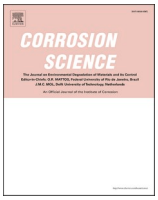
Corrosion Science 214 (2023) 110964

Contents lists available at ScienceDirect

Corrosion Science

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



Acoustic emission as a reliable technique for filiform corrosion monitoring on coated AA7075-T6: Tailored data processing

C. Abarkane [a], A.M. Florez-Tapia [b], J. Odriozola [b], A. Artetxe [b], M. Lekka [a], E. García-Lecina [a], H.-J. Grande [a,c], J.M. Vega [a,d,*]

[a] CIDETEC, Basque Research and Technology Alliance (BRTA), Paseo Miramón 196, 20014 Donostia-San Sebastián, Spain
[b] VICOMTECH, Basque Research and Technology Alliance (BRTA), Mikeletegi Pasealekua, 57, 20009 Donostia-San Sebastián, Spain
[c] University of the Basque Country (UPV/EHU), Advanced Polymers and Materials: Physics, Chemistry and Technology Department, Avda. Tolosa 72, 20018 Donostia-San Sebastian, Spain
[d] Departamento de Ingeniería Química y de Materiales, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040, Spain

ARTICLE INFO

Keywords:
Aluminium
Interfaces
Paint coatings

ABSTRACT

Acoustic emission (AE) was used for in-situ filiform corrosion (FFC) monitoring on coated AA77075-T6. The analysis of AE data using DBSCAN as clustering algorithm (validated by Bhattacharyya Coefficients' evaluation) has revealed the presence of three clusters (out of four) related to phenomena involved in the FFC mechanism: metal-coating interface delamination due to opening (tensile), sliding (shear) and mixed mode enclosing both previous ones. The peak frequency was found to be the most relevant descriptor for clustering by using Random Forest classifier, and the correlation with the dominant frequencies range was validated obtaining the Power Spectrum Density of the AE signals.

1. Introduction

Nowadays, aluminium alloys are still having a prominent role as a lightweight material for aircraft's design, although its protection by coatings is a must. The maintenance, repairs, and operations (MRO) costs turns to be a decisive matter, triggering the development of new strategies to reduce their corresponding budget (the corrosion costs of United States Department of Defense were $10.18 billion in financial year 2018 compared to $8.97 billion in 2017 [1,2]). In this sense, structural health monitoring (SHM) systems based on non-destructive techniques have proven to enable the in-service detection of damage, allowing to rationalize the maintenance efforts, i.e., reducing the level and timing of inspection. Acoustic emission technique (AE) is a powerful tool for real-time monitoring on full-scale elements (e.g., damage on real structures with different materials) [3–7]. Concerning to corrosion damage on structures, studies have been mainly focused on reinforced concrete [6,8,9], stress corrosion cracking and corrosion fatigue [10,11].

Coated metals have been scarcely explored, where filiform corrosion (FFC) typically occurs undermining the substrate-coating interface which could evolve to a more severe corrosion (e.g., nucleation site for fatigue corrosion or further flaking of the coating). Usually, in-situ scanning Kelvin probe (SKP) experiments are able to develop FFC under atmospheric conditions [12,13]. The FFC mechanism was shown to depend on the relative susceptibility of alloy phases to undergo anodic dissolution [12]. In order to minimise FFC, corrosion inhibitors are employed in coatings formulation to minimise corrosion-driven organic coating disbondment [13]. This corrosion phenomenon was already explored by Ruggeri and Beck using AE [14]. It was observed that FCC events were mainly linked to hydrostatic pressure, mechanical prying, anodic undermining, or more precisely, as a combination of them. Further investigations highlight that the formed corrosion products are triggering the failure of coated steel [15] or coated aluminium interfaces [16]: e.g., the mechanical lifting (i.e., disbondment or delamination) of the coating. Therefore, although AE is a suitable technique to monitor FCC, the signature of the AE signals is not defined yet to implement a reliable SHM strategy for in-service structures.

One of the main limitations of the AE data is the classification of the AE signal with the aim to find out the AE signature [17]. AE data depends on the properties of the material, the sensor characteristics, the

* Corresponding author at: Departamento de Ingeniería Química y de Materiales, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040, Spain
E-mail address: jevega@ucm.es (J.M. Vega).

https://doi.org/10.1016/j.corsci.2023.110964
Received 20 October 2022; Received in revised form 3 December 2022; Accepted 7 January 2023
Available online 9 January 2023
0010-938X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

noise and the nucleation/evolution of the damage event (e.g., cracks) [18]. Moreover, the acoustic sources are usually affected by one another, and thus, distinguishing the different stages of damage using a traditional AE analysis method is not straightforward. Therefore, the classification of a collection of recorded AE signals during corrosion must be addressed using different approaches. One of the most popular one is based on clustering algorithms [19–21]. Cluster analysis includes three main steps: i) to extract the AE characteristic parameter, ii) to select the clustering algorithm, and iii) the validation of the defined clusters [19,20].

This methodology has been widely used for the classification of AE signals: from supervised classification methodologies [11,22–24] to unsupervised ones. Different algorithms such as Artificial Neural Network (ANN) [25] and especially K-Means [25–28] have been used to identify the different mechanisms of damage such as pitting [22,24,26] or stress corrosion cracking [11,23,25,27,28]. In addition to this K-means algorithm, lately the DBSCAN algorithm is also being widely used in this field [29,30]. K-means is an unsupervised clustering algorithm that groups objects into *k* groups based on their characteristics. Clustering is performed by minimising the sum of distances (usually quadratic) between each object and the centroid of its cluster. Although the algorithm is suitable for datasets with globular and similar-size clusters, good results have been achieved for clustering pitting corrosion signals [31], analysing the corrosion on pure magnesium [32], classifying acoustic emission signals of stainless steel during stress corrosion [27] and characterizing the stress corrosion cracking on stainless steel [33]. On the other hand, DBSCAN is a density-based clustering method that groups the points which are closer, based on the distance (usually Euclidean). Besides, points that are in low density regions are identified as outliers. The DBSCAN algorithm is particularly effective for datasets of multiple-size clusters or data overlapped by noise, providing solid results in relation to concrete structures monitoring [7], rail steel health analysis [29] and investigation of defects in roll contacts [30]. Therefore, in this paper, given that the data structure of AE signals associated to FFC has not been fully explored yet, both K-means [34] and DBSCAN [35,36] can be explored.

In this study, FFC was triggered on coated AA7075-T6 and in-situ monitored by AE sensors. Features such as the potential of the metal/coating interface and the topography of the filament were also obtained by SKP, together to the morphology of the attack by confocal and optical microscopes. Regarding to the AE results, descriptors and signals were analysed by clustering algorithms (i.e., K-Means and DBSCAN) and time-domain and frequency-domain analysis, respectively. Prior to use the algorithms, Principal Component Analysis (PCA) were used to reduce the dimensionality of the AE descriptors of interest.

2. Material and methods

Experiments were carried out by using plates of AA7075-T6 aluminium alloys (their chemical composition is depicted Table 1). Specimens were subjected to alkaline degreasing using Bonderite® C-AK 18 for 3 min, followed by acid pickling by immersion in nitric acid (30% v/v) for 2 min before painting. MAPAERO P60 was used as a primer and MAPAERO F70 as a topcoat. Both are epoxy-based coatings, with the primer layer containing strontium chromate as a pigment acting as corrosion inhibitor. The final layer of primer and topcoat coatings have reached 20 μm and 25 μm thickness, respectively.

2.1. Experimental details

Two coated plates of 100 × 75 mm^2 and 1 mm thick were used as specimens to promote FFC. An artificial defect (cross scratch of 35 mm length) was created manually by using a scalpel blade (Fig. 1(a)) [37]. The corrosion process was activated in the bare aluminium after exposure to vapour of HCl (2% v/v) for 1 h (2% v/v), followed by 24 h of exposure to a Copper-Accelerated Acetic Acid-Salty Spray (CASS) chamber [38]. Finally, the samples were located into the Scanning Kelvin Probe (SKP) chamber during 850 h and 670 h, for S1 and S2 specimens, respectively.

2.1.1. Scanning Kelvin Probe

Once the specimen was activated (pre-corroded) to trigger FFC, the AE sensor was coupled to it before placing into the SKP. The SKP, supplied by Wicinski-Wicinski GbR, was used to measure the Volta potential of the coating-metal interface and the topography at the surface. The relative humidity in the chamber was kept at high humidity (85%) and a constant temperature of 23 ºC. Prior each experiment, the SKP needle (Ni-Cr one) was calibrated with a saturated $Cu/CuSO_4$ (320 mV Vs. Standard Hydrogen Electrode (SHE)). Besides, the SKP was equipped with a long working distance camera Dino-Lite (Digital Microscope) to optically monitor the filaments.

Line-scan measurements (from 22 to 30 mm length) were performed in such a way that FFC's filaments were crossed by the needle of the SKP. The measurement was set at 100 points/mm (step size of 10 μm) and the distance needle-surface was kept constant during the experiment. Several filaments were chosen to perform the line-scans at different distance with time, starting from distance called 0 (i.e. D0–0, at which the filament was already formed and evolving) to distance n (e.g. D0–6) as it is shown in Fig. 1(b). Fig. 1(c) shows the whole set-up, consisting of a specimen placed on the plate inside the environmental chamber of the SKP probe: the needle and the acoustic sensor are monitoring the potential/topography and the acoustic emission signals, respectively.

2.1.2. Acoustic emission (AE)

The acoustic emission (AE) is a non-destructive technique for material diagnosis based on the capture of the elastic waves that come from an acoustic source when a physical change occurs in the field of the deformations (e.g. delamination). The accumulated energy due to such deformation is released, partially, in the form of elastic energy [39], which spreads in all directions as a volume wave (P and S modes). The acoustic wave, when reaching the surface (interface), becomes a surface wave (Rayleigh) that, in turn, is converted into an AE signal by the piezoelectric sensor.

