Application of ACF-wavelet feature extraction for classification of some artificial PD models of power transformer

Vahid PARVIN DARABAD*
Department of Electrical Engineering, Faculty of Engineering, Golestan University, Gorgan, Iran

Received: 13.08.2017 • Accepted/Published Online: 11.07.2018 • Final Version: 29.11.2018

Abstract: In this paper, 7 different artificial partial discharge (PD) models of power transformer defect types are built in a high voltage laboratory, and then PD signals are recorded in the whole power frequency cycle to develop a preliminary PD knowledge base. A specific technique is used to extract single PD events from the recorded knowledge base. By application of wavelet transform, different decomposition levels of extracted PD signals are investigated. Feature extraction is performed by application of statistic moments on the autocorrelation function of PD signal decompositions. “Best” features are selected based on the F-test. To evaluate the performance of features, a naive Bayes classifier is applied to selected features. Results show that nearly 100 percent accuracy in PD discrimination is achieved using multivariate-multinomial distribution estimation of the feature space. A case study is carried out to show the contribution of the proposed approach. The combined PD detection-classification system proposed in this paper can serve as a complement to conventional PD monitoring systems.

Key words: Autocorrelation function, wavelet, partial discharge, power transformer

1. Introduction

Power transformer condition monitoring improves plant economy and increases availability and useful service life [1, 2]. Online condition monitoring of the transformers and other power system equipment provides the opportunity for continuous operation service, which leads to early detection of problems and possible remedial action and so increases the lifetime expectancy of the transformers [3]. The components that play major roles in the life expectancy of a transformer and other power system equipment are its insulations [4]. Nowadays, due to advancement in technology, most of the condition monitoring techniques can be implemented online. Among these techniques, online PD measurement is one of the most efficient ones, which attracted great interest due to its high sensitivity and fast reflection of insulation conditions [5–7]. For many years PD measurement has been employed as a nondestructive tool for assessing the integrity of the insulation of power system equipment. PD is the main sign of manufacturing defects and insulation deterioration due to electrical, thermal, and mechanical stresses. Monitoring, detection, and trending of PD activities can keep the HV components from internal breakdown through the identification of the existing defects. Different processes are involved in PD measurement and they can be categorized into 3 categories: first, PD detection; second, PD localization; and third, identification of defect types causing PD. This study focuses on the 3rd class of these processes, which deals with separation of different types of PD sources. Some of the first PD classification methods were presented in [8–10]. PD databases have grown up very fast in different parts of the power system industry, while

*Correspondence: v.parvin@gu.ac.ir

© TÜBİTAK
doi:10.3906/elk-1708-17
there are not enough expert people for analyzing the databases. This issue motivated academics to develop artificial expert systems capable of working with databases and extracting useful information using data mining approaches. Generally, to achieve PD source identification, a knowledge base is derived from raw data by feature extraction methods like PRPD patterns [11, 12], FFT [13], statistical analysis [14], cepstral features [15], or wavelet patterns [16, 17], and then a decision making system such as neural networks [18, 19], SVM [20], PSO [19, 21], the Bayes theorem, ANFIS [22], k-means and fuzzy c-means [23], or SOM [24] interprets the knowledge base in a meaningful way to discriminate different PD sources. Methodologies depend on the application and equipment under monitoring; introduction of a new knowledge base still helps experts to make decisions with greater reliability. Besides that, wavelet analysis showed great potential in pattern recognition applications [25, 26]; this motivated authors to evaluate wavelets. To cover the mentioned concerns, as a first step, different PD defect models are built in a high voltage laboratory (HV lab), and then PD current signals are recorded, and in the next step, features are extracted from ACF-wavelet coefficients of the recorded PD data. Employing the ACF results in a normalized feature space, which makes the knowledge base correlate with only the shape of the PD signal, not its amplitude. In the third step a Naive Bayes classifier is applied to feature spaces and results are shown in the fourth step.

2. PD model setup

Seven different PD models, as shown in Figure 1, are built in a high voltage lab, and the description of each PD model is as follows:

a) Air corona: built by using a needle as a high voltage electrode in a distance of 10 mm to a plane electrode with diameter of 22 cm, while tip-radius of needle is 170 μm.

b) Discharges in oil: this model is constructed by using a needle-plane electrode configuration; distance between needle and plane electrode is 10 cm and diameter of plane electrode is 25 cm.

c) Surface discharge: built by putting Plexiglas of 3.7 mm thickness and diameter of 22 cm between high voltage and ground electrode.

d) Sphere cavity in solid insulation: built by creating a hole inside a solid insulation; diameter of cavity is between 1 and 2 mm and thickness of test sample is between 5 and 7 mm.

e) Sharp metallic point in solid insulation: built by putting a 3-cm needle in an epoxy resin sample of 3.5 cm in thickness.

f) Metal particles in oil: built by putting tiny iron filings with approximate length of 1 cm in transformer oil.

g) Cylindrical cavity in solid insulation: a cylindrical hole is created in the epoxy resin dielectric sample; the cavity is close to the surface of the test object and it has 1 mm height and 4 mm diameter.

