Identifying Harmonic Attributes from On-line Partial Discharge Data

V. M. Catterson, *Member, IEEE,* S. Bahadoorsingh, *Member, IEEE,* S. Rudd,
S. D. J. McArthur, *Senior Member, IEEE,* S. M. Rowland, *Senior Member, IEEE*

Abstract—Partial discharge (PD) monitoring is a key method of tracking fault progression and degradation of insulation systems. Recent research discovered that the harmonic regime experienced by the plant also affects the partial discharge pattern, questioning the conclusions about equipment health drawn from PD data. This paper presents the design and creation of an on-line system for harmonic circumstance monitoring of distribution cables, using only PD data. Based on machine learning techniques, the system can assess the prevalence of the 5th and 7th harmonic orders over the monitoring period. This information is key for asset managers to draw correct conclusions about the remaining life of polymeric cable insulation, and prevent overestimation of the degradation trend.

Index Terms—Monitoring, Dielectric breakdown, Partial discharge, Power quality, Epoxy resin insulation, Power system harmonics.

I. Introduction

CONDITION monitoring has the potential to pre-empt catastrophic failure of plant by identifying insulation degradation trends, thus allowing the planning of appropriate maintenance and replacement. One approach is to monitor partial discharge (PD) behavior, which allows the diagnosis and tracking of defects through changes to the pattern of PD activity. Recently it has been determined that power quality can also affect PD behavior, questioning the conclusions drawn about insulation health from field PD data captured when power quality is unknown. A technique is needed to separate power quality effects from insulation aging effects within the data.

Defects in cable polymeric insulation weaken the insulation's dielectric strength favoring PD initiation. Sustained PD activity may eventually lead to breakdown and is a common progressive indicator of insulation health.

Over the years, much research has investigated the potential for diagnosing defect types and tracking the degradation trend using PD data. With an accurate analysis of the PD activity, the progress towards full breakdown can be used to decide the optimal time to remove the asset from service, maximizing asset use while mitigating risk.

Particular harmonic orders, notably the 5th and 7th, can occur in distribution networks with high enough amplitude

to increase the peak voltage. Not only can this initiate PD activity early and more rapidly age insulation systems, but the distorted voltage waveform has been shown to influence the PD pattern characteristics [1], [2]. Asset managers may be attributing PD changes to fault progression, which may truly be due to changes in the harmonic regime experienced by the plant.

This paper describes the design and implementation of an on-line system for harmonic circumstance monitoring of distribution cables. Circumstance monitoring supplements condition monitoring, which assesses the current state of health of the plant. Specifically, this paper presents the training and testing of machine learning techniques for identifying the presence of the 5th and 7th harmonic orders purely from PD data. This environmental information allows for a better understanding of apparent trends in plant condition, and changes in the operating environment can be separated from true changes in plant health, ultimately giving a better estimate of the risk of continued service.

The following two sections give background on harmonics and PD, respectively. Section IV summarizes the methodology for designing the harmonic monitoring system, followed by sections on the selection of machine learning techniques, experimental set-up and data capture, training of harmonic order classifiers, and testing with new data. Section IX presents the software design and implementation of a practical field system for harmonic analysis using PD data. Finally, Section X concludes the paper.

II. Harmonics on the Distribution Network

Distribution companies have an interest in power quality, and in particular in meeting the limits prescribed by various national and international standards (e.g., IEEE 519, G5/4, IEC 1000-2-2, IEC 1000-3-6). In recent years, as the number and types of non-linear loads on the system have increased, harmonic distortion has become more likely to occur [3]. Harmonics in distribution systems are primarily generated by non-linear loads in end-user facilities, which act as harmonic current sources injecting harmonic currents at their point of connection. Harmonic sources include power electronic devices, switch-mode power supplies, arcing devices such as fluorescent lights, and saturable devices such as transformers and rotating machines [4]. Such loads impact power quality in densely-loaded and electrically weaker areas of the network.

Harmonics can be described as time domain occurrences modeled as frequency components that are integer multiples of

the fundamental power frequency (50 Hz or 60 Hz). Harmonic pollution can be characterized using different attributes or indices [3], such as:

- The harmonic orders present.
- The total harmonic distortion (THD), which is the ratio of harmonic magnitude to the fundamental.
- Waveshape factor K_s, which is a measure of the steepness of the derivative of the composite waveform.
- The magnitude of individual harmonic components.

The harmonic regime is dynamic, since it is caused by the connection and disconnection of certain loads to the network. Therefore quantification of harmonic pollution cannot be a single assessment, but tends to examine average conditions and dominant regimes, such as the 95% cumulative probability level [5].

Some harmonic orders are of greater interest than others, being particularly associated with certain devices, or particularly prevalent. Adjustable speed drives employing a six-pulse rectifier are known to cause the 5th, 7th, 11th, and 13th harmonics, while those that use a twelve-pulse rectifier produce the 11th, 13th, 23rd, and 25th [4]. Consequently, wind farms have been linked to increases in the 5th and 7th order harmonics [6]. Switched-mode power supplies are associated with all odd harmonic orders from the 3rd to 15th [4]. Surveys of networks in various countries show a predominance of the 5th harmonic [5], while the potential danger of large neutral currents caused by the multiples-of-three triplen harmonics make them a particular concern. Since the lower orders are generally of greater amplitude than higher orders, the most significant harmonics are considered to be the 5th and 7th [7]. THD of the voltage waveform is typically reported as between 1% and 6% [5].

A key problem with harmonic distortion of the source voltage waveform is that it can increase both the peak voltage and the voltage rate of change. This puts extra stress on equipment insulation, leading to overheating and accelerated aging, and this effect is hidden from the asset manager by lack of knowledge of the harmonic regime. The impact of harmonics on asset lifetime is only just beginning to be quantifiable, with research investigating the relationship between harmonic distortion and insulation degradation such as PD.