In this case, the AE generated during the progress of FFC filaments was captured by a wide-band piezoelectric sensor, a PKBBI model, which integrates a preamplifier of 26 dB_{AE}, and whose acoustic response (sensitivity) is ranged between 20 and 600 kHz (Fig. 2), thus covering the domain of the corrosion phenomena, or to be more precise, the corrosion-related mechanisms as: mechanical delamination (e.g., FFC), crack initiation and growth (e.g., stress corrosion cracking, SCC), Hydrogen/Oxygen evolution as bubbles (e.g., pitting), etc. The sensor was clamped (tightening pressure < 4 Pa) and coupled to the specimen's surface with silicone grease. Prior each test, the coupling of the AE sensor was verified by means a Hsu-Nielsen test, according to common AE standards.

The AE signals were processed in real time with an acquisition system, Micro-SHM, and the data was visualized with the software AEWin. Both, sensor and acquisition system, are from the brand Physical

Table 1
Chemical composition of 7075 aluminium alloy.

	Si [%]	Fe [%]	Cu [%]	Mn [%]	Mg [%]	Cr [%]	Zn [%]	Ti [%]	Others [%]	Al
Al7075	0.069 ± 0.005	0.10 ± 0.02	1.33 ± 0.04	0.032 ± 0.002	2.52 ± 0.07	0.18 ± 0.02	5.77 ± 0.13	0.026 ± 0.003	< 0.15	Balance

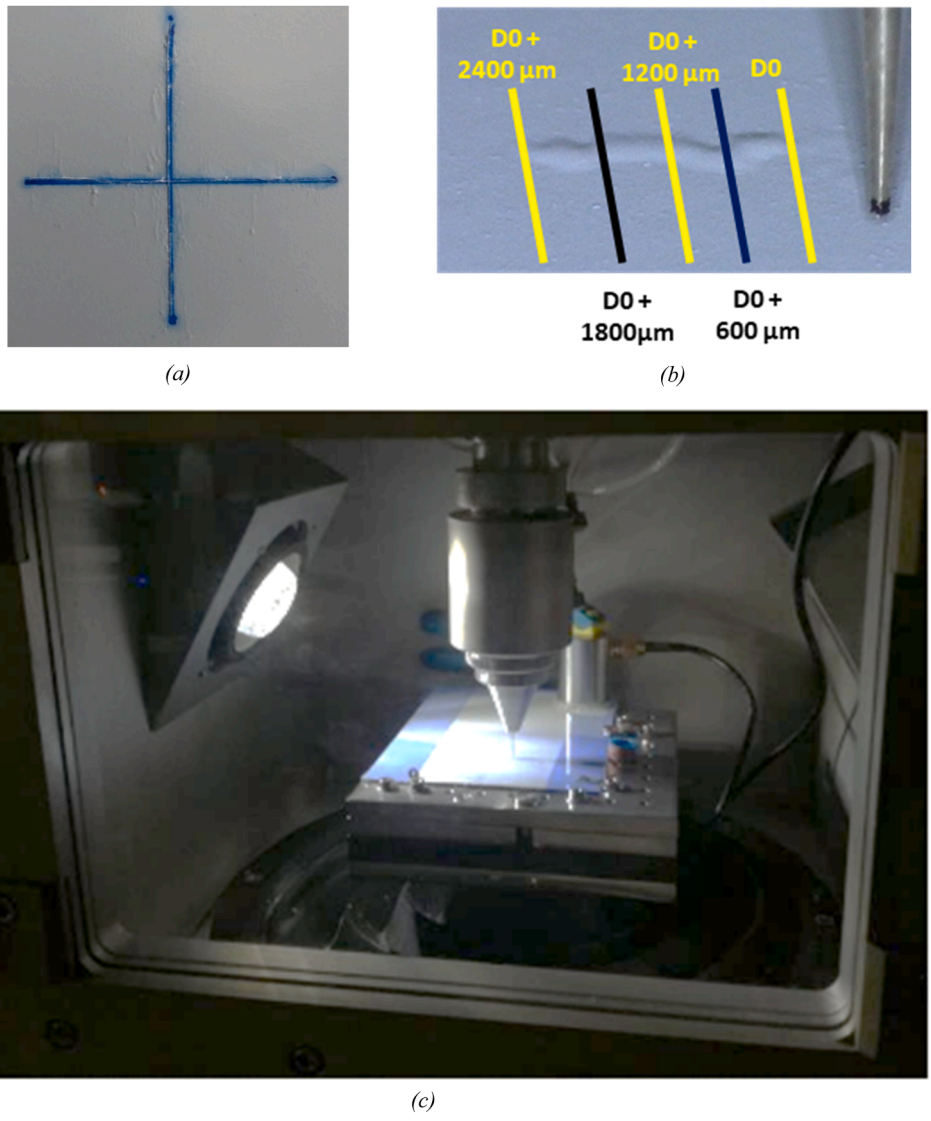


Fig. 1. : (a) Specimen S2 after exposure to the SKP, having a cross scratch as an artificial defect, (b) scheme of the scanning lines at the surface of the specimen (S1) following the progress of the filament with time by SKP, and (c) experimental set-up during environmental exposure inside the Scanning Kelvin Probe and AE monitoring.

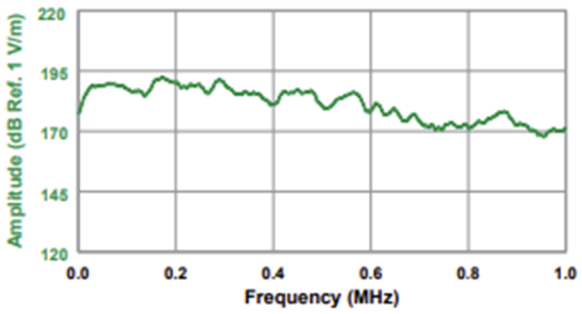


Fig. 2. : Sensitivity curve of wide-band sensor PKBBI.

Acoustic Corporation (PAC) provided by Mistras. The data acquisition system was set with a threshold of 25 dB_{AE} based on background noise measurements, while the sample rate was established at 2 MHz, and the waveforms were recorded with 4k of length. Regarding the AE timing parameters, the values were set as PDT = 700 µm, HDT = 1500 µm, HLT = 1500 µm and Max. Duration = 100 ms, following standard acquisition criteria for metallic material.

3. Calculation

Clustering techniques have been used to analyse the AE results. Taking into account that AE data is not labelled, and no defined dependant variable is considered, unsupervised classification methods are suitable to be chosen as clustering technique.

3.1. Preliminary analysis of the AE descriptors

AE analysis is conventionally focused on the descriptors of the AE signal, traditionally called AE parameters, which are extracted from the waveform of the signal (temporal domain) as well as from the frequency spectrum (frequency domain). AEwin software has provided up to 20 descriptors but some of them were deemed negligible for the analysis (e.g. Channel or Parametric inputs (not used)). Moreover, the dendrogram of correlation has shown that Signal Strength parameter was redundant, giving the same information as the Energy (correlation > 0.99), so it was discarded. Thus, a sorted list of the 16 relevant descriptors have been included in Table 2.

Table 2
AE descriptors recorded during the experiments.

#	Descriptor name	#	Descriptor name
1	Risetime [μs]	9	Initiation Frequency [kHz]
2	Counts-to-Peak	10	Absolute Energy [aJ]
3	Counts	11	Partial Power 1 [%]
4	Energy [J]	12	Partial Power 2 [%]
5	Duration [μs]	13	Partial Power 3 [%]
6	Amplitude [dB]	14	Partial Power 4 [%]
7	Average Frequency [kHz]	15	Frequency Centroid [kHz]
8	Reverberation Frequency [kHz]	16	Peak Frequency [kHz]

During the next step, Principal Component Analysis (PCA) has been used as pre-processing procedure before clustering. PCA was performed on AE descriptors (Table 2) in order to define/extract the uncorrelated features [40]. PCA constructs a set of uncorrelated directions that are ordered by their variance and is based on the following basic assumption: directions of large variance are the result of structure in those directions [41].

Prior to apply PCA, the value of each descriptor was normalized to N (0,1). The selected principal components where the j most representative eigenvectors. In this work, j was defined as the minimum number of eigenvalues that corresponded together to more than 95% of the standard deviation of the data set, $j = 6$. Table 3 shows the features that have been selected to apply the clustering method. The remaining principal components could be neglected. Finally, the data was expressed in the j-principal components base.

Once the relevant features were selected, the two clustering methods were evaluated: the K-Means and the DBSCAN algorithms.

3.2. Clustering methods

Despite clustering is considered one of the most important methods for unsupervised learning, an agreement has not been reached yet to define it [42]. A classic definition enclose the following points [43]:

1. Instances, in the same cluster, must be similar as much as possible.
2. Instances, in the different clusters, must be different as much as possible.
3. Measurement for similarity and dissimilarity must be clear and have practical meaning.

A variety of clustering algorithms can be separated based on the criteria used to create the groups [42], where the most common are the partition-based clustering, the distribution-based clustering, the density-based clustering and the centroid-based clustering. In this work, the analysis is focused over the partition and density-based clustering, the K-Means and the DBSCAN algorithms, respectively. The former one is based on the centre of data points as the centre of the corresponding cluster, whereas the latter one is based on considering regions with high density of the data space to belong to the same cluster.

3.2.1. K-means algorithm

One of the most popular and user-friendly unsupervised machine learning algorithms is K-Means clustering [44]. A target number k is defined, which refers to the number of centroids that are in the dataset, being a centroid, the imaginary or real location representing the centre of a cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.

Table 3
Descriptors associated to the first 6 PCs.

#	Descriptor name	#	Descriptor name
1	Risetime [μs]	4	Absolute Energy [aJ]
2	Duration [μs]	5	Frequency Centroid [kHz]
3	Amplitude [dB]	6	Peak Frequency [kHz]

The approach k-Means follows is to solve the problem called *Expectation-Maximization*. The E-step is assigning the data points to the closest cluster. The M-step is computing the centroid of each cluster. The objective function is:

$$J(x) = \sum_{k=1}^{K}\sum_{i=1}^{n} w_{ik}\left\|x_i^{(k)} - \mu_k\right\|^2 \tag{1}$$

Where $w_{ik} = 1$ for data point x_i if it belongs to cluster k; otherwise, $w_{ik} = 0$. Also, μ_k is the centroid of $x_i^{(k)}$ cluster.