2.1. PD measurement configuration

Figure 2 shows the configuration of the PD measuring system. High voltage is applied through a protection fuse to prevent short circuit current due to probable breakdown of insulation resistance of test samples. PD current pulses, which are captured by HFCT, are recorded by a 600 MHz 10GS/s digital oscilloscope. Frequency
Figure 1. Structure of PD models.

Sampling rate is fixed at 500 Ms/s. HFCT is used to pick up PD current pulses, and the pick point and 50-Hz phase location of PD pulses are used to make PRPD patterns; now a system is developed that shows conventional PRPD patterns and at the same time single PD pulses can be addressed in the PRPD pattern. Test setups are in Faraday cage, so background noise is limited to measurement equipment such as oscilloscope baseline noise. The average signal to noise ratio (SNR) is 40 dB; single PD pulses are extracted from recorded PD data. Length of each PD pulse is 1 μs (512 points). Table 1 shows the voltage range applied for each PD model. PRPDs of PD models are shown in Figure 3; as can be seen, different PD models show different PRPD patterns in HV lab conditions. Figure 4 shows examples of PD signals captured from different PD models. As can be seen in the figure, different PD models produce different PD waveforms; this means they have the potential to show different features.

3. Application of discrete wavelet transform and autocorrelation functions

Many researchers [26–30] showed that wavelet analysis is a popular method of extracting effective and critical features to judge about the type of signals, because it provides both the frequency and spatial locality. A brief explanation of a wavelet is as follow. If \( f(t) \) is the signal characteristics, the discrete wavelet transformation of \( f(t) \) will be:
**Figure 2.** PD Measurement and recording setup.

**Figure 3.** PRPD of PD models.

**Table 1.** Structure of proposed PD diagnosis system.

<table>
<thead>
<tr>
<th>PD model</th>
<th>Applied voltage (kV)</th>
<th>Total extracted PD signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air corona</td>
<td>6</td>
<td>12,059</td>
</tr>
<tr>
<td>Discharge in oil</td>
<td>19</td>
<td>1955</td>
</tr>
<tr>
<td>Surface discharge</td>
<td>9</td>
<td>1337</td>
</tr>
<tr>
<td>Sphere cavity</td>
<td>12</td>
<td>13,396</td>
</tr>
<tr>
<td>Metallic point in solid insulation</td>
<td>11</td>
<td>1851</td>
</tr>
<tr>
<td>Metal particles in oil</td>
<td>12</td>
<td>7320</td>
</tr>
<tr>
<td>Cylindrical Cavity</td>
<td>11</td>
<td>1399</td>
</tr>
</tbody>
</table>
where $a = 2^{-m}$ and $b = n \cdot 2^{-m}$ are real constant parameters (called dyadic analysis), $m$ and $n$ are scaling and shifting factors, respectively, and $\psi(\cdot)$ is the mother wavelet. In this research, the maximum decomposition (scaling) level is 9, so for each PD signal there is 1 approximation and 9 details. To make features comparable for each single PD signal, the autocorrelation function (ACF) of each detail and approximation of a single PD signal is calculated and used as the feature space. Given $x^k = [x^k_1, x^k_2, \ldots, x^k_N]$ as a $1 \times N$ vector of the $k$th input, the ACF would be also a $1 \times N$ vector, $ACF^k = [ACF^k_1, ACF^k_2, \ldots, ACF^k_N]$, whose elements are calculated as follows:

$$ACF^k_j = \sum_{i=1}^{N} x^k_i x^k_{i+j} \sum (x^k_i)^2.$$ (2)
The important aspect of a normalized ACF is that similar signals have similar ACFs even if they have different energy content; another aspect of the ACF is that statistical features that are dependent on the starting and ending points of a signal can be extracted in the same condition for each PD signal, so using the ACF makes statistical features comparable for PD classification.

In this approach 55 types of mother-wavelet functions (Table 2) are applied to PD signals and different feature extraction methods are extracted from the ACF of details and approximation of each PD signal, and then “Best” features resulting from “Best” decompositions, which consist of the corresponding mother-wavelet and its related detail or approximation ACF, are explored.

