

Bahadoorsingh, S. and Rowland, S. M. and Catterson, V. M. and Rudd, S. E. and McArthur, S. D. J. (2010) The role of circumstance monitoring on the diagnostic interpretation of condition monitoring data. In: IEEE International Symposium on Electrical Insulation 2010 (IEEE ISEI), 6-9 June 2010, San Diego, USA.

<http://strathprints.strath.ac.uk/26477/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk>) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge. You may freely distribute the url (<http://strathprints.strath.ac.uk>) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk

The Role of Circumstance Monitoring on the Diagnostic Interpretation of Condition Monitoring Data

S. Bahadoorsingh¹, S. M. Rowland¹, V. M. Catterson², S. E. Rudd² and S. D. J. McArthur²

¹The University of Manchester,
School of Electronic and Electrical Engineering,
PO Box 88, Manchester, M60 1QD, United Kingdom

²University of Strathclyde
Institute for Energy and Environment,
Glasgow, G1 1XW, United Kingdom

Abstract - Circumstance monitoring, a recently coined term defines the collection of data reflecting the real network working environment of in-service equipment. This ideally complete data set should reflect the elements of the electrical, mechanical, thermal, chemical and environmental stress factors present on the network. This must be distinguished from condition monitoring, which is the collection of data reflecting the status of in-service equipment. This contribution investigates the significance of considering circumstance monitoring on diagnostic interpretation of condition monitoring data. Electrical treeing partial discharge activity from various harmonic polluted waveforms have been recorded and subjected to a series of machine learning techniques. The outcome provides a platform for improved interpretation of the harmonic influenced partial discharge patterns. The main conclusion of this exercise suggests that any diagnostic interpretation is dependent on the immunity of condition monitoring measurements to the stress factors influencing the operational conditions. This enables the asset manager to have an improved holistic view of an asset's health.

I. INTRODUCTION

Power systems are evolving. The strong thrust to meet increased energy production targets from renewable and distributed generation sources have seen alternative locations on the ageing power networks [1]. With the gradual rise in electricity demand and technological advancements heralding a steady increase in the number of controllable and efficient consumer power electronic devices, the load characteristics also change. The connection of these non-linear loads may lead to locations of high harmonic content and reduced power quality not experienced previously.

In addition to these concerns, power network asset managers have other new issues including requirements to minimize energy loss, optimizing environmental performance, changing weather patterns and minimizing health and safety risks. In a deregulated competitive market, such considerations must be balanced for optimal returns to shareholders [2]. These mature networks boast a large proportion of the asset base approaching or exceeding the original design-life [1]. Thus, there is an urgent need for improved plant and network diagnostics.

Circumstance monitoring was born from this evolution of the traditional power system. Dynamic electrical, mechanical, thermal, chemical and environmental stress factors influencing the working conditions of the insulation systems have led to questioning of the applicability of current life models and diagnostic assessments. As an example, harmonic influence on partial discharge (PD) activity [3] can lead to possible misinterpretation of the dominant ageing mechanism often derived from artificial intelligence techniques. However, such techniques are developed solely on a database of the unpolluted power frequency. If changes in partial discharge activity are a result of an unmonitored change in power quality (circumstance monitoring), overestimation of the insulation's ageing state will occur resulting in inappropriate asset management decisions taken. The uppermost layer of the developed asset management framework [1] in Fig. 1 illustrates the two essential streams of data which an asset manager must critically assess to determine the state of the insulation (condition monitoring) before making decisions.



Fig. 1. Crucial information for an assets manager acquired from circumstance and conditioning monitoring

This contribution investigates the impact of harmonic influenced electrical treeing partial discharge activity subjected to a series of machine learning techniques.

II. METHODOLOGY

Point-plane LY/HY5052 epoxy resin samples using 3 μm radius conditioned hypodermic needles and a 2 mm gap were employed. The capture of electrical treeing PD activity from a total of 11 samples containing electrical trees has been outlined in [4] using the experimental facility described in

[5]. All samples were subjected to the seven composite waveforms, outlined in Table I at a constant peak voltage of either 10.8 or 14.4 kV peak. φ is defined as the phase difference between the harmonic frequencies and the fundamental (zero in all cases here). K_s is defined as a measure proportional to the rms derivative of the composite waveform [6, 7]. The total harmonic distortion (THD) index [8] is also provided.