It is a minimization problem of two parts. The E-step consists of assigning the data point x_i to the closest cluster judged by its sum of squared distance from cluster's centroid:

$$\frac{\partial J}{\partial w_{ik}} = \sum_{k=1}^{K}\sum_{i=1}^{n}\left\|x_i^{(k)} - \mu_k\right\|^2 \Rightarrow w_{ik} = \left\{\begin{array}{c} 1 \text{if} k = argmin_j\left\|x_i^{(k)} - \mu_k\right\|^2 \\ 0 \text{otherwise.} \end{array}\right\} \tag{2}$$

The M-step is recomputing the centroid to each cluster to reflect the new assignment:

$$\frac{\partial J}{\partial \mu_{ik}} = 2\sum_{i=1}^{n} w_{ik}\left(x_i^{(k)} - \mu_k\right) = 0 \Rightarrow \mu_k = \frac{\sum_{i=1}^{m} w_{ik}x_i^{(k)}}{\sum_{i=1}^{m} w_{ik}} \tag{3}$$

K-Means algorithm is good in capturing structure of the data if clusters have a spherical-like shape. However, it does not provide optimum results when clusters have a complicated geometric shape.

3.2.2. Density-based spatial clustering of applications with noise (DBSCAN)

As it was mentioned above, one limitation of clustering algorithms based on spherical-shaped clusters is the presence of noise and outliers. The DBSCAN, a density-based clustering algorithm, was the first one used to address this drawback. It was designed to cluster data of arbitrary shapes in the presence of noise in special and non-spatial high dimensional databases [35]. According to DBSCAN, the neighbourhood of a given radius (ε), for each object into a cluster must contain at least a minimum number of objects (MinPts), which means that the cardinality of the neighbourhood has to exceed certain threshold.

- ε defines the size and borders of each neighbourhood. $\varepsilon(0)$ is a radius, if ε is too small, then large part of the data will be considered as outliers; whereas if it is too large, the clusters will merge, and the majority of the data points will be in the same clusters. The ε-neighbourhood of x is given by [47]:

$$N_\varepsilon(x) = B_d(x,\varepsilon) = \{y \vee \delta(x,y) \leq \varepsilon\} \tag{4}$$

  where δ(x,y) is the distance between two points x, and y, the ε-neighbourhood of x, N_ε (x).
- MinPts: the density threshold. A point will be considered dense if there are at least the value of MinPts points in its ε-neighbourhood: these are the *core points*. A *border point* has ε-neighbourhood that contains < MinPts points, but it belongs to the ε-neighbourhood of another core point. If a point is not a *core point* or a *border one*, it is a *noise point*, or an outlier. A heuristic way to choose MinPts value:

$$MinPts = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{5}$$

where P_i is the number of points in the ε-neighbourhood of point i, and n is the number of points in the dataset [45].

3.2.3. Evaluation methods

In contrast to supervised learning, where there is some ground truth available to evaluate the model's performance, clustering analysis does

not have a solid evaluation metric. Moreover, since *k*-Means requires *k* as an input and does not learn it from data, the number of clusters is undefined for a specific problem. However, there are some metrics that may give some insights [46] such as the Elbow method, the Silhouette analysis and the Bhattacharyya Coefficient.

3.2.3.1. Elbow method. The Elbow method is able to distinguish the hypothetical number of clusters for a certain dataset. Initially, $k = 2$ is specified as the optimal cluster number *k*, and then it keeps increasing *k* in order to distinguish the potential optimal cluster (number *k* corresponding to the plateau) [47]. The goodness of the *k* number is based on the sum of squared distance (SSE) between data points and their assigned clusters' centroids. The optimal *k* corresponds to the point where SSE starts to flatten out and forming an elbow.

3.2.3.2. Silhouette analysis. There are several methods to evaluate clustering results, such as the Rand index [48], adjusted Rand Index [49], the distortion score [50] and the Silhouette index. However, the most appropriated method for clustering among them is the Silhouette index because it does not need a training set to evaluate the clustering results [51]. Thus, it can be used to determine the degree of separation between clusters. For each sample:

- Compute the average distance from all data points in the same cluster (a_i).
- Compute the average distance from all data points in the closest cluster (b_i).

$$s = \frac{b_i - a_i}{\max(a_i, b_i)} \in [-1, 1] \Longrightarrow \left\{ \begin{array}{l} 0 : \text{sample very close to neighboring clusters} \\ 1 : \text{sample far from the neighboring clusters} \\ -1 : \text{sample is assigned to the wrong cluster} \end{array} \right\} \quad (6)$$

Therefore, an optimal clustering can be obtained for large Silhouette score, *s* values, ideally close to 1.

3.2.3.3. Bhattacharyya coefficient. A different approach to study the clustering results is to determine the degree of dissimilarity between their distributions, where several indices have been suggested in literature [52–56]. Despite being introduced for different purpose, in general, such indices are increasing when the two distributions involved move apart. Therefore, an index with this property is measuring divergence of one distribution from another. A general method based on that feature (i.e., generating measures of divergence) has been already discussed in literature [57].

That is the case of the Bhattacharyya Distance which can be used for *p*-variate normal populations (i.e., case under study in this paper). If the variance-covariance matrices $\Sigma_0 = \Sigma_i = \Sigma$ are not equal (having also different means μ_0 and μ_1), the Bhattacharyya Distance between two probability density functions for multi-normal variables can be expressed as:

$$d_i = \frac{1}{2}\ln\frac{|V_i|}{\sqrt{|\Sigma_0\Sigma_i|}} + \frac{1}{8}(\mu_0 - \mu_i)^T V_i^{-1}(\mu_0 - \mu_i) \quad (7)$$

Where $V_i = (\Sigma_0 + \Sigma_i)/2$ [58].

The Bhattacharyya Coefficient can determine the overlapping between two statistical samples (an approximate measurement). Such coefficient can determine the relative closeness of two samples of interest, and it is closely related to the Bhattacharyya Distance, Eq. (8):

$$BC = e^{-d_i} \quad (8)$$

The Bhattacharyya coefficient becomes 0 without overlapping whilst higher values are expected with the overlap of the two sample's members within it. It is considered that two distributions are significantly equal if $BC \geq 0.95$, and significantly different if $BC \leq 0.05$.

4. Results and discussion

4.1. FFC progress

Two types of filaments were monitored in detail using the SKP, one far from the artificial defect (i.e., scratch) of the specimen S1 (Fig. 3(a)) and the other close to the scratch in specimen S2 (Fig. 3(b)). Fig. 4 is showing the evolution of the filament for 3 days, where it can be seen the different distance to the scratch. The filament from specimen S1 is already more than 90 mm far from the scratch whilst the filament from specimen S2 is around 40 mm away.

In order to obtain complementary information to the optical one, the potential and the topography of the coatings were measured by SKP. Fig. 5 is showing the value of both in an already formed filament (line D0) which can be easily detected by a change in the topography centred at 2150 μm (around 30 μm height). In contrast, the potential difference within the filament (−0.66 V_{SHE}) is quite similar to the one of the surrounding intact coatings (from −0.64 to −0.66 V_{SHE}). It seems that such potential values are common in other intact areas of the coating (i.e., absence of filaments) measured at different distances from the initial filament (from 600 μm up to 2400 μm). It can be explained due to chromate-bearing pigments does not provide any depression of the potential on the intact coated aluminium alloys [59,60]. Indeed, the potential of the unexposed coating (prior to CASS chamber) is already showing a potential value around − 0.75/− 0.8 V_{SHE}. Therefore, although the depression of the potential is commonly used to follow the progress of the anodic head throughout the intact coating [61], the topography has been used here instead.

Fig. 6 is showing the evolution of the potential and height for both filaments of interest (Fig. 4). Independently of the filament, the potential is showing a value below − 0.60 V in all cases. The line scan for S1 (Fig. 6(a)) at 0 h, prior to an occurrence of the head, is showing a constant potential of − 0.64 V and a flat topography of the surface. Once the exposure time is increased to 14 h, the head has already occurred, having a height around 8 μm. If the time of exposure is increased (up to 70 h), the height of the filament keeps increasing. However, the potential does not show any significant change and a similar profiles with time can be observed.

This behaviour can also be found in Fig. 6(b) for specimen S2. In this case, the line scan at 0 h is already showing the presence of the head centred at 1050 μm. Regarding to the potential, there is a plateau around − 0.64 V_{SHE} within filament (950–1150 μm) which remains from 1150 to 1950 μm. On the other hand, slightly higher potential values (−0.60/−0.63 V_{SHE}) are obtained from 0 to 700 μm. The most important feature seems to be the potential depression (around 25 mV) that occurs in the boundary filament/intact coating (centred at 750 μm). Indeed, if the time of exposure is increased (from 14 up to 70 h), such potential depression (20/25 mV range) seems to happen at both boundaries filament/intact coating. It could be explained by the presence of local anodes at the edge, where fresh surface can be attacked when the filament is growing laterally due to delamination (i.e. larger width). Further research is still needed to provide a conclusive statement about this.

At the end of the exposure to the SKP chamber, confocal images were taken to observe the overall corrosion underwent by the coating. The FFC filaments were mainly developed to on the rolling direction of the aluminium alloy (i.e. vertical position of the images). Apparently, not only filaments are observed but also different damage as blisters (Fig. 3). Therefore, such phenomena together FFC have to be considered during the discussion of the AE results as potential sources of acoustic signals.

As a summary, although large potential differences were not observed during FFC between zone such as the head/tail/intact coating for this metal/coating configuration, the topography measurements by SKP were used to follow the evolution of the filaments with time.

