Signal analysis and feature generation for pattern identification of partial discharges in high-voltage equipment

O. Perpiñán, M.A. Sánchez-Urán, F. Álvarez, J. Ortego, F. Garnacho

Electrical Engineering Department, EUITI-UPM
Ronda de Valencia 3, 28012 Madrid, Spain.

Abstract

This paper proposes a method for the identification of different partial discharges (PD) sources through the analysis of a collection of PD signals acquired with a PD measurement system. This method, robust and sensitive enough to cope with noisy data and external interferences, combines the characterization of each signal from the collection, with a clustering procedure, the CLARA algorithm.

Several features are proposed for the characterization of the signals, being the wavelet variances, the frequency estimated with the Prony method, and the energy, the most relevant for the performance of the clustering procedure.

The result of the unsupervised classification is a set of clusters each containing those signals which are more similar to each other than to those in other clusters. The analysis of the classification results permits both the identification of different PD sources and the discrimination between original PD signals, reflections, noise and external interferences.

Email address: oscar.perpinan@upm.es (O. Perpiñán)

The methods and graphical tools detailed in this paper have been coded and published as a contributed package of the R environment under a GNU/GPL licence.

Keywords: partial discharge, wavelet variance, prony method, feature generation, box cox transformation, clustering, visualization tools, R-project

---

Nomenclature

α_k Damping factor of the prony decomposition.

B Backward shift operator.

CLARA Clustering Large Applications.

Δt Sampling time of a signal.

$\widetilde{\mathcal{D}}_j$ jth level detail of an MRA with a DWT.

DWT Discrete Wavelet Transform.

$\widetilde{\mathcal{H}}_j^D$ The squared gain function associated with the wavelet filter at scale j.

λ_{J_0} Scale J_0 of a DWT.

MRA Multiresolution analysis.

$\nu_{X,j}^2$ Wavelet variance of the scale τ_j.

ω_k Radian frequency of the prony decomposition.

PAM Partitioning Around Medoids clustering method.

PD Partial Discharge

$r\nu^2_{X,j}$ Proportion of variance for each wavelet variance

SDF Spectral density function.

σ^2_X Variance of the time series X_t.

$\widetilde{\mathcal{S}}_{J_0}$ J_0th level smooth of an MRA with a DWT.

$S_j(f)$ Spectral density function of a time series of the DWT coefficients at scale j.

$S_X(f)$ Spectral density function of a time series X_t.

τ_j Scale j of a DWT.

$\widetilde{\mathbf{W}}$ N dimensional vector of DWT coefficients.

$\widetilde{\mathcal{W}}$ $N \times N$ real-valued matrix defining the DWT.

$\mathbf{X}$ N dimensional vector containing a real-valued time series.

X_t Real-valued time series.

z_k Complex exponentials decomposing a real-valued time series.

1. Introduction

In order to enhance the reliability of high voltage equipment, partial discharge (PD) tests have been carried out for more than 20 years. PD analysis has been used to asses the condition of the insulation of power systems with the aim to detect premature degradation in different dielectrics.

For an accurate diagnosis it is important to have a good sensitivity in the acquisition process and powerful processing tools, especially when more than one PD source is present in the test object and when the measurement is performed under noisy conditions and in the presence of electronic interferences.

The present paper deals with the identification of different PD sources combining the individual characterization of each captured signal and the unsupervised classification of the set of signals. This method provides an efficient interpretation and evaluation of the phase resolved PD patterns obtained in the PD measurement.

The method is built on three main steps:

1. A partial discharge measuring system filters and crops an electrical time series. This system is able to filter noisy data and to detect reflected signals. After this step only a group of signals is retained and provisionally labeled as original partial discharges (section 2).
2. Each signal is described with a set of features useful to divide the group of signals in different clusters, and to decide if a signal should be regarded as a partial discharge or if it can be wiped out as noise or as external interference (section 3).
3. The matrix of features which describes the collection of signals is the input of a clustering or unsupervised classification method (section 4). The result of the clustering procedure is a set of different clusters each grouping those signals which are similar in the sense of the distance computed with the features. The analysis of the characteristics of each cluster allows for the identification of different PD sources present in

the test object.

Different statistical connections between PD sources and time series patterns have been repeatedly used to link an specific discharge pattern to a PD source [1], commonly using statistical descriptors *of the collection* (or parts of it) as suitable features. Several authors [2, 3, 4] divide the voltage cycle into phase windows, determine different distributions and classify the patterns with statistical operators summarizing these distributions. In [5] the collection is characterized with several features such as phase position, magnitude or shape, while [6, 7] use two fractal features of the *whole image* for the pattern recognition.

Others authors [8, 9, 10] propose several signal descriptors to characterize *each* PD signal, although these individual features are not used for clustering techniques . The authors of the present paper, describe in [11] a parametric method where the waveform of each signal is modelled with the combination of two functions. These functions are defined with the main frequency and with several waveform parameters of the signal. The result of this parametrization is the input of a clustering method.

The present development also focuses on the characterization of *each* of the signals captured by the data acquisition system. However, instead of using a parametric approach, the model structure is not specified a priori but is instead mostly determined from data. This paper proposes a collection of features to describe the PD data relying on few assumptions. From the set of features to be detailed in section 3 the wavelet variance [12], and the frequency estimated with the Prony method [13, 14, 15] must be highlighted.

The combination of these features and the clustering algorithm (section

4) constitutes a powerful tool for a better interpretation of phase resolved PD patterns, grouping the measured and filtered signals in different clusters. The improvement in the interpretation of the acquired PD data is very useful to discriminate different PD sources present in high voltage equipment and thus more reliable test results can be achieved.

2. Measuring system and denoising procedure

Throughout the document several examples are included to illustrate the concepts and tools. These examples make use of a dataset comprising a collection of 9955 signals acquired from a partial discharge offline test on a high-voltage cable.

The experimental arrangement in our laboratory (figure B.1) is composed of three sections of XLPE power cable 12/20 kV with two joints. The length of the test sample is 1500 m. A resonant high voltage generator energizes the test sample at the nominal voltage of the cable with a frequency around 50 Hz.

This test sample presents surface discharges, and also corona from the aerial connection. Besides, the resonant generator injects interferences due to the commutation of the insulated-gate bipolar transistors (IGBTs). Consequently, these interferences are synchronized with the sinusoidal signal from the generator (figure B.1).

The PD pulses are captured with a non-intrusive sensor (a high frequency current transformer, HFCT) placed at the beginning of the power cable. This HFCT sensor is connected to the PD measuring device. The measuring device includes a denoising filter tool and is able to separate the reflected signals

from the original PD pulses. Reflections are identified using a time window after each original PD signal. The time size of this window is adjusted according to the length of the circuit and to the amplitude of the PD signal. Those pulses received inside the time window related to the original PD signal with lower amplitude are classified as reflections. These pulses are removed before the clustering procedure to improve its performance [11, 16].

The figure B.2a displays a phase resolved partial discharge pattern of this dataset previous to the filtering procedure. The energy values have been previously transformed with the techniques described at section 3.5. Part of the signals have been labeled as reflections as a result of the previous analysis. The figure B.2b displays the original dataset *without* the reflections. This figure will be compared with the results of the clustering method (figure B.3).