TABLE I
PROPERTIES OF THE SEVEN TEST WAVEFORMS [4].

| Waveform ID # | Harmonic Composition + 50 Hz | | | K_s | THD % |
|---------------|------------------------------|--------------------------|-----------|-------|-------|
| | Harmonic Order | % Magnitude per Harmonic | φ | | |
| 1 | 3 | 40.0 | 0 | 1.56 | 40.0 |
| 7 | 1 | 100 | 0 | 1.00 | 0.00 |
| 8 | 5 | 5.00 | 0 | 1.03 | 5.00 |
| 9 | 7 | 5.00 | 0 | 1.06 | 5.00 |
| 11 | 7 | 17.8 | 0 | 1.60 | 17.8 |
| 12 | 5,7,11,13,23,25 | 3.20 | 0 | 1.60 | 7.85 |
| 13 | 5,7,11,13,23,25 | 2.00 | 0 | 1.27 | 5.00 |

These harmonic polluted waveforms defined in Table I can be characterized by a number of attributes, including harmonic sets, presence of particular components and THD. If a machine learning-based classifier can be trained to correctly identify these attributes from PD pattern features, it supports the argument that changes to the pattern are caused by particular effects traceable to the waveform distortion.

This research aimed to train accurate classifiers for identifying certain waveform attributes. The accuracy of each classifier can be used to draw conclusions about the effects of particular attributes, such as the relative importance of particular harmonic orders. This section describes the process of classifier selection and training.

A. Feature Selection

The acquisition hardware [5] records phase position and apparent charge of each PD, grouped into bursts of activity. Each burst may contain a varied number of discharges captured in the cyclical acquisition window of 80 ms. This raw data must be pre-processed into a regularized form known as a feature vector for use by a classifier. Moving from the raw data to a set of pattern features reduces the size of the dataset and the computational complexity of training a classifier. More importantly, it helps avoid the "curse of dimensionality" [9], where the performance of a machine learning technique tends to degrade as the number of features increases. For this reason, the features chosen for the feature vector should characterize the data adequately for the classification problem at hand, while also limiting the size of the feature vector.

The literature on automated analysis of PD data reports a variety of features, which can be grouped into three classes. The most established are statistical features based on the distribution of discharges across the positive and negative voltage half-cycles. This includes mean pulse height, pulse count, and number of peaks in pulse amplitudes [10, 11].

A second approach is to split the voltage cycle more finely. Brown *et al* suggest six phase segments [12]; three per half cycle corresponding to the rising positive segment, the positive peak, the falling positive segment, and falling, peak, and rising segments of the negative half cycle. For each of these phase segments, statistical features of mean, standard deviation, and kurtosis are calculated.

A more recent suggestion defined features based on a human expert's analysis of a PD pattern [13]. Termed descriptors, these features are labels rather than continuous-valued, and include phase position (with sample values 'zero-crossing' or 'peaks') and shape (e.g. 'chopped sine' or 'knife-blade'). Both voltage half-cycles are assigned labels, giving two features per descriptor. A third feature is defined from the descriptor symmetry, which takes one of two values: 'symmetrical' or 'asymmetrical'.

However, bearing in mind the desirability of as small a feature vector as possible, it would be inappropriate to use all reported features for this classification problem. A set of 47 potential features was defined, including statistical features of the half cycles, statistical features of six phase segments, and pattern descriptors. For each classification problem described below, the Information Gain Ratio (IGR) was used to rank the importance of each feature, allowing subsets of informative features to be used as feature vectors, and is calculated by:

$$I(c,f) = [H(c) - H(c|f)] / H(f) \quad (1)$$

For a classification c , a feature f , and H is entropy. The numerator is Mutual Information (MI) between the class and the feature, which is also a standard measure of the importance of features; however MI tends to bias towards features with many possible values. Since the set of 47 potential features contains a wide range of types of feature, from continuous-valued statistics to some descriptors with only two possible labels, MI tends to promote the statistical features over the descriptors. IGR gives better consideration to the discriminating abilities of the descriptors.

Once the features have been ranked according to their IGR, a subset must be selected. In a very few cases a feature would have an IGR of zero, indicating the feature gave no information about the classification problem and could easily be discounted. For example, when identifying the presence of the 5th harmonic, the kurtosis of the positive half cycle has IGR=0. This does not mean that this feature should be permanently discounted, as it has non-zero IGR when identifying the presence of the 7th harmonic. This example highlights the information provided by each feature is problem-specific, and no single representative feature vector can be defined for all PD analysis problems.

