International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 7, No. 4 (2009) 1–23

World Scientific
www.worldscientific.com

WAVELET PACKET DECOMPOSITION FOR THE IDENTIFICATION OF CORROSION TYPE FROM ACOUSTIC EMISSION SIGNALS

GERT VAN DIJCK

Laboratorium voor Neuro- en Psychofysiologie
Computational Neuroscience Research Group
Katholieke Universiteit Leuven, Herestraat 49
B-3000 Leuven, Belgium
gert.vandijck@med.kuleuven.be

MARTINE WEVERS

Departement Metaalkunde en Toegepaste Materiaalkunde
Materiaalgedrag en Niet-destructieve Evaluatie
Katholieke Universiteit Leuven, Kasteelpark Arenberg 44
B-3001 Heverlee, Belgium
martine.wevers@mtm.kuleuven.be

MARC M. VAN HULLE

Laboratorium voor Neuro- en Psychofysiologie
Computational Neuroscience Research Group
Katholieke Universiteit Leuven, Herestraat 49
B-3000 Leuven, Belgium
marc.vanhulle@med.kuleuven.be

Received 31 October 2008
Revised 1 March 2009

Corrosion causes a degradation of the structural integrity of petrochemical plants, nuclear power plants, ships, bridges and other constructions containing steel with the consequence that people and the environment may be exposed to dangerous situations. The detection of corrosion and the prediction of the type of corrosion are studied in this article by means of the acoustic emission technique. We use a wavelet packet decomposition to compute features from the acoustic emission signals. The basis functions with the highest discriminative power are selected according to the highest pair-wise Kullback–Leibler divergence between distributions of wavelet coefficients. It is proven that the pair-wise Kullback–Leibler divergence used in the local discriminant basis algorithm requires class conditional independence of the wavelet coefficients. Several classification algorithms using the most discriminative wavelet coefficients are compared for the prediction of three types of corrosion and the absence of corrosion.

Keywords: Information theory; pattern recognition; wavelets.

AMS Subject Classification: 94A15, 68T10, 65T60

1. Introduction

Acoustic emission (AE) is the elastic wave propagation resulting from the rapid release of energy within a material. According to Shaikh *et al.*[16]: "The acoustic emission phenomenon is the result of transient elastic wave propagation generated by a rapid release of energy within a material due to changes in local stress and strain fields. These elastic waves propagate over a wide range of frequencies ranging from audible frequencies to frequencies in MHz range". However, this definition fails to recognize that there exists also continuous emission besides burst (transient) emission.[21] An advantage of the acoustic emission technique among non-destructive testing techniques is that it allows to monitor the structure or material continuously and hence damage can be detected when it occurs.[21] A second advantage is that structures and tools need not to be taken out of service for testing.

These advantages have an impact on how planning of inspections can be organized nowadays. Traditionally, inspection of structures such as petrochemical plants,[22] occurs periodically, e.g., every six months or every year. This inspection is then often performed visually or by means of testing techniques which may require the installation to be taken out of service temporarily. However, such periodic inspection is not adapted to the damage state of the plant with the possible consequence that damage may occur immediately after inspection. A further degradation of structural integrity is then allowed until the next inspection. Therefore, the acoustic emission technique has been used for monitoring in applications where possibly large damage to people and environment may occur and hence anticipation to the damage needs to be fast. Especially in critical applications where corrosion is the source of damage, the acoustic emission technique has been used extensively, e.g., in petrochemical plants,[22] nuclear power plants,[3] offshore installations[12] and ship hull structures.[19]

1.1. *Acoustic emission in corrosion*

This article focuses on the prediction of common types of corrosion that occur in chemical plants[22]: uniform corrosion, pitting and stress corrosion cracking (SCC). In fact, one should take into account the possibility that there is no corrosion process active and this forms a fourth class to be predicted. Corrosion phenomena lead to a redistribution of the energy within a material and therefore become a potential source for acoustic emission activities. Different processes lead to the emission of acoustic activity in corrosion[23]: breakdown of thick oxide film, crack growth, fracture or decohesion of precipitate and inclusions at crack tip, hydrogen gas evolution, metal dissolution, plastic deformation by slip or twin at crack tip and stress induced martensitic transformation among others. These processes are illustrated in Fig. 1.

In Winkelmans,[22] it was observed that absence of corrosion and uniform corrosion are characterized by continuous emission, while stress corrosion cracking and pitting are characterized by a burst type emission. The fact that under conditions

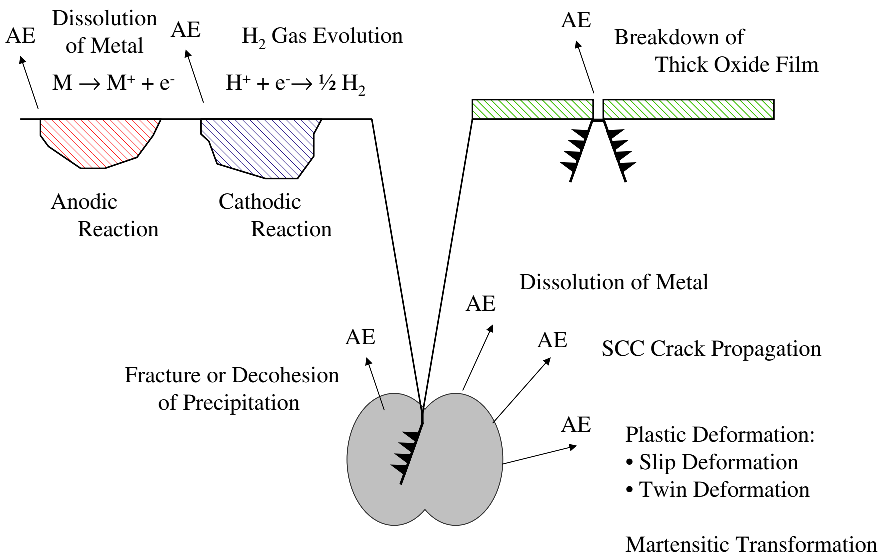

Fig. 1. Sources of acoustic emission in corrosion: Breakdown of thick oxide film, crack propagation, decohesion or fracture of precipitation, dissolution of metal, H_2 gas evolution, martensitic transformation and plastic deformation. Figure adapted from Yuyama.[23]

of uniform corrosion only a very limited number of events (bursts) can be detected as opposed to non-uniform corrosion and intense localized corrosion such as in pitting and SCC is also supported in Seah *et al.*[15] and Jaubert.[6] Example signals of absence of corrosion and uniform corrosion are shown in Fig. 2. As can be seen both signals are of continuous type emission.

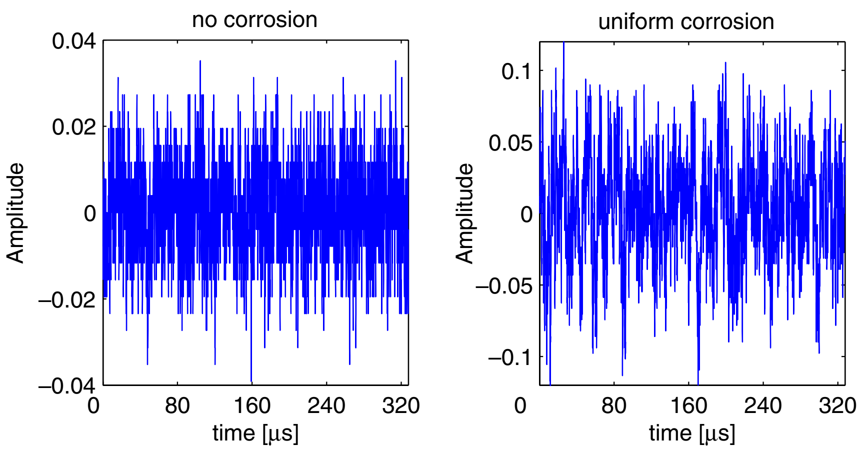

Fig. 2. Example signal of absence of corrosion (no corrosion) on the left. Example signal of uniform corrosion on the right. Both AE signals are continuous type emission signals.

