However, in cases where more than one PD source exists, the obtained results using phase resolved information or pulse sequence analysis (PSA) are of less use for PD type identification. In this case, analysis in the frequency domain using the frequency spectrum and/or wavelet analysis is an effective method for discriminating and locating between different PD sources. The frequency domain is useful in locating and discriminating between different PD sources because the captured signal from the sensor is a convolution of the original signal at the PD source and the transfer function of the equivalent circuit from the source to the coupling sensor. Time domain analysis is an important tool to represent the stochastic, statistical and physical characteristics of the PD event and type. Therefore, a potential approach to discriminate between different PD types, sources or locations is to combine both frequency and time domain analysis [2].

In this paper, the use of a PD on-line condition monitoring system which consists of a wide bandwidth sensor, a digital oscilloscope and a personal computer to assist the automatic PD source identification has been assessed. The obtained raw measurement data were pre-processed using wavelet decomposition. The data obtained from detail level 3 were then processed using correlation analysis. The obtained pre-processed results have then been further analyzed by using accepted approaches, such as phase resolved techniques. A machine learning approach, namely the support vector machine (SVM) was also used to identify the PD sources and results indicate that this approach can identify different PD sources from raw measurement data.

EXPERIMENTAL ARRANGEMENT

In order to generate partial discharge measurement data from different sources and ensure similarity with signals obtained from on-line PD measurements associated with power transformers a simple experiment has been designed. One potential PD measurement point is at the bushing tap of high voltage apparatus such as large auto-transformers [3], [4]. The discharge current flowing to earth can be measured at the bushing tap point using a radio frequency current transducer (RFCT) and this approach has been applied to on-line PD monitoring of power transformers in the field [4].

PD Measurement System

The experimental model is based on models of PD signal sources being coupled to a bushing core bar and the current flowing to earth measured at the tap point using a RFCT as shown in Figure 1.

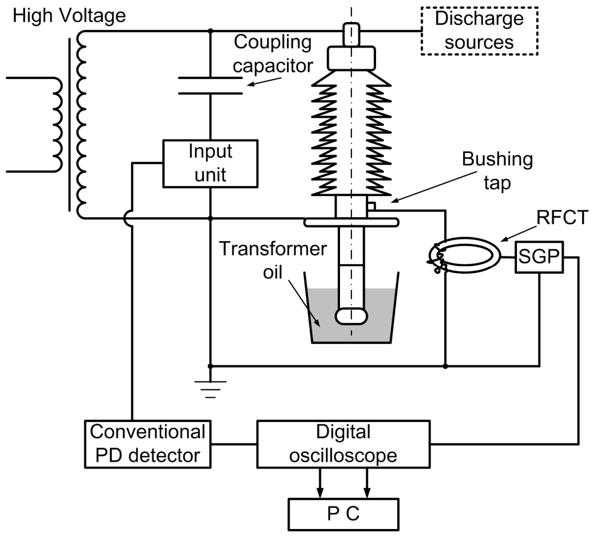


Figure 1 Experiment setup

The 60 kV bushing, model 60HC755, has a 235 pF nominal capacitance, and is PD free under its standard application condition. As a sensor used in this investigation, the clamp-type split core RFCT EMCO model 93686-5 has a measurable frequency range from 10 kHz to 200 MHz. A digital oscilloscope, Tektronix DPO7254 with a bandwidth of 2.5 GHz and 400 MSample memory was used to display, analyse and store the obtained signals. A Robinson conventional PD detector Type 5 Model 700 with 40 kHz - 300 kHz band-pass response was used for calibration for quantifying the apparent charge and generating suitable training data. The trained SVM was then tested using data obtained from the RFCT measurement.

PD Source Models

Three different artificial PD sources were studied: corona discharge with distant earth, surface discharge in air and internal void discharge in oil, sources are as shown in Figure 2. Each artificial PD model has two different arrangements which are used to generate training data and testing data respectively.