As a result, a second stage of feature selection was required. C4.5 trees were trained on subsets of the ranked list of features, with the percentage accuracy of the classifier used to select the best feature set. The C4.5 algorithm [14] is a supervised machine learning technique which generates a decision tree for classification. Each node in the tree is a rule

derived from the training data set, leading to leaf nodes that represent different classes. Once the tree is trained, new data can be classified by traversing the tree.

This meant for each classification problem (i.e. identifying the presence of a specific harmonic order) different groups of top ranked features were tested and compared. If the classifier using all features with $IGR > 0.02$ out-performed all other classifiers, that indicates this feature vector contains the optimal mix of features for this classification problem.

B. Data Mining

Preliminary data mining consisted of training a classifier to identify the waveform ID (see Table I) from individual PD patterns. This classifier would not be particularly useful when assessing new data, such as PD data captured from an in-service cable, since it is unlikely that one of the test waveforms would match exactly the network harmonics. Rather, the aim of training this classifier sought to confirm that waveform attributes can affect the dataset, and therefore further classifier training could proceed.

For the case of waveform ID identification, the classes are the labels 1, 7, 8, 9, 11, 12, and 13. A C4.5 tree was trained using 10-fold cross validation, and Table II shows the resulting confusion matrix. The overall accuracy of the tree was 59.02%.

TABLE II
WAVEFORM ID CLASSIFIER CONFUSION MATRIX

| Waveform ID # | Classified as | | | | | | |
|---------------|---------------|-----|-----|-----|-----|-----|-----|
| | 1 | 7 | 8 | 9 | 11 | 12 | 13 |
| 1 | 968 | 9 | 1 | 9 | 9 | 0 | 3 |
| 7 | 4 | 558 | 144 | 123 | 65 | 72 | 78 |
| 8 | 2 | 149 | 598 | 104 | 22 | 79 | 67 |
| 9 | 13 | 110 | 111 | 565 | 216 | 27 | 52 |
| 11 | 11 | 92 | 44 | 197 | 525 | 70 | 67 |
| 12 | 1 | 84 | 79 | 55 | 79 | 531 | 190 |
| 13 | 4 | 82 | 96 | 73 | 86 | 191 | 532 |

While the final accuracy figure may seem rather low, the confusion matrix allows investigation of the sources of inaccuracy. There is relatively high confusion between waveforms 7, 8, and 9, which represent the waveforms with lowest levels of THD. Confusion also occurs between waveforms 9 and 11, both of which contain the 7th harmonic and fundamental components only. Thirdly, waves 12 and 13 show relative confusion, and both waves contain the same seven harmonic orders. In comparison, waveform 1 is the most accurately identified. It is the only waveform containing the 3rd harmonic, characterized by the highest THD magnitude.

This suggested that individual harmonic polluted waveform attributes each contribute to the distortion of the PD pattern. Since the waveforms with similar levels of THD cause confusion, it is not the set of harmonic orders alone that affects PD. However, confusion occurs between the duplicate sets of harmonic orders, suggesting that THD is not the only influence on PD either.

To conclude, preliminary data mining shows that the PD pattern is affected to some extent by the level of harmonic distortion and the set of harmonic orders present.

C. Classifier training

Three particular attributes of the harmonic polluted waveform were targeted for classification. They were:

- classification of the set of harmonics
- identification of the 5th harmonic
- identification of the 7th harmonic

The harmonic magnitude and THD were not considered for this initial analysis. Classifiers were trained for each of these three tasks. This section reports on the best feature vector for each task, the accuracy of the final classifier, and any conclusions drawn from classifier performance.

1) *Classification of the set of harmonics*: The intended output of this training is a classifier that can identify the set of harmonics polluting the test waveform by considering features of the PD pattern. Specifically, an accurate classifier would associate PD generated under waveform 1 with the 3rd harmonic; waveform 7 with the fundamental frequency; waveform 8 with the 5th harmonic; waveforms 9 and 11 with the 7th harmonic; and waveforms 12 and 13 with the grouping of 5th, 7th, 11th, 13th, 23rd, and 25th harmonics (hereafter labelled 'multiple'). This gives five classes.