III. PARTIAL DISCHARGE

Partial discharge (PD) is a localized phenomenon caused by weakening of the dielectric strength of insulation, allowing charge to partially bridge the insulation gap between conductors. PD is a symptom of degrading insulation health, but also damages the insulation further, causing progressive damage [8]. Different types of defects can cause PD, including rolling particles and metallic protrusions in liquid or gaseous insulation [9], as well as voids and electrical trees in solid insulation, all of which can be introduced to the plant during manufacture, commissioning, or over the life in service.

Monitoring of PD behavior can give information about the current health of the insulation system. There can be two aims of monitoring: identifying that PD is occurring; or diagnosing the type of defect causing the PD. The first aim is somewhat
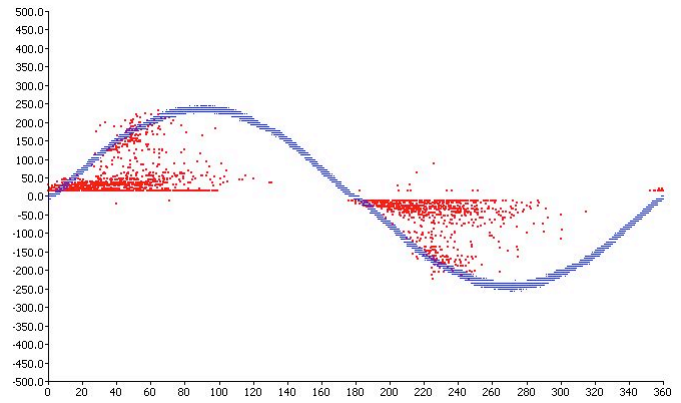

Fig. 1. Phase-resolved plot of PD activity and excitation voltage waveform.

simpler, and can be achieved using a handful of common techniques, including dissolved gas analysis for transformers [8] and acoustic monitoring [10]. This gives an indication of whether the insulation has deterioration, but without diagnostic information about the source of the PD it is difficult to define a plan of maintenance to manage potential problems.

One tool employed for defect diagnosis is based on the timing of discharges relative to their phase position on the excitation voltage waveform. This requires high resolution monitoring, techniques for which include radio frequency (RF) monitoring of the plant [11] or substation [12]; or electrical monitoring of apparent partial discharge, using the IEC 60270 technique [13]. PDs captured with these techniques are cross-referenced with the voltage waveform to construct a phase-resolved plot, illustrated in Fig. 1.

Different features of the phase-resolved pattern correspond to different source defects [14]. This knowledge has been exploited to produce intelligent systems for defect diagnosis, starting with data-driven techniques such as neural networks [15], k-means clustering and rule induction [16]; and more recently in knowledge-based expert systems [17] and support vector machines [18].

After diagnosing the cause of the PD, some quantification of the severity and urgency of the problem is required. Increases in the regularity and amplitude of discharges are generally indicative of fault progression consistent with deterioration of insulation integrity. Changes to the occurrence of discharges relative to phase are generally ascribed to the manifestation of a second defect, as the characteristics of the first defect are unlikely to change. As a result, the rate of insulation deterioration is gaged by the rate of change of discharge amplitude or regularity of activity.

However, this assumes that the conditions experienced by the item of plant are relatively constant, and specifically that the excitation voltage is a pure power frequency, either 50 Hz or 60 Hz. Recently it has been established that harmonics present in the source voltage can influence the observed PD pattern [1], [2]. As a result, it may be that changes to the pattern of PD activity are due to variations in the harmonic content of the voltage waveform, and not as previously assumed due to fault progression.

To continue using PD analysis as a tool for insulation health

assessment, the harmonic content of the voltage experienced by the plant under study must be considered. Harmonic monitoring equipment could be installed, but specialized units may be considered too expensive for their limited storage capabilities [5]. Since harmonics are expressed through the PD data, data mining makes it possible to extract facets of the harmonic regime experienced by the plant from the PD data itself. This is an attractive solution, since no further equipment is required for data capture; instead, further analysis is performed on the data for circumstance and condition monitoring.

IV. METHODOLOGY

Given the influence of harmonic pollution on the PD pattern, asset managers are in need of a tool for assessing the harmonic regime present. Condition monitoring technologies employ the assumption that no harmonics are influencing the plant under study, but knowledge of the harmonic circumstances would help decide if changing PD activity is due to plant health or harmonics.

The literature shows that the 5th and 7th harmonic orders are likely to be most prevalent, with the potential to distort the PD pattern. The aim of this work was to build a system for identifying the presence of the 5th and 7th harmonics from PD data, and report to engineers which harmonics occurred during the monitoring period and how frequently.

The core of this system is two classifiers, one for identifying each of the harmonic orders of interest, and a report generation system for presenting statistics about the regimes experienced.

The next section discusses the choice of machine learning technique for classification. In order to train the classifiers, experimental data had to be collected from defects in polymeric insulation. The following sections describe the experimental set-up, classifier training, and testing with fresh data. After that, the end-to-end, on-line system is presented.

V. MACHINE LEARNING FOR HARMONIC ANALYSIS

There is a vast array of different machine learning techniques available, from the widely known multi-layer perceptron (MLP) neural network to more esoteric types such as Gaussian mixture models, K* clustering, and support vector machines. These techniques have strengths and weaknesses for different types of tasks, but one key differentiating factor is whether a technique is for classification or regression; that is, whether it can identify the class or group an instance belongs to, or whether it is for estimating a function.