3. Features generation

The goal of the feature generation is to discover compact and informative representations of the obtained data. Since our subsequent interest is to break the collection of signals into useful groups, the features should lead to large between-class distance and small within-class variance in the feature vector space. This means that features should take distant values in the different classes and closely located values in the same class [17]. Moreover, due to the large amount of data, there is a compromise between accurate representation and computational complexity.

In a clustering context with no previous class labels for patterns, the feature selection method involves a trial-and-error process: various subsets of features are selected, the resulting patterns clustered, and the output

evaluated using a validity index [18].

Several features have been previously proposed for partial discharges [1], commonly using statistical descriptors *of the collection* (or parts of it) as suitable features.

Several authors [2, 3, 4] divide the voltage cycle into phase windows representing the phase angle axis and determine three distributions: the pulse count distribution, which represents the number of observed discharges in each phase window as a function of the phase angle; the mean pulse height distribution, which represents the average amplitude in each phase window as a function of the phase angle; the maximum pulse height distribution, showing the maximum discharge magnitude in each window. Later on, these distributions are summarized with statistical operators as skewness and kurtosis, whose results are used for the classification procedure.

In [5] various descriptors as phase position, magnitude, shape, inception symmetry, pulse distribution, range, density and magnitude consistency are related, while [6, 7] use two fractal features (fractal dimension and lacunarity) of the *whole image* for the pattern recognition.

Others authors propose several signal descriptors to characterize *each* partial discharge signal in the context of PD tests on power transformers and gas-insulated substations with UHF sensors [8, 9, 10], although these individual features are not used for clustering techniques. The authors of the present paper describe in [11] a parametric method where the waveform of each PD pulse is modelled with the combination of two functions. These functions are defined with the main frequency of the PD pulse and with several waveform parameters. The results of the model are classified with a

clustering method and an artificial neural network.

The present development also focuses on the characterization of *each* of the signals captured by the data acquisition system. However, instead of using a parametrical approach, the model structure is not specified a priori but is instead mostly determined from data. This paper proposes a collection of features to describe the PD data relying on very few assumptions. It must be highlighted that the increased robustness of this non-parametric approach usually requires a large sample size to draw useful conclusions. This drawback is not a problem since PD measurements usually provide large datasets.

As previously stated, the feature selection method involves a trial-and-error process: encouraging results have been obtained combining the clustering procedure detailed at section 4 with these features[1]:

- Wavelet variances at different scales (subsection 3.1).
- Frequency and damping factors (subsection 3.2).
- Zero-crossing rate (subsection 3.3).
- Energy (subsection 3.4).
- Range (subsection 3.4).
- Maximum location (subsection 3.4).
- Length or number of samples (subsection 3.4).

These features are a compromise between accurate representation and computational complexity. The most computationally expensive features are

[1]These features can be computed with the `analysis` function of the `pdCluster` package

the wavelet variance and the frequency and damping factors, but they are also the most informative for the clustering techniques. By means of variable importance measures the candidate features can be compared with respect to their impact in the clustering results. An analysis of variable importance with random forests [19, 20] identifies the wavelet variances, frequency and energy as the most relevant features for the clustering method.

3.1. Wavelet variance

The theoretical development of this section follows the book of Percival and Walden [21].

Let $\mathbf{X}$ be an N dimensional vector containing the real-valued time series $X_t : t = 0, \ldots, N-1$. The Discrete Wavelet Transform (DWT) of $\mathbf{X}$ is a transform given by $\widetilde{\mathbf{W}} = \widetilde{\mathcal{W}}\mathbf{X}$ [21]. It is a linear filtering operation which produces a set of time-dependent wavelet and scaling coefficients which are related to variations over a set of scales.

The time series $\mathbf{X}$ can be recovered from the DWT with a multiresolution analysis (MRA). In this analysis $\widetilde{\mathcal{D}}_j$ is jth level detail and $\widetilde{\mathcal{S}}_{J_0}$ is the J_0th level smooth. The jth level detail is associated to a scale $\tau_j \equiv 2^{j-1}$ while the J_0th level smooth is related to the scale $\lambda_{J_0} \equiv 2^{J_0}$.

$$\mathbf{X} = \widetilde{\mathcal{W}}^T\widetilde{\mathbf{W}} = \sum_{j=1}^{J_0} \widetilde{\mathcal{W}}_j^T\widetilde{\mathbf{W}}_j + \widetilde{\mathcal{V}}_{J_0}^T\widetilde{\mathbf{V}}_{J_0} \equiv \sum_{j=1}^{J_0} \widetilde{\mathcal{D}}_j + \widetilde{\mathcal{S}}_{J_0} \tag{1}$$

Besides, the energy decomposition of $\mathbf{X}$ can be defined *only* in terms of the wavelet and scaling coefficients:

$$|\mathbf{X}|^2 = |\widetilde{\mathbf{W}}|^2 = \sum_{j=1}^{J_0} |\widetilde{\mathbf{W}}_j|^2 + |\widetilde{\mathbf{V}}_{J_0}|^2 \tag{2}$$

The spectral analysis with the spectral density function, designed to work with stationary processes, cannot be directly used on nonstationary processes [22]. As an alternative, the wavelet variance, defined both for stationary and nonstationary processes with dth order stationary backward differences, provides a consistent summary of the information contained in the spectral density function (SDF) on a octave band basis [21, 12].

Using the energy decomposition of the DWT (equation 2), the contribution of each scale to the variance of the time series X_t, $\sigma_X^2 = \frac{1}{N}|\mathbf{X}|^2 - \overline{X}^2$ can be determined with[2]

$$\sigma_X^2 = \frac{1}{N}\sum_{j=1}^{J_0}|\widetilde{\mathbf{W}}_\mathbf{j}|^2 \tag{3}$$

since $\frac{1}{N}|\widetilde{\mathbf{V}}_{\mathbf{J_0}}|^2 \simeq \overline{X}^2$.

Suppose that X_t is a stochastic process whose dth order backward difference, $Y_t = (1-B)^d \cdot X_t$, is a stationary process (where d is a nonnegative integer and B is the backward shift operator defined by $BX_t = X_{t-1}$ and $B^k X_t = X_{t-k}$). Then, the result of the DWT of X_t, if an adequate wavelet filter is used, is a set of coefficients time series, $\widetilde{W}_{j,t}$, which can be regarded as stationary processes whose SDF can be defined by:

$$S_j(f) = \widetilde{\mathcal{H}}_j^D(f) \cdot S_X(f) \tag{4}$$

where $\widetilde{\mathcal{H}}_j^D$ is the squared gain function associated with the wavelet filter at scale j [21].

[2]The limits of this sum must be corrected to include only those coefficients which are not subject to to reflection conditions and then obtain an unbiased estimator of the variance. However, for ease of exposition, the equation includes the whole set of coefficients.