The feature selection procedure resulted in a vector of 38 features, comprising a mixture of all three types of feature. A C4.5 tree was trained using 10-fold cross validation, giving an overall accuracy of 77.5%. The confusion matrix is shown in Table III, showing a strong diagonal indicating an accurate classifier. Of each set, a classification of the fundamental frequency is inaccurate most often.

TABLE III
HARMONIC SET CLASSIFIER CONFUSION MATRIX

| Harmonic Set | Classified as | | | | |
|-----------------|-----------------|-------------|-----------------|-----------------|----------|
| | 3 rd | Fundamental | 5 th | 7 th | Multiple |
| 3 rd | 971 | 7 | 0 | 10 | 4 |
| Fundamental | 9 | 613 | 158 | 89 | 149 |
| 5 th | 2 | 165 | 687 | 50 | 92 |
| 7 th | 11 | 75 | 41 | 1687 | 235 |
| Multiple | 4 | 152 | 78 | 265 | 1537 |

2) *Identification of the 5th harmonic*: Determining the presence of the 5th harmonic is a binary classification task, where an accurate classifier would give a positive response for data from waveforms 8, 12, and 13; and a negative response for data from waveforms 1, 7, 9, and 11. In contrast to the previous classifier, the accuracy of this classifier should indicate whether or not there is some particular change in the PD pattern caused by the 5th harmonic only; i.e. the presence of the 5th harmonic can be identified amongst multiple harmonic orders and where it exists in isolation, and its absence can be determined from the fundamental and other harmonic order waveforms.

A feature vector of 37 features was most accurate, and the resulting C4.5 tree is 83.8% accurate. The confusion matrix

(Table IV) shows approximately equal numbers of false positives and false negatives. These results highlighted an increased asymmetrical distribution of incorrect classifications (Table V). Most errors occurred when diagnosing the absence of the 5th harmonic from the fundamental wave.

TABLE IV
PRESENCE OF 5th HARMONIC CLASSIFIER CONFUSION MATRIX

| Presence | Classified as | |
|----------|---------------|------|
| | False | True |
| False | 3482 | 577 |
| True | 570 | 2462 |

TABLE V
ERRORS IN 5th HARMONIC CLASSIFIER

| Waveform ID # | Description | Misclassifications | Percentage |
|---------------|----------------------|--------------------|------------|
| 1 | NO 5 th | 2 | 0.20% |
| 7 | NO 5 th | 306 | 29.3% |
| 9 | NO 5 th | 147 | 13.4% |
| 11 | NO 5 th | 122 | 12.1% |
| 8 | 5 th only | 201 | 19.7% |
| 12 | Multiple | 163 | 16.0% |
| 13 | Multiple | 206 | 19.4% |

3) *Identification of the 7th harmonic:* As with the 5th harmonic, this is a binary classification task. The expected outcome is for waveforms 9, 11, 12, and 13 to give a positive response; and waveforms 1, 7, and 8 to give a negative response. Feature selection highlighted 25 features as the optimal set: significantly smaller than for either of the previous classification tasks. After training using 10-fold cross validation, a C4.5 tree had 90.9% accuracy. The confusion matrix is shown in Table VI, showing approximately equal numbers of false positives and negatives. More detailed inspection of the incorrect classifications shows the highest accuracy from waveforms 1 and 11, then 9 and 12, and less certainty from waveforms 7, 8, and 13 (Table VII). These latter sets of waveforms have lowest levels of THD, suggesting that the presence or absence of the 7th harmonic is easier to determine at higher levels of wave distortion.

TABLE VI
PRESENCE OF 7th HARMONIC CLASSIFIER CONFUSION MATRIX

| Presence | Classified as | |
|----------|---------------|------|
| | False | True |
| False | 2680 | 326 |
| True | 322 | 3763 |

TABLE VII
ERRORS IN 7th HARMONIC CLASSIFIER

| Waveform ID # | Description | Misclassifications | Percentage |
|---------------|----------------------|--------------------|------------|
| 1 | NO 7 th | 12 | 1.20% |
| 7 | NO 7 th | 194 | 18.6% |
| 8 | NO 7 th | 120 | 11.8% |
| 9 | 7 th only | 89 | 8.14% |
| 11 | 7 th only | 19 | 1.89% |
| 12 | Multiple | 70 | 6.87% |
| 13 | Multiple | 144 | 13.5% |