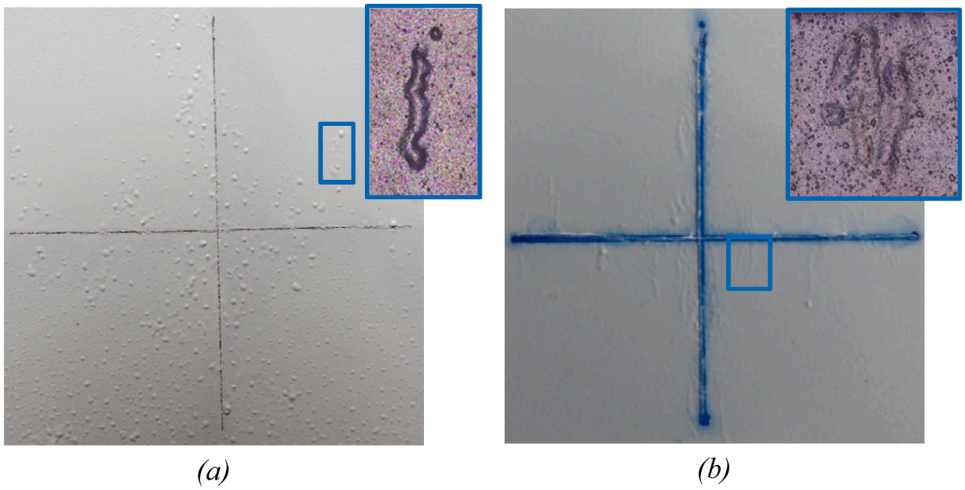


Fig. 3. : Photograph and confocal image (zoom of the delimited area) at the end of the SKP test of (a) specimen 1, S1, and (b) specimen 2, S2, for AA7075 with P60 + F70 coating system.

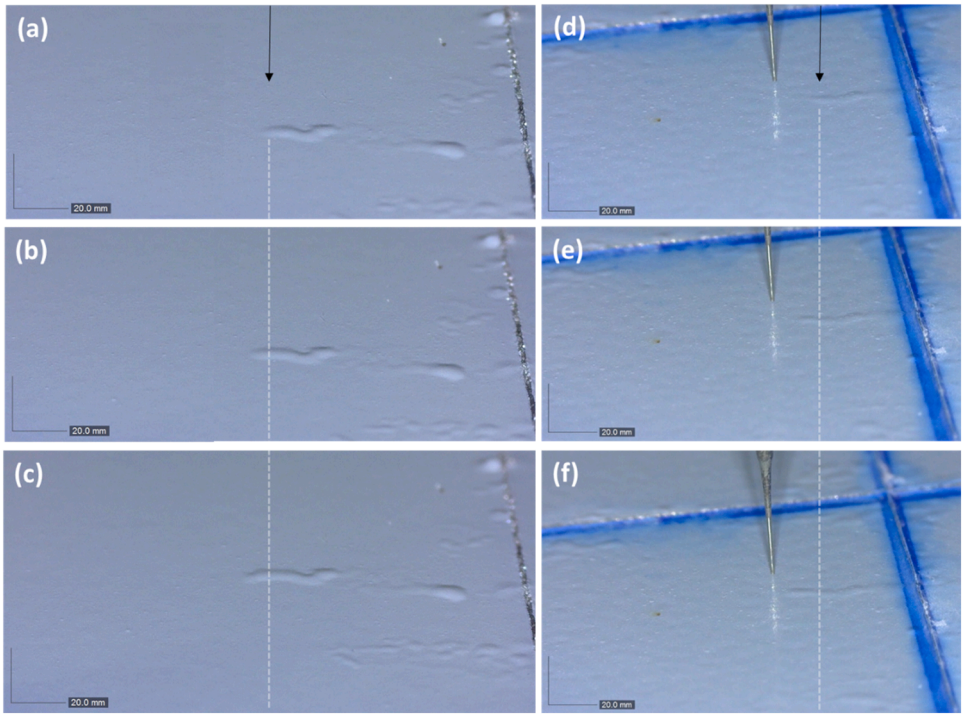


Fig. 4. : Movement of two different FFC filament from specimen S1 ((a), (b) and (c)) and specimen S2 ((d), (e) and (f)) at different time of monitoring by SKP: (a) & (d) beginning of the measurement, (b) & (e) 1 day, (c) & (f) 3 days.

4.2. AE data analysis

Different clustering algorithms were used and compared during the analysis of the AE data. The aim was to assign the different clusters to the proper physicochemical event which is taking place during the filiform corrosion test.

4.2.1. Clustering approach

In order to obtain the optimum k value (to be used for the K-Means algorithm in each experiment), the Elbow method was used (Fig. 7(a)), and the Silhouette Score was calculated (Fig. 7(b)). As it can be observed in Fig. 7(a), the sum of squared distance (SSE) between data points and their assigned clusters' centroids does not provide enough information to apply the Elbow method (there is not a clear elbow bend, specially referring to S1), although it does provide some hints that the optimum K would be around K= 4–6. However, combined with the Silhouette score, it can be stated that the best option for both S1 and S2 is K= 5, where the Silhouette score has a local maximum, in Fig. 7(b).

Regarding to the DBSCAN algorithm, there are two parameters to be determined: ε and the MinPts. A sweep of both parameters is made to see which combination obtains the highest Silhouette Score values: $\varepsilon = [0.001 - 0.2]$, and $\min Pts = [50 - 250]$. Table 4 is showing the results, where the number of clusters obtained, K, is the same as those defined for the K-Means algorithm.

The results of S1 and S2 using both algorithms, K-Means and DBSCAN, are shown in Fig. 8 and Fig. 9, where the clusters are drawn on the Principal Components (PC0 and PC1). Although groups can clearly be distinguished in both cases, the division obtained with DBSCAN

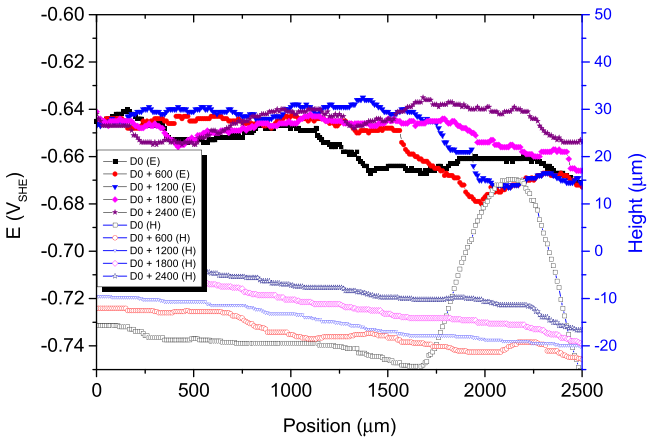


Fig. 5. Potential and height vs. the position as a function of different distances (from 600 µm up to 2400 µm). from an existing filament (line D0).

coincides with this group distinction, whilst a more arbitrary division was obtained with K-Means.

In order to confirm the findings using DBSCAN algorithm, the Bhattacharyya Coefficients (BC) were calculated for both algorithms (Table 5 and Table 6). It can be observed that the distributions of the clusters obtained from the DBSCAN algorithms are more separated that the clusters obtained from the K-Means algorithm (i.e., lower BCs are generally obtained using the former algorithm).

From these results, it can be stated that the DBSCAN algorithm obtains a better separation among the clusters for both experiments. This is due to the fact that the underlying assumptions on the shape of the clusters (e.g., clusters are spherical, equally sized, equally dense and not contaminated by noise) are not met for the K-Means algorithm.

4.2.2. *Clusters analysis*

4.2.2.1. Robustness of the clustering. Clustering results obtained by DBSCAN have been listed in Table 7 as a range of the AE descriptors values within a specific cluster. The intervals were set by including significant data (i.e., considering only values between the percentiles 10 and 90 in order to exclude marginal data), avoiding large intervals and simplifying the data complexity for analysis. In order to explore the robustness of the clustering method, the results form S1 and S2 have been compared. It can be observed that the CL1 of S1 and the CL1 of S2 have similarities in their frequency bands, especially in their frequency centroid and their peak frequency. This similarity also occurs between clusters CL2 of S1 and S2, clusters CL3, as well as clusters CL4. The cluster CL0, conversely, was characterized by wide range of values for all AE descriptors (see Table 7). The high dispersion of the values suggests that cluster CL0 gathers outliers coming from the rest of clusters, so it was excluded from the subsequent analysis to not distort the results.

To verify such similarities, the Bhattacharyya Coefficients (BC) were

Table 4
Parameters used for DBSCAN algorithm.

Specimen	ε	minPts	Silhouette Score	K
S1	0.06	28	0.595	5
S2	0.06	130	0.605	5

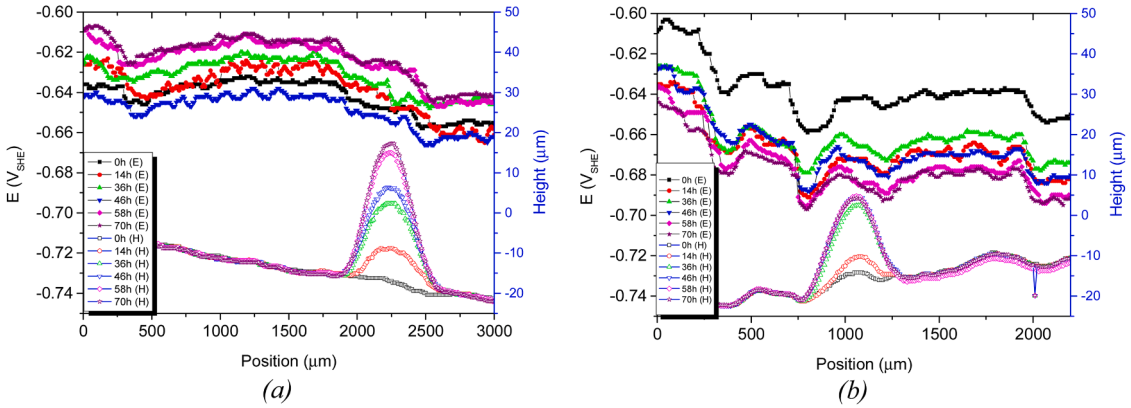


Fig. 6. : Potential and height vs. the position varying the time of exposure: (a) filament far from the artificial defect, (b) filament close to the artificial defect.