If we denote with $\nu^2_{X,j}$ the wavelet variance of the scale τ_j, since the variance of a stationary process is equal to the integral of its SDF, then:

$$\nu^2_{X,j} = \int_{-1/2}^{1/2} S_j(f)\mathrm{df} = \int_{-1/2}^{1/2} \widetilde{\mathcal{H}}^D_j \cdot S_X(f)\mathrm{df} \tag{5}$$

Therefore, the variance of $\mathbf{X}$ is

$$\sigma^2_X = \sum_{j=1}^{J_0} \nu^2_{X,j} \tag{6}$$

The wavelet variance of each scale represents the contribution to the SDF in the correspondent octave. The width of the octave of the scale τ_j is $1/(2^{j+1}\Delta t)$, where Δt is the sampling time of the signal, and the frequency band of this scale is $1/(2^{j+1}\Delta t) \leq f \leq 1/(2^j \Delta t)$. The average value of SDF over this interval is:

$$\overline{S}_{X,j} = 2^{j+1} \cdot \Delta t \cdot \int_{\frac{1}{2^{j+1}\Delta t}}^{\frac{1}{2^j \Delta t}} S_X(f)\mathrm{df} \tag{7}$$

Thus, the wavelet variance can be used as an estimator of the average value of the SDF [21, 12]:

$$\nu^2_{X,j} = \frac{\overline{S_{X,j}}}{2^j \Delta t} \tag{8}$$

The number of decomposition levels are limited by the length of the signal. A minimum common number of levels must be defined for the collection of signals, since they are different in length (for example, the first four levels of wavelet variance).

Finally, the wavelet variance depends on the total variance of the signal (equation (6)). Since another feature is measuring a similar value (energy, subsection 3.4) the final wavelet variance feature[3], $r\nu^2_{X,j}$, is the proportion of variance for each wavelet variance:

$$r\nu^2_{X,j} = \frac{\nu^2_{X,j}}{\sigma^2_X} \tag{9}$$

3.2. Frequency and damping factor: the Prony method

The classical frequency estimation methods, based on periodograms and FFT algorithms, have a low computational cost. However, they share several drawbacks, namely aliasing, resolution bias error or picket-fence effect and leakage [23]. On the other hand, with a slight increase in the computational cost, a parametric method is able to obtain high-accuracy frequency estimates working with a relatively short data set [24]. Among these parametric methods the Prony's method is a good compromise between accuracy and computational effort.

A high voltage cable can be represented using the transmission line model, with an infinite series of distributed resistance, inductance, capacitance and conductance components. The interaction of a partial discharge phenomenon with such an structure results in a complex combination of attenuation, dispersion and reflection effects. Under these conditions, a clean partial discharge signal can be regarded as a finite combination of damped complex exponentials.

[3]This feature can be computed with the `wavVarPD` function of the `pdCluster` package.

In this context, the Prony's method allows for the estimation of frequency, amplitude, phase and damping components of the signal [13, 14, 15]. Let $y(t)$ be a time series, and y_n the nth element of its discrete counterpart. If the Prony assumption holds for y_n, then it can be decomposed with L complex exponentials z_k:

$$y_n = \sum_{k=1}^{L} a_k z_k^n \tag{10}$$

with a_k as complex constants, and $n \in (1, 2, \ldots, N)$ where N is the number of samples of the time series.

The equation (10) can be solved with a weighted sum of a set of L past values of the signal:

$$y_n = \sum_{i=1}^{L} -c_i y_{n-i} \tag{11}$$

Combining and rearranging equations (10) and (11) the following expression results:

$$\sum_{k=1}^{L} a_k \sum_{i=0}^{L} c_i z_k^{n-i} = 0 \tag{12}$$

where $c_0 = 1$. Since $z^{n-i} = z^{n-L} \cdot z^{L-i}$ this equation can be rewritten as:

$$\sum_{k=1}^{L} a_k z_k^{n-L} \cdot \sum_{i=0}^{L} c_i z_k^{L-i} = 0 \tag{13}$$

A solution for the equation (13) is

$$\sum_{i=0}^{L} c_i z_k^{L-i} = 0 \tag{14}$$

with $c_0 = 1$. Therefore, the z_k complex exponentials that decompose the signal y_n with the equation (10) are the roots of the polynomial of the equation (14), whose coefficients c_i are the same of the sum of past values (equation (11)).

The z_k complex exponentials can be expressed as $z_k = \exp(\alpha_k + i\omega_k)$. Here α_k are the damping factors and ω_k the radian frequencies. Thus, the Prony's method is able to provide the amplitude and phase (module and argument of the complex coefficients a_k), and the damping factors and radian frequencies for each of the signals composing the time series.

The procedure for the Prony's method[4] is as follows:

1. With a fixed number for L, solve the linear equation (11). The solution is the set of coefficients c_i of the polynomial of the equation (14).
2. Solve the polynomial of the equation (14). The solution is the set of complex exponentials z_k, and therefore the set of damping factors α_k and radian frequencies ω_k.
3. The complex amplitudes a_k are the solution of the system defined by the equation (10). This is an overdetermined system since $N > L$. An approximate solution can be obtained via a least squares method.

As an example of the procedure, the figure B.4 shows a signal (black line) approximated by several Prony estimations with different number of complex exponentials.

A drawback of this procedure is that the number of complex exponentials L must be chosen *a priori*. It is possible to improve the method to decide the number of components with a singular value decomposition (SVD)[5] and to cope with noisy signals [15]. However, the procedure as previously exposed with L being high enough is recommendable because:

[4] This method is implemented under the `prony` function of the `pdCluster` package.

[5] http://en.wikipedia.org/wiki/Singular_value_decomposition

- The signals are previously denoised with a wavelet filtering method [11, 16].
- Our main interest are only the two or three most important components of the signal, being those with the highest relative energy of the set.

3.3. Zero-crossing rate

The zero-crossing rate[6] is defined as [17]:

$$\text{zcr} = \frac{1}{2 \cdot N} \sum_{n=1}^{N} |\text{sign}(x_n) - \text{sign}(x_{n-1})| \tag{15}$$

where the sign operator is defined as:

$$\text{sign}(x_n) = \begin{cases} 1 & x_n \geq 0 \\ -1 & x_n < 0 \end{cases} \tag{16}$$

This is a basic feature that can be computed easily and provides information about the oscillation speed of the signal.

3.4. Other features

There are other remarkable features included in the set: energy, range, location of the maximum value and length of the signal.

The energy is a very simple feature that can be calculated with:

$$E_x = \sum_{n=1}^{N} x_n^2 \tag{17}$$

The range, $|\max(x) - \min(x)|$, the length of the signal, N, and the maximum location are easily computable features that do not deserve additional explanations.

[6] This feature can be computed with the `nZC` function of the `pdCluster` package.

3.5. Transformations

This set of features provides a matrix of values whose distributions functions are strongly positively skewed. Thus, before using this matrix with the clustering algorithm a transformation is recommended [17].

The matrix of features must be conveniently cleaned to wipe out possible outliers[7] and transformed with the family of the Box-Cox power functions[8] [25]. These functions create a rank-preserving transformation and are recognized as a useful data pre-processing technique used to stabilize variance and make the data more normal distribution-like.

Let x be the original feature and x_λ the transformed feature. Then:

$$x_\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases} \tag{18}$$

where λ is calculated for each feature with the Box-Cox method [25].

As an example, the figure B.5 displays the histograms of the zero-crossing rate for the partial discharge dataset before and after a Box-Cox transformation. Two groups can be clearly distinguished after the transformation.

4. Clustering

The subsequent step in this analysis is the unsupervised classification or clustering of the signal data. There is a wide variety of clustering techniques [1, 18]. The CLARA algorithm (Clustering Large Applications) [26], a member of the partitioning methods has been chosen. This method is a variant of

[7] The function `filterPD` of the `pdCluster` package implements this functionality.

