Electric Power Systems Research 81 (2011) 644–659

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Partial discharge signal denoising with spatially adaptive wavelet thresholding and support vector machines

Hilton de Oliveira Mota [a,*], Leonardo Chaves Dutra da Rocha [a,1], Thiago Cunha de Moura Salles [b], Flávio Henrique Vasconcelos [c,2]

[a] Department of Computer Science, Federal University of São João del-Rei, Visconde do Rio Branco Ave., Colônia do Bengo, São João del-Rei, MG, 36301-360, Brazil
[b] Department of Computer Science, Federal University of Minas Gerais, 6627 Antônio Carlos Ave., Pampulha, Belo Horizonte, MG, 31270-901, Brazil
[c] Department of Electrical Engineering, Federal University of Minas Gerais, 6627 Antônio Carlos Ave., Pampulha, Belo Horizonte, MG, 31270-901, Brazil

ARTICLE INFO

Article history:
Received 5 August 2009
Received in revised form 14 October 2010
Accepted 21 October 2010
Available online 16 November 2010

Keywords:
Partial discharges
Wavelet transform
SVM
Denoising
On-site measurements

ABSTRACT

In this paper an improved method to denoise partial discharge (PD) signals is presented. The method is based on the wavelet transform (WT) and support vector machines (SVM) and is distinct from other WT-based denoising strategies in the sense that it exploits the high spatial correlations presented by PD wavelet decompositions as a way to identify and select the relevant coefficients. PD spatial correlations are characterized by WT modulus maxima propagation along decomposition levels (scales), which are a strong indicative of the their time-of-occurrence. Denoising is performed by identification and separation of PD-related maxima lines by an SVM pattern classifier. The results obtained confirm that this method has superior denoising capabilities when compared to other WT-based methods found in the literature for the processing of Gaussian and discrete spectral interferences. Moreover, its greatest advantages become clear when the interference has a pulsating or localized shape, situation in which traditional methods usually fail.

1. Introduction

Partial-discharges (PDs) are electric discharges that occur inside the insulation of high voltage equipments usually due to the presence of cavities or contaminants resulting from failures on the manufacturing process, mechanical stress or ageing [1]. Due to their repetitive nature and the confinement to small regions, PDs may originate localized damages which, in the long-term, can propagate throughout the dielectric and cause a breakdown. Therefore, indiscriminate PD occurrence can lead to catastrophic results with both economical and safety implications and its monitoring is one of the most important activities performed by energy generation and transmission companies.

Manufacture and commissioning PD tests are usually straightforward since these can be performed in laboratories containing appropriate arrangements to reduce interferences, like shielded walls, proper grounding and all sorts of filtering mechanisms [2]. Nevertheless, operation stage monitoring is much more difficult to realize due to the interferences usually encountered in high voltage facilities, which include signals from radio and communication systems, switching mechanisms, power electronics, rotating machines and PDs generated by other sources [3]. Due to these impairments, on-site PD measurements demand the use of more elaborate signal processing techniques. In the last few years several methods have been proposed in the literature to address the problem, including both analog and digital approaches [4–9], but it seems that a definitive solution is not available yet. The fact is that when performing on-site PD measurements one is faced with such a diversity of signal waveshapes and sources of interference that it is a hard task to achieve a general solution to the problem.

More recently the wavelet transform (WT) was recognized as a powerful tool for PD processing due to its potential to adapt to inhomogeneous and time-localized signals. Several authors have made use of this characteristic in order to develop denoising methods that frequently outperform those based on the Fourier theory [10–19]. In their seminal work Shim et al. [10] showed the effectiveness of the discrete wavelet transform (DWT) for the processing of power cable PD signals corrupted by Gaussian white noise. In [11,12] the authors improved the technique by introducing PD-fitted methods to determine the threshold levels and presented results also for power cable measurements. In [3] the authors proposed a PD denoising method based on multi-resolution decomposition (MRD) and showed its superiority when compared to linear filters when

* Corresponding author. Present address: Paulo Piedade Campos st. no. 680/202-B, Estoril, Belo Horizonte, MG, 30455-250, Brazil. Tel.: +55 32 3373 3985.
E-mail addresses: hmota@ufsj.edu.br, hilton@cpdee.ufmg.br (H.d.O. Mota), lcrocha@ufsj.edu.br (L.C.D.d. Rocha), tsalles@dcc.ufmg.br (T.C.d.M. Salles), fvasc@ufmg.br (F.H. Vasconcelos).
[1] Tel.: +55 32 3373 3985.
[2] Tel.: +55 31 3409 4837.

0378-7796 © 2010 Elsevier B.V. Open access under the Elsevier OA license.
doi:10.1016/j.epsr.2010.10.030

processing PD from a point-plane arrangement. In [13] the authors performed a comparison of ten PD filtering techniques, including linear, time-frequency, adaptive, statistically based filters and wavelets. The processing of signals from a point-plane arrangement indicated that WT-based strategies give superior results. In [15] the authors showed the efficacy of the Wavelet Packets decomposition to the denoising of PD signals obtained from gas-insulated switchgears and high-voltage generators. In [16,17] the authors proposed an empirical method to determine threshold levels and reported successful results for the processing of PDs from power cables. In [18] the authors proposed the Second Generation Wavelet Transform as a way to better fit the shape of the wavelet to PD signals and, thus, improve its decorrelation from noise. They demonstrated its efficacy by the processing of signals from high-voltage power cables. In [19] the authors employed an alternative approach to the PD denoising by applying the Wavelet Packets decompositions to the images of phase-resolved distributions. They demonstrate its effectiveness in the processing of signals from high voltage motors and generators.

In spite of the advances seen in the recent years, there are still many open questions regarding to the WT-based PD processing methods. For example, there is no consensus about the best way to choose the wavelet family and the number of decomposition levels, mainly with characteristics fitted to partial discharges. The investigated cases are restricted to white and harmonic noises, nevertheless there is no adequate coverage of pulsating interferences which are recognizably more difficult to process due to their similarities with the PDs shape and frequency spectra [3]. Also, there is a lack of automatic adjustment methods to improve system robustness and reduce dependencies on operator skills. It is therefore important to continue the investigations to answer these and other open questions.

In this paper a novel PD denoising method is proposed. It relies on the fact that partial discharge WT decompositions have a very distinct aspect, with coefficients clustering in the same relative positions in each decomposition level (scale). It is shown that, by determining these positions, it is possible to separate the noise-related from PD-related coefficients, thus allowing a denoising procedure. The method is based on the translation-invariant discrete wavelet transform (TI-DWT) [20], a variant of DWT that avoids the loss of data caused by decimation. Spatial correlations are characterized by DWT local modulus maxima propagation along scales [21]. Denoising is performed by identification of noise-related maxima lines using a classifier based on a support vector machine (SVM) [22–24]. It is shown that the method offers improved denoising results for the processing of white and harmonic noises when compared with other linear and WT-based denoising strategies. Moreover, its greatest advantages become clear when the noise has a pulsating or localized shape, situation in which traditional strategies fail completely. The paper is organized as follows: Section 2 presents the characteristics of PD signals and their wavelet decompositions, which form the basis for the proposed method. Section 3 presents the approach to establish the spatial correlations of PD data. Section 4 presents the method used to separate the PD and noise-related coefficients. Section 5 introduces and justifies the use of the translation invariant discrete wavelet transform. Section 6 presents the results for several case studies. Section 7 presents the conclusions.

2. Partial discharge signals and wavelet decomposition characteristics

Partial discharges reflect on the external circuitry as the circulation of low-amplitude fast current pulses whose shape and temporal parameters depend on factors like the type and location

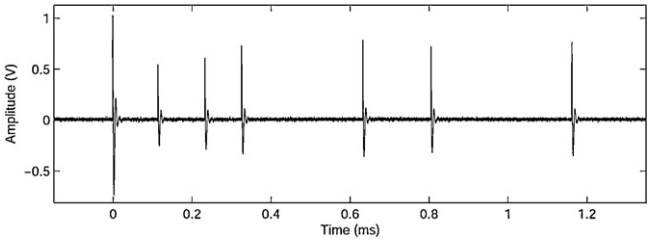

Fig. 1. Partial discharge pulses generated in air at atmospheric pressure.

of the defect, the size of the equipment under test, the distance between the discharge site and the sensor and the type of the measurement system [1]. The pulse amplitudes are usually in the range of micro- to milli-volts, what makes them very difficult to detect under noisy conditions.

PD measurement systems are classified as narrow, wide and ultra-wide band depending on their pass-band and center frequency [1]. Ultra-wide band measurements usually concern with temporal analysis and, thus, can be performed only in situations in which the pulses can propagate without attenuation or distortion until reaching the sensor (e. g. power cables and gas insulated systems). Narrow and wide-band systems are composed by band-pass measuring circuits syntonized to a specific frequency, what makes them less sensitive to noise. In this work the attention was restricted to the last systems since these are the most employed under on-site noisy conditions.

Fig. 1 shows a typical narrow-band PD signal obtained from a point-plane electrode arrangement in air at atmospheric pressure. When this signal is decomposed using the wavelet transform the result has a very peculiar shape, characterized by coefficients clustering at the same relative positions in each decomposition level. This can be visualized in Fig. 2 as a set of coefficients bounded by dashed lines, which delimits the *cone of influence* of the PD. Therefore PDs WT decompositions present a strong spatial correlation and this characteristic can be used as a criterion for their identification.

In this work the characterization of spatial correlations was accomplished by a procedure based on WT local modulus maxima propagation theory [21]. This procedure relies on the fact that the largest coefficients, which are the ones usually seeked by other WT denoising methods, are also in the cone of influence of the PD. But instead of performing the analysis along the entire time-scale plane or on a level-by-level basis, this approach investigates the

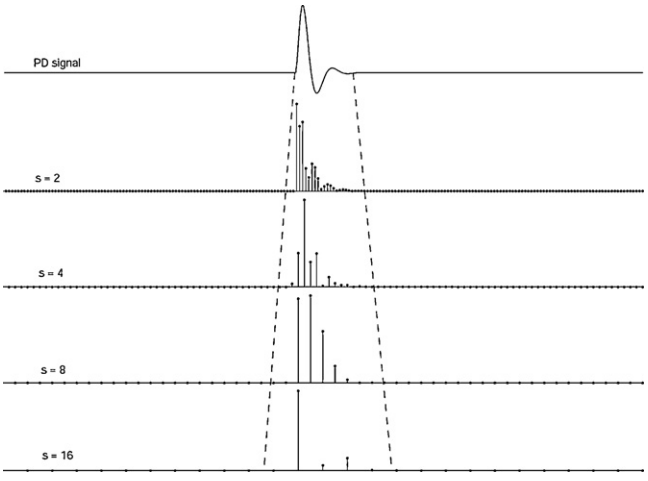

Fig. 2. WT coefficients influenced by a PD pulse. Top figure: time-domain PD signal. Bottom figures: wavelet coefficients modulus at scales 2–16.