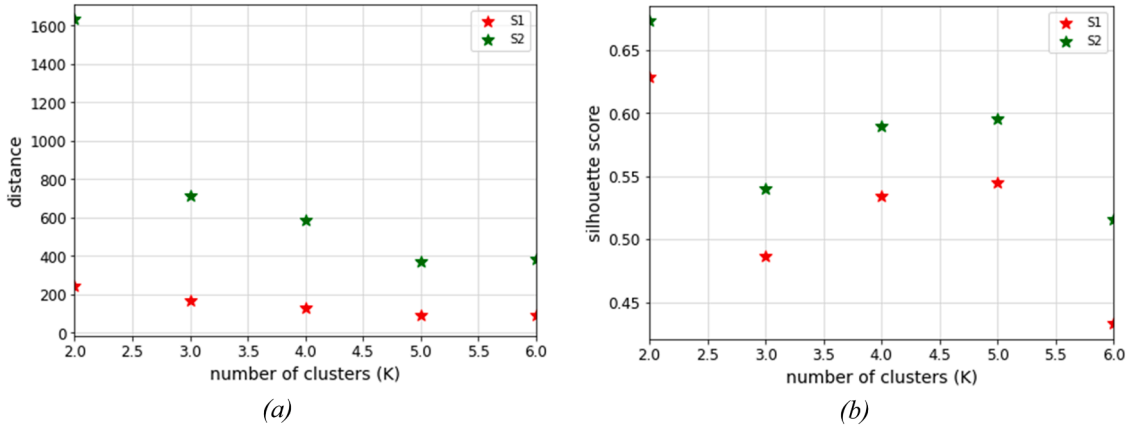


Fig. 7. : (a) Distance between clusters, and (b) Silhouette score for different K for S1 and S2.

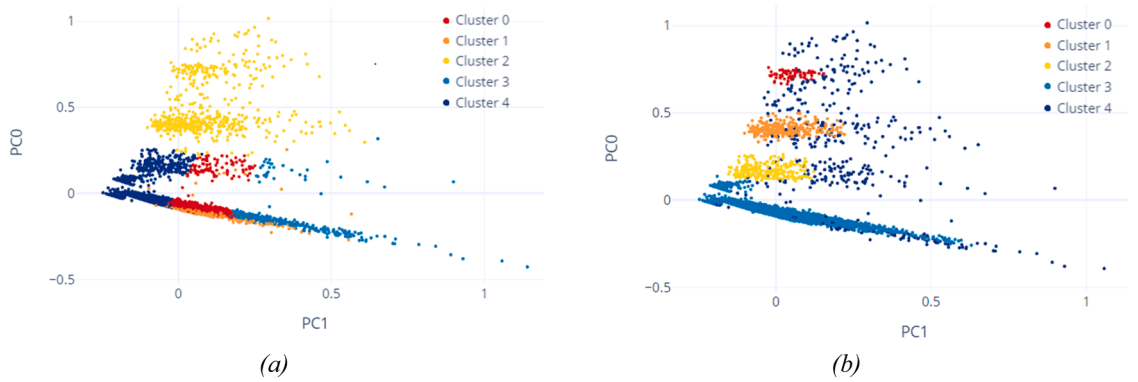


Fig. 8. : S1. Clusters obtained using the (a) K-Means algorithm, and (b) DBSCAN algorithm, represented the first principal components.

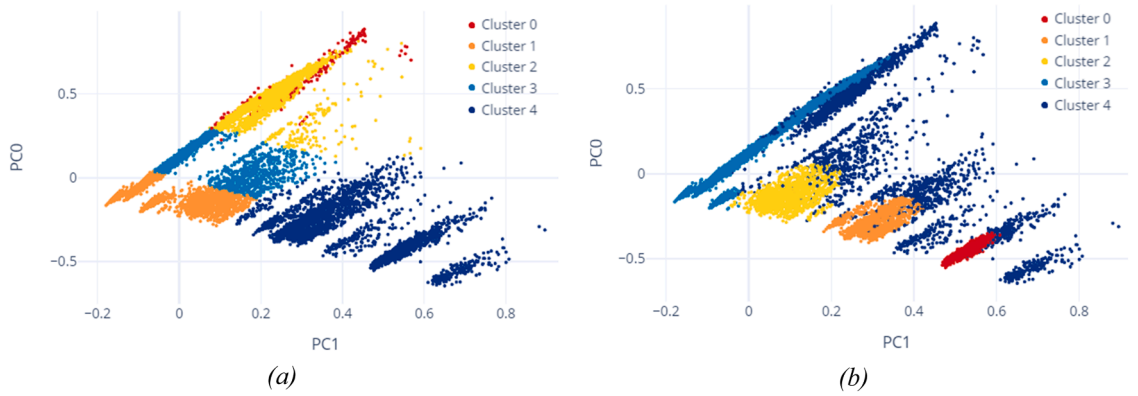


Fig. 9. : S2. Clusters obtained using the (a) K-Means algorithm, and (b) DBSCAN algorithm, represented the first principal components.

Table 5
Bhattacharyya Coefficients between clusters of S1, for (a) K-Means and (b) DBSCAN algorithms.

(a)						*(b)*					
CL0	1					**CL0**	1				
CL1	0.29	1				**CL1**	0.1	1			
CL2	0.31	0.18	1			**CL2**	0.001	0.09	1		
CL3	0.46	0.22	0.78	1		**CL3**	0.39	0.51	0.04	1	
CL4	0.42	0.18	0.37	0.65	1	**CL4**	0.51	0.4	0.04	0.74	1
	CL0	**CL1**	**CL2**	**CL3**	**CL4**		**CL0**	**CL1**	**CL2**	**CL3**	**CL4**

Table 6
Bhattacharyya Coefficients between clusters of S2, for (a) K-Means and (b) DBSCAN algorithms.

(a)						*(b)*					
CL0	1					**CL0**	1				
CL1	0.66	1				**CL1**	0.21	1			
CL2	0.72	0.77	1			**CL2**	0.29	0.12	1		
CL3	0.68	0.85	0.84	1		**CL3**	0.38	0.15	0.68	1	
CL4	0.55	0.6	0.66	0.72	1	**CL4**	0.35	0.14	0.7	0.8	1
	CL0	**CL1**	**CL2**	**CL3**	**CL4**		**CL0**	**CL1**	**CL2**	**CL3**	**CL4**

calculated between clusters of different experiments (Table 8). BC values (from 0.87 to 0.97) reveal that clusters along the two experiments are significantly equal, especially for the ones close to 0.95 or above.

Interestingly, this finding is not only observed for the AE descriptors previously identified by PCA but also in the waveform of the AE signals. Fig. 10 shows the waveforms in the frequency domain, where similar frequency spectrums of the clusters are found. It can be observed that signals from similar clusters on specimens S1 and S2, e.g. clusters CL1 (Fig. 10(a) and Fig. 10(b)), show dominant frequencies at the same range, i.e. 10–30 kHz. Moreover, not only similar dominant frequencies bands were obtained but also secondary frequency bands also match for each couple of clusters (CL1(S1) vs. CL1(S2), etc.).

4.2.2.2. Relevance of the AE descriptors. In order to explore which AE descriptor (here called variable) was more relevant for clustering, results from Fig. 10 and Table 7 have been compared. Although the frequency seems to be one of the most relevant variables to distinguish between clusters, a depth analysis is required. Taking into account that, for the clustering process (i.e., identifying the location of the object in the right cluster), high similarity between a centroid and its objects is likely more

Table 7
AE descriptors values resulting from the DBSCAN clustering.

	Rt. [μs]	F.C. [kHz]	Dur. [μs]	Amp. [dB]	A.E. [aJ]	P.F. [kHz]
S1						
CL0	19.3–2082	186–277	215 – 6581	26–44	1.8 – 127.0	15–300
CL1	19 – 731	219–256	60 – 3867	25–29	0.45–23.46	13–15
CL2	5 – 88	236–268	32 – 296	25–33	0.27–4.58	74–105
CL3	8 – 124	245–274	53 – 607	26–36	0.5–9.5	167–190
CL4	3.3 – 127	269–286	46 – 425	26–33	0.5–4.1	294–302
S2						
CL0	15–9453	100–268	365–19112	33–58	5–20293	13–291
CL1	31–2928	100–251	98 – 5598	25–32	0.66–68.98	13–15
CL2	10 – 248	218–265	48 – 946	25–33	0.33–10.77	74–113
CL3	8 – 290	238–272	67 – 1206	26–35	0.46–12.40	167–203
CL4	9 – 258	256–283	77 – 1473	26–35	0.54 – 11.80	289–296

Table 8
Bhattacharyya Coefficients between clusters of the experiments S1 and S2.

CL1	0.87	0.39	0.44	0.34
CL2	0.67	0.94	0.73	0.65
CL3	0.63	0.71	0.94	0.7
CL4	0.57	0.63	0.66	0.97
S2	**CL1**	**CL2**	**CL3**	**CL4**
S1				

important than a variable that has low similarity, such variables can be ranked. In this sense, the DBSCAN algorithm is not able to provide a raking about which variable is more relevant for each cluster, and therefore, Random Forest classifier was used to overcome this issue. Random Forest is an ensemble classifier, consisting of multiple decision trees and trained using randomly selected feature subspaces [62]. This method builds multiple decision trees at training phase. Often, a pruning process is applied to reduce both, tree complexity and training data overfitting. In order to predict the class of a new instance, this instance is put down to each of these trees; each tree gives a prediction (votes) and the class having most votes over all the trees of the forest will be selected (majority voting). The algorithm uses the bagging method [63], where each tree is trained using a random subset (with replacement) of the original dataset. In addition, each split uses a random subset of features.

Once the events are labelled through clustering, the algorithm proceeds to train the classifier, which, in return, will rank the importance of the variables when constructing the classifier as it is shown in Table 9. The importance ranking of the variables, computed by the classification algorithm, is then compared with the most important features for the clustering algorithm. Independently of the specimen, the importance of the variable coincides in both cases, noting also that the peak frequency (PF) is more important (i.e., twice) than the next variable (i.e., Amplitude) in both experiments (Table 9, columns S1 (a) and S2 (a)).