[8] Available with the `transformPD` function of the `pdCluster` package.

the Partitioning Around Medoids algorithm, PAM, which searches for k representative objects, called *medoids*, among the objects of the data set. These *medoids* are computed such that the total dissimilarity of all objects to their nearest medoid is minimal. The CLARA algorithm is more indicated for large datasets since it does not store the entire dissimilarity matrix. Moreover, it is possible to construct a quasi-fuzzy technique with consecutive calls to this algorithm.

The figure B.3 shows an energy phase resolved pattern[9] using a combination of colours and transparency to show membership and density of points in figure B.3a, and encoding the distance to the medoid of each cluster with colors in figure B.3b. Those points labeled as reflections (figure B.2b) were removed before entering the clustering procedure and are not shown in the figures.

The figure B.3b displays a separated phase resolved PD pattern for each cluster. These individual phase resolved patterns are the signatures generated either by the defects present in the test object or by external interferences. In most cases, the analysis of these patterns permits the identification of the defects and to discriminate different interferences sources or unfiltered background noise.

The relation of each cluster with the defects of the test sample arrangement (section 2) can be deduced from figures B.3. Cluster no.2 comes from surface discharges (asymmetrical behaviour in both semicycles of the voltage waveform and an arched shape in the positive semicycle) and cluster no.5 is

[9]The function `claraPD` and the method `xyplot` of the package `pdCluster` implement this functionality.

related to the corona effect (flat pattern in the negative half cycle shifted to the left side of the crest in the voltage waveform). Cluster no.3 is the result of the interference from the IGBTs of the voltage generator (pulses occur at fixed points located over the zero crossings of the voltage waveform). Finally, clusters no.1 and no.4 are noise data and reflections (flat patterns all along the period of the voltage waveform, no pulses synchronized with the applied voltage) which could not be removed by the filtering procedure (section 2) [11, 16].

The figure B.6 displays the density estimates for each feature and cluster. Colours of figures B.3a and B.6 identify the same cluster as determined by the CLARA algorithm. It must be noted that the x-axis shows values previously transformed with the techniques of section 3.5.

This figure is a useful tool to understand the information provided by each feature and the results of the clustering algorithm. As previously said (section 3), the analysis of variable importance with random forests [19, 20] identifies the wavelet variances, frequency and energy as the most relevant features for the clustering method. This figure explains graphically these results: these features generate density functions whose shape and location are noticeable different for each cluster.

The frequency feature produces sharp, unimodal and distant functions. Clusters 2 and 5 (surface and corona defects) are located at the upper extreme, clusters 1 and 4 (noise and reflections) at the lower extreme, and cluster 3 (IGBTs) at the middle.

The wavelet variances mimic this behaviour although with flatter and more diffuse functions. It is interesting to note that the relative position

between the functions of wavelet variances of clusters 2 and 5 change from the first to the fourth scale.

These two clusters are clearly separated with the energy feature, with sharp functions which almost do not intersect. The function of the cluster 3 is located at the upper extreme and, again, the functions of clusters 1 and 4 can be found at the lower extreme.

It must be underlined that none of the variables is able to separate all the clusters by itself. A cluster can be more clearly visible using a particular feature (for example, the cluster no.3 is easily identified with the frequency feature) but the rest of the clusters are not so well defined with this same variable. The adequate results are obtained with the cooperation of the set of features. For example, the clusters no.2 and no.5 are the most difficult (and interesting) to separate. Although they intersect in several features, their behaviors are different from feature to feature and so the clustering algorithm is able to distinguish them.

5. Conclusion

This paper proposes a method for the identification of different PD sources combining the characterization of each signal acquired with a PD measurement system, and the unsupervised classification of the collection of signals.

The collection of signals is the output of a filtering and cropping procedure applied to the electrical time series produced by the PD measurement system. Each signal of the collection is described with a set of features using a non-parametric approach. These features, a compromise between accurate representation and computational complexity, have been chosen to combine

adequately with the clustering method, the CLARA algorithm.

A variable importance analysis with random forests identify the wavelet variances, the frequency estimated with the Prony method and the energy, as the most relevant features for the clustering method.

This set of features provides a matrix of values whose distributions functions are strongly positively skewed. Thus, before entering the clustering procedure, the matrix is transformed with the family of the Box-Cox power functions. Groups which were hidden in the original distributions are clearly distinguished after the transformation.

The unsupervised classification procedure generates a set of clusters each grouping those signals more similar in the sense of the distance computed with the matrix of features. The analysis of the classification results, both with a phase resolved pattern graphic and with the estimation of the probability density functions, permits the identification of the different PD sources and to discriminate between original PD signals, reflections, external interferences and noise.

6. Software tool

The methods and graphical tools here exposed have been coded in the R environment [27] and published as a contributed package [28] under a GNU/GPL licence. This package includes a reference manual and an introductory document with examples.

Bibliography

[1] N. Sahoo, M. Salama, R. Bartnikas, Trends in partial discharge pattern classification: a survey, IEEE Transactions on Dielectrics and Electrical Insulation 12 (2) (2005) 248 – 264. doi:10.1109/TDEI.2005.1430395.

[2] E. Gulski, F. Kreuger, Computer-aided recognition of discharge sources, IEEE Transactions on Electrical Insulation 27 (1) (1992) 82 –92. doi:10.1109/14.123443.

[3] A. Krivda, Recognition of discharges: discrimination and classification, Ph.D. thesis, Delft University (1995).

[4] E. Gulski, Computer-aided recognition of partial discharges using statistical tools, Ph.D. thesis, Delft University (1991).

[5] S. Rudd, S. D. J. McArthur, M. D. Judd, A generic knowledge-based approach to the analysis of partial discharge data, IEEE Transactions On Dielectrics and Electrical Insulation 17 (1) (2010) 149–156.

[6] L. Satish, W. Zaengl, Can fractal features be used for recognizing 3-d partial discharge patterns, IEEE Transactions on Dielectrics and Electrical Insulation 2 (3) (1995) 352 –359. doi:10.1109/94.395421.

[7] A. Krivda, E. Gulski, L. Satish, W. Zaengl, The use of fractal features for recognition of 3-D discharge patterns, IEEE Transactions on Dielectrics and Electrical Insulation 2 (5) (1995) 889 –892. doi:10.1109/94.469983.

[8] M. Judd, L. Yang, I. Hunter, Partial discharge monitoring for power transformer using UHF sensors. Part 2: field experience, IEEE Electrical Insulation Magazine 21 (3) (2005) 5 –13. doi:10.1109/MEI.2005.1437603.

[9] M. Judd, L. Yang, I. Hunter, Partial discharge monitoring of power transformers using UHF sensors. Part I: sensors and signal interpretation, IEEE Electrical Insulation Magazine 21 (2) (2005) 5 –14. doi:10.1109/MEI.2005.1412214.

[10] M. D. Judd, Radiometric partial discharge detection, Proceedings of 2008 International Conference On Condition Monitoring and Diagnosis (2007) 1025–1030.