The application considered in this paper—recognition of the 5th and 7th harmonics—is classification. Previous work compared the accuracy of the C4.5 rule induction technique and support vector machines for these tasks, and found there was no significant difference between the results [19]. This concurs with other comparisons of classification for power systems problems, such as that in [20]. For practical purposes, namely the training time and amount of pre-training parameterization required, the C4.5 algorithm was preferred.

The C4.5 technique is an example of a rule-induction algorithm [21]. It 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.

The next section describes how data was collected for training and testing C4.5 trees for the two classification tasks: identifying the 5th and 7th harmonic orders.

VI. DATA COLLECTION

Electrical trees are a particularly damaging type of defect affecting polymeric insulation. After tree initiation, PD activity enhances electrical tree growth, eventually leading to insulation breakdown.

Electrical tree samples were created in LY/HY5052 epoxy resin with point-plane geometry, using $3\mu m$ radius conditioned hypodermic needles with a $2mm$ insulation gap. These samples were subjected to various harmonic regimes, described below, and the resulting PD activity recorded. For each PD, the acquisition hardware captures apparent charge and excitation waveform phase position. Discharges are collected as bursts of activity (hereafter called *PD burst* or *PD pattern*).

The testing facility used for PD capture was previously described in [22], and is summarized here. The raw PD data was collected at a sampling rate of $5Msps$ over periodic $80ms$ windows of continuous capture. The PD instrumentation system monitored and recorded electrical PD activity in accordance with the IEC 60270 standard over the frequency range, $100kHz$ to $400kHz$ [13]. This system (Fig. 2) consisted of a wideband balanced circuit and an amplifier bandpass filter. This arrangement yielded a minimum detectable level of $5.5pC$ at a signal gain of $50pC/V$. A 2nd order Chebyshev filter characterized by higher roll off at the cut off frequencies, relative to other filters including the Butterworth filter, was incorporated. Design considerations for $0.01dB$ ripple were selected to minimize characteristic ripples of the Chebyshev filter near the cut off frequencies. A $\pm 12V$ supply facilitated increased sensitivity at increased signal gain after the amplifier bandpass filter stage, capturing reduced magnitudes of PD. However the trade off was the corresponding reduction in pC range, as the constant power supply swing can result in saturation of the PD phase-resolved plots. Variable signal gain was implemented to mitigate this problem, facilitating manual gain control as the PD magnitude increased.

Using this system, one data set was captured for training the harmonic order classifiers, and another for testing purposes. The test set was not used for training, making it a fair challenge of the classifiers' abilities.

Additionally, a third set of data was collected for testing. This data was captured from a different type of defect: a void (air bubble) in the insulation. A void behaves similarly to an electric tree in producing PD, but it is not a progressive type of fault. Void samples were created in epoxy resin, and data collected using the same hardware as before. This data set was intended to test the classifiers' generalization capabilities, investigating whether the classifiers can identify harmonic attributes from PD data generated by a similar but different type of defect.
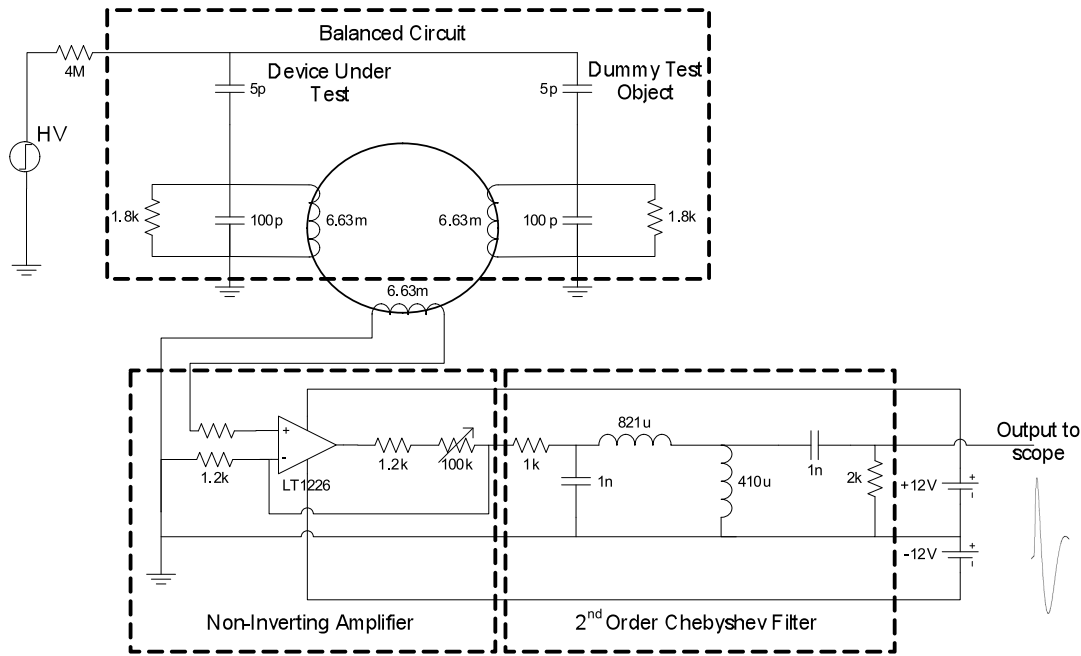
Fig. 2. Implemented balance circuit integrated with the amplifier filter stages.

TABLE I
SEVEN TEST WAVEFORMS EMPLOYED FOR THE CAPTURE OF DATASET 1

| Waveform ID number | Composition + 50Hz | | | THD (%) |
	Harmonic Order	Magnitude (%) per Harmonic	ϕ	
1	3	40.0	0	40.0
7	1	0.0	0	0.0
8	5	5.0	0	5.0
9	7	5.0	0	5.0
11	7	17.8	0	17.8
12	5, 7, 11, 13, 23, 25	3.2	0	7.85
13	5, 7, 11, 13, 23, 25	2.0	0	5.0

A. Data capture

The first experiment aimed to capture data that could be used to train the harmonic order classifiers. Samples were exposed to excitation waveforms that were polluted with different numbers and combinations of harmonic order, with some variance of the THD.