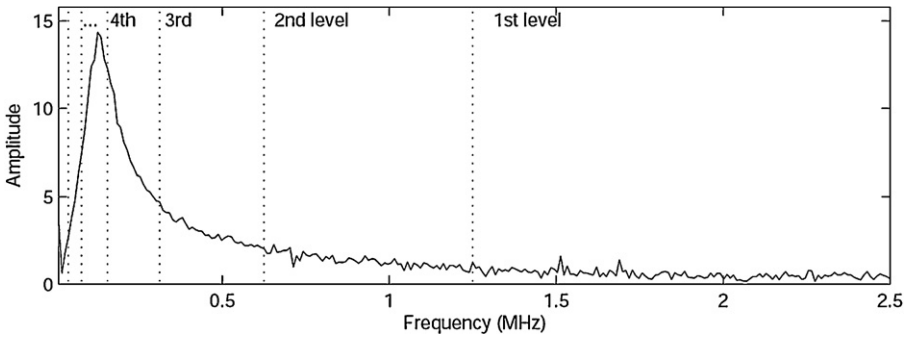

Fig. 3. DWT frequency spectrum dyadic division.

coefficients propagation characteristics between adjacent scales. By these, it separates the coefficients related to PDs from those related to noise.

3. Characterization of spatial correlations by the wavelet transform local modulus maxima

A well-known property of the wavelet transform is its potential to identify the singularities of a function. This is often used to solve problems related to singularity location, as in image border detection and data compression applications [25–28]. It can be proved that the singularities of a function can be localized and characterized by measuring the decay of the local maxima of the WT coefficients modulus, the so-called WT modulus maxima, down to the finest scale available [21]. A WT local modulus maximum is defined as any point (s_0, t_0) in the time-scale plane such that $|WT(s_0, t)| < |WT(s_0, t_0)|$ when t belongs to either a right or left neighborhood of t_0. A maxima line is any curve in the plane which connects modulus maxima points. If a signal $x(t)$ has a singularity at an arbitrary instant t_0 its wavelet transform modulus maxima tend to be located below the cone of influence of $x(t_0)$. Therefore, when the modulus maxima are linked the result is a set of lines that point to the signal singularities as they progress from large to small scales.

With this capability in mind, the characterization of PD spatial correlations was accomplished in two stages. At first, the decomposition level from which the analysis will start must be determined. It is associated to the highest energetic band of the PD frequency spectrum, which can be determined in the following ways.

The discrete wavelet transform (DWT) decomposition reflects in frequency domain as a logarithmic division where the first level coefficients correspond to the high frequency band between π and $\pi/2$, the second between $\pi/2$ and $\pi/4$ and so on (where π represents the Nyquist frequency). As an example, Fig. 3 shows such bands for the spectrum of one of the PD pulses shown in Fig. 1. Since DWT is an energy conserving transform [29], the largest DWT coefficients tend to cluster in the levels corresponding to the most energetic frequency bands of the signal, which are directly related to the center frequency of the measurement system pass-band. Therefore, in a first approach, the initial analysis level can be determined by the knowledge of the measurement system characteristics.

A more systematic procedure can be employed if a clean PD pulse is available or by using a calibration pulse according to IEC-60270 [4]. In these cases, the initial level can be determined by decomposing the pulse and calculating the energy of the coefficients at each level, which is given by

$$E(j) = \sum_{i=1}^{N/2^j} cD^2(i,j), \tag{1}$$

where $E(j)$ is the energy at level j, $cD(i,j)$ is the ith detail coefficient at level j and N is the signal length. The initial level would be the one with the highest energy.

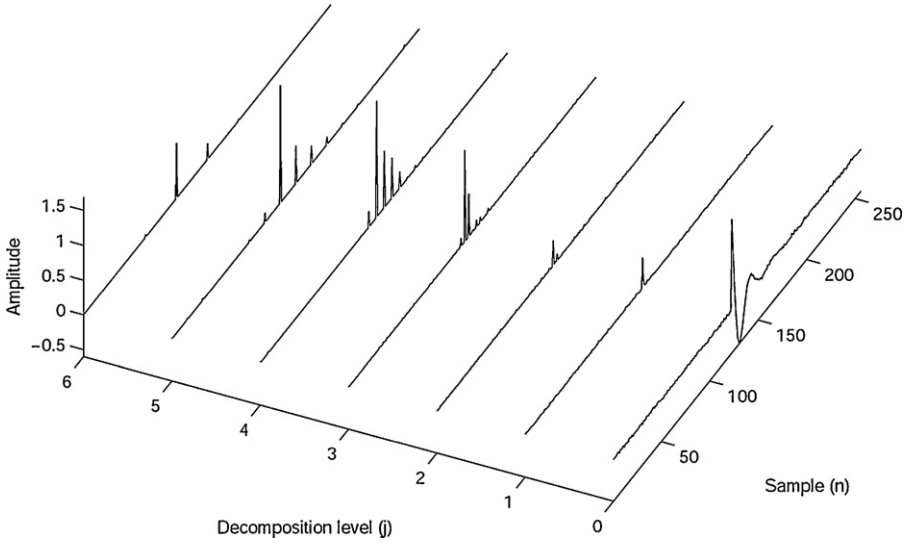

Fig. 4. Six-level DWT modulus distribution for a narrow-band PD pulse. The PD is shown at decomposition level 0.

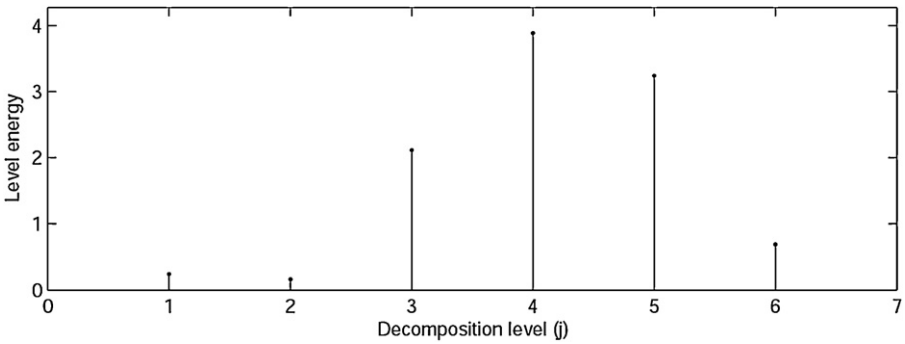

Fig. 5. Signal energy distribution along DWT decomposition levels.

As an example, Fig. 4 shows a six-level DWT decomposition of an isolated PD pulse. It can be seen that the highest coefficients are concentrated in the third, fourth and fifth levels, so it can be expected that the highest energy bands will be located at their corresponding levels. This is confirmed by inspecting Fig. 5, which shows the energy distribution according to Eq (1). Thus, in this case the fourth level would be selected to initiate the analysis.

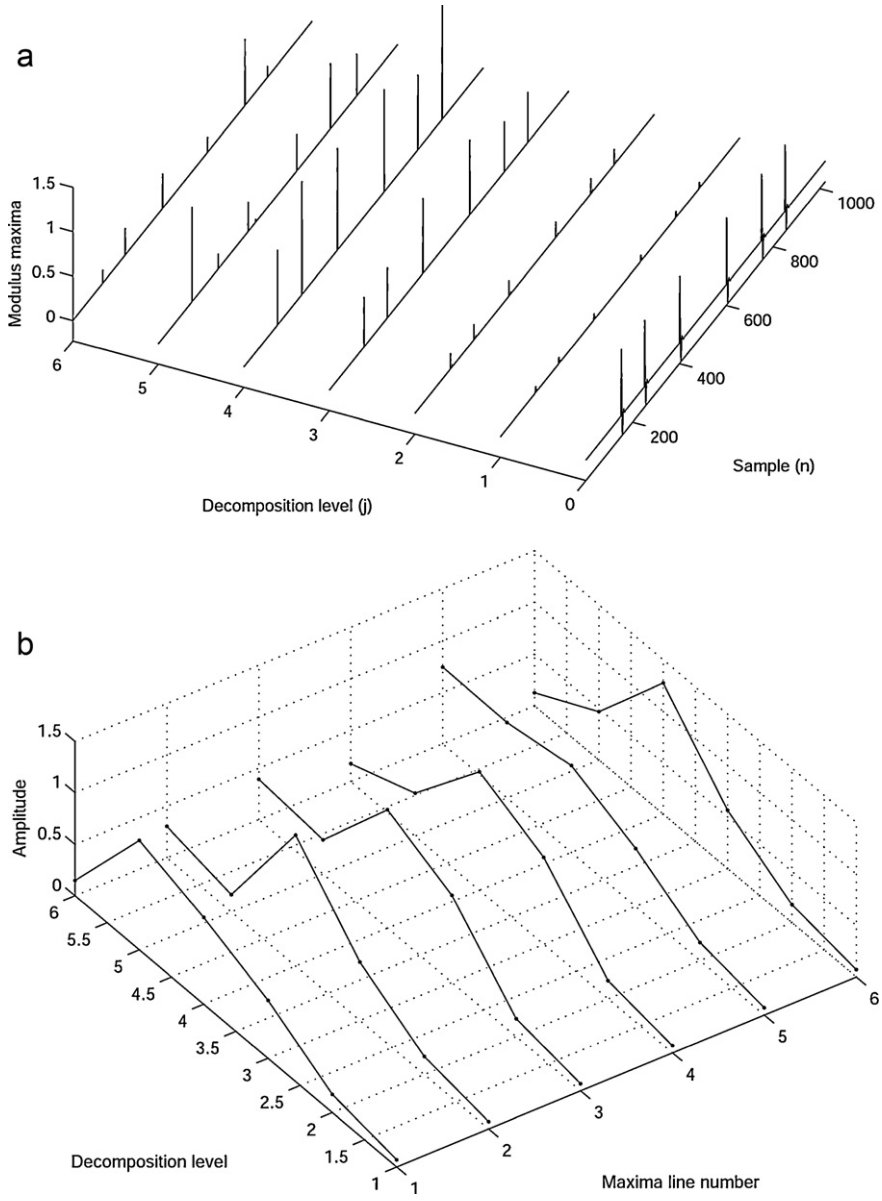

Fig. 6. DWT modulus maxima distribution and maxima lines propagation for a narrow-band PD signal. The signal is shown at decomposition level 0. (a) Local modulus maxima distribution. (b) Modulus maxima lines.