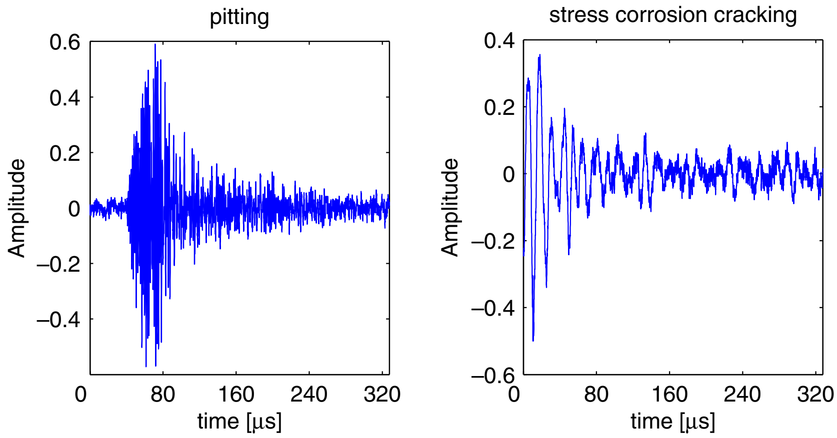

Fig. 3. Example signal of pitting on the left. Example signal of stress corrosion cracking (SCC) on the right. Both AE signals are burst type emission signals.

Example signals of pitting and stress corrosion cracking (SCC) are shown in Fig. 3. As can be seen both signals are of burst type emission.

1.2. *Prediction of type of corrosion*

There are at least two important reasons why industrial experts should try to distinguish between different types of corrosion.

Firstly, pitting and SCC are more harmful types of corrosion compared to uniform corrosion. Uniform corrosion reduces the thickness of the material relatively uniformly, hence taking a long time before holes are formed in the material. On the other hand, pitting causes pits and SCC causes cracks which can grow much faster through the thickness of the material. This may sooner lead to leaks in chemical and nuclear plants. Therefore, occurrence of pitting and SCC AE events should trigger sooner a visual inspection of the installation.

Secondly, the discrimination between different corrosion processes should be performed prior to the quantitative analysis of correlating acoustic emission activity to the corrosion rate. In Seah *et al.*,[15] a quantitative analysis has shown that the count rate (this is defined by the authors as the total number of threshold crossings of AE signals per unit area of the exposed part of the metal sample and per unit time) is correlated with the rate of corrosion measured by means of the weight loss of the metal sample. A quantitative relation between the number of AE events and the number of pits in pitting as well with the pitted area and volume was established in Mazille *et al.*[10] In stress corrosion cracking, a relationship between AE parameters (counts change per unit time and energy change per unit time) and the corrosion speed (change of crack length per unit time) has been established.[16] This shows that in different corrosion processes one can estimate the corrosion speed from AE parameters, although one should first link an AE event to the corresponding corrosion process. Erroneously relating AE events originating from pitting to SCC leads to a wrong estimate of the corrosion speed of SCC and vice versa.

2. Processing Stages

This section describes the different steps for making predictions of the type of corrosion starting from the signal acquisition.

2.1. *Signal acquisition*

We describe briefly the experimental set-up in obtaining the acoustic emission signals. A steel sample is shown by means of the U-shape in Fig. 4. The probe is designed such that the corrosion occurring in the probe is representative for the corrosion occurring in the plant.[22] This means that the probe is made of the same type of steel as the plant and that the probe is exposed to the same environmental conditions: the corrosive medium, temperature and pressure. This is represented in Fig. 4 by means of the input flow that arrives from the plant and the output flow that is guided back to the plant.

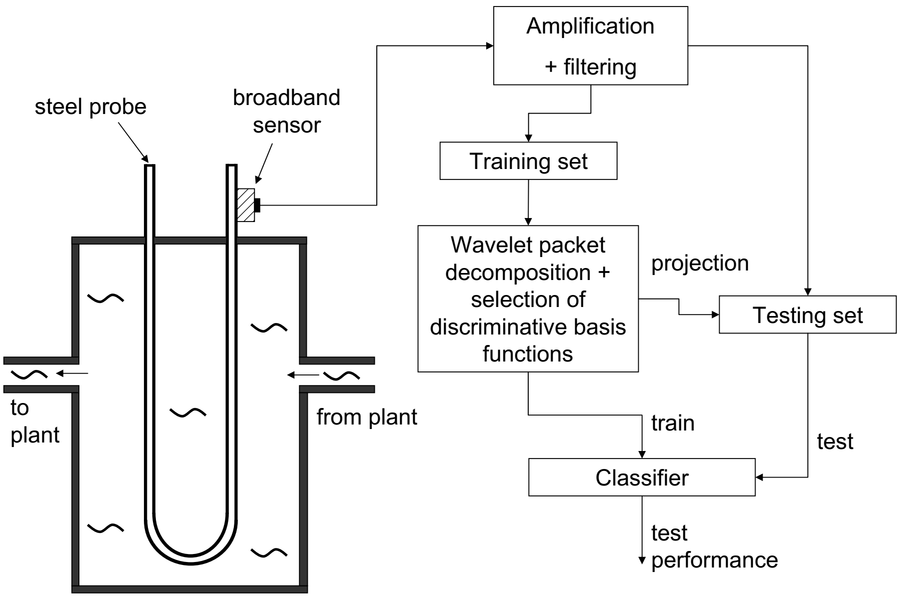

Fig. 4. Processing stages for making predictions of the corrosion type. A steel probe is exposed to the same environmental conditions as the installation. Subsequently, AE signals are amplified and filtered. Features are computed from the signals by means of a wavelet packet decomposition. A classifier is trained based on the selected wavelet coefficients of the training set. Testing signals are projected onto the selected basis functions. Subsequently, the wavelet coefficients of the testing signals are used to test the performance of the system.

The advantage of measuring the corrosion on a reference probe can be seen as follows. The probe is a relatively small: approximately 300 m in height. This means that dampening of the waves when they propagate over such small distances is small. On the other hand when performing measurements on the large installation itself, AE waves may have dampened out before they reach a sensor when there is no sensor in the neighborhood of the AE source. Hence, when a dense configuration of sensors is not used a lot of AE events may be missed. Moreover, due to the large difference in distances that waves may have travelled, AE events can be deformed to different degrees e.g., due to dispersion. This deformation will hamper the recognition of the type of corrosion from the waveforms. Thirdly, installations are often exposed to external sources that can create AE events: e.g., mechanical vibrations, rain drops, etc. These sources may be confounded with AE events originating from corrosion events.

The damage that occurs in the probe can be captured by means of piezoelectric sensors attached to the corroding probe. In order to guarantee a good acoustical transfer from the probe to the sensor, a high vacuum grease (DOW Corning®) is applied between the sensor and the probe. The sensors used[22] here are broadband sensors (B1025, Digital Wave Corporation). This sensor has a guaranteed frequency bandwidth from 50 kHz to 2 MHz and can be used in a temperature range from −50°C to 100°C. Subsequently, the signals are amplified with an amplification factor of approximately 40 dB. The signals are then bandpass filtered between 50 kHz–2 MHz, because outside this range the sensor does not guarantee reliable information. Signals are sampled at 20 MHz or 25 MHz, both sampling rates are safely higher than the Nyquist sampling rate of 4 MHz for signals up to 2 MHz. Before computing any wavelet transform, signals are resampled to 25 MHz if they were sampled at 20 MHz.

2.2. *Feature construction, feature selection and prediction*

The section describes briefly the steps taken to predict the corrosion type. The wavelet packet decomposition and the selection of the basis functions are more thoroughly described in Secs. 3 and 4, respectively.

After signals have been obtained from each of the 4 classes: "absence of corrosion", uniform corrosion, pitting and SCC, the signals are divided into a training set and a testing set as shown in Fig. 4. Next, the wavelet coefficients of all training signals are computed from a wavelet packet decomposition, see Sec. 3. We use the wavelet coefficients as the constructed features. The most discriminative basis functions are selected by means of the wavelet coefficients of the training signals for which the pair-wise Kullback–Leibler divergence is the highest, see Sec. 4. This means that the pair-wise Kullback–Leibler divergence is used as the feature selection criterion. The testing signals are projected on the selected basis functions. Subsequently, a classifier is trained using the selected wavelet coefficients of the training set. In Sec. 5, we will consider different classifiers. Using the wavelet coefficients of the test signals, we assess the performance of the system shown in Fig. 4.