Figure 2a illustrates corona in air with distant ground model which is achieved by suspending a piece of thin aluminum wire from the high voltage conductor. By adjusting the length of the wire, different PD inception voltages can be realized. To simulate surface discharge behaviour, a perspex block was inserted between a pair of planar electrodes, the upper electrode was connected to the high voltage power supply, and the lower electrode was grounded, as shown in Figure 2b. To generate testing surface PD data, the upper planar electrode was replaced by a needle electrode. A void of 5 mm(diameter) × 1 mm(depth) was embedded between two pieces of perspex, which was placed between two symmetric planar electrodes. Again the HV source was connected to the upper electrode and the lower electrode earthed. The whole arrangement was immersed in transformer oil, as shown in Figure 2c. A perspex block containing a smaller void of size 2 mm(diameter) × 1 mm(depth) was also used between the two electrodes to generate internal PD data for testing.

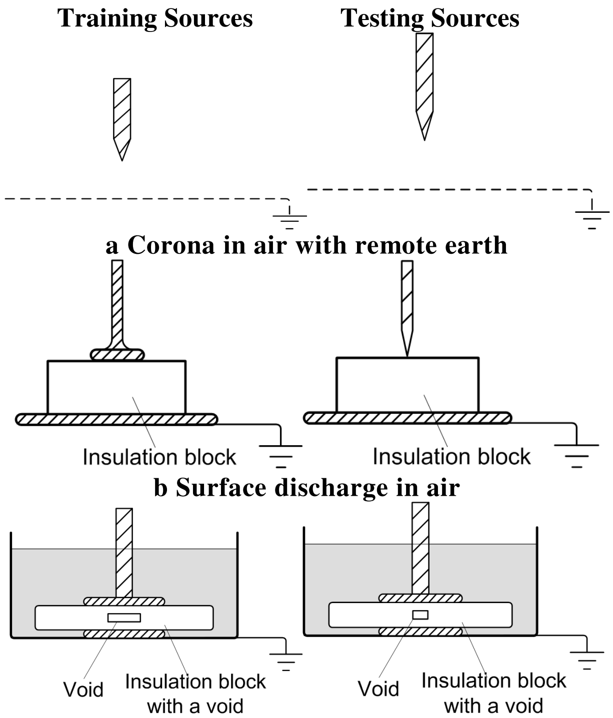


Figure 2 Artificial PD sources

DATA ACQUISITION AND PRE-PROCESSING

Data Acquisition

The signal from the Robinson detector was displayed and sampled via the oscilloscope at 500 kSample/s for 500 power cycles as one acquisition. The output of the RFCT was also connected to the oscilloscope for display and storage. The sampling rate was set to 500 MSample/s to coordinate with the bandwidth of the RFCT for 20 power cycles as one acquisition. Table 1 summarizes the structure of the obtained data.

TABLE 1 Structure of obtained PD data

Sensor	Robinson Detector	RFCT
Sampling rate	500 kSample/s	500 MSample/s
Sampling duration	20 ms	20 ms
Sampling length	10 k points	10 M points
Sampling quantity	500 cycles	20 cycles

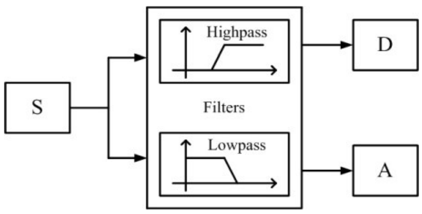
The training data was only obtained from the Robinson detector and each of the three different PD sources were energized at two different voltages. The testing data generated by the other three PD sources was collected via the RFCT and they were tested at two different applied voltages. Table 2 shows the data structure of the training and testing data.

TABLE 2 Summary of training and testing data

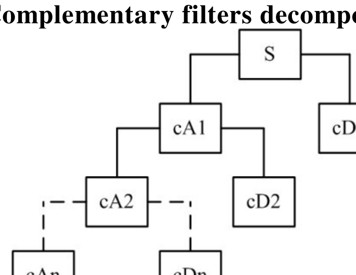
Data type	Training data	Testing data
Sensor	Robinson detector	RFCT
Data quantity	500 cycles × 3 PD sources × 2 applied voltages	20 cycles × 3 PD sources × 2 applied voltages