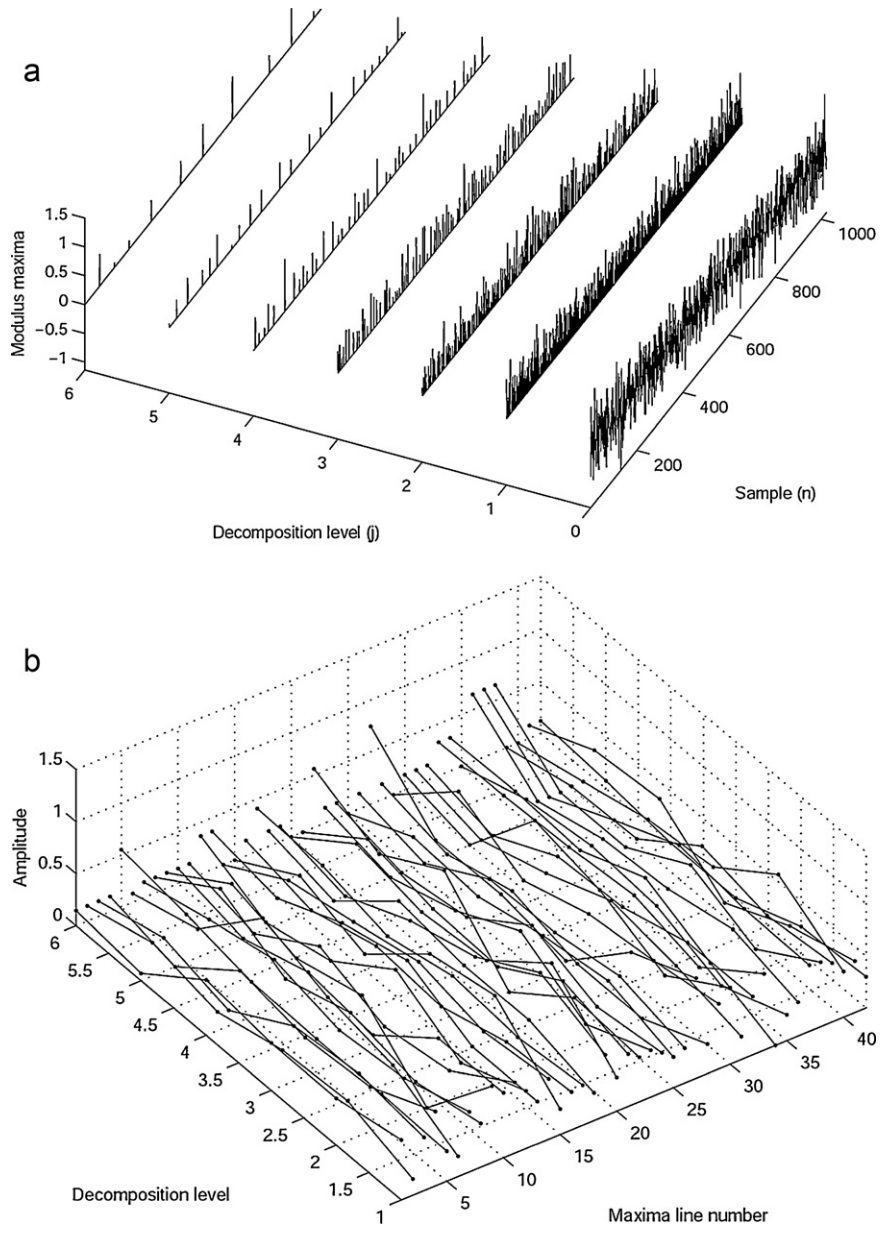

Fig. 7. DWT modulus maxima distribution and maxima lines propagation for a Gaussian white noise. The signal is shown at decomposition level 0. (a) Local modulus maxima distribution. (b) Modulus maxima lines.

In the second stage the local modulus maxima at all levels are determined and the maxima located at the initial level are identified. For each of these, the algorithm evaluates a set of coefficients from the previous as well as the next levels, defined by

$$\Delta_{\text{Previous}} = cD(k, j-1)\Big|_{2i\max-M<k<2i\max+M} \tag{2}$$

$$\Delta_{\text{Next}} = cD(k, j+1)\Big|_{\frac{i\max-M}{2}<k<\frac{i\max+M}{2}} \tag{3}$$

where i max represents the index of the maxima found in level j and M is the support of the wavelet filter. If there is another maximum in this set a connection is established and the procedure continues until all maxima in all decomposition levels are evaluated. The result of this processing is a set of maxima lines crossing along the decomposition levels and linking all the maxima found in the time-scale plane, as shown in Fig. 6.

4. Separation of maxima lines related to PDs and noise

The most remarkable characteristic of the PD maxima lines is that they decrease strongly when propagating from higher to lower decomposition levels, as can be seen in Fig. 6b. This behavior is directly related to the shape of the PD frequency spectrum since its energy is concentrated in a small number of bands. When the signal is corrupted by noise the modulus maxima propagation depends on the noise type and its frequency spectrum. For example, it is a well known fact that the wavelet transform of a random white noise is also a random process [21,29] thus its modulus maxima lines tend to be randomly distributed and present no characteristic shape, as can be seen in Fig. 7. Therefore, the lines related to PDs and to noise have very distinct shapes, which can be used for their identification and separation.

The maxima lines separation was performed by a pattern classifier based on a support vector machine (SVM). The input space was defined as the maxima located at each level, thus the maxima lines

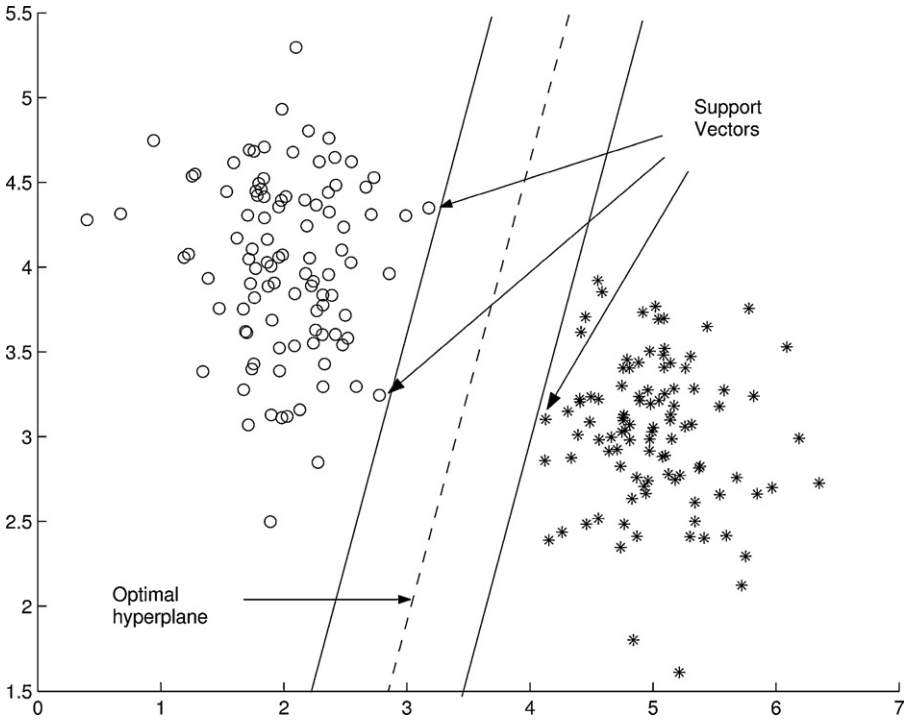

Fig. 8. SVM optimal hyperplane for a two-dimensional case.

were fed directly into the classifier. The machine task is to indicate if a maxima line is related to a PD or to noise. The noise related lines are marked and their coefficients are removed from the set, thus reconstruction is made only with coefficients related to PDs.

The choice of an SVM classifier was based on two characteristics: first, there is a number of previous investigations that indicate that, for a given classification problem, SVM generalization usually either matches or is significantly better than other classification methods [22]. Allied to that, SVMs have much less parameters to be adjusted, thus being less dependent on empirical and ad hoc procedures. In the following we briefly discuss the SVM classifier, as well as the adopted classification procedure.

4.1. An introduction to support vector machines

SVMs are a particular group of supervised learning machines which address the problem of complex pattern classification by mapping the data into a higher dimensional space followed by a linear classification [22,23]. Suppose we have a data set X composed by samples x_i, $i = 1, \ldots, N$, each one defined by n features and belonging to a known class y_i. Thus, we can think of each sample as a vector in an n-dimensional input space. The objective of SVM is, based in this pre-classified (or training) set, find a separation surface with the maximum margin between classes and whose shape is complex enough to fit the data distribution, while providing a good generalization when presented to unknown data. As an example, Fig. 8 shows a two-dimensional input space with a data set categorized in two classes, identified by circles and asterisks. This data can be separated by several planes but only one maximizes the separation margin. This is known as the *optimal hyperplane* pursued by SVM.

In order to find the optimal hyperplane SVMs automatically search for a subset of the training data, called *support vectors*, which lie in the class boundaries. The separating surface is obtained by mapping the data into a high-dimensional *feature space* followed by the construction of a linear hyperplane. When this is mapped back to the input space the result is a complex non-linear separating surface, hopefully capable to separate the data distribution. The solution is obtained through a quadratic optimization problem defined by [30,31]:

$$\min_{w,b,\xi} \left(\frac{1}{2} \cdot \vec{w}^T \cdot \vec{w} + C \cdot \sum_{i=1}^{N} \xi_i \right) \tag{4}$$

subject to

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i)$$

$$y_i[\vec{w}^T \cdot \vec{\varphi}(x) + b] = y_i \left[\sum_{i=1}^{N} \alpha_i y_i \vec{\varphi}^T(x_i) \vec{\varphi}(x) + b \right] \geq 1 - \xi_i, \tag{5}$$

$$\xi_i \geq 0,$$

where x_i is the ith input vector, y_i is its assigned class, N is the size of the training set, w is an weight vector, b is a bias, ξ_i is a measure of separability of data point i, C is a cost defined by the user, α_i is a Lagrange multiplier and $\varphi(x)$ is the set of mapping functions.