1st Reading
June 22, 2009 11:32 WSPC/181-IJWMIP 00306

3. Wavelet Packet Decomposition

We motivate the use of wavelet packet decompositions and provide the necessary background.

3.1. *Feature construction from wavelet packets*

A basic approach to construct features consists in computing some general statistical parameters from time series such as the median, the mean, the standard deviation and higher-order moments. A more thorough approach exists in using templates that can be used to construct features. The prior information about the processes to be predicted is then related to the choice of the templates. However, generic approaches that generate a library of templates, such as wavelet packets, exist.[8] Wavelet packet decompositions (WPD)[8,9] offer a library of templates that have many desired properties. First of all, WPD's are founded on a solid mathematical theory[9] that allows to represent the signals in new bases. The decomposition in a new wavelet packet basis guarantees that no "information" is lost as the original signals can always be reconstructed from the new basis. Secondly, the templates in a wavelet packet decomposition are easily interpreted in terms of frequencies and bandwidths.[9] Thirdly, wavelet packet decompositions are more flexible than the discrete wavelet transform and the Fourier transform. This means that the basis functions that are used in a discrete wavelet transform (DWT) are also available in the wavelet packet decomposition, see Sec. 3.2.

3.2. *Wavelet packet decomposition background*

This section introduces the necessary background to understand feature construction from wavelet packet decompositions. This background is needed in order to understand the feature selection in Sec. 4. We will use the terminology of template and basis function interchangeably. Strictly speaking, a template is a more general terminology, because it does not need to be part of a basis.

We represent a single time series by means of a sequence of observations $x(t)\,:\,x(0), x(1), \ldots, x(N-1)$, where "$t$" refers to the time index and "N" is the number of samples. Time series $x(t)$ can be considered as being sampled from an "N" dimensional distribution defined over an "N" dimensional variable $X(t) : X(0), X(1), \ldots, X(N-1)$, we write this "$N$" dimensional variable in shorthand notation as $X_{0:N-1}$. Features are computed from a wavelet packet decomposition by computing the inner product between the templates and the time series (using a continuous notation, for the ease of notation):

$$\gamma_{i,j,k} = \langle x(t), \psi_i^j(t - 2^i k)\rangle = \int_{-\infty}^{+\infty} x(t)\psi_i^j(t - 2^i k)\, dt. \tag{3.1}$$

A feature, in this case a wavelet coefficient, in the wavelet packet decomposition needs to be specified by the scale index "i", frequency index "j" and time index "k". The coefficient $\gamma_{i,j,k}$ can be considered as quantifying the similarity, by means of the inner product, between time series $x(t)$ and wavelet function $\psi_i^j(t - 2^i k)$ at position $2^i k$ in time. The parameter "i" is the scale index and causes a dilation (commonly called a "stretching") of the wavelet function $\psi^j(t)$ by a factor 2^i:

$$\psi_i^j(t) = \frac{1}{\sqrt{2^i}} \psi^j \left(\frac{t}{2^i} \right). \tag{3.2}$$

It is the parameter "j" that determines the shape of the template. In case we choose the 12-tap Coiflet filter,[13] we obtain the first 8 different templates $\psi^0(t), \psi^1(t), \psi^2(t), \ldots, \psi^7(t)$ shown in Fig. 5. This 12-tap Coiflet filter has been consistently used in the experiments in Sec. 5. The construction of these basis functions can be found in text books.[9]

The shapes of these basis functions also motivates the use of wavelet packet decompositions in our application. With an appropriate scaling and time shift some of the basis functions in Fig. 5 resemble the AE bursts in Fig. 3. Choosing the appropriate template, the scaling factor and the time shift is the task of the feature selection procedure in Sec. 4.

In Fig. 6, we show a graphical representation of the different subspaces that are obtained in a wavelet packet decomposition. In the discrete wavelet transform the only nodes in the tree that are considered are W_1^1, W_2^1, W_3^1 and W_3^0 these subspaces are shaded in grey.

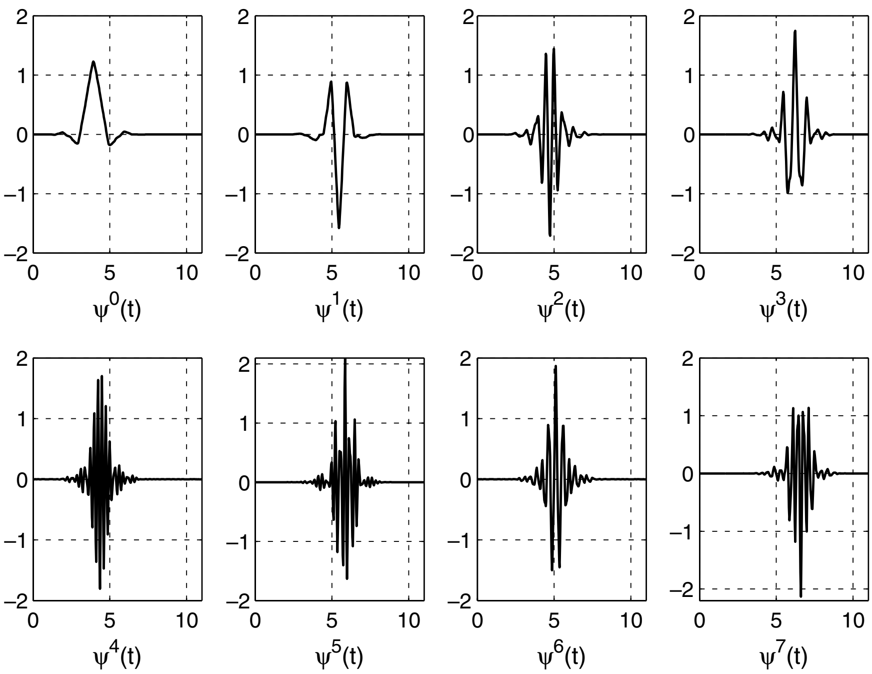

Fig. 5. Templates (wavelet packets) corresponding with the 12-tap Coiflet filter.

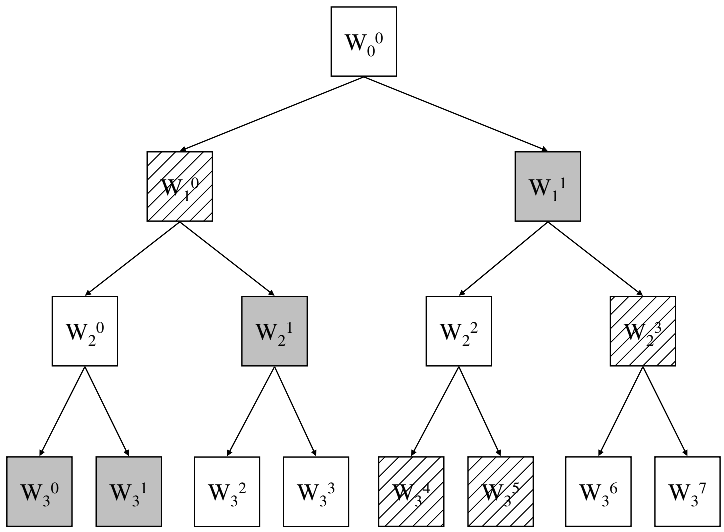

Fig. 6. Library of wavelet packet functions. Different subspaces are represented by W_i^j. Index "i" is the scale index, index "j" is the frequency index. The depth "I" of this tree is equal to 3. Every tree within this tree where each node has either 0 or 2 children is called an admissible tree. Two admissible trees are emphasized, one shaded in grey and one marked with diagonals.