Data Pre-processing

The raw data from the RFCT stored on computer for each power cycle is approximately 100 Mbytes in size. Therefore, some pre-processing procedures must be undertaken to reduce the dimensionality of the data and recover the useful information. Previous research [5], [6] has shown the advantages of wavelet decomposition coefficients on PD signal analysis in both time and frequency domain. Some successful results have been achieved not only when applied to the simulated data but also when applied to field data [2], [7], [8]. The wavelet decomposition process works like a pair of complementary high-pass and low-pass filters, which decomposes the original signal into series of detail and approximate coefficients respectively, as shown in Figure 3a, where S represents the original signal, D represents the detail decomposition coefficients and A represents the approximate decomposition coefficients.



a Complementary filters decomposition



b Iterative decomposition process

Figure 3 The concept of wavelet decomposition coefficients

As an iterative process, the original signal can be decomposed into different levels and each level is half the bandwidth (sampling rate in frequency domain) and half the length (sample number in time domain) than the above level, as shown in Figure 3b, where cA and cD represent the approximate and detail decomposition coefficients respectively and the number after cA or cD represents the decomposition level.

The "symlet" family of order 7 was chosen as the mother wavelet and detail coefficients of level 3 (referred to as sym7D3) were used as the feature output for further processing since good results in PD denoising have been reported [6], [9]. After processing, the data length was reduced to approximate 1/8 of the original. A peak searching algorithm was used on the pre-processed data to extract useful PD pulse details and record the position of the phase occurrence. Pulses were located by comparing measurements with a threshold value which represents the noise level, i.e. the sensitivity of the measurement system. Figure 4 shows an example of an extracted PD pulse.

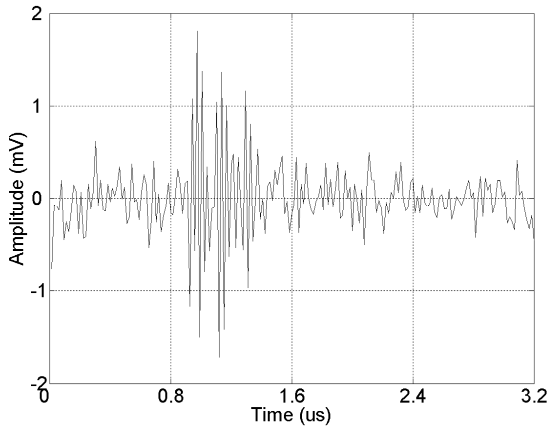


Figure 4 Wavelet decomposition coefficients (sym7D3) of a PD pulse

PD SOURCE IDENTIFICATION

Overview of SVM

As a pattern recognition tool, the support vector machine (SVM) is based on statistical learning theory which has been researched since the 1960s. As an application of statistical learning theory, the SVM was first proposed by V. N. Vapnik in 1995 [10]. This learning machine uses a central concept of SVMs, as well as kernels, for a number of learning tasks [11]. Based on kernel methods, SVMs can be adapted to different tasks and domains by the choice of the kernel function and base algorithm [10]. They represent great advantages in small sample quantity, nonlinear and high dimensionality pattern recognition problems. Successful applications have demonstrated that SVMs can perform as well or better than neural networks in a wide variety of fields, including engineering, information retrieval, and bioinformatics [11].

The SVM is a method for finding functions from a set of labeled training data. Individual sets of measurement data (e.g. discrete PD measurements) can be represented by specific features (e.g. φ-q representation for a PD event). Thus each set of data can be described by a vector whose length/dimension is dependent on the number of features chosen to represent it. The function can be either a classification function or a regression function. SVM earns its name by constructing the solution to the learning problem in terms of a subset of the training data; this subset is referred to as the support vectors (SVs) [12].

Data Normalization

Normalizing or scaling data is very important not only in the application of SVM but also in many other pattern recognition tools such as neural networks. The main advantage of normalization is to avoid attributes existent in greater numeric ranges dominating over those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation using kernel functions. In this investigation, the φ-average q feature vector is normalized to be within the range of 0 and +1.

Kernel Selection

The application of various kernels to PD data has been assessed [13]. It has been found that the Gaussian Radial Basis Function (Gaussian-RBF) kernel is the most effective for PD pulse-like data, where the Gaussian-RBF kernel is defined as:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \quad (1)$$

where $\gamma > 0$ is the kernel parameter controlling the flexibility of classifiers.