### Table 2. Mother wavelet families used in this work.

<table>
<thead>
<tr>
<th>Wavelet family</th>
<th>Wavelets used in the family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daubechies</td>
<td>‘db1’; ‘db2’; ‘db3’; ‘db4’; ‘db5’; ‘db6’</td>
</tr>
<tr>
<td></td>
<td>‘db7’; ‘db8’; ‘db9’; ‘db10’; ‘db11’; ‘db12’</td>
</tr>
<tr>
<td>Symlets</td>
<td>‘sym2’; ‘sym3’; ‘sym4’; ‘sym5’; ‘sym6’</td>
</tr>
<tr>
<td></td>
<td>‘sym7’; ‘sym8’</td>
</tr>
<tr>
<td>Coiflets</td>
<td>‘coif1’; ‘coif2’; ‘coif3’; ‘coif4’; ‘coif5’</td>
</tr>
<tr>
<td>BiorSplines</td>
<td>‘bior1.1’; ‘bior1.3’; ‘bior1.5’; ‘bior2.2’; ‘bior2.4’</td>
</tr>
<tr>
<td></td>
<td>‘bior2.6’; ‘bior2.8’; ‘bior3.1’; ‘bior3.3’; ‘bior3.5’</td>
</tr>
<tr>
<td></td>
<td>‘bior3.7’; ‘bior3.9’; ‘bior4.4’; ‘bior5.5’</td>
</tr>
<tr>
<td>ReverseBior</td>
<td>‘rbio1.1’; ‘rbio1.3’; ‘rbio1.5’; ‘rbio2.2’; ‘rbio2.4’</td>
</tr>
<tr>
<td></td>
<td>‘rbio2.6’; ‘rbio2.8’; ‘rbio3.1’; ‘rbio3.3’; ‘rbio3.5’</td>
</tr>
<tr>
<td></td>
<td>‘rbio3.7’; ‘rbio3.9’; ‘rbio4.4’; ‘rbio5.5’; ‘rbio6.8’</td>
</tr>
<tr>
<td>DMeyer</td>
<td>‘dmey’</td>
</tr>
</tbody>
</table>

#### 3.1. Feature extraction method

Three well-known features, which are called skewness, kurtosis, and energy, are calculated on each ACF of 9 details and 1 approximation of a total of 39,317 PD signals [31];

\[ Sk = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^3}{(\sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2})^3}, \]  

(3)

\[ Ku = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{(\sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2})^2}; \]  

(4)

\[ En = \sum_{i=1}^{N} x_i^2. \]  

(5)

There are totally 10 ACFs for each single PD signal, containing 9 details and 1 approximation, which results in a \(1 \times 30\) feature vector.

Implementation of the mentioned feature extraction by 55 different mother-wavelets over the PD input space obtains a \(39,713 \times 30 \times 55\) knowledge database.

3105
3.2. F-test

To find the best features for classification algorithms, two statistic parameters are calculated within: variance $S_w^2$, which is the variance of features for each class, and between variance $S_b^2$, which is the variance between the centers (means) of all classes. “Best features” have the highest relation of $S_b^2$ to $S_w^2$ [31]:

$$S_w^2(i) = \sum_{P=1}^{7} \frac{N^P - 1}{N^P} \sum_{k=1}^{N^P} \left[ FV^k_i - \frac{1}{N^P} \sum_{k=1}^{N^P} FV^k_i \right]^2,$$

(6)

$$S_b^2(i) = \sum_{P=1}^{7} N^P \left[ \frac{1}{N^P} \sum_{k=1}^{N^P} FV^k_i - \frac{1}{N} \sum_{k=1}^{N} FV^k_i \right]^2,$$

(7)

$$F(i) = \frac{S_b^2(i)}{S_w^2(i)} \times \frac{N - 7}{6},$$

(8)

where $P$ is the set of PD classes, $N$ is the number of all samples, $N^P$ is the number of samples in class $P$, and $FV^k_i$ is the $i$th feature of the $K$th sample inside class $P$. Table 3 shows the top 10 features and their corresponding wavelets for PD classification.

Table 3. Top 10 features and their corresponding wavelets for PD classification.