[11] M. Sánchez-Urán, J. Ortego, F. Álvarez, O. Perpiñán, E. Puelles, R. Moreno, D. Prieto, D. Ramos, New procedure to determine insulation condition of high voltage equipment by means of PD measurements in service, in: CIGRE 2012, Vol. D1-309, 2012.
URL http://www.cigre.org/content/download/16814/679705/version/1/file/D1_309_2012.pdf

[12] D. P. Percival, On estimation of the wavelet variance, Biometrika 82 (3) (1995) 619–631.
URL http://staff.washington.edu/dbp/PDFFILES/wavevar.pdf

[13] R. Kumaresan, D. Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise, IEEE Transactions

on Acoustics, Speech and Signal Processing 30 (6) (1982) 833 – 840. doi:10.1109/TASSP.1982.1163974.

[14] J. Hauer, C. Demeure, L. Scharf, Initial results in prony analysis of power system response signals, IEEE Transactions on Power Systems 5 (1) (1990) 80 –89. doi:10.1109/59.49090.

[15] R. Kumaresan, D. Tufts, L. Scharf, A prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models, Proceedings of the IEEE 72 (2) (1984) 230 – 233. doi:10.1109/PROC.1984.12849.

[16] M. Sánchez-Urán, G. F., J. Ortego, J. Moreno, J. Álvarez, O. Perpiñán, J. Gonzalo, J. Vallejo, PD monitoring systems of H.V. cable systems by means of powerful digital tools to discriminate noise and to perform efficient PD diagnosis, in: 8th International Conference on Power Insulated Cables, 2011.
URL http://jicable.org/2011/index.php

[17] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Fourth Edition, Academic Press, 2009.

[18] A. K. Jain, M. N. Murty, P. J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323. doi:http://doi.acm.org/10.1145/331499.331504.
URL http://doi.acm.org/10.1145/331499.331504

[19] L. Breiman, Random Forests, Machine Learning 45 (1) (2001) 5–32.

doi:10.1023/A:1010933404324.
URL http://dx.doi.org/10.1023/A:1010933404324

[20] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis, Conditional variable importance for random forests, BMC Bioinformatics 9 (1) (2008) 307. doi:10.1186/1471-2105-9-307.
URL http://www.biomedcentral.com/1471-2105/9/307

[21] D. Percival, A. T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press, 2006.

[22] M. B. Priestley, Evolutionary spectra and non-stationary processes, Journal of the Royal Statistical Society. Series B (Methodological) 27 (1965) 204–237.

[23] A. Girgis, F. Ham, A quantitative study of pitfalls in the fft, IEEE Transactions on Aerospace and Electronic Systems AES-16 (4) (1980) 434 –439. doi:10.1109/TAES.1980.308971.

[24] I. Santamaría, C. Pantaleón, J. Ibañez, A comparative study of high-accuracy frequency estimation methods, Mechanical Systems and Signal Processing 14 (5) (2000) 819 – 834. doi:DOI:10.1006/mssp.2000.1321.

[25] G. E. P. Box, D. R. Cox, An analysis of transformations, Journal of the Royal Statistical Society. Series B (Methodological) 26 (2) (1964) pp. 211–252.
URL http://www.jstor.org/stable/2984418

[26] A. Struyf, M. Hubert, P. Rousseeuw, Clustering in an object-oriented

environment, Journal of Statistical Software 1 (4) (1997) 1–30.
URL http://www.jstatsoft.org/v01/i04

[27] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2011).
URL http://www.R-project.org

[28] O. Perpinan, M. A. Sanchez-Uran, pdCluster: Partial Discharges Clustering (2011).
URL http://pdcluster.r-forge.r-project.org/

[29] D. B. Carr, R. J. Littlefield, W. L. Nicholson, J. S. Littlefield, Scatterplot matrix techniques for large N, Journal of the American Statistical Association 82 (398) (1987) pp. 424–436.
URL http://www.jstor.org/stable/2289444

[30] W. S. Cleveland, The elements of graphing data, AT&T, 1994.

[31] E. R. Tufte, The visual display of quantitative information, Graphic Press, 2001.

[32] E. R. Tufte, Envisioning information, Graphic Press, 1990.

[33] M. Friendly, D. Denis, The early origins and development of the scatterplot, Journal of the History of the Behavioral Sciences 41 (2) (2005) 103–130. doi:10.1002/jhbs.20078.

Appendix A. Graphical exploratory tools

This section provides information about graphical tools suitable for the exploration of PD datasets. These large multivariate datasets are not adequately represented with conventional approaches. Our first proposal combines the conventional phase resolved patterns with the hexagonal binning, and the second proposal shows relation between variables with enhanced scatterplot matrices built upon hexagonal binning.

The display of a large number of points in a scatter plot produces hidden point density, long computation times for selected enhancement operations, and slow displays. These problems can be circumvented with the estimation and representation of points densities. A common encoding uses gray scales, pseudo colors or partial transparency. An improved scheme encodes density as the size of hexagon symbols inscribed within hexagonal binning region [29].

The figure B.7 shows a phase resolved partial discharge pattern build with a hexagonal binning representation. Darker blues are used for bins with tens of points, while lighter blues denote those bins where less than ten points are found. Clusters of points are clearly visible.

The scatterplot matrices [30] are based on the technique of *small multiples* [31, 32]: small, thumbnail-sized representations of multiple images displayed all at once, which allows the reader to immediately, and in parallel, compare the inter-frame differences. A scatterplot matrix is a display of all pairwise bivariate scatterplots arranged in a $p \times p$ matrix for p variables. Each subplot shows the relation between the pair of variables at the intersection of the row and column indicated by the variable names in the diagonal panels [33].

The figure B.8 shows an enhanced version of this tool using a feature matrix from a collection of partial discharges:

- The diagonal panels include the univariate density function estimation for each feature.
- The lower panels include a local polynomial regression fitting (*loess*) [30] to ease the identification of relations between variables.
- Each panel display represent densities of points with a hexagonal binning, as described above.

With this dataset, several connections can be observed. The wavelet variances are linearly related between them. Another linear relation can be found between the range and the energy. The frequency, the zero-crossing rate and the wavelet variances are also linearly related although with a higher dispersion. These variables are connected with the damping factor but in a non-linear fashion.