III. DISCUSSION OF RESULTS

The results of this investigation are promising, showing classifier accuracies of 77.5% to 90.9%. This suggests there is enough regularity in the PD patterns to identify attributes of the harmonic content of the voltage waveform, and therefore different attributes of the resultant harmonic influenced waveforms cause particular changes to the behaviour of PD. Further research attempted to improve the accuracy of attribute identification by investigating other classifier techniques. In particular, recent research on the use of Support Vector Machines (SVMs) for PD data analysis has suggested accuracies exceeding 99% are possible [15]. On the other hand, comparative research into ship-board fault diagnosis suggests SVMs did not perform better but in some cases more poorly than C4.5 [16]. SVMs were explored as an alternative to C4.5 for two of the three classification problems described above: harmonic set classification and identification of the 7th harmonic. While accuracies could be improved by a few percentage points, extremely good results could not be achieved. This section describes the approach to SVM training, and possible reasons for the disparity in the literature.

A. Training

A Support Vector Machine is a supervised machine learning technique for classification [17]. It works by converting input data into higher dimensional space using a particular function called a kernel function, with the aim of separating different classes by a hyperplane in the higher dimensional space. The learning task is to find the hyperplane between classes. While there is no definitive methodology for appropriate training of SVMs, there are some generally agreed best practices to follow [18]. For a classification task that may be non-linear, the Radial Basis Function (RBF) kernel is recommended. This is parameterized by two values: the error cost C , and the RBF width parameter, γ .

Before training, input data must be normalized and appropriate values for C and γ chosen. There are no heuristics reported for selecting these values, therefore an exhaustive search technique called grid-search is recommended [18]. This involves training SVMs using all pairs of values of C and γ between certain limits (rough grid search); finding the area of parameter space that shows most accuracy; then training SVMs using pairs of more finely-grained values within that region (fine grid search).

For rough grid search, the search space was:

$$C = 2^{-5}, 2^{-3}, \dots, 2^{15}; \gamma = 2^{-15}, 2^{-13}, \dots, 2^3$$

Once the SVM with highest accuracy was determined, that $\langle C=2^c, \gamma=2^g \rangle$ pair was the source of fine grid search:

$$C = 2^{(c-0.5)}, 2^{(c-0.25)}, 2^c, 2^{(c+0.25)}, 2^{(c+0.5)}, \\ \gamma = 2^{(g-0.5)}, 2^{(g-0.25)}, 2^g, 2^{(g+0.25)}, 2^{(g+0.5)}$$

The most accurate SVM resulting from fine grid search became the final classifier.

Classification of the harmonic set found the best parameters to be $C=2^{12.25}, \gamma=2^{-3.75}$. The overall accuracy of the classifier was 81.10%, giving an improvement of 3.6% over C4.5. For

identification of the presence of the 7th harmonic, the best parameters were found to be $C=2^{4.75}$, $\gamma=2^3$. The overall accuracy was 91.21%, an improvement of 0.31% over C4.5.

B. Results

Clearly, the accuracies of SVM and C4.5 for this study are very similar. This concurs with the result reported in [16] for network fault diagnosis. Given there is little difference in accuracy, C4.5 is preferable for practical reasons. First, the process of grid search to parameterize the SVM takes significantly longer than setting up a C4.5 training run. Secondly, the rules induced by C4.5 are human-readable and can be studied for insight into the classification problem, whereas the support vectors resulting from a trained SVM are a comparatively obscure coding of the knowledge. However, the disparity between accuracy reported in [15] and results given here merits further discussion. Three possible sources of difference are considered below.

The classification problem addressed is different in both cases. Despite both utilizing PD data, this research attempts to identify harmonic attributes whereas [15] identifies defect type. Harmonic attribute identification is significantly different and perhaps more challenging than defect type identification. However, other research into defect classification by machine learning does not suggest it is a particularly easy task [19]. The sizes of training data sets are different in each case. While [15] uses 30 examples of five defects at three different voltage levels, giving datasets of 90 patterns per class, the research here uses between 999 and 1094 patterns for each test waveform. This order of magnitude difference may result in a more diverse set of patterns, increasing classification difficulty, but potentially more realistic. It is also possible that the cross validation training coincidentally selected a favourable test set with significantly higher accuracy than other sets, which is due to a quirk of the data rather than a facet of the technique.