Eleven samples containing electrical trees experienced seven composite waveforms sequentially, described in Table I. Peak voltage was constant at 10.8 or 14.4 kV, and the duration of each waveform was 5 minutes or less. The sequence was repeated, meaning that data capture lasted for a maximum of (7 x 5) x 2 = 70 minutes per sample (see Fig. 3). The total set of data comprises 7091 PD bursts, collectively termed Dataset 1.

Next, a second set of data was captured from different samples. This was intended as a test set for the classifiers, and therefore the conditions were varied from the previous experiment. The waveforms listed in Table I were applied individually to samples for up to 30 minutes, with a peak voltage of 14.4 kV. Nine samples in total were used: one sample for each of waves 1, 7, 8, 11, and 12; and two samples

each for waves 9 and 13, in order to capture approximately equal numbers of PD bursts from each waveform (since waves 9 and 13 resulted in fewer PD bursts per sample, more samples were required). The total number of PD patterns recorded was 9195, called Dataset 2.

Finally, a third data set was captured from void defect samples. Since this set was also being used for test, conditions were varied to exercise the classification capabilities. Fourteen void samples experienced increasing levels of THD for one minute intervals, following the sequence 0% (fundamental only), 3%, 5%, 8%, 10%, 15%, and 18% THD. This sequence was applied eight times, giving a capture time of (7 x 1) x 8 = 56 minutes per sample (see Fig. 3). Of the fourteen samples, four were exposed to the fundamental and 5th harmonic only; three were exposed to the fundamental and 7th harmonic only; three experienced the fundamental, 5th, and 7th harmonics; and four experienced the grouping of many harmonics, specifically the 1st, 5th, 7th, 11th, 13th, 23rd, and 25th orders. In total, 9217 PD patterns were captured, forming Dataset 3.

B. Feature selection

The raw data associated with one burst of PD activity is not very amenable to machine learning, due to its size and the irregular number of discharges. A standard machine learning approach is to transform raw data into a feature vector: a list of pertinent features or attributes of the raw data, which characterize the PD pattern. Not only does this reduce the size of the dataset and hence reduce the computational complexity of training a learning technique; it also helps avoid the "curse of dimensionality" [23], a phenomenon where the performance of a machine learning technique tends to degrade as the dimensionality of the input data increases. Consequently, the number of features chosen for the feature vector should be
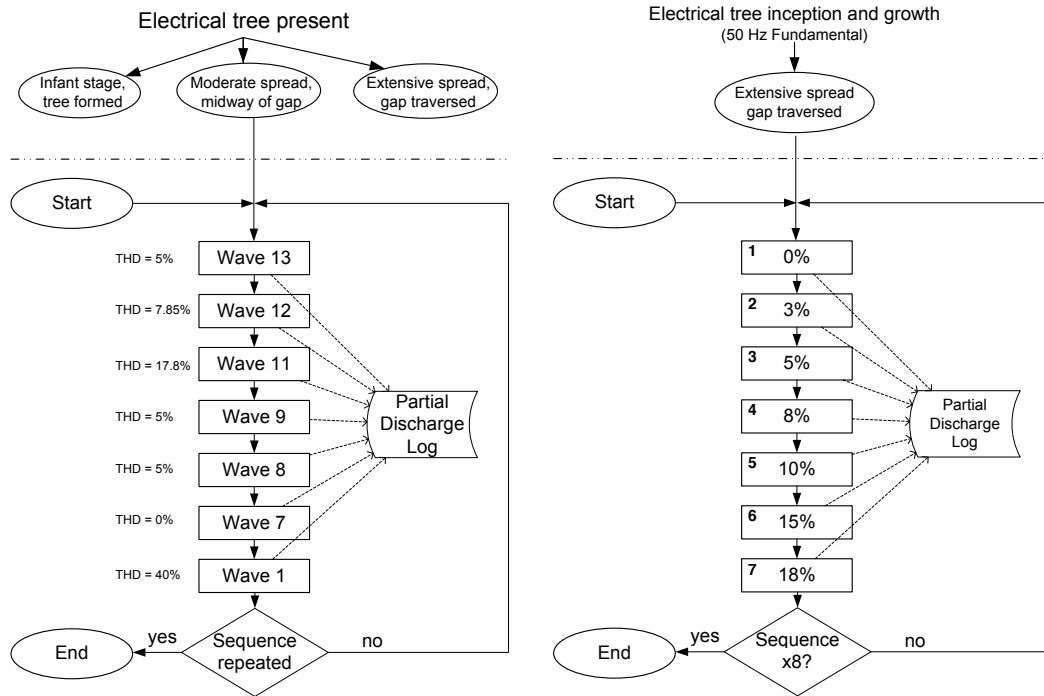
Fig. 3. Testing sequences for Dataset 1 (left) and 3 (right).

minimized, while giving enough detail to learn differences and patterns associated with the task at hand.

Three broad types of PD pattern features were identified from the literature through previous work [19]. The first and most established are statistical features calculated from the positive and negative voltage half-cycles, such as mean pulse height, pulse count, and the number of peaks in pulse amplitudes [15], [17]. Secondly, smaller phase windows can be used for statistical calculations, namely mean, standard deviation, and kurtosis of discharge amplitude calculated for six equal-width windows [24], or mean calculated for 100 phase windows [18].