The first three subspaces are spanned by the functions $\{\psi_1^1(t-2k)\}_{k\in\mathbb{Z}}$, $\{\psi_2^1(t-2^2k)\}_{k\in\mathbb{Z}}$ and $\{\psi_3^1(t-2^3k)\}_{k\in\mathbb{Z}}$, respectively. Subspace W_3^0 is spanned by $\{\psi_3^0(t-2^3k)\}_{k\in\mathbb{Z}}$. So in the discrete wavelet transform the signals are only analyzed by means of the time translated functions of $\psi_3^0(t)$ ($\psi_0^0(t)$ is called the scaling function and is shown as the first template in Fig. 5) and dilated and time translated functions of $\psi_0^1(t)$ (this function is called the mother wavelet function and is shown as the second template in the top row of Fig. 5). The division in subspaces in Fig. 6 also corresponds to a tiling of frequency space.[9] In Fig. 6, only two bases are shown: the grey shaded basis corresponds with the discrete wavelet transform, the basis marked with diagonals is chosen arbitrarily and is one of the possible bases in the wavelet packet decomposition. The basis marked with diagonals puts more emphasis on a finer analysis of the higher frequency part of the signals.

Retaining any binary tree in Fig. 6, where each node has either 0 or 2 children, leads to an orthonormal basis for finite energy functions, denoted as $x(t) \in L^2(\mathbb{R})$:

$$\int_{-\infty}^{+\infty} |x(t)|^2\, dt < \infty. \tag{3.3}$$

Such a tree is called an admissible tree. If the leaves of this tree are denoted by $\{i_l, j_l\}_{1\le l\le L}$ the orthonormal system can be written as:

$$W_0^0 = \oplus_{l=1}^{L} W_{i_l}^{j_l}. \tag{3.4}$$

This means that the space W_0^0, which is able to represent the input space of the time series, can be decomposed into orthonormal subspaces $W_{i_l}^{j_l}$. For reasons of completeness, it should be mentioned that some conditions (conjugate mirror filter[9]) should be satisfied in order to form an orthonormal system. The mathematical details are out of scope for this paper, we emphasize here the use of the wavelet packet decompositions as a feature construction method based on a library of templates.

It is should be noted that a full wavelet packet decomposition yields to many features. In cases where one can assume that the exact time location "k" of the template is of no importance, one can, e.g., consider an average or the energy of wavelet coefficients over time for each possible combination of the scale index "i" and the frequency index "j". This will lead to less features to be selected from. Here, we will consider the full complexity of the problem, when the exact time location of the template can be of importance, and consider all coefficients from a full wavelet packet decomposition to be selected from. A full wavelet packet decomposition leads to $N*(\log_2 N+1)$ features. This can be seen as follows. From Fig. 6, it can be noted that the number of subspaces at a certain scale "i" is determined by the scale index "i". The number of subspaces at scale "i" is equal to 2^i. Therefore the frequency index "j" at a certain scale "i" will be an integer from $[0, 2^i-1]$, indicating the starting position of the subspace at scale "i". As can be seen from Eq. (3.1) at scale "i" the inner products are computed at discrete time positions $2^i k$. Therefore at scale 0, we obtain "N" (length of the signals) coefficients: $\gamma_{0,0,0}, \ldots, \gamma_{0,0,N-1}$. At the next scale "$i$" $= 1$ we obtain "$N/2$" coefficients in each subspace i.e., $\gamma_{1,0,0}, \ldots, \gamma_{1,0,N/2-1}$ and $\gamma_{1,1,0}, \ldots, \gamma_{1,1,N/2-1}$. At the highest frequency resolution, "i" $= \log_2 N$ and we obtain coefficients: $\gamma_{logN,0,0}, \ldots, \gamma_{logN,N-1,0}$. Hence at each scale there are "N" coefficients and in total there are $\log_2 N + 1$ different scale levels. This leads overall to $N * (\log_2 N + 1)$ different coefficients to select from. The variables that can be associated with the coefficients $\gamma_{i,j,k}$ are further denoted by capitals $\Gamma_{i,j,k}$.

4. Selection of Basis Functions

In this section, we will consider the selection of the most discriminative basis functions $\psi_i^j(t - 2^i k)$ in order to make a prediction about the target variable "y" (the corrosion class). The target variable is a class variable taking values $1 \cdots \#C$, where $\#C$ is the total number of classes. An outline of the Local Discriminant Basis algorithm[13] is provided and we reveal some limitations in this algorithm.

4.1. *Motivation basis function selection*

The basis functions are not selected directly, but indirectly by means of the coefficients $\gamma_{i,j,k}$. The selection of a coefficient $\gamma_{i,j,k}$ implies that the basis function $\psi_i^j(t - 2^i k)$ should be selected. With this basis function we can identify an associated frequency band[9] as well as the time localization "k" of the frequency band.

Because one can interpret the basis functions intuitively in terms of these frequency bands, the method of choice is a feature subset selection procedure rather than a feature extraction procedure such as principal component analysis (PCA), linear discriminant analysis (LDA) or maximization of mutual information (MMI).[18] Feature subset selection procedures select features, while feature extraction procedures make combinations of the coefficients in a linear or nonlinear way. This would imply that the newly constructed features contain contributions from possibly many basis functions, which makes the interpretation cumbersome.

4.2. *Local discriminant basis algorithm*

The new local discriminant basis (LDB) algorithm[14] is summarized. We assume that we are given a set of training signals $\mathbf{x}_j$ and, for each one of them, we are given the associated target class c_j: $\{\mathbf{x}_j, c_j\}$.

Step 0. Expand each training signal into a time-frequency dictionary D: this involves the computation of all coefficients $\gamma_{i,j,k}$ for each training signal and assumes that we choose a particular conjugate mirror filter[9] in advance that will define the templates.

Step 1. Estimate the class conditional probability density functions (PDF's) for each wavelet coefficient variable, $\Gamma_{i,j,k}$, in the dictionary. Superscript "y" refers to the class label, with $y = 1, 2, \ldots, \#C$ and $\#C$ is the total number of classes. These PDF's were estimated by means of the averaged shifted histograms method (ASH) as in Saito *et al.*[14]

Step 2. For each wavelet coefficient variable, $\Gamma_{i,j,k}$, compute the discriminant measure $\delta_{i,j,k}$. The computational cost of this procedure is $\mathrm{O}((N+1)\log_2 N)$. Many discriminant measures can be used in practice. We use the symmetric relative entropy, Eq. (4.2), as in Saito *et al.*[14] The relative entropy for $\Gamma_{i,j,k}$ between two classes, $y = 1$ and 2, can be computed as[4]:

$$D(\hat{p}^1(\Gamma_{i,j,k}), \hat{p}^2(\Gamma_{i,j,k})) \triangleq \int \hat{p}^1(\gamma_{i,j,k}) \log \frac{\hat{p}^1(\gamma_{i,j,k})}{\hat{p}^2(\gamma_{i,j,k})}\, d\gamma_{i,j,k}. \tag{4.1}$$

Because this discriminant measure is, in general, not symmetric, a symmetric version is obtained as:

$$\begin{aligned}\delta_{i,j,k} &= D^S(\hat{p}^1(\Gamma_{i,j,k}), \hat{p}^2(\Gamma_{i,j,k}))\\ &= D(\hat{p}^1(\Gamma_{i,j,k}), \hat{p}^2(\Gamma_{i,j,k})) + D(\hat{p}^2(\Gamma_{i,j,k}), \hat{p}^1\Gamma_{i,j,k})).\end{aligned} \tag{4.2}$$

When more than two classes are considered, $\delta_{i,j,k}$, is defined as the sum over all $(\#C.(\#C-1))/2$ pairs of different classes as:

$$\begin{aligned}&D^S_{Pair}(\hat{p}^1(\Gamma_{i,j,k}), \hat{p}^2(\Gamma_{i,j,k}), \ldots, \hat{p}^c(\Gamma_{i,j,k}))\\ &\quad = \sum_{m=1}^{\#C-1} \sum_{n=m+1}^{\#C} D^S(\hat{p}^m(\Gamma_{i,j,k}), \hat{p}^n(\Gamma_{i,j,k})).\end{aligned} \tag{4.3}$$

Step 3. Evaluate the discriminant power of each basis $B \in D$ (the dictionary) and obtain the best basis Ψ for which the discriminant power is maximal:

$$\Psi = \underset{B \in D}{\operatorname{argmax}} \sum_{(i,j,k) \in B} \delta_{i,j,k}. \tag{4.4}$$

Hence, one searches for the indices (i, j, k) such that the associated basis functions form a basis B. This corresponds also with the search for an admissible tree in Fig. 6 with the largest discriminant power.