Appendix B. Figures

List of Figures

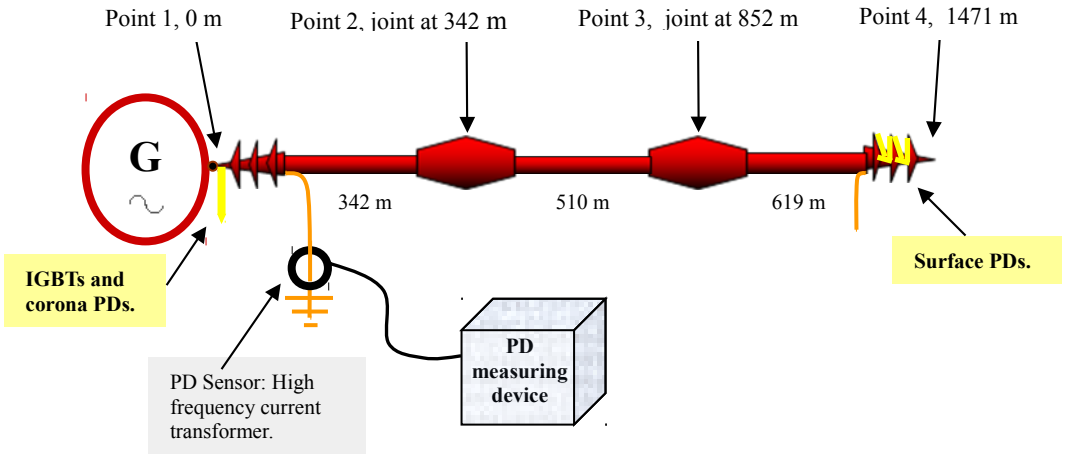
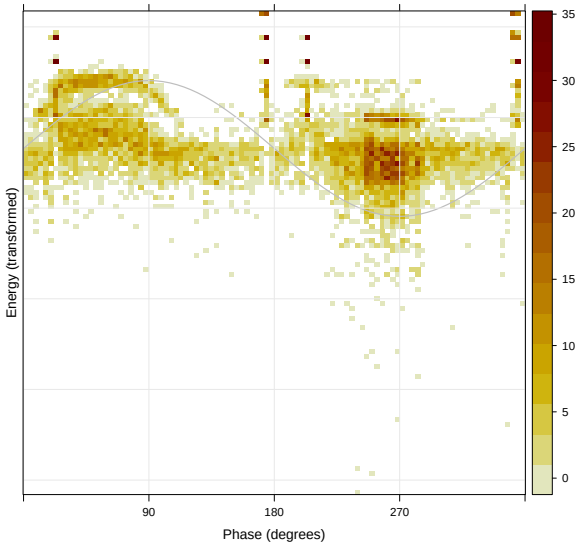
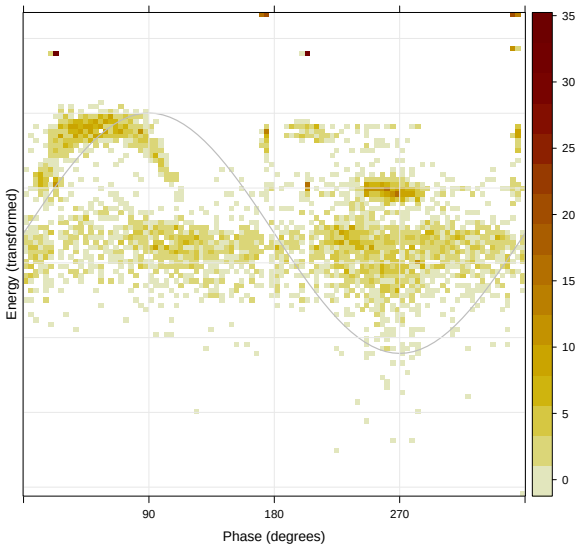


Figure B.1: Layout of the experimental arrangement used for the partial test on a high-voltage cable. The resonant high voltage generator, source of electronic interferences due to IGBTs, is located at point 1. There is also a corona defect at this point 1, and a surface defect is present at point 4.

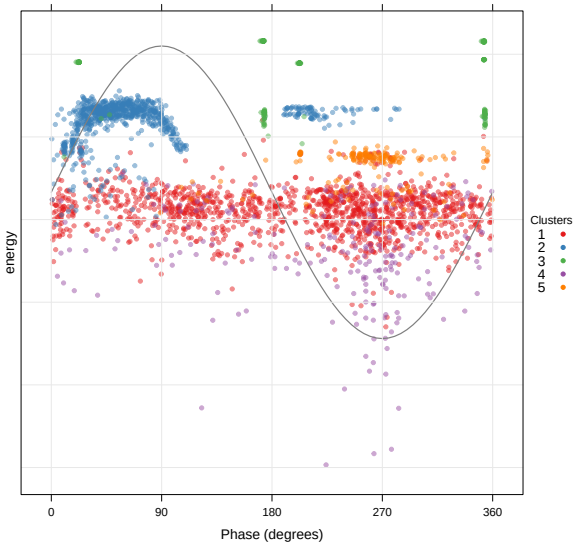


(a) With reflections

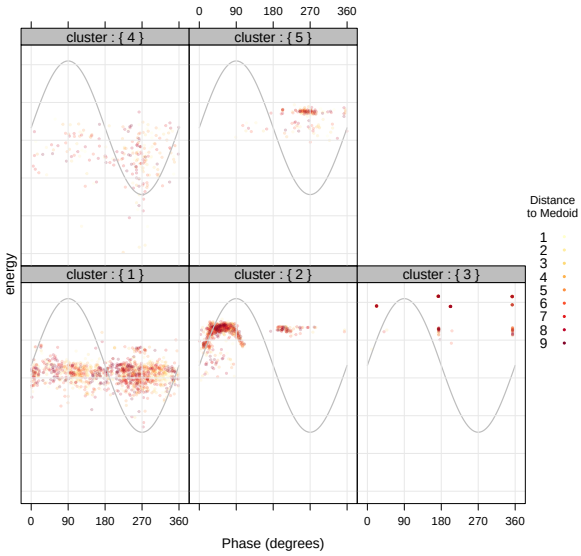


(b) Without reflections

Figure B.2: Phase resolved partial discharge pattern of the dataset. The y-axis shows the energy feature as described in the section 3.4. This feature has been previously transformed with the tools detailed at section 3.5. The color encoding represents the number of samples per bin.



(a) Five different clusters determined by the CLARA algorithm represented together. Colours indicate membership to a cluster and intensity denotes density of points in a region.



(b) Separated phase resolved PD pattern for each cluster. Colours indicate distance to the medoid of each cluster (9 is near, 1 is far).

Figure B.3: Phase resolved partial discharge pattern. The energy values have been previously transformed (section 3.5). Reflected signals (figure B.2b) are not shown in these figures.