The third type of feature is the pattern descriptor, which aims to mimic the way a human expert would view and describe a PD pattern [25]. These features often take labelled values, rather than being numeric, and identify pattern attributes such as the phase position of discharges (e.g. 'on the zero-crossing points' or 'on the peaks') and shape (e.g. 'chopped sine' or 'knife blade'). These descriptors are calculated for each half cycle, then the half cycles are compared for symmetry, giving up to three features per descriptor (e.g. positive shape, negative shape, and shape symmetry).

Since this paper considers PD in the presence of harmonics, which, to the best of our knowledge, has not previously been investigated with machine learning techniques, these reported feature vectors were not automatically assumed to be appropriate to the task. One large feature vector containing 47 features drawn from all three types was defined, and a parameter selection process was developed to identify the best mixture of features for each of the two classification tasks.

Taking each classification task in turn, the 47 features were ranked in order of their importance using the Information Gain

Ratio:

$$IGR(c, f) = \frac{(H(c) - H(c|f))}{H(f)} \quad (1)$$

for a classification c, a feature f, and entropy H. Subsets were selected from the top of this ranked list, and used to train C4.5 trees. The best subset, and hence the best feature vector, was the set of features that gave the most accurate C4.5 tree. The effect of this was to maximize the accuracy of the classification while limiting the size of the feature vector at the same time.

The 47 features and the best subsets for classifying the presence of the 5th harmonic and the presence of the 7th harmonic are shown in Table II.

VII. CLASSIFIER TRAINING

After selecting appropriate parameters, C4.5 trees were trained for identifying the presence of the 5th and 7th harmonics. Trees were trained using 10-fold cross-validation, and the accuracy results are reported below.

A. 5th harmonic

The first C4.5 tree is a binary classifier identifying the presence or absence of the 5th harmonic. Trained using the 7091 patterns of Dataset 1, an accurate classifier should give a positive response for data captured under waveforms 8, 12, and 13; and a negative response for data from waveforms 1, 7, 9, and 11.

The C4.5 tree is 83.8% accurate at this task. The confusion matrix is shown in Table III, where the diagonal corresponds to accurate classifications. Deeper investigation reveals varying error rates for different source waveforms (Table IV), with the

TABLE II
47 FEATURES AND THE BEST SUBSET FOR EACH TASK

Feature name	5th	7th	Feature name	5th	7th	Feature name	5th	7th
Phase 0–60° mean	X		Phase 0–60° std. deviation	X		Phase 0–60° kurtosis		
Phase 60°–120° mean	X	X	Phase 60°–120° std. deviation	X	X	Phase 60°–120° kurtosis	X	
Phase 120°–180° mean	X	X	Phase 120°–180° std. deviation	X	X	Phase 120°–180° kurtosis	X	X
Phase 180°–240° mean			Phase 180°–240° std. deviation	X	X	Phase 180°–240° kurtosis		
Phase 240°–300° mean	X	X	Phase 240°–300° std. deviation	X	X	Phase 240°–300° kurtosis	X	
Phase 300°–360° mean	X		Phase 300°–360° std. deviation			Phase 300°–360° kurtosis	X	
Positive half mean	X	X	Positive half std. deviation	X	X	Positive half kurtosis		
Negative half mean	X	X	Negative half std. deviation			Negative half kurtosis		
Positive half bias			Positive half inception phase	X	X	Positive half extinction phase	X	X
Negative half bias	X	X	Negative half inception	X	X	Negative half extinction	X	X
Positive half distribution	X	X	Positive half no. peaks in amplitude	X	X	Positive half phase position	X	X
Negative half distribution	X	X	Negative half no. peaks in amplitude	X		Negative half phase position	X	X
Positive half shape			Positive half range	X		Positive half density	X	X
Negative half shape	X		Negative half range	X		Negative half density	X	X
Shape symmetry			Range symmetry			Density symmetry	X	X
Total PDs in burst	X	X	Density ratio of half cycles	X				

TABLE III
PRESENCE OF THE 5TH HARMONIC CONFUSION MATRIX

Presence	Classified as	
	False	True
False	3482	577
True	570	2462

TABLE IV
ERRORS IN THE 5TH HARMONIC CLASSIFIER

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

classifier highly accurate at identifying the absence of the 5th harmonic from waveform 1, but least accurate at identifying its absence from waveform 7.

B. 7th harmonic

Identifying the presence of the 7th harmonic is also a binary classification task. Trained using Dataset 1, an accurate classifier would identify its presence from data captured under waveforms 9, 11, 12, and 13; and give a negative response to data from waveforms 1, 7, and 8.

This C4.5 tree is 90.9% accurate (confusion matrix in Table V). Considering accuracy by source waveform (Table VI), this tree appears most accurate at the higher THD levels of waves 1 and 11, then 9 and 12, and least certain on the lower THD levels of 7, 8, and 13.

TABLE V
PRESENCE OF THE 7TH HARMONIC CONFUSION MATRIX

Presence	Classified as	
	False	True
False	2680	326
True	322	3763

TABLE VI
ERRORS IN THE 7TH HARMONIC CLASSIFIER

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

VIII. TEST

The training process described in the previous section resulted in two classifiers for harmonic orders. This section presents the accuracy of testing with Datasets 2 and 3. It should be noted that "accuracy" can result from a positive or negative identification by the classifier. For example, data captured in the presence of the 5th harmonic and fundamental only would be *accurately* classified in the following cases:

- If the 5th harmonic classifier gives a *positive* response (5th is present, classifier says 5th is present),
- If the 7th harmonic classifier gives a *negative* response (7th is not present, classifier says 7th is not present).

In all cases, classifier accuracy would ideally tend towards 100%.

A. Dataset 2: Electrical trees

As described in Section VI, a second dataset was captured from electrical tree samples and a different experimental regime from the training data set. The aim was to assess the capabilities of the classifiers when applied to blind data that had not been seen by the classifiers before. The results are shown in Fig. 4.