<table>
<thead>
<tr>
<th>Hierarchical order</th>
<th>Wavelet family</th>
<th>Feature no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(rbior3.1)</td>
<td>(skewness of ACF of d3)</td>
</tr>
<tr>
<td>2</td>
<td>(db2)</td>
<td>(energy of ACF of d6)</td>
</tr>
<tr>
<td>3</td>
<td>(sym2)</td>
<td>(energy of ACF of d6)</td>
</tr>
<tr>
<td>4</td>
<td>(rbior1.5)</td>
<td>(energy of ACF of d4)</td>
</tr>
<tr>
<td>5</td>
<td>(rbior3.5)</td>
<td>(energy of ACF of d6)</td>
</tr>
<tr>
<td>6</td>
<td>(rbior3.7)</td>
<td>(energy of ACF of d6)</td>
</tr>
<tr>
<td>7</td>
<td>(rbior3.3)</td>
<td>(skewness of ACF of d3)</td>
</tr>
<tr>
<td>8</td>
<td>(rbior3.1)</td>
<td>(kurtosis of ACF of d1)</td>
</tr>
<tr>
<td>9</td>
<td>(rbior1.3)</td>
<td>(energy of ACF of d4)</td>
</tr>
<tr>
<td>10</td>
<td>(bior3.3)</td>
<td>(energy of ACF of d4)</td>
</tr>
</tbody>
</table>

3.3. Application of naive Bayes classifiers to differentiate PD classes

The naive Bayes classification framework is provided by a simple theorem of probability [32] called the Bayes theorem:

$$P(C = c_k | X = x) = P(C = c_k) \frac{P(X = x | C = c_k)}{P(X = x)}.$$

(9)

By assuming that all possible events fall into exactly one of $N_c$ classes $[c_1, \ldots, c_k, \ldots, c_N]$:

$$P(X = x) = \sum_{j=1}^{N_c} P(X = x | C = c_j) P(C = c_j),$$

(10)
where \( C = c_1, \ldots, c_k, \ldots, c_N \) is a random variable whose values are the classes and \( X \) is a random variable whose values are feature vector \( x = [x_1, x_2, \ldots, x_D] \). The goal is to find the conditional probability that a PD fault belongs to class \( c_j \) given that the feature vector is \( x \), \( P(C = c_j | X = x) \). The Bayes rule specifies how this conditional probability can be computed from the conditional probabilities of seeing particular feature vectors for each class, \( P(x|c) \), and the unconditional probability of seeing each class. By simplifying notations, we have:

\[
P(c_j|x) = P(c_j) \frac{P(x|c_j)}{P(x)}.
\] (11)

If \( P(c_j|x) \) is known, classification would be done with many different measures, but estimation of \( P(c_j|x) \) directly is difficult. According to the Bayes rule, estimation of \( P(x|c_j), P(c_j) \) and \( P(x) \) leads to an estimate of \( P(c_j|x) \).

There are a variety of options to estimate the \( P(x|c_j) \), such as maximum likelihood (ML) parameter estimation, maximum a posteriori (MAP) probability estimation, Bayesian inference, maximum entropy (ME) estimation, and the expectation maximization (EM) algorithm, which have their own benefits and costs. In this research the PDF is estimated by two different density estimations (normal and multivariate-multinomial) over feature input spaces using the ME algorithm. Results of classification are shown in Figure 5. From Figure 5a, classification improvement does not show an absolute ascending trend when normal distribution is used, but as can be seen in Figure 5b, nearly 100% accuracy is achieved using multivariate-multinomial distribution. In order to find the reason why normal distribution is not suitable for this application, a histogram of “Best features” is shown in Figure 6. The shape of histograms can be seen in Figure 6, and estimation of the histograms is obviously more precise by using multinomial distribution compared to using the normal function. That’s why multivariate-multinomial is suitable for classification; this can justify the inaccuracy of the naive Bayes classifier in the case of using a normal distribution function in this application.

4. A case study on small-scale transformers

In order to show the contribution of the proposed method with conventional PRPD approach, a case study is performed on three small-scale 1 kVA, 220 V/1000 V transformers, which are built according to TraMer and B&C diagnostics, as shown in Figure 7. The dielectric strength between LV and HV windings is more than 15 kV. One of these models (T1) has an artificially delaminated paper insulation; in the second one (T2), a sharp metallic point is placed within its paper insulation; and the third one (T3) is kept without any defect as a control sample. Figure 8 shows the schematic model of these artificial defects. PD measurement was carried out using an oscilloscope, which was set to sequence capturing in the power frequency window; an example of captured PD data is illustrated in Figure 9. PRPD patterns are developed using peak detection of PD samples. Two scenarios are considered for this case study: Scenario 1: All three models are tested individually. Scenario 2: A corona model is put in parallel with defected transformers.