Training of the SVM

As summarized in Table 2, the data used to train the SVM were obtained from the Robinson detector. The processed φ q n pattern using the data from Robinson detector has 200 phase windows in one power cycle in order to generate the two dimensional histograms and three dimensional φ-q-n pattern. The selection of 200 phase windows per power cycle is based on the output data characteristic of the measurement system and is a compromise between the output resolution and data processing speed.

Clustering and Identification

The wavelet decomposition coefficients obtained contain both frequency and time domain information and represent characteristics of PD pulses from different sources. Therefore, the wavelet decomposition coefficients (in this paper sym7D3) can be used as a potential feature parameter in distinguishing between different PD sources. While clustering the PD pulses (sym7D3) from the same source into a group the phase occurrence of the pulse is also recorded. After this process, the time domain information that represents stochastic, statistical and physical characteristics of PD types can be also obtained. In this paper, φ-average q results have been obtained. Applying the SVM to the φ-average q results, identification of different PD types is therefore possible.

Clustering using correlation analysis. As a commonly used operator in probability theory and statistics, the correlation coefficient represents the strength and direction of a linear relationship between two variables. In general applications, it can be used to measure the similarity between two different variables. The correlation coefficient R between two n dimensional variables X and Y (referred to x_i and y_i, i=1, 2, …, n) is defined in (2):

$$R\left(X,Y\right) = \frac{\sum x_i y_i - n\bar{X}\bar{Y}}{\left(n-1\right)\sigma_X \sigma_Y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \quad (2)$$

where $\bar{X}$ and $\bar{Y}$ are the means of X and Y, σ_X and σ_Y are the standard deviations of X and Y. The obtained R is in the range of $-1 \le R \le +1$.

The correlation coefficient function used in this investigation can return a matrix of P-values for testing the hypothesis of no correlation. Each P-value is the probability of obtaining a correlation that tends to zero. If P is small, for example less than 0.05, the correlation R is significant having a magnitude of at least 0.95. The P-value is computed by transforming the correlation to create a t statistic having n-2 degrees of freedom, where n is the dimension of the input vector. The confidence bounds are based on an asymptotic normal distribution of (3), with an approximate variance equal to 1/(n-3).

$$0.5\log\frac{1+R}{1-R} \quad (3)$$

Some guidelines for interpretation of a correlation coefficient have been developed. However, it is accepted that all pre-defined criteria are arbitrary and dependent on the specific application. For example, a correlation coefficient of 0.9 may represent a very low correlation but a coefficient of 0.1 in another application may represent a very strong correlation. Therefore, the correlation coefficient used as the criteria to evaluate the correlation between different PD sources must be carefully considered.

By using the non-linear transform, the obtained P is more representative than the correlation coefficient R in this application. One RFCT testing data set containing the three PD sources is used to evaluate the selection of the P value. The energy spectrum of the wavelet decomposition coefficients defined as

$$E_s = S^2 \quad (4)$$

This has been found to be more characteristic for representing the PD activities than the decomposition coefficients themselves. Figure 5 shows the mean energy spectra of the wavelet decomposition coefficients of the three PD sources. The energy unit is in the range of an arbitrary normalized unit. A P value of 0.5, representing a correlation confidence of 50% achieves 100% clustering accuracies for the three PD sources when they are tested individually. Therefore this value may be a suitable threshold for further applications.