4.2. SVM parameterization and training

SVM use demands the definition of the cost C and the parameters related to the kernel function $K(x, x_i)$. The cost is a user defined regularization parameter that controls a compromise between the complexity of the machine and the acceptable number of non-separable points. The kernel function, defined by

$$K(x, x_i) = \vec{\varphi}^T(x_i) \vec{\varphi}(x) \tag{6}$$

is a linear combination of mapping functions used to avoid the construction of the high-dimensional space explicitly. There are several kernels proposed in the literature and, depending on the choice, it is possible to construct more or less complex surfaces in the input space. In this work three kernels were investigated: the linear, polynomial and radial-basis functions (RBFs), defined respectively by

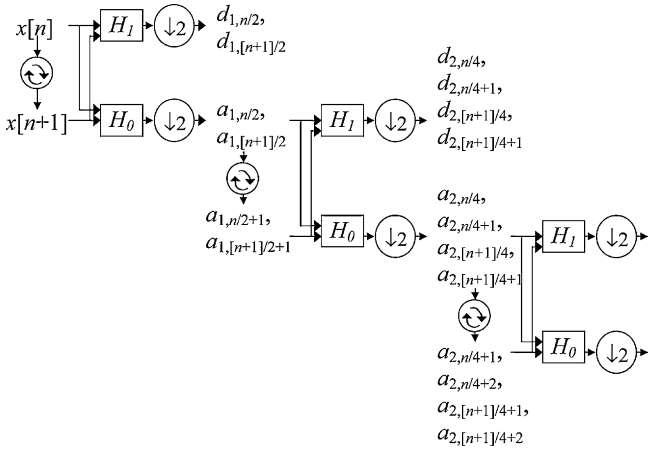

Fig. 9. Filter bank for a 3-level TI-DWT decomposition.

[23]:

$$K_l(x_i, x_j) = x_i^T \cdot x_j \tag{7}$$

$$K_p(x_i, x_j) = (\gamma \cdot x_i^T \cdot x_j + r)^d \tag{8}$$

$$K_{\mathrm{RBF}}(x_i, x_j) = \exp(-\gamma ||x_i^T - x_j||^2) \tag{9}$$

where γ, r and d are the function parameters. The goal was to evaluate the degree of non-linearity of the input data. The RBF kernel offered more robust results in all cases evaluated, thus was chosen as the preferred one.

As can be seen in (9) the RBF kernel demands the definition of γ, which controls the width of the Gaussian functions that perform the input-output mapping. The optimal pairs (C, γ) for each case were defined by a cross-validation procedure with a grid search [32]. The cross-validation is a procedure used to avoid overfitting and, thus, to obtain a better generalization. It consists of dividing the pre-classified input data in two groups, named training and validation. The training group is used to adjust the parameters and the validation group is used to verify the generalization performance. At each pass the training and validation groups are switched and the procedure is repeated until an error limit is obtained or a pre-defined number of realizations is reached. To accelerate the procedure, during training the parameters (C, γ) were grown exponentially in the vicinities of initial values defined by the user. This allows the identification of regions in the (C, γ) plane which offer the best results in terms of generalization. Once these are identified, a finer grid search was performed only in these regions, resulting in the optimum (C, γ) pair. After optimization, the SVM structure was fixed and then used to classify data from new signals.

4.3. Training SVMs for PD modulus maxima identification

In this work the SVM training was performed as follows. During the development phase only artificially generated data were considered. Several combinations of PD and noise were used, but the training has always been made in the same way: a noise-free PD signal and a sample of one type of noise were separately decomposed and manually classified. This data were used during training and, after tuning the parameters, the machine was used to classify other signals composed by that type of PD and noise.

For the evaluation of measured signals the procedure was as follows. The target PD signal was obtained by calibration pulses injected in the system before energization or, when possible, by clean PD pulses obtained with the equipment turned on. A sample of noise was acquired just before the PD inception level. These data were manually classified and used for training. After the parameter definition, the machine was used to classify new signals containing both PD and noise. In all cases, the data has been scaled to fit in the range $[-1, +1]$. Scaling is an important step for pattern classification because it avoids the dominance of attributes with greater numerical ranges, the influence of outliers and the numerical difficulties related to limited number representation [32].

5. The translation-invariant discrete wavelet transform

During evaluation the method has shown to be very sensitive to the position of the PD pulses, what caused random pulse losses even for small displacements. This effect is associated to the decimation performed by DWT since it can eliminate high-energy coefficients in intermediary decomposition levels, thus sometimes impeding the modulus maxima tracking. The problem was solved by using the translation-invariant DWT (TI-DWT) [20].

TI-DWT is a non-orthogonal wavelet transform that compensates DWT decimation by introducing circular shifts in each decomposition level. This is shown schematically in Fig. 9, where $x[n]$ represents the discrete-time signal, H_1 and H_0 are high and low-pass complementary wavelet filters and $d_{j,k}$ and $a_{j,k}$ are the kth detail and approximation coefficients at level j, respectively. The circular arrows represent circular shifts which, when applied to an approximation vector $a_{j,k}$ generates $a_{j,(k+1)\mathrm{modulo}(N/2\hat{}j)}$. When a length N signal is decomposed to J levels TI-DWT generates an N by J matrix containing all the possible circularly shifted DWTs. Reconstruction is made by systematically averaging every pair of approximations, what is similar to averaging all decompositions.

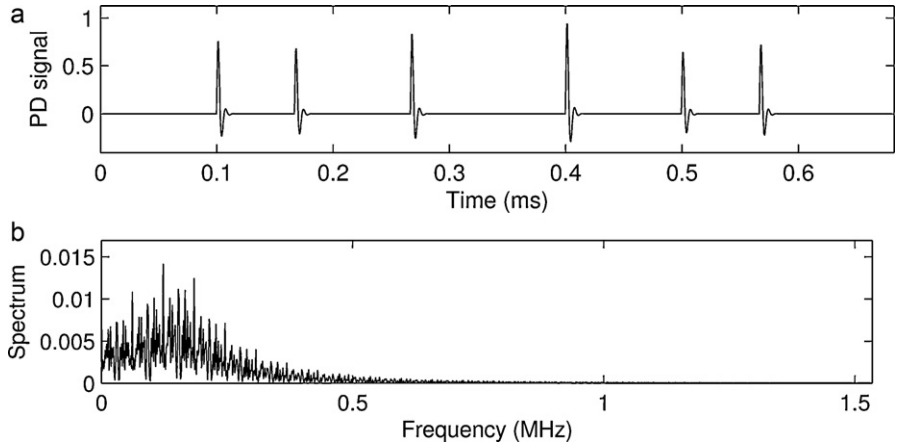

Fig. 10. PD pulse train evaluated. (a) PD signal. (b) Frequency spectrum.

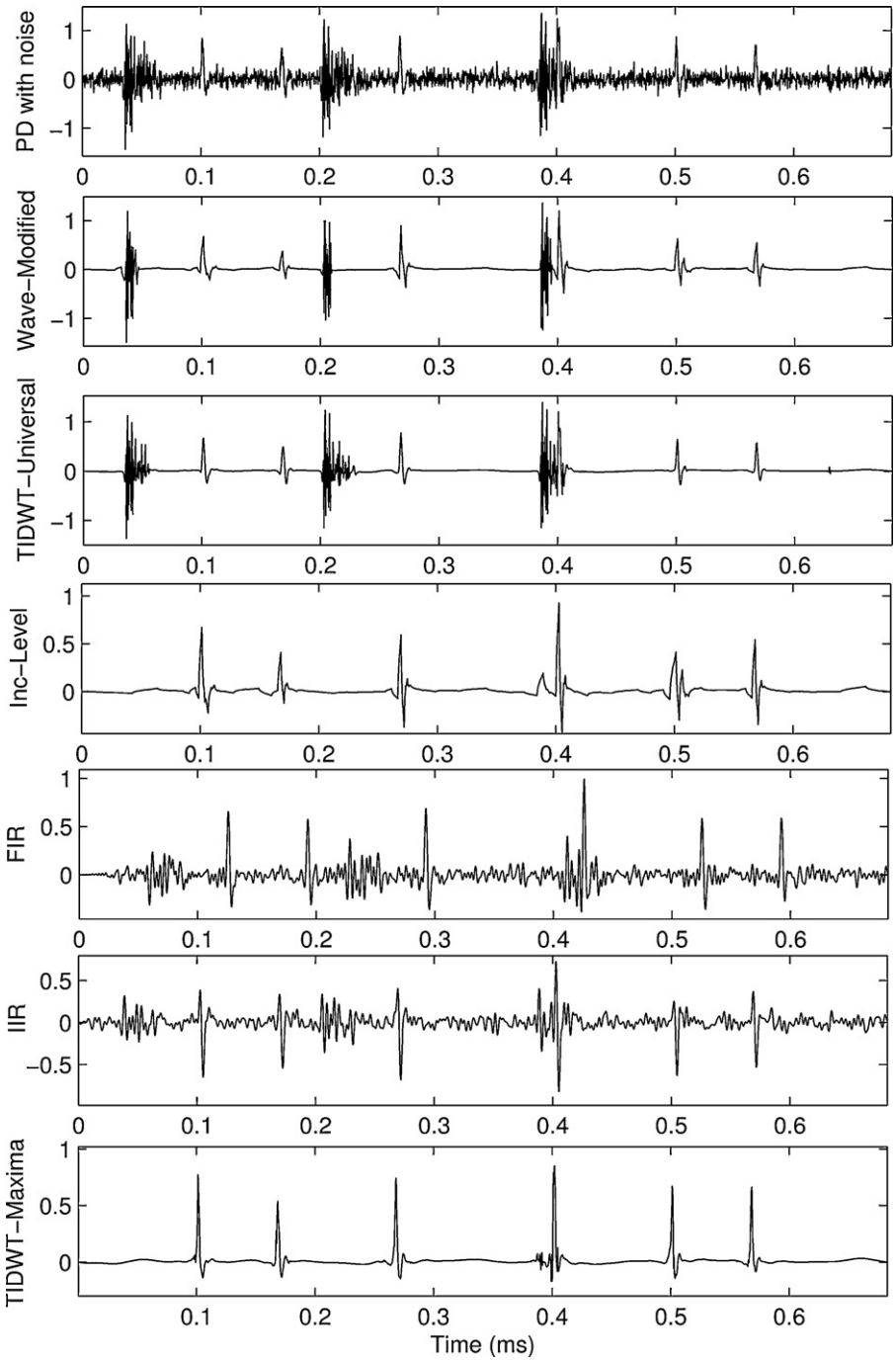

Fig. 11. Denoising results for signal 1. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

Both decomposition and reconstruction are efficiently performed in $O(N\log_2(N))$ operations for a length $N = 2^J$ vector.