In order to find out the relevance of each AE descriptor in the FFC process, random forest classifier has been used again once CL0 and CL1, the two clusters related to outliers and noise (discussed below: Section 4.2.3), respectively, have been removed. Results from the histogram have been depicted in Table 9 (columns S1 (b) and S2 (b)) which reveals that the Peak Frequency remains as the most important parameter to distinguish the clusters. In contrast to this finding, the peak frequency was not a relevant AE descriptor (the last one out of 7) in the case of stress corrosion cracking (SCC) using also a wideband sensor [23]. Moreover, random forest (which properly predicted 80 out of 82 AE signals) was showing that the rise time was playing the most significant role for SCC, followed by duration, average frequency and absolute energy. The two first descriptors were found to be the best to distinguish between the dislocations and crack signals [23]. However, rise time does not play a crucial role as discriminatory descriptor for FCC (either considering four clusters or only three (from CL2 to CL4)). Therefore, it seems that SCC and FFC could be easily distinguished if both corrosion forms are monitored by AE, where their key descriptors are not necessary the same.

If the importance of the remaining descriptors is compared considering either four clusters (columns S1 (a) and S2 (a)) or three (columns S1 (b) and S2 (b)), the following differences are observed: (i) the importance of the amplitude is hindered in both tests (especially in S1) which could indicate that such descriptor is mainly linked to the noise/outliers clusters rather than to the FCC ones, (ii) the duration becomes to be a key parameter only for S1, and (iii) the absolute energy seems to have a large influence in both tests, but especially in S2. Then, the peak frequency + absolute energy + duration are having the largest contribution (above 80%) to classify the waveform for both tests (Table 9). It was found that absolute energy and duration could be used to identify pitting corrosion (covered morphology) on stainless steel. Wu et al. found that the absolute energy was a key AE descriptor to explore pitting corrosion on AISI 304, where two type of clusters were identified: the low-energy cluster (having mainly continuous signals and related to the hydrogen-bubble evolution inside the pits) and the high-energy one (having mainly burst signals attributed to the rupturing of the pit covers during the pit growth) [22]. On the other hand, they also stablished a morphological descriptor for pitting corrosion derived from the AE duration-amplitude plots: long-duration AE signals were obtained when occluded pits (generated under low-concentration NaCl solutions) whilst short-duration burst signals were obtained when open pits with ruptured covers occurred (under high-concentration NaCl) [64]. Although further research has to be done to understand the differences between S1 and S2 tests (e.g., using other AE features, such as wavelet parameters), the findings in literature about pitting corrosion seems to be aligned with the weight of the absolute energy and the duration in the clustering".

In order to confirm one of the main findings of this work, i.e., the peak frequency as the most discriminant descriptor, the peak frequency vs. time was plotted for both experiments. It can be clearly observed the similarities between the clusters of S1 (Fig. 11(a)) and S2 (Fig. 11(b)), where clusters have been arranged according to the DBSCAN range values (see Table 7) and the frequency spectrum (Fig. 10).

As result, it can be argued that the peak frequency of the AE signals seems to be a key feature to characterize and to distinguish the different mechanisms taking place during FFC. At this point, if the peak frequency (PF) is considered the most relevant descriptor for clustering (random forest analysis, [62]), the following questions can arise concerning the correlation between the dominant frequencies of the AE signals and the PF: Does the PF match the range of the dominant frequencies? How can be assured that the dominant frequencies of a signal from cluster CL2 (from 74 to 113 kHz in Table 7) is truly ranged between 60 and 120 kHz?

In order to answer these questions, the Power Spectrum Density (PSD) of the AE signals was calculated and evaluated. The PSD describes how the power of a signal or time series is distributed over frequency, and is defined as follows:

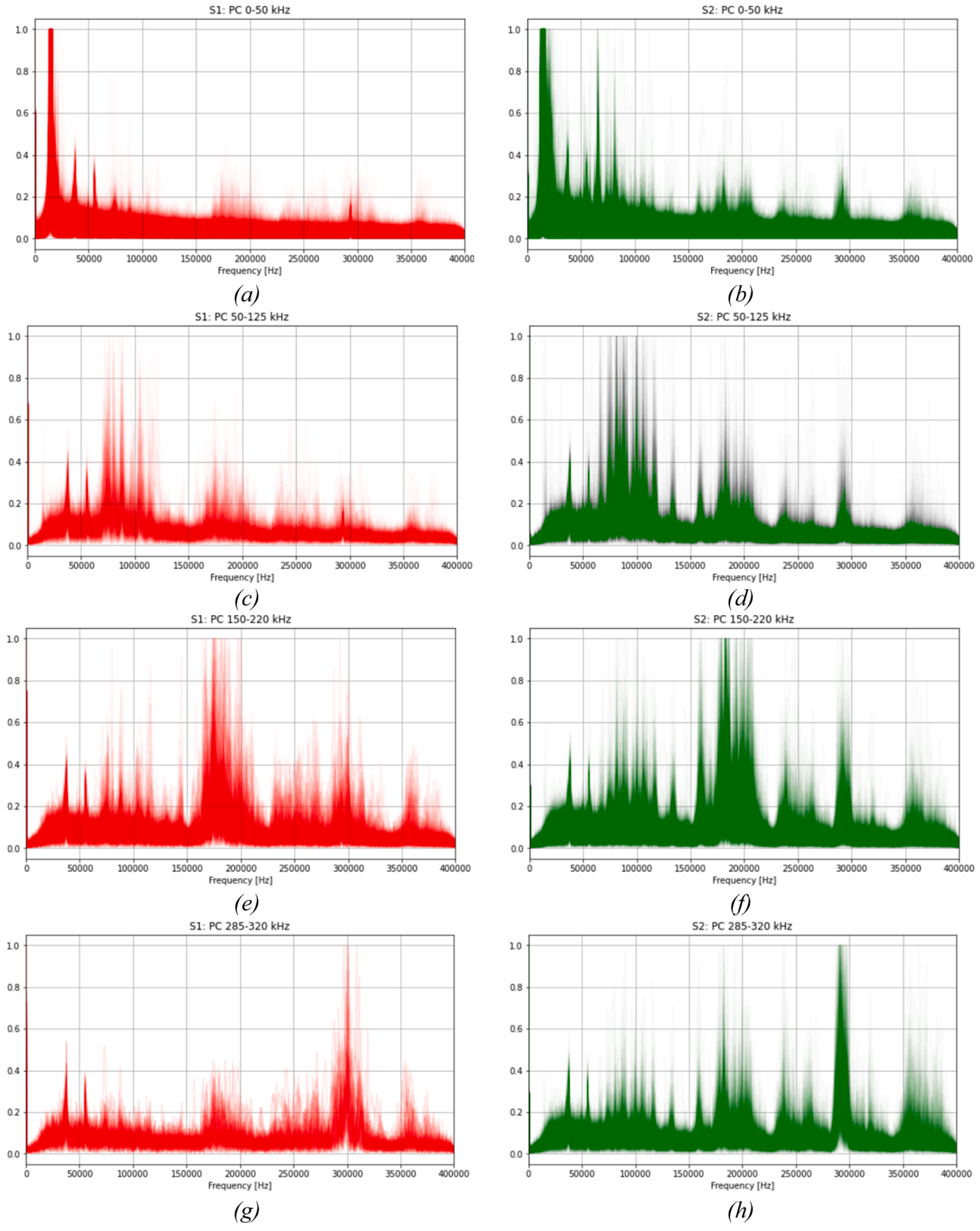


Fig. 10. : Comparison of the FFTs of the waveforms belonging to the CL1s (a) of S1 and (b) of S2, the CL2s (c) of S1 and (d) of S2, the CL3s (e) of S1 and (f) of S2, the CL4s (g) of S1 and (h) of S2.

Table 9
Importance of the experiments according to the Random Forest algorithm, used with labels created by DBSCAN clusterization for S1 and S2 tests: (a) considering all clusters (from CL0 to CL4), (b) taking into account only clusters related to FFC (CL2 to CL4).

Descriptor	S1		Descriptor	S2	
	(a) Importance all clusters (CL0 to CL4)	(b) Importance on clusters related to FFC (CL2 to CL4)		(a) Importance all clusters (CL0 to CL4)	(b) Importance on clusters related to FFC (CL2 to CL4)
Peak Freq. [kHz]	0.39	0.31	**Peak Freq. [kHz]**	0.51	0.49
Amplitude [dB]	0.21	0.05	**Amplitude [dB]**	0.23	0.11
Duration [μs]	0.17	0.29	**Duration [μs]**	0.10	0.01
Risetime [μs]	0.14	0.09	**Risetime [μs]**	0.10	0.01
Freq. Centroid [kHz]	0.05	0.05	**Freq. Centroid [kHz]**	0.04	0.07
Absolute Energy [aJ]	0.04	0.21	**Absolute Energy [aJ]**	0.02	0.31

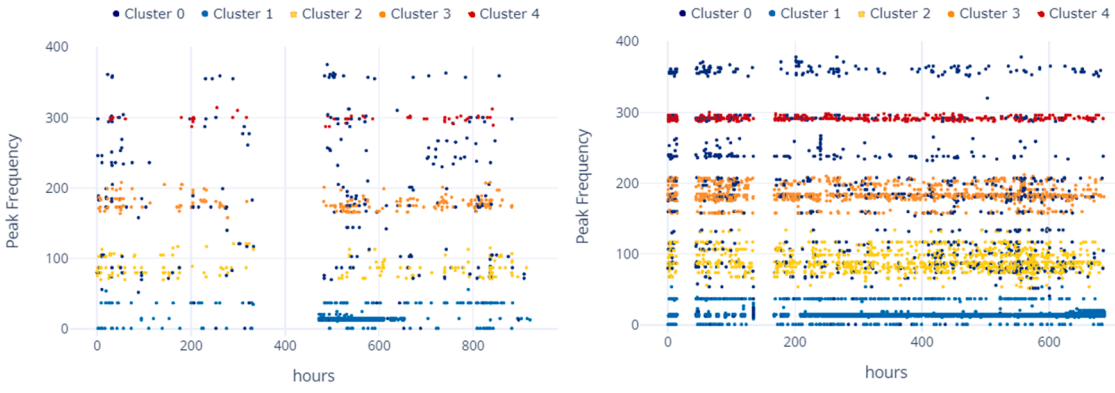


Fig. 11. : Clusters obtained by the DBSCAN algorithm represented in the Peak Frequency in time, (a) S1 and (b) S2.

$$S_{xx}(f) = \lim_{T\to\infty}\frac{1}{T}|\hat{x}_T(f)|^2 \tag{9}$$

Where the period T is centred about some arbitrary time $t = t_0$. On the other hand, $\hat{x}_T(f)$ is the ordinary Fourier Transform of $x_T(t) = x(t)\cdot\omega_T(t)$, and $\omega_T(t)$ is unity within the arbitrary period and zero elsewhere.