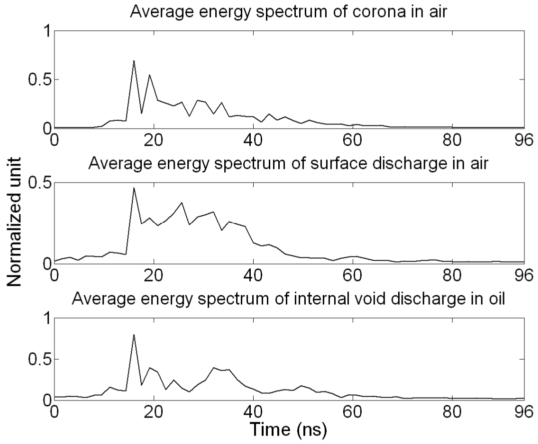
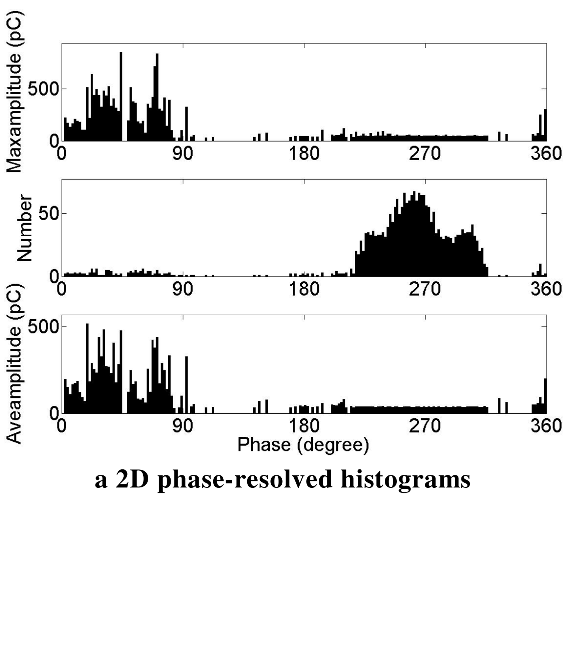
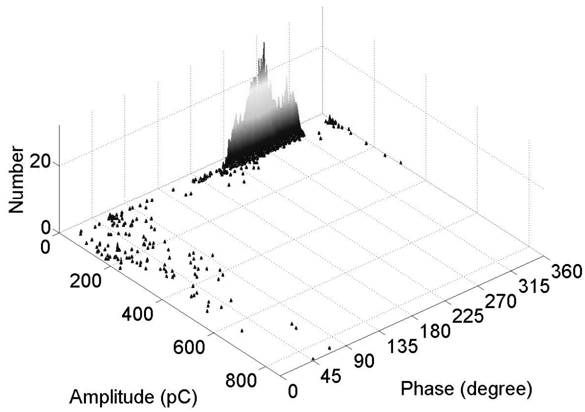


Figure 5 Energy spectra of the wavelet decomposition coefficients

The testing data as summarized in Table 2 are manually combined to simulate the simultaneous multiple PD sources. The obtained processed phased resolved φ-q-n patterns are shown in Figure 6. Before discriminating the multiple PD sources, the well trained SVM was applied to test the single PD sources of each type. The extracted sym7D3 pulses were compared with the reference pulse of each group in sequence using correlation coefficient P value. If the obtained P value is small than the preset expected value (in this case P=0.5) the pulse is categorized to the current pulse group and the reference pulse representing that group is updated by averaging with the new pulse. Otherwise the pulse is categorized as a new group and used as the first reference pulse in that group.



a 2D phase-resolved histograms



b 3D φ-q-n pattern

Figure 6 Phase resolved analysis of simulated three PD sources

SVM identification. The obtained φ-average q information was then used as the feature vector for SVM identification. The SVM was trained using the data set obtained from the Robinson detector. The SVM identification results using correlation coefficient grouped data for single PD source are shown in Table 3. The numbers in the lower left corners represent the cycle numbers of each single PD source for testing. The numbers in the upper right corners are the identified cycle numbers. Each PD source consists of 20 cycles data. For corona discharge, 20 cycles were correctly classified and 1 cycle is misclassified to internal discharge in oil. For surface discharge in air, 23 cycles were identified. Among them, 22 cycles were correctly classified and 1 cycle is misclassified to internal discharge. For internal discharge in oil, all 20 cycles testing data were classified successfully.

Table 3 Correlation analysis based SVM identification results (single source)

Identification type / Testing type	Corona	Surface	Internal
Corona in air	20 **20**	0 0	1 0
Surface discharge in air	0 0	22 **20**	1 0
Internal discharge in oil	0 0	0 0	20 **20**

For multiple PD sources, there are 20 cycles of testing data which consist of three types of different PD: corona in air, surface discharge in air and internal void discharge in oil. Different from the identification result for the unprocessed data without using correlation analysis, which classified the 20 cycles data as surface discharge, the correlation coefficient based SVM identified the correlation analysis grouped data as three types: corona in air, surface discharge in air, internal void discharge in oil, as shown in Table 4.