After decomposition the modulus maxima analysis was separately performed for each shifted DWT obtained from TI-DWT. The filtering consists on the generation of a matrix containing significance levels for each coefficient. The matrix is initialized with zeros and every time a coefficient survives the filtering process its significance is increased by an amount given by

$$S_{cD}(i,j) = S_{cD}(i,j) + 2^{j-J} \tag{10}$$

where $S_{cD}(i, j)$ is the significance of the coefficient i at level j. After analysis the coefficients are multiplied by their significances and the signal is reconstructed by the Inverse translation-invariant DWT (TI-IDWT).

6. Results and discussion

6.1. Damped oscillatory pulses

6.1.1. Parameters

The effectiveness of the method was first evaluated with a damped oscillatory waveform defined by [6,10]:

$$PD_{Osc}(t) = A(e^{-\alpha t} - e^{-\beta t})\sin(\omega t), \tag{11}$$

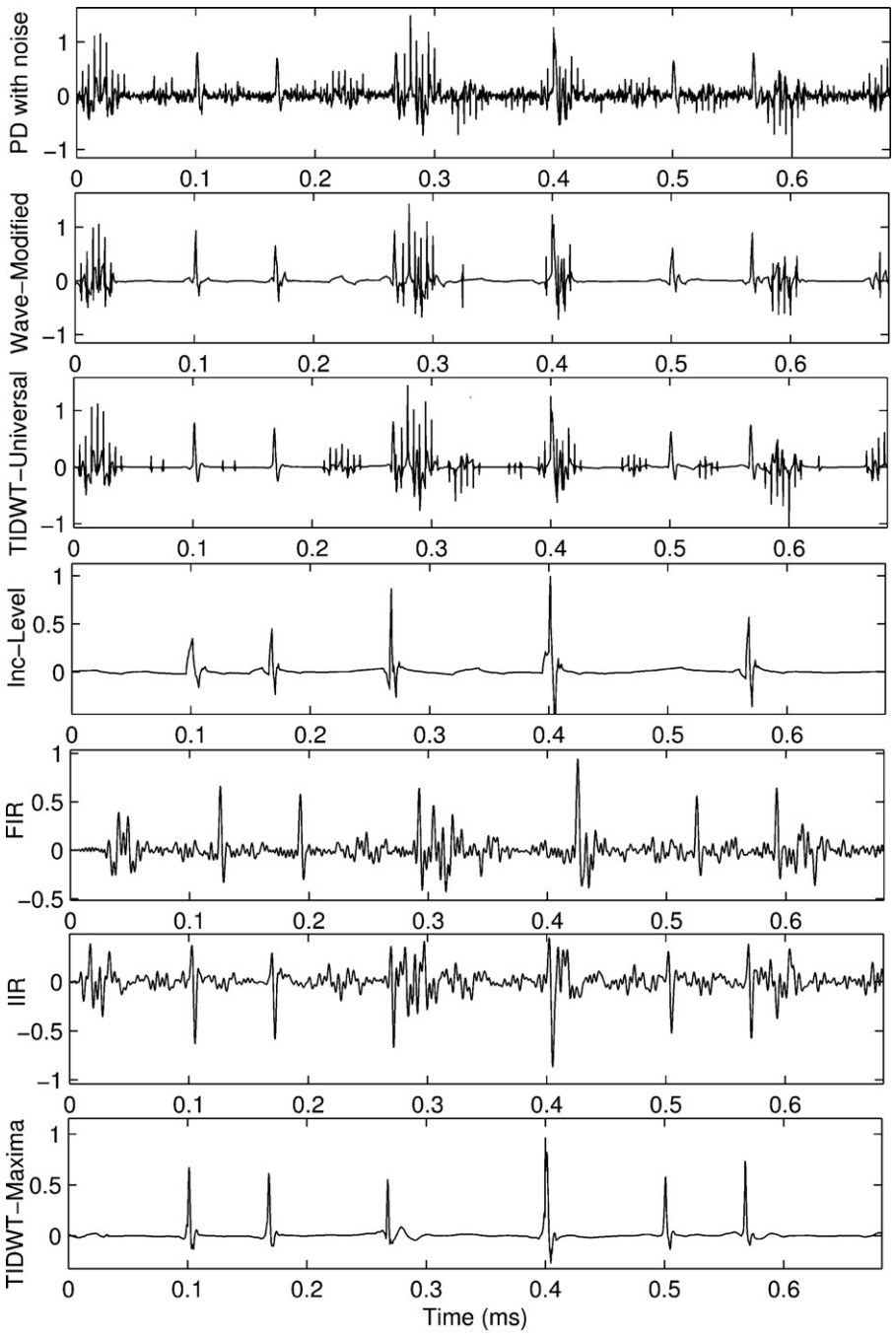

Fig. 12. Denoising results for signal 2. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

where A represents the pulse amplitude, α and β the damping factors and ω the oscillatory frequency. A train of pulses with different spacing and amplitudes was used to take the signal variability into consideration.

Three types of noise usually encountered during on-site measurements were considered. Thermal and background noises were modeled as Gaussian distributions with zero mean and different standard deviations. Sinusoidal and harmonic noises were represented by amplitude modulated (AM) sine waves defined by [3]:

$$n_{AM}(t) = \sum_{c=1}^{n} A_c \left(1 + \sum_{m=1}^{k} \mu_m \sin(\omega_m t)\right) \cos(\omega_c t), \tag{12}$$

where A_c and ω_c are the amplitude and frequency of the carrier waves, μ_m is the modulation index and ω_m the frequency of the mth modulating wave, respectively. The following values were used during simulation: $A_c = 1$, $\omega_c = 50$, 125, 250, 400, 600, 800, 1000, 1200 and 1400 kHz, $\mu_m = 0.4$, $\omega_m = 1$, 5, 8, 11, 15 and 20 kHz.

Pulsating interferences were modeled as Gaussian noise multiplied by exponential, Gaussian and rectangular windows with variable amplitudes, widths and spacing. The three window functions are defined by

$$W_{\mathrm{EXP}}(t) = A(e^{-\alpha t} - e^{-\beta t}); \tag{13}$$

$$W_{\mathrm{GAUSS}}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2}; \tag{14}$$

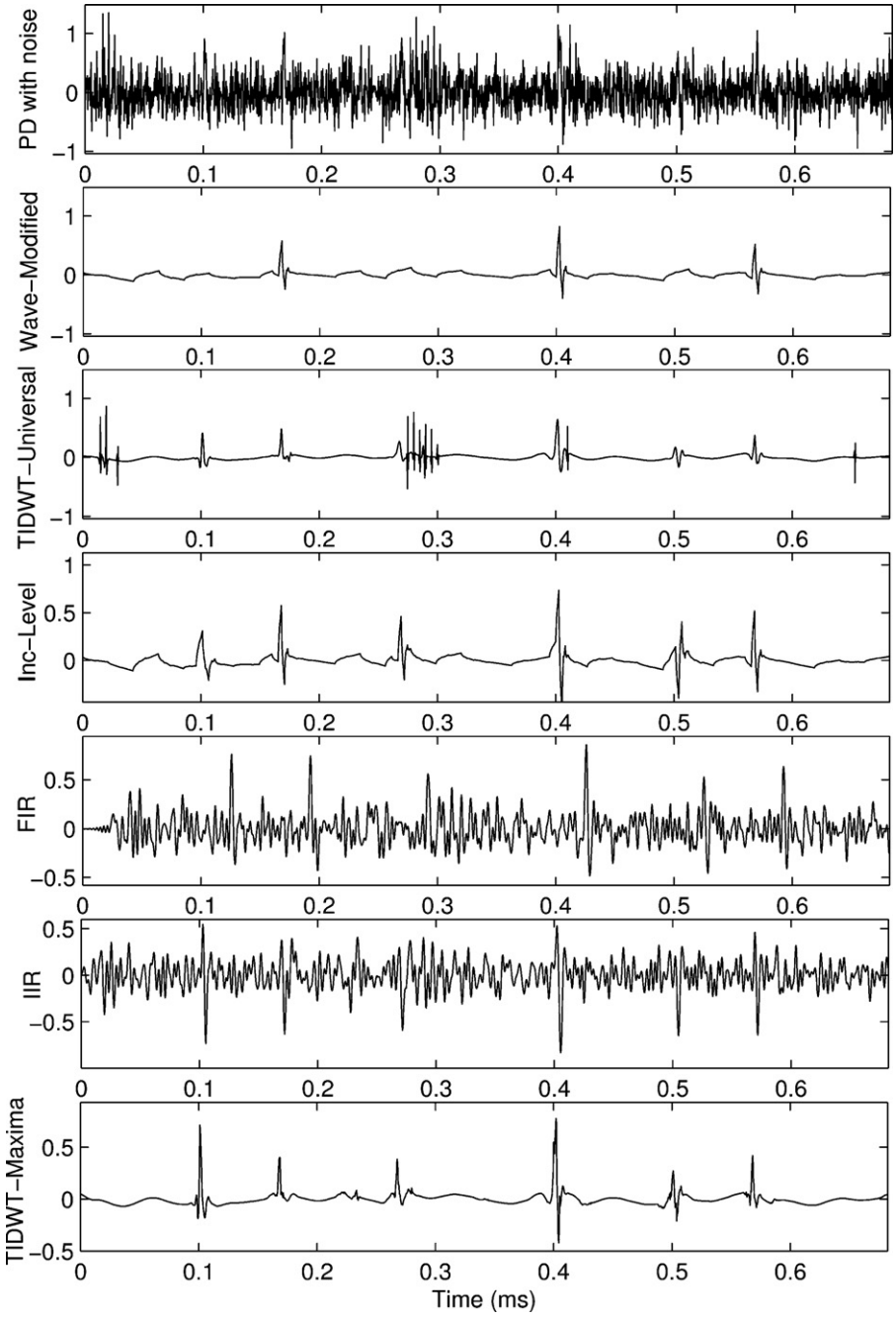

Fig. 13. Denoising results for signal 3. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

$$W_{\mathrm{RECT}}(t) = \begin{cases} 1, & t_0 \le t \le t_1 \\ 0, & \text{otherwise} \end{cases}; \tag{15}$$

where A, α and β have the same meaning as in Eq. (11), yet with different values. σ represents the standard deviation of the Gaussian window and t_0 and t_1 are the discontinuity points of the rectangular window.