In this case, the PSD was calculated for the AE signals coming from cluster CL2, associated to the opening interface delamination. In particular, the weight of PSD at different ranges of 60 kHz R1, R2, …, R7 (see Fig. 12), called Spectral Energy (SE), was evaluated. In this way, if the Spectral Energy (SE) of R2 (60–120 kHz) is 20% higher than the SE of the rest of ranges (Eq. (10)), then the dominant frequencies are considered located in range 2, i.e. between [60 − 120] kHz.

$$\frac{SE_2}{SE_1} > \frac{1.2\wedge SE_2}{SE_3} > \frac{1.2\wedge SE_2}{SE_4} > \frac{1.2\wedge SE_2}{SE_5} > \frac{1.2\wedge SE_2}{SE_6} > \frac{1.2\wedge SE_2}{SE_7} > 1.2 \tag{10}$$

The results have shown that the 97.09% of the AE signals from S1 and the 95.03% from S2 have the dominant frequencies of CL2 in the 60 – 120 kHz range. Therefore, the PF parameter was consistent with the dominant frequencies of the AE signals.

4.2.3. *Clusters assignation*

According to the confocal images and pictures (Fig. 3), both specimens underwent different visual damage in the metal/coating interface. However, not only the same number of clusters were obtained but also they were significantly equal (Fig. 11) for both specimens. Therefore, once clustering has been successfully implemented, the different clusters

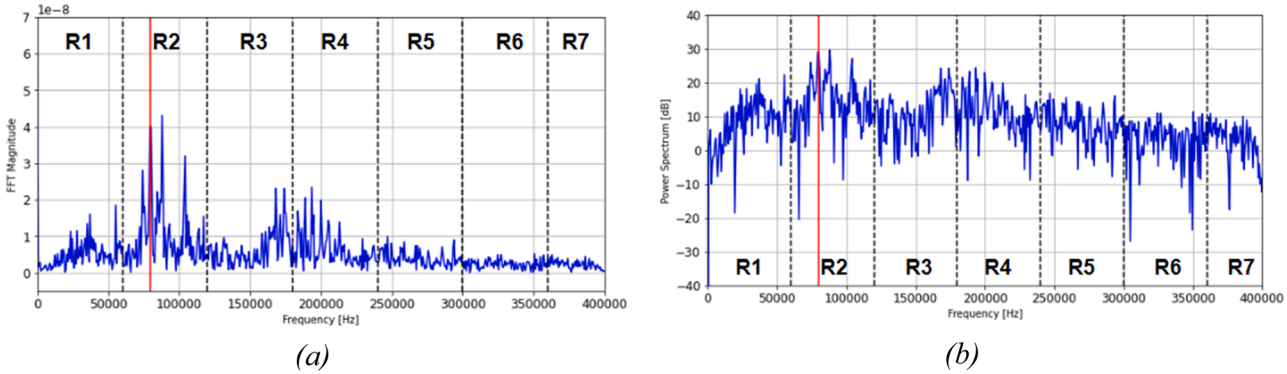


Fig. 12. : (a) FFT of a waveform and (b) its equivalent PSD belonging to CL2, with a Peak Frequency of 59 kHz.

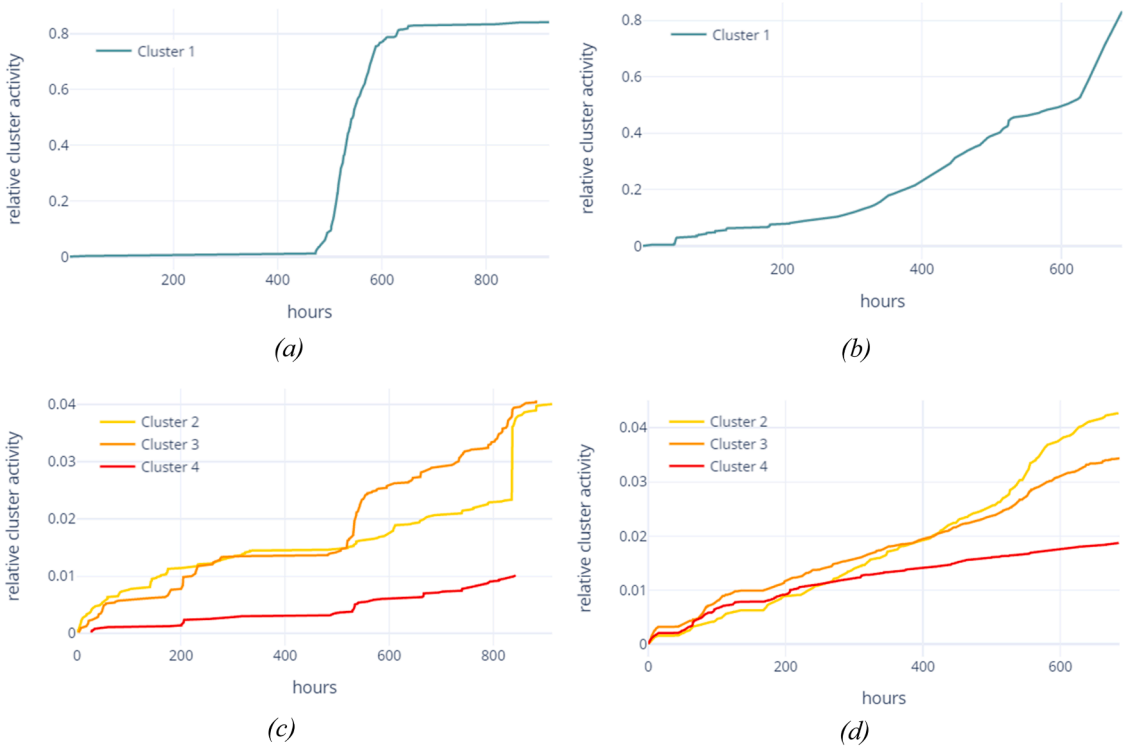


Fig. 13. : Relative activity of the CL1 of (a) S1 and (b) S2, and relative activity of the CL2, CL3 and CL4 of (c) S1 and (d) S2.

are going to be identified.

The cluster CL1 is characterized by low and specific values of peak frequency in the audible range (13–15 kHz, see Table 7) and its signals reached very high durations (S1: 3866.8 μs, S2: 5597.5 μs) and energy values (23.46 aJ, 68.98 aJ) comparing with the rest of clusters (CL2–4). Moreover, this cluster is gathering more than 80% of data (Fig. 13) while FFC phenomena is expected to release low rate of AE activity. All these features suggest that cluster CL1 includes AE signals of the ambient noise in the laboratory (probably coming from electromagnetic noise of SKP or other laboratory equipment). It should be noted that this cluster appears for a limited period of time during the experiment S1, in contrast with S2 where this cluster took place the whole experiment (Fig. 13). This erratic nature of cluster CL1 agrees with the assignation to noise sources.

Clusters CL2–4 show similar values between them in terms of duration, rise time and amplitude; they are of transient type as they exhibit low duration/rise time but a relatively higher amplitude. Nevertheless, if comparing those clusters between experiments S1 and S2, in the latter one some signals reached higher durations and rise time, but keep exactly the same frequency content (peak frequency and frequency centroid) when evaluating cluster by cluster. In this sense, it is known that every mechanism release AE signals of a characteristic frequency content, which is substantially affected by propagation only in dispersive medium [65] (not the case of metallic substrates). Therefore, it could be concluded that clusters CL2–4, even though they show slight differences in the time domain, they still being of transient type and come from the same mechanisms for both tests.

Regarding the time domain, transient AE signals were observed in previous works, as for example the mechanical disbonding at the interface mild steel/lacquer reported by Callow and Scantlebury [66]. In fact, based on the mechanical component of the disbondment, early works in this domain suggested that the delamination substrate/paint is a possible AE source [67] coming, in some extent, from the pressure on the substrate/coating during FFC tests [68,69] or mechanical tests [15]. Tsuru et al. [70] validated this hypothesis and concluded that some AE signals were coming from the metal/resin bonds after failure (i.e., disbondment) due to the tensile stress. The latter is released there, partially, as a vibration energy and spreads through metal substrate.

Considering the clusters CL2–4 in the frequency domain, the values displayed by signals of cluster CL2 (PF: 74–113 kHz) are in agreement with the finding from Yao et al. [71], where the dominant frequencies of the opening interface cracks during delamination (tensile mode) on thermal barrier coatings (TBC) was found in the range of 80–100 kHz. Further works on TBC [72] have evidenced that sliding interface cracks (shear mode) occurring during the delamination process are characterized by AE signals of frequencies ranged in 270–300 kHz, which coincides with the signals of cluster CL4 (PF: 289–302 kHz).

Finally, the cluster CL3, corresponding to AE signals of peak frequency in the range 167–203 kHz, could be associated with a mixed mode of tensile (opening) and shear (sliding) modes. At this point, it should be noted that the delamination mode on coating systems, unless they are tested under specific standards to evaluate pure tensile or shear modes, they are rather of mixed mode on real processes. The lift angle at the tip/edges of the FFC filament, which depends on factors as the quality of adhesion [16] or the height of the filament, implies that the stresses transferred by the lift force to the interface metal-coating are

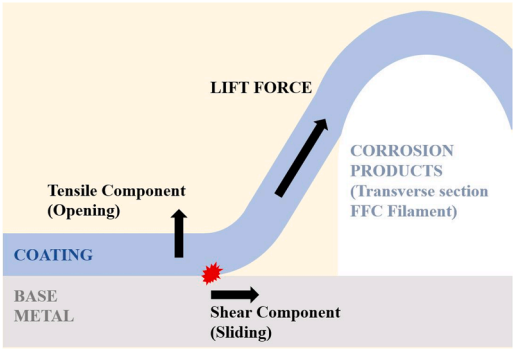


Fig. 14. : Theoretical decomposition of the lift force on tensile and shear components at the edge of the filament during the FFC progress.

decomposed in tensile and shear components (Fig. 14).