4.1. Results of Scenario 1

Figure 10 shows the PRPD patterns of T1 and T2. Background noise level is determined by T3, which in this case is 5 pC at 30 kV test voltage, which is the voltage level at which T1 and T2 are tested. This level is considered as a threshold to T2 and T3 PRPD analysis to form a pure PRPD pattern as illustrated in Figure 10. T1 and T2 patterns are somewhat similar; however, they are separable by comparing the slope of patterns at
180 degrees in the power frequency window. The proposed approach is applied to single PD activities and a database of about 50,000 single PD activities is built for both T1 and T2. Table 4 shows the results of this approach. Results show that a majority of votes are for sphere cavity defect for T1 and sharp metallic point in solid insulation for T2, which complies with the PRPD pattern method.

**Table 4.** Results of PD classification using proposed approach for each scenario.

<table>
<thead>
<tr>
<th>PD database</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Air corona</td>
<td>2633</td>
<td>2197</td>
</tr>
<tr>
<td>Discharges in oil</td>
<td>13</td>
<td>112</td>
</tr>
<tr>
<td>Surface discharge</td>
<td>653</td>
<td>1329</td>
</tr>
<tr>
<td>Sphere cavity</td>
<td>45,322</td>
<td>2</td>
</tr>
<tr>
<td>Sharp metallic point</td>
<td>211</td>
<td>47,031</td>
</tr>
<tr>
<td>Metal particles in oil</td>
<td>197</td>
<td>97</td>
</tr>
<tr>
<td>Cylindrical cavity</td>
<td>1181</td>
<td>1262</td>
</tr>
<tr>
<td><strong>Total no. of PD</strong></td>
<td><strong>50,210</strong></td>
<td><strong>52,030</strong></td>
</tr>
</tbody>
</table>

### 4.2. Results of Scenario 2

Scenario 1 is repeated except that a point-plane configuration is also energized in parallel with test samples. The corona is intentionally so severe that it disturbs the PD patterns of T1 and T2. PD pattern of the point-plane configuration and superimposed PD patterns of T1 and T2 are shown in Figure 11. As mentioned before, corona activities are set to be so severe that separation of internal PD patterns is difficult. Similar to the previous
subsection, the proposed approach is employed on PD records and derived single PD pulses are used to develop a database of about 50,000 single PD activities. Table 4 shows the results of Scenario 2 using the proposed approach. It is seen that “air corona” and “sphere cavity in solid insulation” classes have major votes for the T1 model, and “air corona” and “sharp metallic point in solid insulation” have major votes for T2.

5. Summary and conclusions

In this research a new feature extraction approach based on the autocorrelation function of wavelet decomposition on single PD events was developed. Best features were selected based on tests. Classification was carried
out using normal and multivariate-multinomial distribution functions; previous works of the authors [33] show accuracies of less than 91% when just two of the best features are employed in classification, and to achieve
Figure 10. A sample of PD patterns inside T1 (a) and T2 (b).

Figure 11. Point-plane PD pattern, T1 and T2 PD patterns superimposed with corona: (a) point-plane PD pattern, (b) T1 mixed PD pattern, (c) T2 mixed PD pattern.

100% accuracy, at least 6 of the top features should be used, while results of this work showed that classification performance using the multivariate-multinomial distribution function reached 100% by employing the 3 first top features.

The advantage of the proposed method is that the data volume would considerably decrease because just single PD activities are extracted from each raw PD data, so the memory of the recording system can be easily purged. In this method, there is no need to fix a narrow window for the PD detection system as it is obvious using a narrow fixed window, may lead to loss of potentially important PD data.
The advantage of ACF application on PD wavelet components is to put these components in the same scale for all PD class data. This solves the problem of extraction of statistical features that are very dependent on starting and ending points of time domain signals. As the main contribution of this work, internal PD activities can be distinguished in noisy condition, e.g., in the existence of a corona; it can cover the concern of image processing-based PD pattern recognition techniques. Using PRPD patterns and ACF-wavelet feature extraction solved the problem of uncertainty of statistical features; for instance, in [15], statistical features were just reliably used for making sure the detected signal is not noise; however, in this work, it is shown that by application of the ACF, statistical features can be employed for detection of PD sources.

A case study was carried out to show the contribution of the proposed approach to conventional PRPD pattern analysis. From the economic point of view, the proposed PD diagnosis system, composed of both hardware and software, can be installed beside conventional PD monitoring systems in the field, trained during operation lifetime as shown in Figure 12, and considered as a double check when a PD alarm shows up. Using an AD card besides conventional PD measuring systems enables the system to get deep insight about insulation behavior of power system equipment.

![Figure 12. Structure of proposed PD diagnosis system.](image_url)

**References**