6.1.2. Numerical evaluation

To evaluate the errors introduced by the denoising procedure the following indices have been used:

(a) Signal to reconstruction error ratio

SRER is defined as

$$\mathrm{SRER\,(dB)} = 10\log_{10}\frac{\sum_{n=1}^{N} X^2[n]}{\sum_{n=1}^{N}(X[n]-Y[n])^2}, \tag{16}$$

where $X[n]$ is the PD reference signal, $Y[n]$ the denoised signal and N the number of samples. It determines the effectiveness of the denoising method by relating the original signal energy to the reconstruction error energy.

(b) Cross-correlation coefficient

The cross-correlation is defined as [33]:

$$R_{XY}(r) = \sum_{n=0}^{N-r-1} X[n]Y[n+r] \tag{17}$$

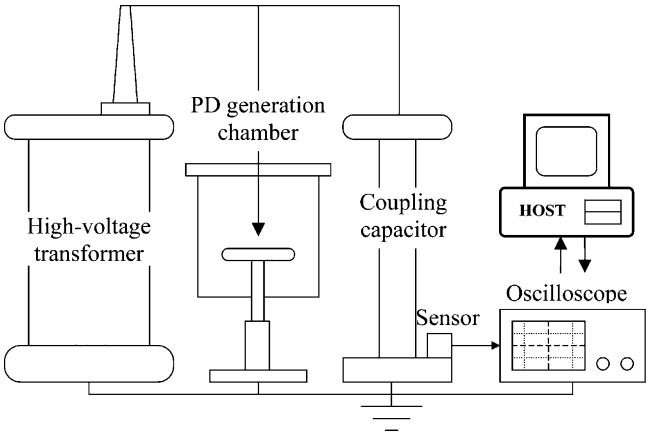

Fig. 14. PD generation and measurement system.

where $X[n]$ is the original reference signal, $Y[n]$ the denoised signal and N the number of samples. R_{XY} indicates the degree of similarity between the original and the denoised signals.

(c) Pulse amplitude distortion

PAD is determined by

$$\mathrm{PAD}\,(\%) = \frac{|X_{\mathrm{MAX}} - Y_{\mathrm{MAX}}|}{X_{\mathrm{MAX}}} \times 100 \tag{18}$$

where X_{MAX} is the amplitude of the reference PD pulse and Y_{MAX} the amplitude of the denoised pulse. In cases involving several pulses a mean value was computed.

In general a good denoising strategy should result in a high signal to reconstruction error ratio, a high cross correlation coefficient and a low pulse amplitude distortion.

6.1.3. Comparisons with other methods

As a way to evaluate the effectiveness of the method the results were compared with linear filters and three other wavelet-based PD denoising procedures. The linear filters are band-pass finite and infinite impulse response (FIR, IIR) filters adjusted to the frequency spectrum of the PD signal. The first wavelet-based method is Waveshrink with the scale-based thresholding variation proposed by Ma et al [11,12]. The second is Waveshrink with Universal rule and hard thresholding applied to TI-DWT decomposition, as proposed in [20]. The third method is the "inception level" procedure proposed in [3]. For further reference the methods and their identifications are listed in Table 1.

Table 1
Denoising techniques evaluated.

Method	Name	Description
A	Wave-Modified	Waveshrink + modified level-based thresholding.
B	TIDWT-Universal	Waveshrink + Universal threshold applied to TI-DWT.
C	Inc-Level	The "inception level" method.
D	FIR filter	Equiripple FIR filter with 30–300 kHz pass band.
E	IIR filter	Chebychev type 2 IIR filter with 30–300 kHz pass band.
F	TIDWT-Maxima	The modulus maxima method proposed in this work.

6.1.4. Results and discussion

The test signals are composed by several combinations of PD pulse trains and noises. They were processed under the same conditions: all decompositions were performed with Daubechies orthogonal 4-tap wavelet filters and 6 levels. The choice for the wavelet filter was based on the cross correlation approach proposed in [11]. For brevity, in this paper we report only what can be considered the worst cases.

The original PD signal is shown in Fig. 10 along with its frequency spectrum. It is composed by six PD pulses generated according to (11) with $\alpha = 4.5 \times 10^5$, $\beta = 14 \times 10^5$, $\omega = 2\pi 150 \times 10^3$. The pulses have varying amplitudes A and times-of-occurrence. The signal is characteristic of a narrow-band measurement system with a pass-band of about 30–300 kHz. The sampling frequency is 3 MHz.

The first test signal and its denoised versions are shown in Fig. 11. The PD is corrupted by a moderate white noise with zero mean and variance 0.1 and a large amplitude pulsating noise generated with an exponential window, according to (13). The signal to noise ratio (SNR) was used to measure the extent of noisiness, as defined by

$$\mathrm{SNR}\,(\mathrm{dB}) = 10 \log_{10} \frac{\sum_{i=1}^{N} s^2(i)}{\sum_{i=1}^{N} n^2(i)}, \tag{19}$$

where $s(i)$ is the original PD signal, $n(i)$ is the noise and N is the number of samples. SNR for signal 1 is −5.93 dB. The first and second noise bursts are isolated from the PD pulses but the third is superimposed to the PD located at 0.4 ms.

The numerical results for signal 1 are shown in Table 2. It can be seen that the methods wave-modified and TIDWT-Universal, which are based on uniform or scalar thresholding, are efficient for elimination of the Gaussian part of the noise but fail for the pulsating part since these are not easily characterized by statistical measures. The resulting signals for these methods remained with a high-energy noise component leading to low signal to error ratios and correlation coefficients. TIDWT-Universal offered a slight better reconstruction which reflected in a greater correlation coefficient.

The results for the linear filters are worse than those obtained for wavelets methods. In spite of their band-pass characteristic, both FIR and IIR reconstructions remained with a large noise component. This is due to the superposition of PD and noise frequency spectra. The IIR response resulted in a pulse amplitude distortion (PAD) of more than 100% due to the polarity inversion caused by its steeper borders. Both filters resulted in low SRERs and R_{xy}.

The methods Inc-Level and TIDWT-Maxima resulted in a good noise reduction while maintained all the PD pulses. TIDWT-Maxima offered a softer reconstruction, which resembles better the original PD shapes, while Inc-Level result is more similar to the wavelet itself. This led to a greater correlation coefficient and signal to error ratio for TIDWT-Maxima. Also, its better reconstruction reflected in the lowest pulse amplitude distortion of all methods, which is about 10%.

The second test signal is shown in Fig. 12. The same PD signal was corrupted by a moderate white noise with zero mean, variance 0.1 and a larger amplitude AM sinusoidal noise generated according to (12). SNR is −5.09 dB. The numerical evaluation is shown in Table 3. Again, wave-modified and TIDWT-Universal failed in the elimination of the localized parts of noise. FIR and IIR results presented a very strong noise component and a large pulse amplitude distortion, again due to the superposition of frequency spectra. Inc-Level removed most of the noise but also eliminated the PD located at 0.5 ms. It also introduced a considerable amplitude distortion of about 36%. Again, TIDWT-Maxima reconstruction resembles better the original PD shapes, what reflects in the greatest signal to error ratio, correlation coefficient and the lowest pulse amplitude distortion.

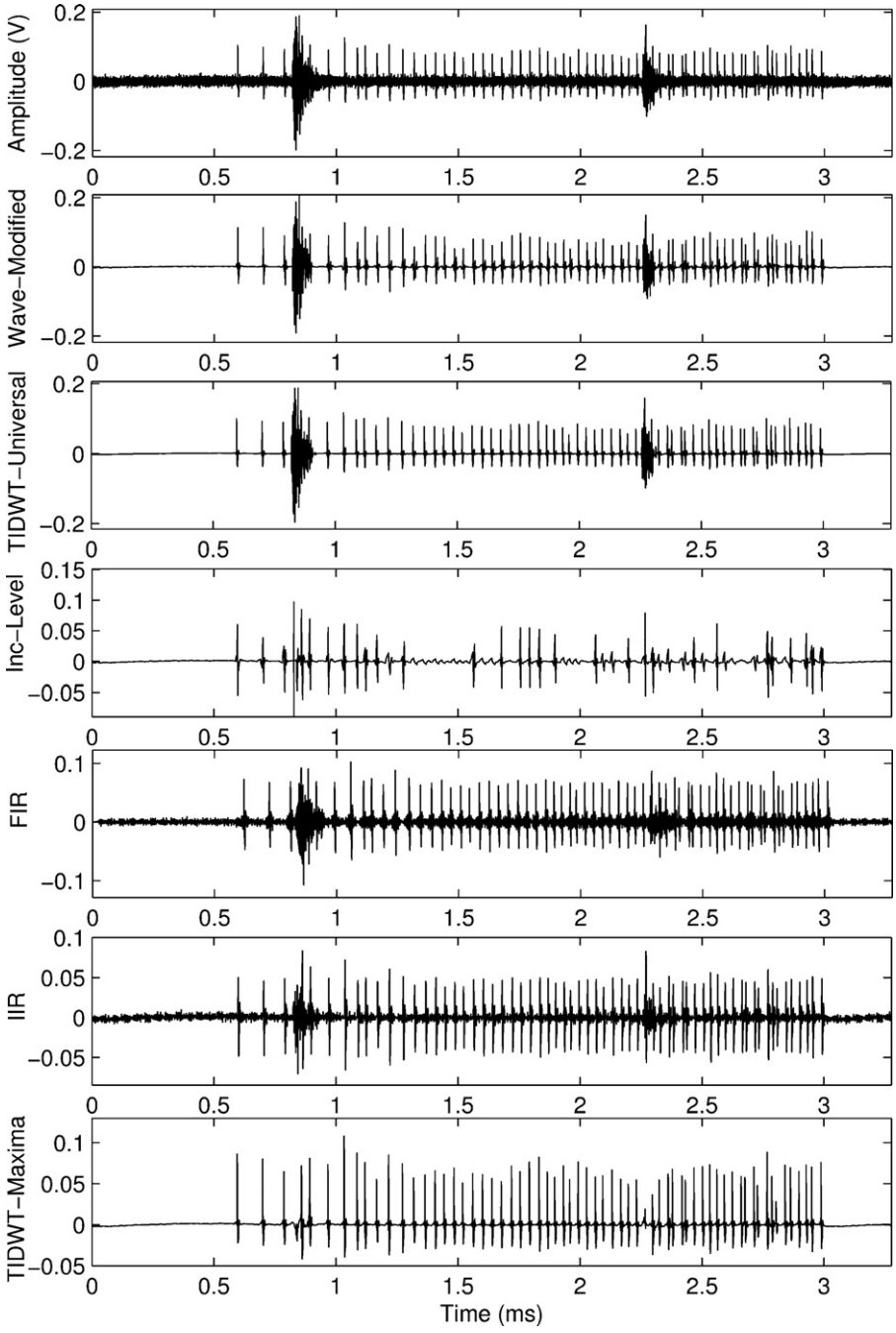

Fig. 15. Measured PD signal corrupted by random and pulsating noises. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

The third test signal is shown in Fig. 13. It is corrupted by the same AM noise of the previous example, but now added to a high amplitude Gaussian noise with zero mean and variance 0.25. The PD pulses are completely buried by noise and the SNR is −9.53 dB. The numerical evaluation is shown in Table 4. Since this case is an extreme situation several denoising methods lost pulses, thus their signal to reconstruction errors are very low. Note that even the white noise fitted methods resulted in poor reconstructions, in spite of the main noise component having a Gaussian distribution. Again, Inc-Level and TIDWT-Maxima offered the best results with the later giving a superior SRER and correlation coefficient.