Step 4. Select "m" basis functions, $\psi_i^j(t - 2^i k)$, from Ψ corresponding to the "m" largest $\delta_{i,j,k}$. The number of basis functions "m" to be retained is not defined in Saito *et al.*[14] Therefore, we perform experiments for "m" ranging from 1 to 50 basis functions.

Step 5. Construct classifiers with features derived from the "m" basis functions. In the construction of new features from the wavelet coefficients, one can exploit prior knowledge about the differences between the different processes. Suppose, e.g., that the energy within frequency bands is important to distinguish the different processes and not the exact time location, one can then construct features such as the sum of squares of the wavelet coefficients within each frequency band: $\sum_k \gamma_{i,j,k}^2$. We did not make such assumptions and use the "m" coefficients, $\gamma_{i,j,k}$, rather than deriving new features from the coefficients. Experiments with different classifiers are performed.

4.2.1. *Best basis from the dictionary*

Performing an exhaustive search over all possible bases in the dictionary D, see Eq. (4.4), is computationally infeasible. This is due to the fact that the number of possible bases that can be selected from a wavelet packet tree grows exponentially with the length "N" of the signal. This can be easily seen as follows. It is proven in Mallat[9] that the number of bases, denoted by B_I, in a wavelet packet binary tree (as shown in Fig. 6) of depth I satisfies:

$$2^{2^{I-1}} \leq B_I \leq 2^{\frac{5}{4}2^{I-1}}. \tag{4.5}$$

The maximal depth of a wavelet packet tree is equal to $\log_2 N$. Filling this out in the lower and upper bound for the number of bases leads to:

$$2^{\frac{1}{2}N} \leq B_I \leq 2^{\frac{5}{8}N}. \tag{4.6}$$

Hence, the number of possible bases increases exponentially with the length "N" of the signal. Note that in Eq. (4.4) the discriminant power of a basis is written as the sum of the discriminant powers of its coefficient variables. The discriminant power for a particular node (i, j) in the binary tree can thus be computed by summing over the time indices in that node:

$$D_{i,j} = \sum_{k=0:N/2^i-1} \delta_{i,j,k}. \tag{4.7}$$

The search for an optimal basis in practice is performed as follows[9,13]:

Step 3.1. Set $A_I^j = W_I^j$ and $\Delta_I^j = D_{I,j}$ for $j = 0, \ldots, 2^I - 1$ with "I" the maximal depth of the tree.

Step 3.2. Determine the best subspace A_i^j for $i = I - 1, \ldots, 0,\ j = 0, \ldots, 2^i - 1$ by the following rule:

$$\text{set: } \Delta_i^j = D_{i,j}. \tag{4.8}$$

$$\text{if: } \Delta_i^j \geq \Delta_{i+1}^{2j} + \Delta_{i+1}^{2j+1}. \tag{4.9}$$

$$\text{then: } A_i^j = W_i^j. \tag{4.10}$$

$$\text{else: } A_i^j = A_{i+1}^{2j} \oplus A_{i+1}^{2j+1} \text{ and } \Delta_i^j = \Delta_{i+1}^{2j} + \Delta_{i+1}^{2j+1}. \tag{4.11}$$

It can be proven that the above algorithm leads to the best basis, see Proposition 9.5, p. 403 in Ref. 9. Intuitively, this can easily be seen as follows. We initialize the best basis with the 2^I subspaces from the bottom of the tree at depth "I", i.e., W_I^j for $j = 0, \ldots, 2^I - 1$. Subsequently, it is tested at the next higher level in the tree, i.e., we go from "$i + 1$" to "i", whether a better basis from the nodes at level "i" can be found than the best basis found so far. If the discriminant power of node (i, j) is higher than the best subtree below that node, i.e., $\Delta_i^j \geq \Delta_{i+1}^{2j} + \Delta_{i+1}^{2j+1}$, this node replaces the underlying best subtree, i.e., $A_i^j = W_i^j$. If node (i, j) has a lower discriminant power, we simply keep the best subtree found so far, i.e., $A_i^j = A_{i+1}^{2j} \oplus A_{i+1}^{2j+1}$. At every iteration level "i" it is clear one disposes of the best basis considered over depths "I" till "i". This process is repeated until the top node of the tree is reached. The final test is to compare the discriminant power of the basis formed by the top node of the tree, W_0^0, with the discriminant power of the basis of the best subtree under the top node.

4.3. *Restrictions of the LDB algorithm*

In Step 3, the algorithm searches a basis Ψ for which the discriminant power is maximal. However, the total discriminant power in Step 3 is computed as the sum of the discriminant measures of each of the coefficients in a basis B: $\sum_{(i,j,k)\in B} \delta_{i,j,k}$. The additive property of the discriminant powers of coefficients in a basis leads to a very rapid search for the basis with the highest discriminant power. It easily seen that an optimal basis can be found in $O(N)$ comparisons, with "N" the length of the signal, see Ref. 9. However, one has to question which "limiting" assumptions need to be made to obtain this additive property for the symmetric relative entropy. This may reveal a weakness in the LDB algorithm. In the following theorem we proof that one requires that the coefficient variables are class conditional independent for each basis. This means that the coefficient variables need to be independent when conditioned on each class for each basis. This result was derived in Van Dijck.[20]

Theorem 4.1. *The full symmetric relative entropy based on the N-dimensional class conditional probability density functions (PDF's), $\hat{p}^y(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})$, is equal to a sum of marginal symmetric relative entropies if the coefficient variables for each basis are class conditional independent, i.e.,* $\forall B, y{:}\hat{p}^y(\{\gamma_{i,j,k}\}_{(i,j,k)\in B}) = \prod_{(i,j,k)\in B}\hat{p}^y(\gamma_{i,j,k})$.

Proof. We provide a proof for the case there are 2 classes, i.e., "y" $= 1$ or "y" $= 2$. The proof for more than 2 classes is straightforwardly obtained from this proof. First, let us denote the N-dimensional class conditional PDF for Class 1 and Class 2 for basis B, respectively, as: $\hat{p}^1(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})$ and $\hat{p}^2(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})$.

We write the symmetric relative entropy for basis B based on the N-dimensional class conditional PDF's as:

$$\begin{aligned}
&D^S(\hat{p}^1(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}), \hat{p}^2(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}))\\
&\quad \triangleq \int \hat{p}^1(\{\gamma_{i,j,k}\}_{(i,j,k)\in B}) \log \frac{\hat{p}^1(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})}{\hat{p}^2(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})} \prod_{(i,j,k)\in B} d\gamma_{i,j,k}\\
&\qquad + \int \hat{p}^2(\{\gamma_{i,j,k}\}_{(i,j,k)\in B}) \log \frac{\hat{p}^2(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})}{\hat{p}^1(\{\gamma_{i,j,k}\}_{(i,j,k)\in B})} \prod_{(i,j,k)\in B} d\gamma_{i,j,k}.
\end{aligned} \tag{4.12}$$

It is this discriminant measure that tells the full "truth" about the discriminant power of basis B. Hence, bases should in fact be compared based on this discriminant measure. If we assume class conditional independence of the $\gamma_{i,j,k}$ for each of the classes, we have for Class 1:

$$\hat{p}^1(\{\gamma_{i,j,k}\}_{(i,j,k)\in B}) = \prod_{(i,j,k)\in B} \hat{p}^1(\gamma_{i,j,k}), \tag{4.13}$$

and for Class 2:

$$\hat{p}^2(\{\gamma_{i,j,k}\}_{(i,j,k)\in B}) = \prod_{(i,j,k)\in B} \hat{p}^2(\gamma_{i,j,k}). \tag{4.14}$$