The 5th harmonic classifier performs well overall, at 69.18% accuracy. The graph shows the highest accuracy for waveform 1, meaning that 96% of the time this classifier correctly said there was no 5th harmonic present. The lowest accuracy was on waveform 12, where the classifier was no better than chance at saying the 5th harmonic was present in amongst the grouping of many harmonic orders.
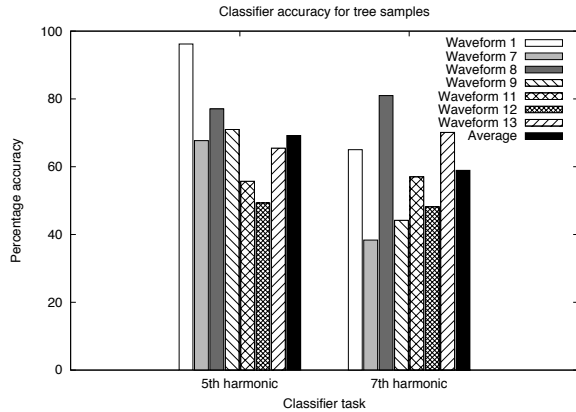
Fig. 4. Accuracy of classifiers on Dataset 2, split by test waveform.

Fig. 5. Accuracy of classifiers on Dataset 3, split by test waveform.

Considering the 7th harmonic classifier, the overall performance was more modest at 58.92% average accuracy. The classifier was good at saying the 7th harmonic was absent from waves 1 and 8, and good at saying the 7th was present in wave 13, but less accurate than a random guess on waves 7, 9, and 12. Interestingly, waves 12 and 13 have the same grouping of harmonic orders but different THD levels, and the classifier finds it significantly more easy to work with the lower THD data of wave 13. This is rather surprising, as engineering intuition suggests the features of each harmonic order would be more dominant at higher levels of distortion. However, a similar relationship exists with waves 9 and 11, and in this case it is the higher-THD wave 11 that is more easily classified.

In short, these results suggest that the harmonic order classifiers generalize well to Dataset 3. Both have average accuracies below their cross-validation accuracies, but above the levels of chance that would indicate the results were noise. This is the expected behavior of a well-trained classifier.

B. Dataset 3: Voids

The third set of data was intended to stretch the classifiers further, by applying them to data generated by a different type of defect from the training data. If the classifiers performed well on this data, we could be reasonably sure that they had truly learned how harmonic attributes are expressed in PD data; whereas if they performed poorly, it may be due to differences between the patterns generated by void and tree defects. The results are shown in Fig. 5.

The 5th harmonic classifier is poor on average, with 40.81% accuracy. However, this figure hides high accuracy at identifying the absence of the 5th harmonic from the fundamental and 7th polluted waveforms. Countering this is extremely low accuracy at identifying the presence of the 5th in groupings of harmonic orders. This contrasts with the electrical tree data results, where moderate to high accuracy was seen on groupings of harmonics. Overall it seems this classifier has significant difficulties with some of the void data.
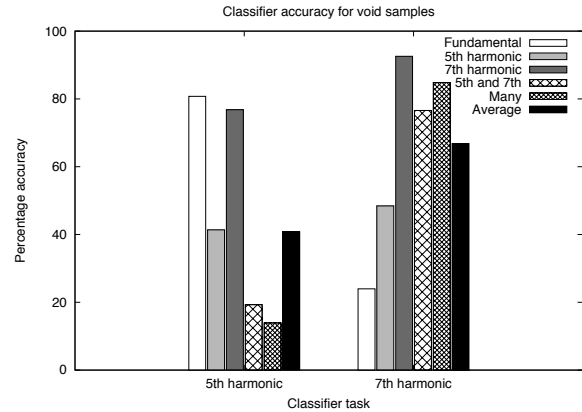
However, the 7th harmonic classifier performs better on average, with 66.83% accuracy. It gives good results at identifying the presence of the 7th harmonic, either in groups of other orders or with the fundamental alone, which is in line with the electrical tree data results. It performs most poorly at identifying the absence of the 7th from the fundamental waveform, which is also the case for the electrical tree data. The only discrepancy is in identifying the absence of the 7th from the waveform polluted by the 5th only, with chance results on the void data but high accuracy on the electrical trees.

In short, the 5th harmonic seems to present itself in the void data differently from the electrical tree data, as the average accuracies show discrepancy between the two test data sets. However, the 7th harmonic appears to present in largely the same way in PD data from both types of defect, as the average accuracies differ by less than 10%. In both cases, the classifiers show good accuracy at identifying harmonics from PD generated by electrical trees.

IX. ON-LINE SYSTEM

Given the successful training and testing of classifiers described above, an on-line system was created to perform harmonic analysis from PD in the field. This section discusses some of the design choices made to create such a system.

The classifiers operate on a single PD pattern, deciding if the 5th or 7th harmonics are present based on an $80ms$ snapshot of activity. In practice, PD monitoring takes place over a period of time, so there is no need to restrict an on-line system to consider only one pattern.

Indeed, the harmonic regime is also variable over time with the connection and disconnection of particular loads. An on-line system for harmonic circumstance monitoring should collect information on the trends of different harmonics, presenting asset managers with a report on the changing regimes experienced over the period of monitoring.