Table 2
Numerical evaluation for signal 1.

	Denoising method					
	A	B	C	D	E	F
SRER (dB)	−3.10	−3.20	3.15	−3.98	-2.63	8.92
R_{XY} (%)	47.32	52.35	72.84	55.73	27.38	93.84
PAD (%)	81.22	78.20	16.56	15.00	219	10.54

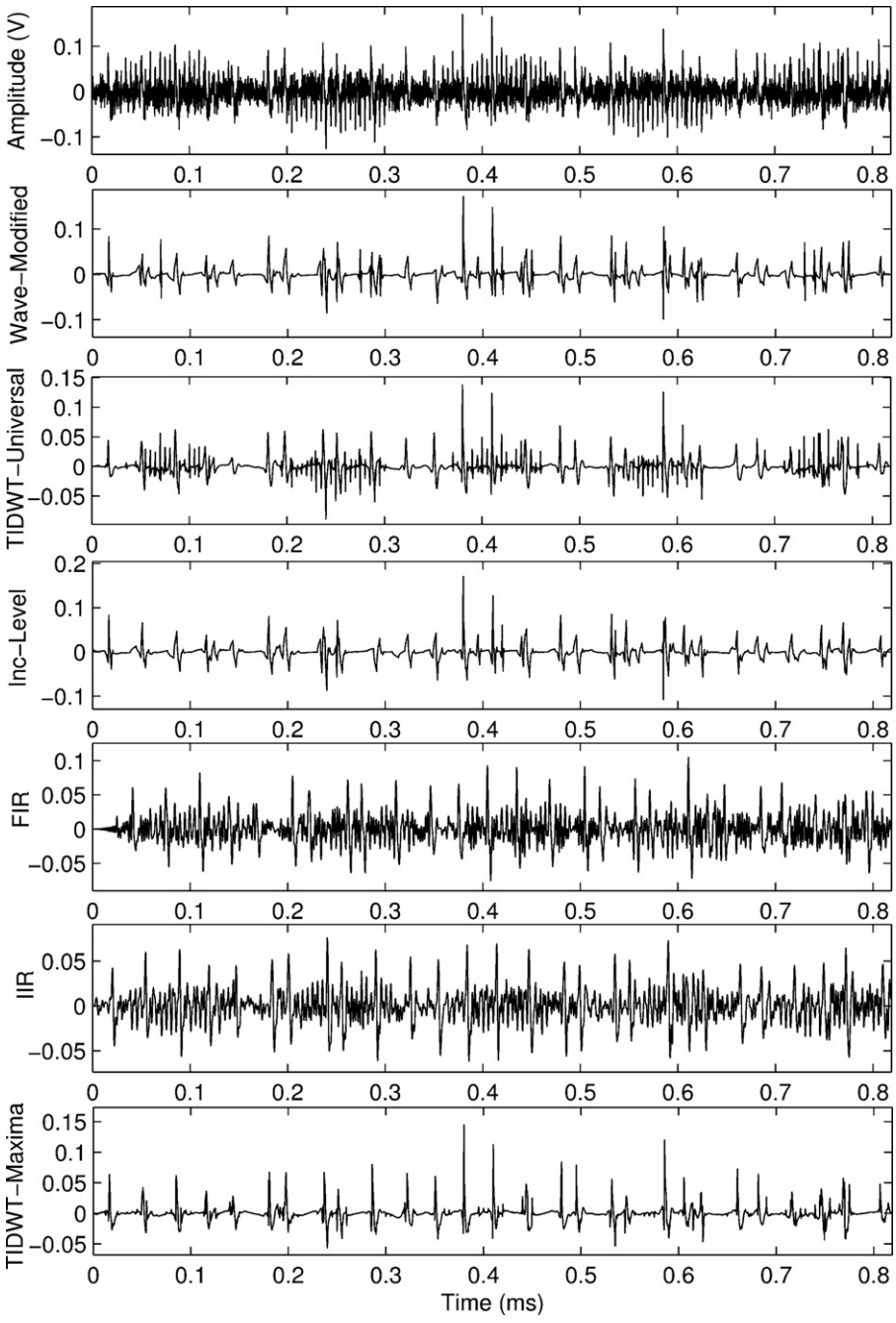

Fig. 16. Measured PD signal corrupted by random and AM noises. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

6.2. Signals from raw PD measurements

In the second stage the technique was evaluated with PD signals obtained from measurements performed in laboratory and in the field. The laboratory measurements were obtained with the arrangement shown in Fig. 14. The system comprises a PD-free high-voltage transformer connected to a PD generation chamber containing a point-plane electrode configuration. The chamber allows the generation of PDs from several electrode and insulating materials configurations, including epoxy resins, SF_6 and oil, but in this work only PDs in air were evaluated. The sensor is an RLC type and is connected to the chamber by a high-

Table 3
Numerical evaluation for signal 2.

	Denoising method					
	A	B	C	D	E	F
SRER (dB)	−1.96	−2.53	3.47	−4.61	−2.77	8.09
R_{XY} (%)	58.65	58.73	74.04	−4.75	32.66	92.03
PAD (%)	23.81	29.10	36.37	12.21	184.15	13.26

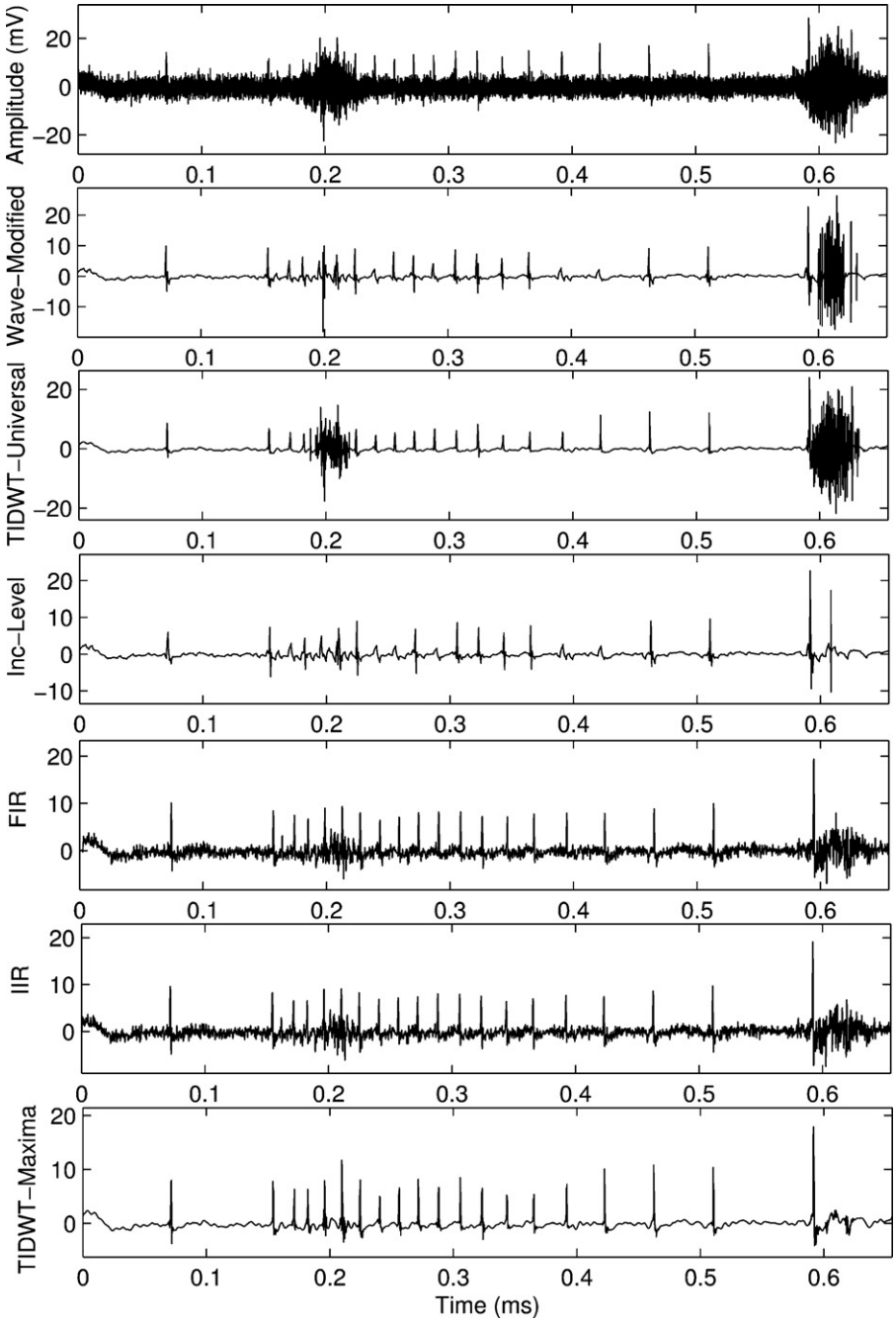

Fig. 17. Measured PD signal from a high-voltage transformer. Top figure: noise-corrupted signal. Bottom figures: denoising results for the evaluated techniques.

voltage coupling capacitor. The sensor output was sent to a digital oscilloscope controlled by a host computer. The measuring system has a 30–400 kHz pass-band and can work with voltages up to 15 kV.