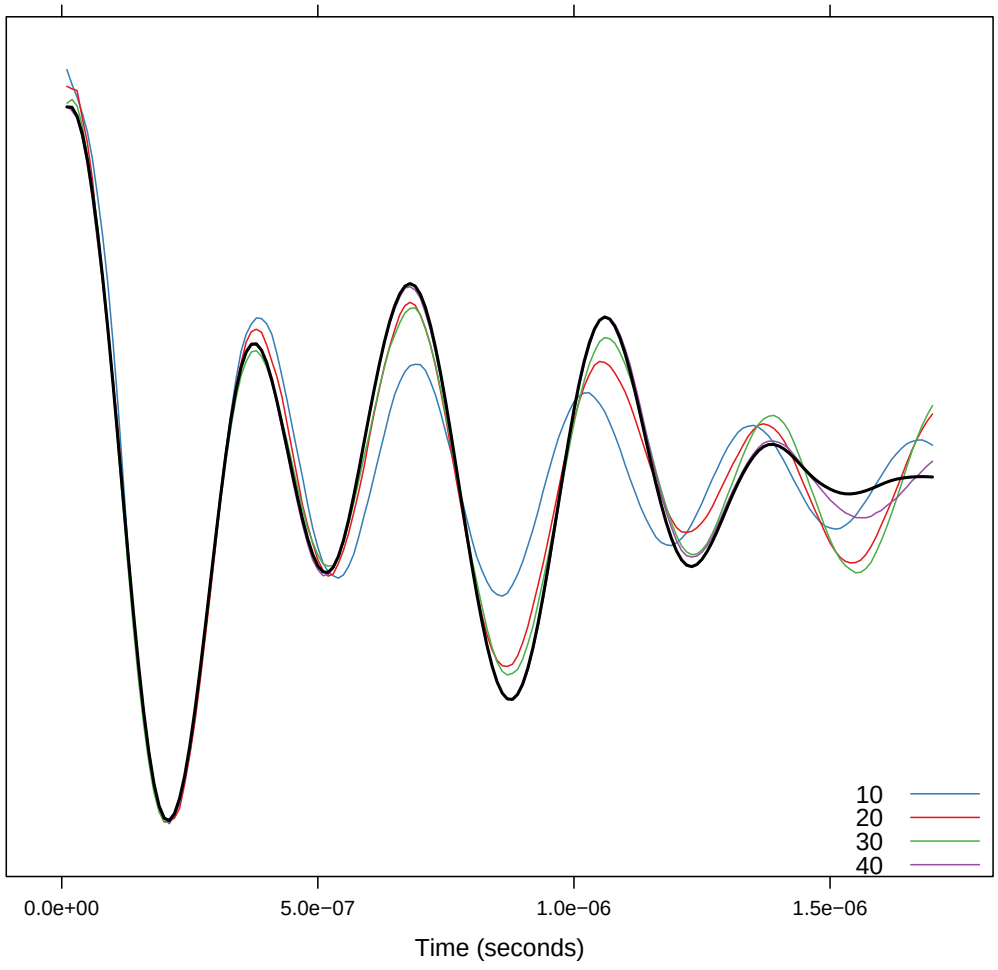


Figure B.4: Prony method with different number of complex exponentials.

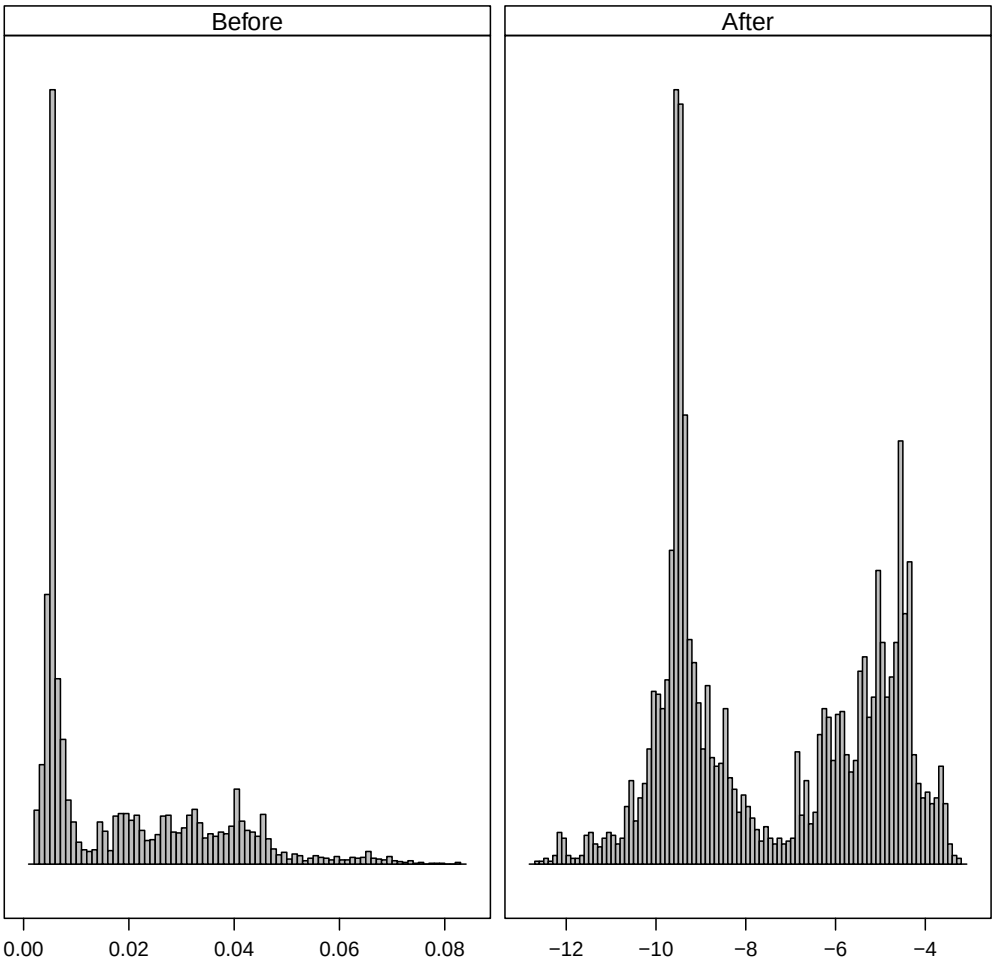


Figure B.5: Histograms for the zero-crossing rate feature of a partial discharge dataset before and after a Box-Cox transformation.