Table 4 Correlation analysis based SVM identification results (multiple sources)

PD type	Testing cycles	Identification cycles	Weights
Corona in air	20	20	28.2%
Surface discharge in air	20	22	31.0%
Internal discharge in oil	20	29	40.8%

CONCLUSION

The application of SVM for PD identification has been investigated in this paper. A prototype algorithm for multiple PD sources classification has also been developed and assessed.

The feasibility of using a wide bandwidth sensor and a digital oscilloscope equipped with massive storage memory system to detect and analyse partial discharge information has been investigated.

The use of more than one feature parameter (for example phase resolved information and wavelet decomposition coefficients) reveals a good performance for multiple PD source identification. The information in time domain can be used to determine the PD types and the frequency or time frequency domain information can be used to clustering different PD sources.

An abundant database of training samples and proper training processes are both of great importance to SVM based PD identification. An approach using correlation analysis based SVM has been assessed and some satisfactory automatic classification and identification results have also been obtained. However, the performance is restricted by the limitation of correlation analysis and SVM. Therefore, some potential improvements on this method, for example seeking different feature vectors, unsupervised algorithms and improved machine learning techniques could be developed from this initial study in the future.

ACKNOWLEDGEMENTS

This work is funded through the EPSRC Supergen V, UK Energy Infrastructure (AMPerES) grant in collaboration with UK electricity network operators working under Ofgem's Innovation Funding Incentive scheme; - full details on http://www.supergen-amperes.org

References

1 *G. C. Stone, "Partial Discharge Diagnostics and Electrical Equipment Insulation Condition Assessment", IEEE Transactions on Dielectrics Electrical and Insulation, Vol. 12, pp. 891-903 (2005).*

2 *A. Kyprianou, P. L. Lewin, V. Efthimiou, A. Stavrou and G. E. Georghiou, "Wavelet packet denoising for online partial discharge detection in cables and its application to experimental field results", Journal of Measurement Science and Technology, Vol. 17, pp. 2367-2379 (2006).*

3 *J. Fuhr, "Procedure for Identification and Localization of Dangerous PD sources in Power Transformers", IEEE Transactions on Dielectrics Electrical and Insulation, Vol. 12, pp. 1005-1014 (2005).*

4 *B. Han, "Partial Discharge Monitoring of Power Transformers", PhD Thesis, University of Southampton (2004).*

5 *L. Hao, P. L. Lewin and S. J. Dodd, "Comparison of Support Vector Machine Based Partial Discharge Identification Parameters", Proceedings of the IEEE International Symposium on Electrical Insulation (2006).*

6 *L. Hao, P. L. Lewin and S. J. Dodd, "Extraction of PD Signals from an Electro-optic Modulator Based PD Measurement System", Proceedings of the IEEE Conference on Electrical Insulation and Dielectric Phenomena (2006).*

7 *C. S. Chang,, J. Jin, C. Chang, T. Hoshino, M. Hanai and N. Kobayashi, "Separation of Corona Using Wavelet Packet Transform and Neural Network for Detection of Partial Discharge in Gas-Insulated Substations", IEEE Transactions on Power Delivery, Vol. 20, pp. 1363-1369 (2005).*

8 *L. Yang and M. D. Judd, "Recognising multiple partial discharge sources in power transformers by wavelet analysis of UHF signals", IEE Proceedings of Science Measurement and Technology, Vol. 150, pp. 119-127 (2003).*

9 *X. Zhou, C. Zhou and B. G. Stewart, "Comparisons of Discrete Wavelet Transform, Wavelet Packet Transform and Stationary Wavelet Transform in Denoising PD Measurement Data", Proceedings of the IEEE International Symposium on Electrical Insulation, (2006).*

10 *V. N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag (1995).*

11 *B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press (2002).*

12 *J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge Univ. Press (2004).*

13 *L. Hao, P. L. Lewin and Y. Tian, "Partial Discharge Discrimination Using a Support Vector Machine", Proceedings of the 14th International Symposium on High Voltage Engineering (2005).*