Then filling out in $D^S(\hat{p}^1(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}), \hat{p}^2(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}))$ we obtain:

$$\begin{aligned}
&D^S(\hat{p}^1(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}), \hat{p}^2(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}))\\
&\quad = \int \left(\prod_{(i,j,k)\in B} \hat{p}^1(\gamma_{i,j,k})\right) \log \frac{\displaystyle\prod_{(i,j,k)\in B} \hat{p}^1(\gamma_{i,j,k})}{\displaystyle\prod_{(i,j,k)\in B} \hat{p}^2(\gamma_{i,j,k})} \prod_{(i,j,k)\in B} d\gamma_{i,j,k}\\
&\qquad + \int \left(\prod_{(i,j,k)\in B} \hat{p}^2(\gamma_{i,j,k})\right) \log \frac{\displaystyle\prod_{(i,j,k)\in B} \hat{p}^2(\gamma_{i,j,k})}{\displaystyle\prod_{(i,j,k)\in B} \hat{p}^1(\gamma_{i,j,k})} \prod_{(i,j,k)\in B} d\gamma_{i,j,k}.
\end{aligned} \tag{4.15}$$

Writing the logarithm of a product as a sum of logarithms and integrating out variables not appearing within the logarithm. This can be further written:

$$\sum_{(i,j,k)\in B} \int \hat{p}^1(\gamma_{i,j,k}) \log \frac{\hat{p}^1(\gamma_{i,j,k})}{\hat{p}^2(\gamma_{i,j,k})}\, d\gamma_{i,j,k} + \sum_{(i,j,k)\in B} \int \hat{p}^2(\gamma_{i,j,k}) \log \frac{\hat{p}^2(\gamma_{i,j,k})}{\hat{p}^1(\gamma_{i,j,k})}\, d\gamma_{i,j,k} = \sum_{(i,j,k)\in B} \delta_{i,j,k}. \tag{4.16}$$

So, we conclude that under the condition of class conditional independence, the high-dimensional symmetric relative entropy for basis B can be written as the sum of $\delta_{i,j,k}$:

$$D^S(\hat{p}^1(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B}), \hat{p}^2(\{\Gamma_{i,j,k}\}_{(i,j,k)\in B})) = \sum_{(i,j,k)\in B} \delta_{i,j,k}. \tag{4.17}$$

□

Of course when such assumptions of class conditional independence are needed; the complex dependencies between the wavelet coefficients within a basis are not taken into account. Therefore the approximation of the full symmetric relative entropy by $\sum_{(i,j,k)\in B} \delta_{i,j,k}$ may be inaccurate. This occurs e.g., if coefficients within a basis are dependent.

A second restriction of the LDB algorithm is present in Step 4. Once a basis has been obtained, the basis functions are ordered according to the descending order of the individual discriminant measures $\delta_{i,j,k}$ of the basis functions. Hence, the information that is present in the previous coefficient variables $\Gamma_{i,j,k}$ is not taken into account when selecting the next coefficient variables. The result is that the first two features have individually a high discriminant power, but it may be that these features are strongly dependent (or correlated in narrower sense) and they are not necessarily the best set of two features.

It has been shown[20] that these restrictions can be avoided by using high-dimensional estimators of the dependency between the target variable "y" and the wavelet coefficient variables $\Gamma_{i,j,k}$. However, taking coefficient dependencies into account using high-dimensional estimators the additive property can not be used and therefore the search for an optimal basis becomes computationally infeasible. In that case, one has to restrict the search to a set of discriminative basis functions which do not necessarily compose a basis.

5. Experimental Results

The different experimental conditions to obtain signals from different corrosion phenomena are described in Sec. 5.1. In Sec. 5.2 we show the performances for six different classification algorithms in distinguishing absence of corrosion, uniform

Table 1. Steels, corrosive medium and number of different experiments considered. Data was selected from (20).

Type of Corrosion	Material	Corrosive Medium + Conditions	Number of Experiments (Number of Time Series)	Total Number of Experiments per Class (Number of Time Series)
Absence of Corrosion	1.0038	NaOH 20 weight% + NaCl 3 weight% 80°C	1(99)	4(197)
	1.4541	$CaCl_2$ 40 weight% 85°C	3(98)	
Uniform Corrosion	1.0038	H_3PO_4 10 weight% $T_{environment}$	6(194)	6(194)
Pitting	1.4541	brackish water + $FeCl_3$ 1 weight% 45°C	9(214)	9(214)
Stress Corrosion Cracking	1.0038	$Ca(NO_3)_2$ 60 weight% 105°C	9(58)	10(205)
	1.4541	$CaCl_2$ 40 weight% 85°C	1(147)	

corrosion, pitting and stress corrosion cracking. Section 5.3 shows the performances for the distinction between three classes: absence of corrosion + uniform corrosion, pitting and stress corrosion cracking.

5.1. *Experimental conditions*

Two types of steel are considered that are regularly used as construction material[22]: carbon steel and stainless steel. The carbon steel considered here is: number 1.0038 (German Material Number), name S235JRG2 (DIN EN 10025) or RSt 37-2 (DIN 17100). The stainless steel considered here is: number 1.4541 (German Material Number), name X6CrNiTi18-10 (DIN EN 10088-2) and similar to AISI 321. In Table 1 all materials and experimental conditions are summarized.

The number of different experiments for the material-environment combination (the environment is the combination of a corrosive medium and a temperature) is shown in the fourth column. The total number of time series obtained from these experiments is indicated in brackets. The signals for each experiment were collected over several days of measuring. The acoustic emission data set contains 197 time series of "no corrosion", 194 time series of uniform corrosion, 214 time series of pitting and 205 time series of SCC. The time series have been assigned a corrosion class label by an expert based on the observation of the damage of the material and on the experimental conditions. Each time series consists of "N" $= 1024$ samples. This leads to $N * (\log_2 N + 1) = 11{,}264$ coefficients to be selected from.

5.2. *Results for four class problem*

In the validation of the different algorithms, we adopt a 10-fold cross-validation. This implies that in Fig. 4, 10 different training sets and 10 different testing sets are considered. We compute the test classification performance on the sets that have not been considered in the selection or the training of the classifiers. We let "m" range from 1 to 50 coefficients. The results for each classifier are summarized in Fig. 7.

Experiments were performed with six different classifiers:

- Support Vector Machine (SVM): The "libSVM"[2] C-support vector classifier is used with a radial basis function (RBF) kernel. The kernel parameter γ is set to

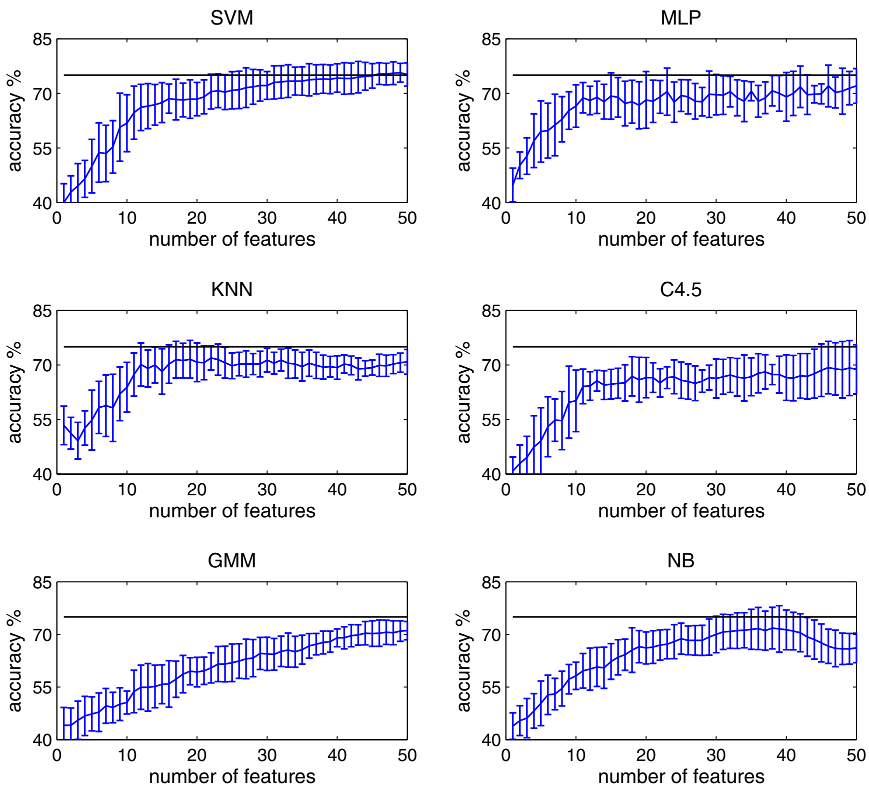

Fig. 7. Classification test performances for distinguishing absence of corrosion, uniform corrosion, pitting and stress corrosion cracking. The highest performance is achieved with the support vector machine classifier. The horizontal line in each figure indicates a 75% classification performance.