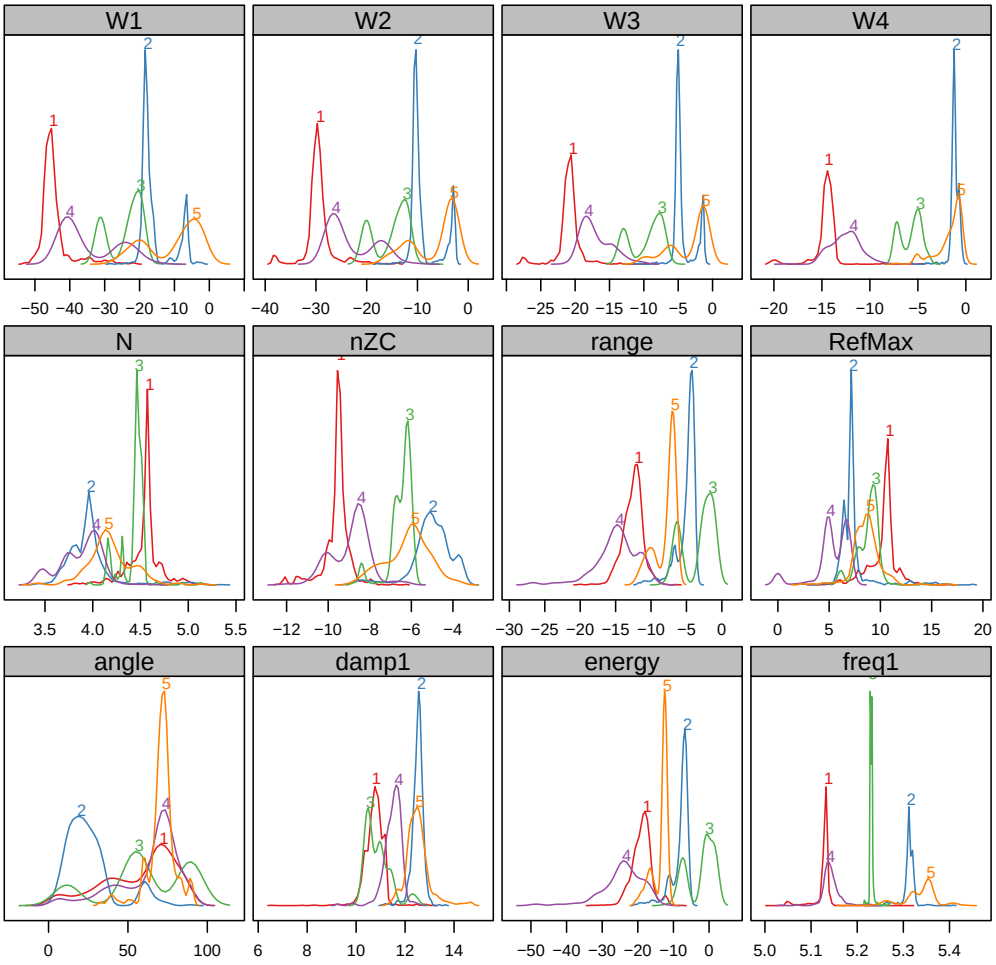


Figure B.6: Density estimates for each feature and cluster for a partial discharge dataset: `RefMax` is the maximum location, W_i is the wavelet variance corresponding to the i-th level, `N` is the number of samples of the signal, `nZC` is the zero-crossing rate, `freq1` and `damp1` are the frequency and the damping factor. Values of the x-axis are the result of the Box-Cox functions.

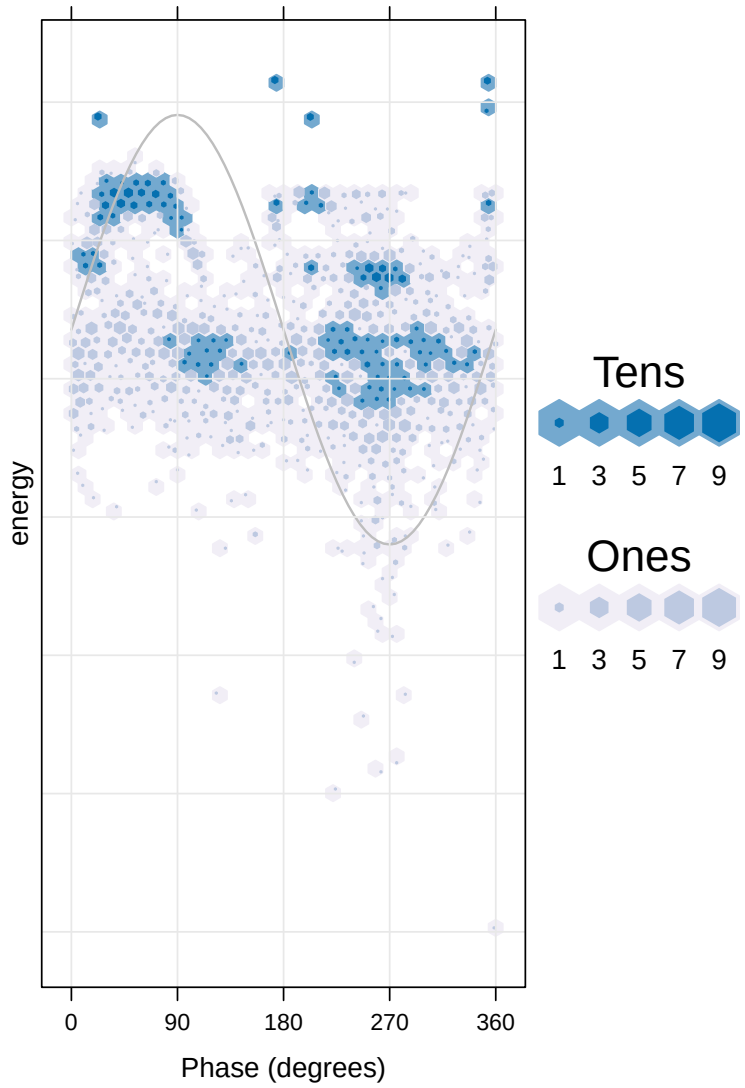


Figure B.7: Phase resolved partial discharge pattern with hexagon binning.

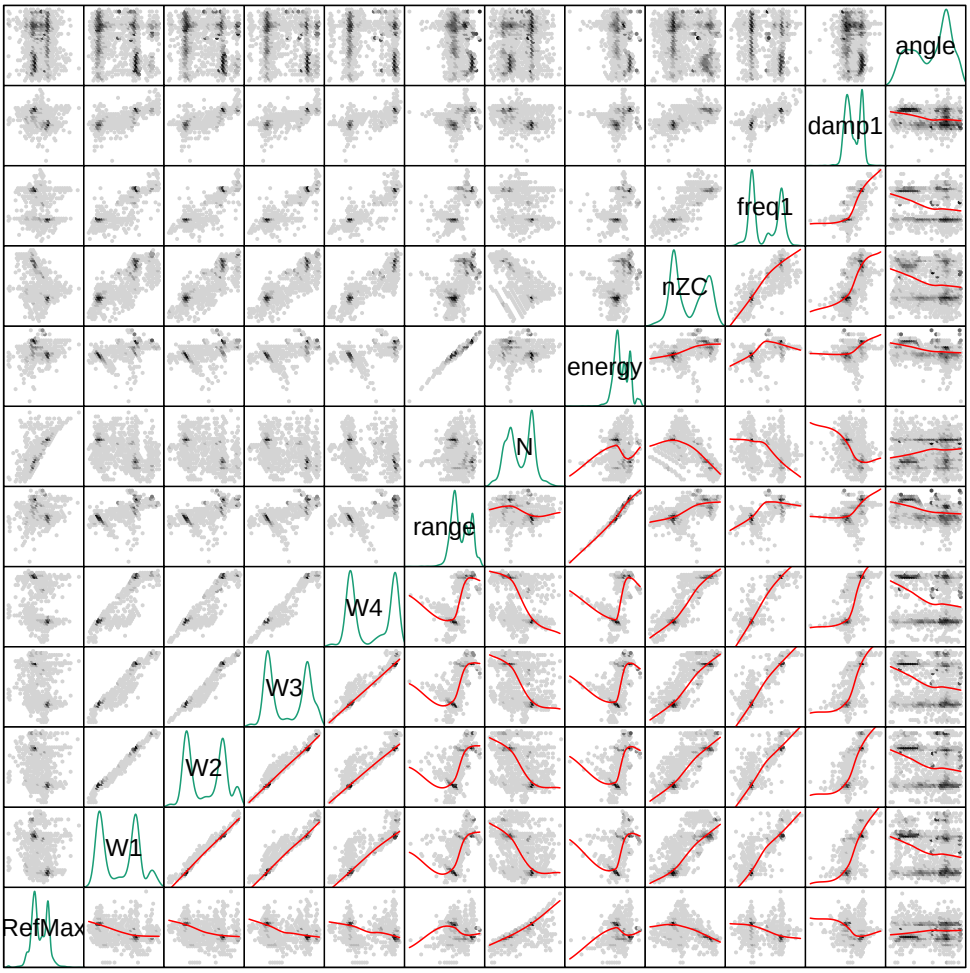


Figure B.8: Scatterplot matrix of a collection of partial discharges, where all the variables are confronted together with their kernel density estimations in the diagonal frames. `RefMax` is the maximum location, W_i is the wavelet variance corresponding to the i-th level, `N` is the number of samples of the signal, `nZC` is the zero-crossing rate, `freq1` and `damp1` are the frequency and the damping factor.