The interferences were introduced on purpose during the laboratory test. Several combinations of PD and noise were evaluated. The first measured signal is shown along with its denoised versions in Fig. 15. It is composed by a set of PD pulses from the negative voltage cycle corrupted by a small amplitude AM noise and a pulsating noise. The latter was generated by a switching voltage converter placed at the vicinities of the measurement set. The signal has 16,384 samples and the sampling frequency is 5 MS/s.

Table 4
Numerical evaluation for signal 3.

	Denoising method					
	A	B	C	D	E	F
SRER (dB)	0.74	1.54	1.59	−5.98	−4.96	3.87
R_{XY} (%)	47.51	60.61	61.18	−1.36	22.90	77.09
PAD (%)	59.27	71.73	34.11	13.38	190.45	35.92

The FIR and IIR filters were adjusted to match the pass-band of the system. The FIR structure is a 246-tap equiripple model and the IIR is a 10-tap Chebychev type-II. SVM training was performed in the following way: signals containing noise-free PDs and only noise were previously measured and used as a training set. After tuning, the machine was used to classify the data from the PD corrupted signal. As can be seen, the methods based on scalar and uniform thresholding were efficient for elimination of the AM noise but failed in the elimination of the pulsating part. The method Inc-Level introduced severe distortions and pulse losses. FIR and IIR filters also caused large distortions and did not remove all the noise. TIDWT-Maxima allowed the best overall PD recovery, even for the pulses located approximately at 0.8 ms, which were completely covered by noise.

The second example is shown in Fig. 16. The signal is composed by PDs from the negative cycle corrupted by a large amplitude AM noise and a random noise. The interference was captured with the help of a wire, which act as an antenna, connected to the sensor input. The signal has 8192 samples and the sampling frequency is 5 MS/s. The FIR and IIR filters were the same used in the previous example since there was no change in the system pass-band. SVM training was also performed in the same way.

All the denoising procedures introduced distortions in the reconstruction, which can be explained by the low SNR condition of this case. The methods wave-modified and TIDWT-Universal caused the loss of several PD pulses, thus reducing the measurement sensitivity level. Inc-Level recovered more PDs but caused severe distortions in their shapes. FIR and IIR were unable to eliminate all the noise. TIDWT-Maxima recovered well the PD signal, with the lowest distortion among all the evaluated methods.

Fig. 17 shows the results for the processing of a PD signal from a 33 MVA power transformer. It was measured with a 10 MHz pass-band sensor coupled to the capacitive tap of the bushing. The sensor output was connected to an oscilloscope and the signal was digitized with a sampling frequency of 25 MS/s. The FIR and IIR filters were adjusted based on the signal frequency spectrum as low-pass equiripple and Chebychev type-II with a cutoff frequency of 3 MHz. The data used to train SVM was composed by a set of calibration pulses measured before energizing the equipment and a sample of noise obtained just below the PD inception level.

Again, the methods wave-modified and TIDWT-Universal performed well for filtering the uniform noise, but let pass the bursts located approximately at 0.2 and 0.6 ms. Inc-Level eliminated most of the noise but caused severe distortions in the pulse shapes and amplitudes. The FIR and IIR filters performed better than the previous examples, probably because of the smaller overlapping of PD and noise spectra, but let pass a small amplitude background noise with a concentration near 0.2 and 0.6 ms. Again, TIDWT-Maxima eliminated most of the noise with the lowest pulse distortion.

7. Conclusion

This work showed the effectiveness of a new PD denoising approach based on spatial correlations of WT coefficients along decomposition levels. It was shown that the method is more efficient than the ones based on uniform or scale-based thresholding since it exploits the localized characteristics of the PDs to better separate the data from noise. Moreover, the spatially adaptive thresholding showed to be particularly efficient for the elimination of pulsating or time-localized interferences, which are recognized to be more difficult to eliminate due to their similarities to the PD characteristics.

The translation-invariant DWT provided benefits related to two aspects of PD denoising. Firstly, it eliminated the loss of pulses caused by the decimation. Secondly, it proportionated softer reconstructions, which resemble better the original pulse shapes. This translates in a higher correlation coefficient and lower amplitude attenuation, thus allowing a better determination of the parameters used for PD analysis. The introduction of the significance matrix allowed both decomposition and reconstruction to be made with the lowest computational loads possible. After SVM training, the algorithms take about 3 seconds to process a 32,768 data vector in a 1.7 GHz Pentium-M processor with 2 GB of RAM. This makes them suitable for on-site use.

Acknowledgments

The authors would like to thank the Department of Computer Science – UFSJ and the Departments of Electrical Engineering and Computer Science – UFMG for supporting this work. Also, H.O. Mota would like to thank CAPES, CNPQ and Fapemig for sponsoring this work.

References

[1] R. Bartnikas, Partial discharges: their mechanism, detection and measurement, IEEE Transactions on Dielectrics and Electrical Insulation 9 (5) (2002) 763–808.
[2] E. Kuffel, W.S. Zaengl, High Voltage Engineering—Fundamentals, Pergamon Press Inc., Oxford, 1984.
[3] L. Satish, B. Nazneen, Wavelet-based denoising of partial discharge signals buried in excessive noise and interference, IEEE Transactions on Dielectrics and Electrical Insulation 10 (2) (2003) 354–367.
[4] International Electrotechnical Commission, Partial Discharge Measurements, 2nd ed., CEI/IEC International Standard 270, 1981.
[5] H. Borsi, E. Grockenbach, D. Wenzel, Separation of partial discharges from pulse-shaped noise signals with the help of neural networks, IEE Proceedings on Science and Measurement Technology 142 (1) (1995) 69–74.
[6] V. Nagesh, B.I. Gururaj, Evaluation of digital filters for rejecting discrete spectral interference in on-site PD measurements, IEEE Transactions on Electrical Insulation 28 (1) (1993) 73–85.
[7] U. Kopf, K. Feser, Rejection of narrow-band noise and repetitive pulses in on-site PD measurements, IEEE Transactions on Dielectrics and Electrical Insulation 2 (6) (1995) 1180–1191.
[8] A. Contin, A. Cavalini, G.C. Montanari, G. Pasini, F. Puletti, Digital detection and fuzzy classification of partial discharge signals, IEEE Transactions on Dielectrics and Electrical Insulation 9 (3) (2002) 335–348.
[9] A. Cavalini, A. Contin, G.C. Montanari, F. Puletti, Advance PD inference in on-field measurements. Part I. Noise rejection, IEEE Transactions on Dielectrics and Electrical Insulation 10 (2) (2003) 216–224.
[10] I. Shim, J.J. Soraghan, W.H. Siew, Detection of PD utilizing digital signal processing methods. Part 3. Open-loop noise reduction, IEEE Electrical Insulation Magazine 17 (1) (2001) 6–13.
[11] X. Ma, C. Zhou, I.J. Kemp, Automated wavelet selection and thresholding for PD detection, IEEE Electrical Insulation Magazine 18 (2) (2002) 37–45.
[12] X. Ma, C. Zhou, I.J. Kemp, Interpretation of wavelet analysis and its application in partial discharge detection, IEEE Transactions on Dielectrics and Electrical Insulation 9 (3) (2002) 446–457.
[13] S. Sriram, S. Nitin, K.M.M. Prabhu, M.J. Bastiaans, Signal denoising techniques for partial discharge measurements, IEEE Transactions on Dielectrics and Electrical Insulation 12 (6) (2005) 1182–1191.
[14] X. Zhou, C. Zhou, I.J. Kemp, An improved methodology for application of wavelet transform to partial discharge measurement denoising, IEEE Transactions on Dielectrics and Electrical Insulation 12 (3) (2005) 586–594.
[15] C.S. Chang, J. Jin, S. Kumar, Q. Su, T. Hoshino, M. Hanai, N. Kobayashi, Denoising of partial discharge signals in wavelet packets domain, IEE Proceedings on Science and Measurement Technology 152 (3) (2005) 129–140.
[16] H. Zhang, T.R. Blackburn, B.T. Phung, D. Sen, A novel wavelet transform technique for on-line partial discharge measurements. Part 1. WT de-noising algorithm, IEEE Transactions on Dielectrics and Electrical Insulation 14 (1) (2007) 3–14.
[17] H. Zhang, T.R. Blackburn, B.T. Phung, D. Sen, A novel wavelet transform technique for on-line partial discharge measurements. Part 2. On-site noise rejection application, IEEE Transactions on Dielectrics and Electrical Insulation 14 (1) (2007) 15–22.
[18] X. Song, C. Zhou, D.M. Hepburn, G. Zhang, Second generation wavelet transform for data denoising in PD measurement, IEEE Transactions on Dielectrics and Electrical Insulation 14 (6) (2007) 1531–1537.
[19] M. Florkowski, B. Florkowska, Wavelet-based partial discharge image denoising, IET Generation, Transmission and Distribution 1 (2) (2007) 340–347.
[20] R.R. Coifman, D.L. Donoho, Translation-invariant de-noising, Lecture Notes in Statistics 103 (1995) 125–150.

[21] S. Mallat, W.L. Hwang, Singularity detection and processing with wavelets, IEEE Transactions on Information Theory 38 (2) (1992) 617–643.
[22] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[23] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice-Hall, Upper Saddle River, 1996.
[24] T. Joachims, Making Large-scale Support Vector Machine Learning Practical. Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, 1999.
[25] J.M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, IEEE Transactions on Signal Processing 41 (12) (1993) 3445–3462.
[26] S.G. Chang, B. Yu, M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising, IEEE Transactions on Image Processing 9 (9) (2000) 1522–1531.
[27] B. Vidakovic, C.B. Lozoya, On time-dependent wavelet denoising, IEEE Transactions on Signal Processing 46 (9) (2001) 2549–2554.
[28] J. Xie, D. Zhang, W. Xu, Spatially adaptive wavelet denoising using the minimum description length principle, IEEE Transactions on Image Processing 13 (2) (2004) 179–187.
[29] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed., Academic Press, San Diego, 1998.
[30] L. Hao, P.L. Lewin, Partial discharge source discrimination using a support vector machine, IEEE Transactions on Dielectrics and Electrical Insulation 17 (1) (2010) 189–197.
[31] V.S. Murthy, K. Tarakanath, D.K. Mohanta, S. Gupta, Insulator condition analysis for overhead distribution lines using combined wavelet and support vector machine (SVM), IEEE Transactions on Dielectrics and Electrical Insulation 17 (1) (2010) 89–99.
[32] C. Chang, C. Lin, LIBSVM: a library for support vector machines, [OnLine] Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[33] S.K. Mitra, Digital Signal Processing: A Computer-based Approach, The MacGraw-Hill Companies, Inc., New York, 1998.