Thirdly, different feature vectors are used for classification in each case. [15] considers features relating to phase position, frequency, and wavelet decomposition of the PD pattern, and report the interesting result that all three sets perform equally well on cross validation testing, but when considering generalization to another voltage level, phase-related features perform most poorly. The research reported here focuses on phase-related features, however cross validation is used for training and testing, and therefore this generalization concern should not be an issue.

V. CONCLUSIONS

Acceptable classifier accuracies were achieved with C4.5 and SVM techniques for identifying harmonic attributes from harmonic influenced electrical treeing PD patterns. This provides encouraging evidence to further investigate avenues of identifying the harmonic content present in the excitation voltage. Future testing employing insulation defects with decreased ageing state variation (e.g. voids) applied to different machine learning techniques is essential to determine the most appropriate. Thus diagnostic

interpretation of condition monitoring data with increased immunity to the stress factors influencing the working environment will improve the asset manager's holistic view of an asset's health.

ACKNOWLEDGMENT

This work was 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; - details on <http://www.supergen-amperes.org>

REFERENCES

- [1] S. Bahadoorsingh and S. M. Rowland, "A framework linking knowledge of insulation ageing to asset management " *IEEE Elect Ins Mag*, vol. 24, pp. 38-46, 2008.
- [2] M. Shahidehpour and R. Ferrero, "Time management for assets: chronological strategies for power system asset management," *IEEE Power and Energy Mag*, vol. 3, pp. 32-38, 2005.
- [3] B. Florkowska, "Analysis of partial discharge mechanisms and their patterns for diagnosis of insulation systems," in *IEEE CDPADM*, 1997, pp. 279-282.
- [4] S. Bahadoorsingh and S. M. Rowland, "Application of a hyperbolic function to model harmonic influenced electrical treeing partial discharge patterns," *IEEE Trans on DEI - submitted*, 2009.
- [5] S. Bahadoorsingh and S. M. Rowland, "Investigating the influence of the lubricant coating on hypodermic needles on electrical tree characteristics in epoxy resin," *IEEE Trans on DEI - submitted*, 2009.
- [6] G. C. Montanari and D. Fabiani, "The effect of non-sinusoidal voltage on intrinsic aging of cable and capacitor insulating materials," *IEEE Trans on DEI*, vol. 6, pp. 798 - 802, 1999.
- [7] G. C. Montanari and D. Fabiani, "Searching for the factors which affect self-healing capacitor degradation under non-sinusoidal voltage," *IEEE Trans on DEI*, vol. 6, pp. 319-325, 1999.
- [8] IEEE, "Std 519-1992, IEEE recommended practices and requirements for harmonic control in electrical power systems."
- [9] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
- [10] E. Gulski and A. Krivda, "Neural Networks as a Tool for Recognition of Partial Discharges," *IEEE Trans on EI*, vol. 28, pp. 984-1001, 1993.
- [11] S. M. Strachan, S. Rudd, S. D. J. McArthur, M. D. Judd, S. Meijer, and E. Gulski, "Knowledge-based diagnosis of partial discharges in power transformers," *IEEE Trans on DEI*, vol. 15, pp. 259-268, 2008.
- [12] A. J. Brown, V. M. Catterton, M. Fox, D. Long, and S. D. J. McArthur, "Learning Models of Plant Behaviour for Anomaly Detection and Condition Monitoring," in *Int Conf on Intelligent Systems Applications to Power Systems*, 2007, pp. 1-6.
- [13] S. Rudd, S. D. J. McArthur, and M. D. Judd, "A generic knowledge-based approach to the analysis of partial discharge data," *IEEE Trans on DEI*, vol. 17, Issue 1, pp. 149-156, 2010.
- [14] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [15] L. Hao, P. L. Lewin, and S. J. Dodd, "Comparison of support vector machine based partial discharge identification parameters," in *IEEE ISEI*. Toronto, Canada, 2006, pp. 110-113.
- [16] A. J. Mair, E. M. Davidson, S. D. J. McArthur, S. K. Srivastava, K. Schoder, and D. A. Cartes, "Machine learning techniques for diagnosing and locating faults through the automated monitoring of power electronic components in shipboard power systems," in *IEEE Electric Ship Technologies Sym*, 2009, pp. 469-476.
- [17] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [18] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," in <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [19] S. D. J. McArthur, S. M. Strachan, and G. Jahn, "The design of a multi-agent transformer condition monitoring system," *IEEE Trans on Power Sys*, vol. 19, pp. 1845-1852, 2004.