Additionally, some of the uncertainty of the classifier output can be mitigated by aggregating statistics about the dominant regimes. Various strategies for combining multiple PD patterns
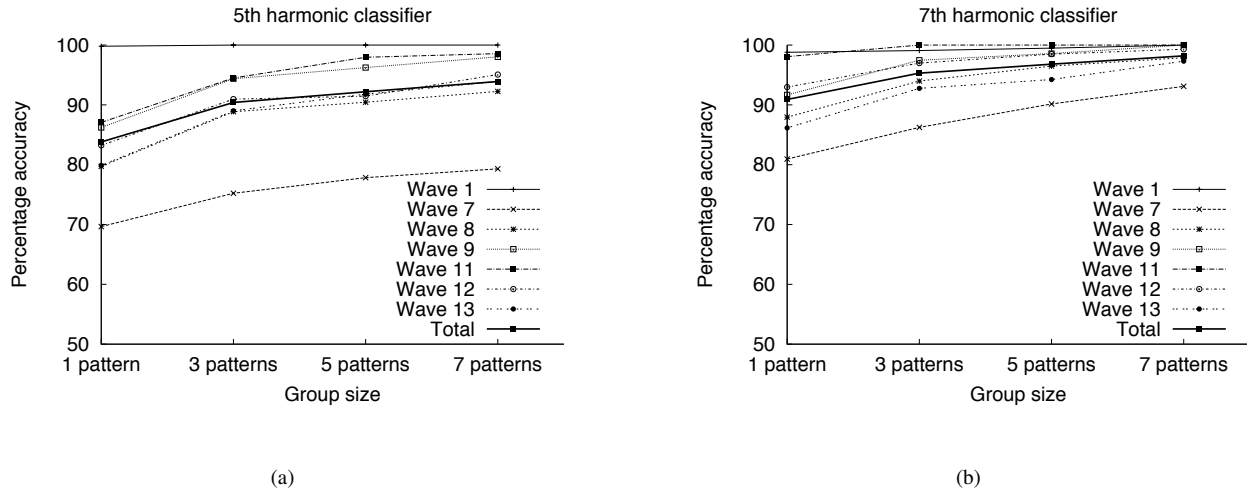
Fig. 6. Accuracy of the classifiers on individual PD patterns, and on groups of three, five, and seven patterns for the training data set.

for classification were considered in [26], with the most practical being to use simple majority voting among a sequence of PD patterns. For example, if three PD patterns are captured under waveform 1 conditions, two of which are classified as no 5th harmonic while one is classified with 5th harmonic, then the majority opinion is that the 5th harmonic is not present. The aggregated output of these classification would be that the 5th harmonic was not present during the monitoring period.

Aggregating multiple patterns can be shown to improve classifier accuracy (see Figs. 6(a) and 6(b)). The training data was grouped into batches of three sequential patterns, five sequential patterns, and seven sequential patterns. These batches were classified by the two harmonic order C4.5 trees, and simple majority voting used to determine the final output for each batch. The results of this indicate that accuracy increases with the number of patterns included in the voting. For identification of the 5th harmonic, accuracy increases from 83.8% when considering individual patterns to 93.9% for groups of seven. For the 7th harmonic, accuracy increases from 90.9% for individual patterns to 98.2% for seven in a row.

While this may suggest majority voting on the entire set of captured PD patterns, this would only identify one dominant regime over the monitoring period. The connection and disconnection of loads can change the harmonic regime rapidly, but this new regime may occur frequently enough to influence many PD patterns. The number of patterns to aggregate becomes a trade-off between the desire for reduction in uncertainty versus the level of detail about the regimes experienced.

The on-line system was built to incorporate the two C4.5 classifiers, aggregation of multiple patterns, and a system for reporting on the harmonic regimes. It stores the aggregated statistics in a database for long term archival, and presents summary and detailed reports to the engineer as needed.

X. CONCLUSIONS

Since recent research identified the effects of harmonic pollution on PD data, a tool is needed for asset managers to determine the harmonic circumstances being experienced by items of plant. This extra information can prevent insulation fault progression from being overestimated. This paper presented the design of components needed to create a system for harmonic circumstance monitoring, identifying the presence of the most prevalent and concerning 5th and 7th harmonics, and the frequency with which they appear on the network.

The system comprises two C4.5 decision trees, one for each of the harmonic orders of interest. The classifications of these trees are aggregated over multiple PD patterns during the period of monitoring, and a report generated about the dominant regimes of that time.

This system could be extended over time to contain more classifiers to identify other harmonic attributes. Preliminary work has been done on assessing the THD [27], but initial test results were surprisingly poor for some defect samples while extremely accurate for others. Further research is needed to understand the pattern of THD classifications, before it can become part of the overall system.

Based on the results reported here, the system as built is ready to be deployed in the field alongside regular PD monitoring, giving engineers a better picture of the true health of polymeric cable insulation.

REFERENCES

[1] B. Florkowski, M. Florkowski, and P. Zydron, "The Role of Harmonic Components on Partial Discharge Mechanism and Degradation Processes in Epoxy Resin Insulation," in *2007 Int. Conf. on Solid Dielectrics*, Winchester, UK, Jul. 2007.

[2] S. Bahadoorsingh and S. M. Rowland, "Modeling of Partial Discharges in the Presence of Harmonics," in *IEEE Conf. on Electrical Insulation and Dielectric Phenomena*, Virginia Beach, USA, 2009, pp. 384–387.

[3] G. T. Heydt, "Electric power quality: a tutorial introduction," *IEEE Computer Applications in Power*, vol. 11, no. 1, pp. 15–19, Jan. 1998.

[4] A. Kusko and M. T. Thompson, *Power quality in electrical systems*. McGraw-Hill, 2007.

[5] V. Gosbell, D. Mannix, D. Robinson, and S. Perera, "Harmonic Survey of an MV Distribution System," in *Proc. AUPEC01*, Perth, Australia, Sep. 2001, pp. 338–342.

[6] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi, *Wind Energy Handbook*. Wiley, 2001, ch. 10.5: Power Quality.