0.05 (default value). The cost factor "C" is set equal to 1 (default value). The classification threshold is set equal to 0 (default value). Pairwise coupling is used for multi-class classification,

- Multilayer Perceptron (MLP): A feed-forward neural network is used with 10 neurons in the hidden layer, a sigmoid activation function for the neurons, a weight-decay factor $\alpha = 0.2$, 10 cycles of the batch training mode and classification threshold equal to 0.5, see Chap. 6 in Ref. 5 for a description of MLP classification,
- k-Nearest Neighbor (KNN): The Euclidean distance is used and "k" is set to 3, see Sec. 4.5.4 in Ref. 5,
- Decision tree C4.5 (C4.5): The C4.5 decision tree from the WEKA package 3.4.1 was chosen, see Sec. 8.4.2 in Ref. 5,
- Gaussian Mixture Model (GMM): The number of Gaussians per class is taken equal to 1 in the experiments, see Ref. 11 for a reference on Gaussian mixture modeling,
- Nave Bayes classifier (NB), see Sec. 2.12 in Ref. 5

The highest test classification performance was obtained with the SVM classifier with an accuracy of 75.7% ± 2.6. We note that in the wavelet literature, especially feature extraction in combination with a feed-forward neural network for prediction is popular.[1,17] Shankar *et al.*[17] named the combination of wavelet feature extraction with a feed-forward neural network: a neuro-wavelet classifier. Wavelet based neural networks and neuro-fuzzy systems were used for time series prediction in Banakar *et al.*[1] However, we have shown here that we obtain a higher classification performance with an SVM classifier. Hence, one should compare test accuracies of different classification paradigms without being biased to the use of feed-forward neural networks.

5.3. *Results for three class problem*

In the second problem only three classes are considered: absence of corrosion + uniform corrosion, pitting and SCC. The results are shown in Fig. 8. Uniform corrosion is a less harmful type of corrosion than pitting and SCC, therefore emphasis in this problem is on an accurate detection of pitting, SCC and the less harmful class of absence of corrosion + uniform corrosion. A confusion between uniform corrosion and absence of corrosion is not punished. In this case the highest classification performance is obtained with the nave Bayes classifier with an accuracy of 97.5% ± 1.8. Hence, a very high accuracy is obtained in distinguishing harmful corrosion classes from the less harmful class absence of corrosion + uniform corrosion.

This result shows that the lower performance in Sec. 5.2 was due to a less successful distinction of absence of corrosion and uniform corrosion, both showing a continuous type of emission. In each of the ten training folds the selected basis functions may differ.

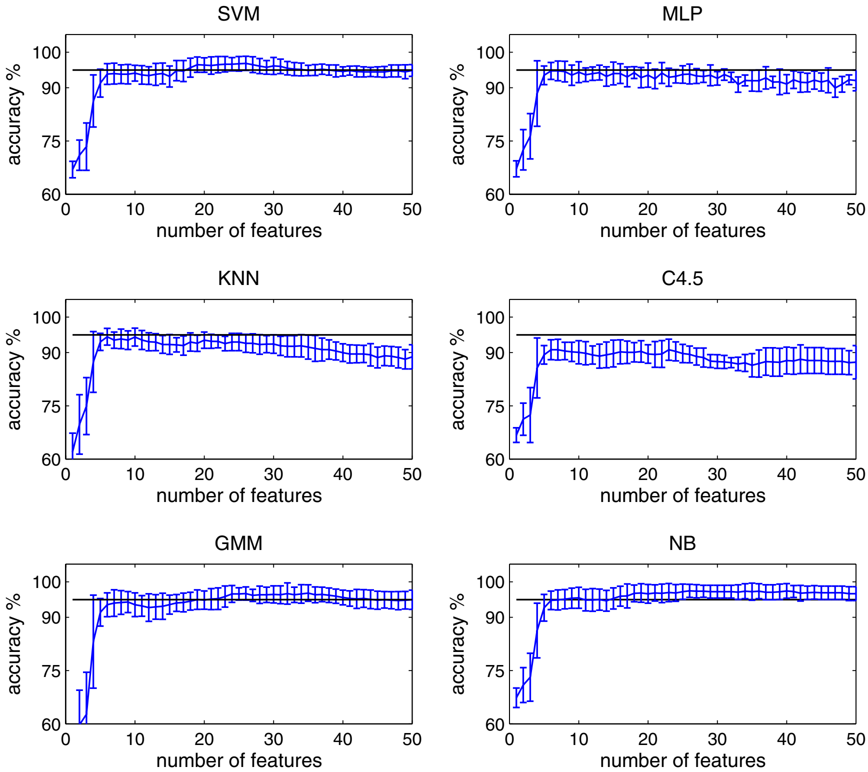

Fig. 8. Classification test performances for distinguishing absence of corrosion + uniform corrosion, pitting and stress corrosion cracking. The horizontal line in each figure indicates a 95% classification performance.

In Fig. 9, we show the wavelet coefficients of the fold, fold number 2, that provided the highest classification accuracy for three features. The highest accuracy for three features was obtained for the multilayer perceptron with an accuracy of 89.2%. For the other classifiers the performance was slightly lower, but without exception the second fold provided for each classifier the highest accuracy.

The coefficients can be related to their corresponding frequency intervals using[8]:

$$\left[g\frac{f_s}{2}2^{-i}, (g+1)\frac{f_s}{2}2^{-i}\right] \tag{5.1}$$

and the center frequency (f_c) as:

$$f_c = \left(g + \frac{1}{2}\right)\frac{f_s}{2}2^{-i}. \tag{5.2}$$

Here, "g" is the Gray order[9] of the wavelet packet and f_s the sampling rate. The first two coefficients $\gamma_{6,0,11}$ and $\gamma_{6,0,12}$, correspond to center frequencies of approximately 97.7 kHz. The third coefficient $\gamma_{5,1,7}$ corresponds with a center frequency

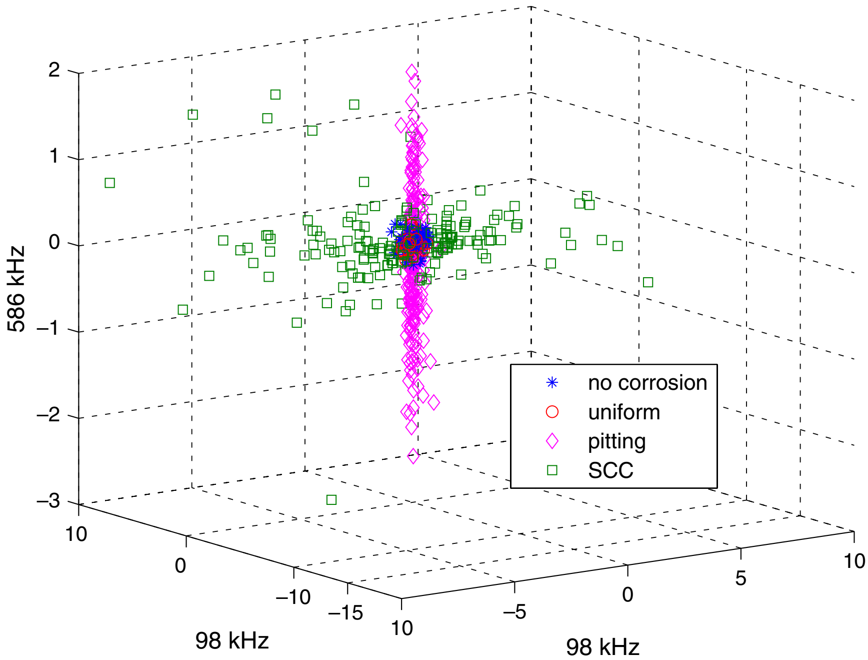

Fig. 9. 3D scatter plot of first three selected wavelet coefficients. SCC signals are indicated by a square "□", pitting signals by a diamond "◇", absence of corrosion by a star "∗" and uniform corrosion by a circle "○". Uniform corrosion and absence of corrosion are not visible due to the small values the coefficients take within frequency bands centered at these frequencies.

of approximately 586 kHz. However, it has to be noted that the coefficients $\gamma_{6,0,11}$ and $\gamma_{6,0,12}$ consider the frequency band of [0, 195.3] kHz around the 97.7 kHz center frequencies, while $\gamma_{5,1,7}$ considers the frequency band [390.6, 781.3] kHz around the 586 kHz center frequency.