[7] A. Asheibi, D. Stirling, and D. Soetanto, "Analyzing Harmonic Monitoring Data using Data Mining," in *Proc. 5th Australasian Data Mining Conference (AusDM2006)*, Nov. 2006.

[8] T. K. Saha, "Review of Modern Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 10, no. 5, pp. 903–917, Oct. 2003.

[9] G. P. Cleary and M. D. Judd, "An Investigation of Discharges in Oil Insulation using UHF PD Detection," in *Proceedings of the 14th IEEE Int. Conf. on Dielectric Liquids (GRAZ)*, Jul. 2002, pp. 341–344.

[10] B. H. Ward, "A Survey of New Techniques in Insulation Monitoring of Power Transformer," *IEEE Electr. Insul. Mag.*, vol. 17, no. 3, pp. 16–23, May/Jun. 2001.

[11] M. D. Judd, L. Yang, and I. B. B. Hunter, "Partial Discharge Monitoring for Power Transformers using UHF Sensors Part 1: Sensors and Signal Interpretation," *IEEE Electr. Insul. Mag.*, vol. 21, no. 2, pp. 5–14, Mar./Apr. 2005.

[12] I. E. Portugues, P. J. Moore, I. A. Glover, C. Johnstone, R. H. McKosky, M. B. Goff, and L. van der Zel, "RF-Based Partial Discharge Early Warning System for Air-Insulated Substations," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 20–29, Jan. 2009.

[13] IEC, "High-Voltage Test Techniques - Partial Discharge Measurements," 2001, document IEC 60270.

[14] CIGRE, "Knowledge Rules for Partial Discharge Diagnosis in Service," Tech. Rep., Apr. 2003, Working Group D1.11, Brochure 226.

[15] E. Gulski and A. Krivda, "Neural networks as a tool for recognition of partial discharges," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 984–1001, Dec. 1993.

[16] S. D. J. McArthur, S. M. Strachan, and G. Jahn, "The Design of a Multi-Agent Transformer Condition Monitoring System," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1845–1852, Nov. 2004.

[17] 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. Dielectr. Electr. Insul.*, vol. 15, pp. 259–268, 2008.

[18] L. Hao and P. L. Lewin, "Partial discharge source discrimination using a support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 189–197, 2010.

[19] S. Bahadoorsingh, S. M. Rowland, V. M. Catterson, S. E. Rudd, and S. D. J. McArthur, "The Role of Circumstance Monitoring on the Diagnostic Interpretation of Condition Monitoring Data," in *Int. Symp. Electrical Insulation (ISEI 2010)*, San Diego, USA, Jun. 2010.

[20] 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 Symposium*, 2009, pp. 469–476.

[21] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[22] 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. Dielectr. Electr. Insul.*, vol. 17, no. 3, pp. 701–708, 2010.

[23] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[24] A. J. Brown, V. M. Catterson, M. Fox, D. Long, and S. D. J. McArthur, "Learning Models of Plant Behavior for Anomaly Detection and Condition Monitoring," *Engineering Intelligent Systems*, vol. 15, Jun. 2007.

[25] S. Rudd, S. D. J. McArthur, and M. D. Judd, "A generic knowledge-based approach to the analysis of partial discharge data," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 149–156, 2010.

[26] V. M. Catterson, S. E. Rudd, S. D. J. McArthur, S. Bahadoorsingh, and S. M. Rowland, "Assessing the effects of power quality on partial discharge behaviour through machine learning," in *7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Stratford-upon-Avon, England, Jun. 2010.

[27] S. Bahadoorsingh, S. M. Rowland, V. M. Catterson, S. E. Rudd, and S. D. J. McArthur, "Interpretation of Partial Discharge Activity in the Presence of Harmonics," in *10th IEEE Int. Conf. on Solid Dielectrics (ICSD)*, Potsdam, Germany, Jul. 2010.

Victoria M. Catterson (M'03) received the B.Eng. (Hons) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 2003 and 2007 respectively. She is a Senior Research Fellow within the Institute for Energy and Environment at the University of Strathclyde. Her research interests include intelligent systems, condition monitoring, and diagnostics.

Sanjay Bahadoorsingh (M'03) was born in Port of Spain, Trinidad. He completed the B.Sc. degree from The University of The West Indies in 2003 and the M.Sc. degree from UMIST in 2005. In 2009, he completed the Ph.D. degree at The University of Manchester. He is currently a lecturer at The University of the West Indies. His areas of research are asset management and dielectric ageing.

Susan Rudd received the B.Sc. (Hons) and Ph.D. degrees from the University of Strathclyde in 2004 and 2010 respectively. She is a Research Assistant in the Institute for Energy and Environment. Her research interests include Knowledge Engineering, intelligent system applications in power engineering and condition monitoring.

Stephen D. J. McArthur (M'93–SM'07) received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1992 and 1996 respectively. He is a Professor in the Institute for Energy and Environment, Director of the British Energy Advanced Diagnostics Centre, and Co-director of the Strathclyde Centre for Applied Intelligent Systems. He currently chairs the IEEE PES Multi-Agent System Working Group. His research interests include intelligent system applications in power, condition monitoring, and agent systems.

Simon M. Rowland (SM'07) was born in London, England. He completed the B.Sc. degree in physics at UEA and the Ph.D. degree at London University. He was awarded the IEE Duddell Premium in 1994 and became a FIEE in 2000. He has worked for many years on dielectrics and their applications. He has also been Operations and Technical Director multinational manufacturing companies. He joined The School of Electrical and Electronic Engineering in The University of Manchester as a Senior Lecturer in 2003 and was appointed Professor of Electrical Materials in 2009. He is Vice-President Admin. of the IEEE Dielectric and Electrical Insulation Society.