From the scatter plot it is clear that pitting shows activity in the higher frequency range, while SCC shows activity in the lower frequency range around 98 kHz. The fact that pitting shows activity in higher frequency ranges could in fact also be observed in Fig. 3, which shows faster oscillations for pitting than for SCC. The observation that pitting is found at higher frequencies in comparison with SCC may be related to several factors. It may be partly due to a difference in frequency content that is excited by the sources of pitting AE signals and SCC AE signals. Another reason for the difference in frequency content may be due to a different distance travelled by the AE waves before they reach the sensor. SCC signals are all initiated at approximately the same position of the probe: the position where maximal stress has been applied to the probes.[22] This position is located approximately at 260 mm from the sensor in Fig. 4 at the bending point of the U-shape. The high frequency components are increasingly damped if a longer distance is traveled by the waves. It is mainly the primary Lamb wave and its lower order harmonics that survive longest.[12] These are typically within the range of 50 kHz to 300 kHz for the thickness of steel plate that are commonly encountered in practice.[12] The 98 kHz of SCC falls within this range.

Note that in the scatter plot the coefficients of uniform corrosion and absence of corrosion are hardly visible. For these classes the activity is rather limited, this can also be seen from the lower amplitude of these signals in Fig. 2 compared to the burst activities in Fig. 3. In most of the other training folds the first three selected wavelet coefficients were $\gamma_{6,0,11}$, $\gamma_{6,0,12}$ and $\gamma_{6,0,10}$. All these coefficients are related to the [0, 195.3] kHz interval with center frequency 98 kHz. As can be seen in Fig. 9, this frequency interval is successful in identifying SCC from the other classes, but is less successful in distinguishing pitting from the other classes. This somewhat disadvantageous ordering of wavelet coefficients in most of the folds is due a to a drawback of Step 4 in the local discriminant basis algorithm. This step orders the coefficients according to their individual discriminant power, it does not take into account the information that is captured by previously selected wavelet coefficients.

6. Conclusion

The wavelet packet decomposition was used to extract features from corrosion acoustic emission signals. The local discriminant basis algorithm was used to search for an optimal basis and to select the most discriminative wavelet coefficients. In a theoretical contribution it was proven that the full symmetric relative entropy criterion reduces to a sum of marginal entropies under the condition of class conditional independence of the wavelet coefficients. Experimentally, it was shown that absence of corrosion, uniform corrosion, pitting and stress corrosion cracking can be distinguished with an accuracy of 75.7% ± 2.6 using a support vector machine classifier. Distinction between the less harmful class of absence of corrosion + uniform corrosion and the harmful classes pitting and stress corrosion classes could be achieved with 97.5% ± 1.8 accuracy using a nave Bayes classifier.

Acknowledgments

Part of this work was performed when the first author was employed at the Departement Metaalkunde en Toegepaste Materiaalkunde, Materiaalgedrag en Niet-destructieve Evaluatie, Katholieke Universiteit Leuven, Kasteelpark Arenberg 44, B-3001 Heverlee, Belgium.

The authors are grateful to Prof. N. Saito, University of California, Davis, for providing the local discriminant basis selection algorithm. We are grateful to Dr. M. Winkelmans for providing data of corrosion experiments.

GVD is supported by the CREA Financing (CREA/07/027) program of the K. U. Leuven. MMVH is supported by research grants received from the Excellence Financing (EF 2005) program of the K. U. Leuven, the Belgian Fund for Scientific Research — Flanders (G.0234.04, G.0588.09), the Interuniversity Attraction

Poles Programme — Belgian Science Policy (IUAP P5/04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), and the European Commission (STREP-2002-016276, IST-2004-027017, and IST-2007-217077).

References

1. A. Banakar, M. F. Azeem and V. Kumar, Comparative study of wavelet based neural network and neuro-fuzzy systems, *Int. J. Wavelets Multiresolut. Inf. Process.* **5** (2007) 879–906.
2. C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines (2001), http://www.csie.ntu.edu.tw/∼cjlin/libsvm.
3. H. Cho and M. Takemoto, Acoustic emission from rust in stress corrosion cracking, in *Proc. of the 26th European Conference on Acoustic Emission Testing*, Berlin, Germany (2004), pp. 605–615.
4. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edn. (John Wiley & Sons, 2006).
5. R. O. Duda, P. E. Hart and D. G. Stork, *Classification*, 2nd edn. (John Wiley & Sons, 2001).
6. L. Jaubert, Étude de la corrosion uniforme d'aciers non alliés et inoxydables: Utilisation conjointe de l'émission acoustique et des techniques électrochimiques, PhD Thesis, Institut National des Sciences Appliquées de Lyon, France (2004).
7. X. Li and R. Du, Monitoring machining processes based on discrete wavelet transform and statistical process control, *Int. J. Wavelets Multiresolut. Inf. Process.* **2** (2004) 299–311.
8. S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989) 674–693.
9. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, 1998).
10. H. Mazille, R. Rothéa and C. Tronel, An acoustic emission technique for monitoring pitting corrosion of austenitic stainless steels, *Corros. Sci.* **37** (1995) 1365–1375.
11. G. McLachlan and D. Peel, *Finite Mixture Models* (John Wiley & Sons, 2000).
12. L. M. Rogers, Crack detection using acoustic emission methods — fundamentals and applications, *Key Eng. Mater.* **293–294** (2005) 33–46.
13. N. Saito and R. R. Coifman, Local discriminant bases and their applications, *J. Math. Imaging Vis.* **5** (1995) 337–358.
14. N. Saito, R. R. Coifman, F. B. Geshwind and F. Warner, Discriminant feature extraction using empirical probability density estimation and a local basis library, *Pattern Recogn.* **35** (2002) 2841–2852.
15. K. H. W. Seah, K. B. Lim, C. H. Chew and S. H. Teoh, The correlation of acoustic emission with the rate of corrosion, *Corros. Sci.* **34** (1993) 1707–1713.
16. H. Shaikh, R. Amirthalingam, T. Anita, N. Sivaibharasai, T. Jaykumar, P. Manohar and H. S. Khatak, Evaluation of stress corrosion cracking phenomenon in an AISI type 316LN stainless steel using acoustic emission technique, *Corros. Sci.* **49** (2007) 740–765.
17. B. U. Shankar, S. K. Meher and A. Ghosh, Neuro-wavelet classifier for multispectral remote sensing images, *Int. J. Wavelets Multiresolut. Inf. Process.* **5** (2007) 589–611.
18. K. Torkkola, Feature extraction by non-parametric mutual information maximization, *J. Mach. Learn. Res.* **3** (2003) 1415–1438.
19. P. Tscheliesnig, Corrosion testing of ship building materials with acoustic emission, in *Proc. of the 26th European Conference on Acoustic Emission Testing*, Berlin, Germany (2004), pp. 29–40.

20. G. Van Dijck, Information theoretic approach to feature selection and redundancy assessment, PhD thesis, Katholieke Universiteit Leuven, Belgium (2008).
21. M. Wevers, Listening to the sound of materials: acoustic emission for the analysis of material behavior, *NDT and E. Int.* **30** (1997) 99–106.
22. M. Winkelmans, Fusie van niet-destructieve onderzoekstechnieken voor corrosiemonitoring in chemische procesinstallaties, PhD Thesis, Katholieke Universiteit Leuven, Belgium (2004).
23. S. Yuyama, Fundamental aspects of acoustic emission applications to the problems caused by corrosion, in *Corrosion Monitoring in Industrial Plants Using Non-Destructive Testing and Electrochemical Methods*, eds. G. C. Moran and P. Labine (American Society for Testing and Materials, 1986), pp. 43–74.