As a summary, it has been shown here that a proper AE data analysis is needed to show/reveal the sensitivity of AE signal to identify different delamination modes in the metal/coating interface, and therefore, to provide new insights in the FFC mechanism. Nevertheless, further research should be done to validate the assignation of clusters CL2, CL3 and CL4 to the different modes of (mechanical) delamination in the interface.

5. Conclusions

FFC has been successfully monitored on coated AA77075-T6 by AE sensors into the SKP chamber. After an initial screening of the AE data (up to six AE descriptors by PCA), DBSCAN algorithm was able to provide four well-differentiated clusters (CL1-CL4) despite having a multivariate complex data set which contains: i) noise, ii) clusters with signals of extreme features, iii) set of clusters with uneven sizes. The goodness of DBSCAN algorithm has been confirmed using metrics such as the Bhattacharyya Coefficient for the clusters related to the FFC (from CL2 to CL4): interface delamination either by opening (CL2) or sliding (CL4), and the mixed mode (CL3) have been successfully identified to be part of the FFC mechanism. Even though the assignation of the clusters to each physical event shall be validated in further research, the peak frequency has been identified as the most relevant classifier in the clustering process, for this research, by Random Forest. This fact was also verified by detecting the range of dominant frequencies, either by observing the frequency domain of the AE signals or by calculating its Power Spectral Density value. Hence, the peak frequency has demonstrated to be the key descriptor to cluster reliably the AE signals of different mechanisms during FFC experiments.

CRediT authorship contribution statement

C. Abarkane: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration. **A. M. Florez-Tapia:** Software, Writing – original draft, Writing – review & editing. **J. Odriozola:** Software, Investigation, Writing – review & editing, Supervision. **A. Artetxe:** Software, Supervision. **M. Lekka:** Methodology, Supervision. **E. García-Lecina:** Resources, Supervision. **H.-J. Grande:** Resources, Supervision. **J.M. Vega:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Acknowledgments

This work has been funded by the European Commission under the Horizon 2020 framework (H2020-CS2-CFP09-2018-02) within the U-Cross project (Grant Agreement 864905) in collaboration with Dassault Aviation, INSA Lyon, Mistras Group, Titania Ensayos y Proyectos Industriales and Université Bourgogne-Franche-Comté. J.M. Vega also would like to acknowledge the support of RYC2021-034384-I by Ministerio de Ciencia e Innovación.

References

[1] Logistics Management Institute (LMI), Estimated Impact of Corrosion on Cost and Availability of DoD Weapon Systems—FY18 Update, Tysons, VA, USA, 2018.
[2] J. Michael Gilmore, Director, Operational Test and Evaluation FY 2016 Annual Report, 2016.
[3] A. Carpinteri, G. Lacidogna, N. Pugno, Structural damage diagnosis and life-time assessment by acoustic emission monitoring, Eng. Fract. Mech. 74 (2007) 273–289, https://doi.org/10.1016/j.engfracmech.2006.01.036.
[4] Kanji Ono, Structural integrity evaluation using acoustic emission, J. Acoust. Emiss. 25 (2007).
[5] D.G. Aggelis, N.-M. Barkoula, T.E. Matikas, A.S. Paipetis, Acoustic structural health monitoring of composite materials: damage identification and evaluation in cross ply laminates using acoustic emission and ultrasonics, Compos. Sci. Technol. 72 (2012) 1127–1133, https://doi.org/10.1016/j.compscitech.2011.10.011.
[6] H.A. Elfergani, R. Pullin, K.M. Holford, Damage assessment of corrosion in prestressed concrete by acoustic emission, Constr. Build. Mater. 40 (2013) 925–933, https://doi.org/10.1016/j.conbuildmat.2012.11.071.
[7] A. Behnia, H.K. Chai, T. Shiotani, Advanced structural health monitoring of concrete structures with the aid of acoustic emission, Constr. Build. Mater. 65 (2014) 282–302, https://doi.org/10.1016/j.conbuildmat.2014.04.103.
[8] A. Zaki, H. Chai, D. Aggelis, N. Alver, Non-destructive evaluation for corrosion monitoring in concrete: a review and capability of acoustic emission technique, Sensors 15 (2015) 19069–19101, https://doi.org/10.3390/s150819069.
[9] S.E. Dunn, J.D. Young, W.H. Hartt, R.P. Brown, Acoustic emission characterization of corrosion induced damage in reinforced concrete, CORROSION 40 (1984) 339–343, https://doi.org/10.5006/1.3593933.
[10] S. Yuyama, T. Kishi, AE analysis during corrosion, stress corrosion cracking and corrosion fatigue processes, J. Acoust. Emiss. 2 (1983).
[11] S. Ramadan, L. Gaillet, C. Tessier, H. Idrissi, Detection of stress corrosion cracking of high-strength steel used in prestressed concrete structures by acoustic emission technique, Appl. Surf. Sci. 254 (2008) 2255–2261, https://doi.org/10.1016/j.apsusc.2007.09.011.
[12] N. Wint, D. Eaves, E. Michailidou, A. Bennett, J.R. Searle, G. Williams, H. N. McMurray, The kinetics and mechanism of filiform corrosion occurring on zinc-aluminium-magnesium coated steel (Ahead of Print), Corros. Sci. (2019), https://doi.org/10.1016/j.corsci.2019.06.028.
[13] C.F. Glover, G. Williams, Inhibition of corrosion-driven organic coating delamination and filiform corrosion on iron by phenyl phosphonic acid, Prog. Org. Coat. 102 (2017) 44–52, https://doi.org/10.1016/j.porgcoat.2016.03.006.
[14] R.T. Ruggeri, T.R. Beck, An analysis of mass transfer in filiform corrosion, Corrosion 39 (1983) 452–465, https://doi.org/10.5006/1.3581907.
[15] F. Mansfeld, M. Kendig, Lifetime Prediction of Organic Coating/Metal Systems, 1986.
[16] H. Leth-Olsen, K. Nisancioglu, Filiform corrosion morphologies on painted aluminum alloy 3105 coil material, Corrosion 53 (1997).
[17] D. Li, W. Yang, W. Zhang, Cluster analysis of stress corrosion mechanisms for steel wires used in bridge cables through acoustic emission particle swarm optimization, Ultrasonics 77 (2017) 22–31, https://doi.org/10.1016/J.ULTRAS.2017.01.012.
[18] A. Farhidzadeh, A.C. Mpalaskas, T.E. Matikas, H. Farhidzadeh, D.G. Aggelis, Fracture mode identification in cementitious materials using supervised pattern recognition of acoustic emission features, Constr. Build. Mater. 67 (2014) 129–138, https://doi.org/10.1016/j.conbuildmat.2014.05.015.
[19] L. Li, S.V. Lomov, X. Yan, V. Carvelli, Cluster analysis of acoustic emission signals for 2D and 3D woven glass/epoxy composites, Compos. Struct. 116 (2014) 286–299, https://doi.org/10.1016/J.COMPSTRUCT.2014.05.023.
[20] M.G.R. Sause, A. Gribov, A.R. Unwin, S. Horn, Pattern recognition approach to identify natural clusters of acoustic emission signals, Pattern Recognit. Lett. 33 (2012) 17–23, https://doi.org/10.1016/J.PATREC.2011.09.018.
[21] L. Calabrese, G. Campanella, E. Proverbio, Noise removal by cluster analysis after long time AE corrosion monitoring of steel reinforcement in concrete, Constr. Build. Mater. 34 (2012) 362–371, https://doi.org/10.1016/J.CONBUILDMAT.2012.02.046.
[22] K. Wu, J.Y. Kim, Acoustic emission monitoring during open-morphological pitting corrosion of 304 stainless steel passivated in dilute nitric acid, Corros. Sci. 180 (2021), 109224, https://doi.org/10.1016/J.CORSCI.2020.109224.
[23] Z. Zhang, X. Wu, J. Tan, In-situ monitoring of stress corrosion cracking of 304 stainless steel in high-temperature water by analyzing acoustic emission waveform, Corros. Sci. 146 (2019) 90–98, https://doi.org/10.1016/J.CORSCI.2018.10.022.
[24] L. Calabrese, M. Galeano, E. Proverbio, D. Di Pietro, F. Cappuccini, A. Donato, Monitoring of 13% Cr martensitic stainless steel corrosion in chloride solution in presence of thiosulphate by acoustic emission technique, Corros. Sci. 111 (2016) 151–161, https://doi.org/10.1016/J.CORSCI.2016.05.010.
[25] L. Calabrese, G. Campanella, E. Proverbio, Identification of corrosion mechanisms by univariate and multivariate statistical analysis during long term acoustic emission monitoring on a pre-stressed concrete beam, Corros. Sci. 73 (2013) 161–171, https://doi.org/10.1016/J.CORSCI.2013.03.032.
[26] R. Piotrkowski, E. Castro, A. Gallego, Wavelet power, entropy and bispectrum applied to AE signals for damage identification and evaluation of corroded galvanized steel, Mech. Syst. Signal Process. 23 (2009) 432–445, https://doi.org/10.1016/J.YMSSP.2008.05.006.
[27] J. Li, G. Du, C. Jiang, S. Jin, The classification of acoustic emission signals of 304 stainless steel during stress corrosion process based on K-means clustering, Anti-Corros. Methods Mater. 59 (2012) 76–80, https://doi.org/10.1108/00035591211210848/FULL/XML.

[28] L. Calabrese, E. Proverbio, A review on the applications of acoustic emission technique in the study of stress corrosion cracking, 1–30, Corros. Mater. Degrad. Vol. 2 (2) (2020) 1–30, https://doi.org/10.3390/CMD2010001.
[29] K. Wang, X. Zhang, S. Song, Y. Wang, Y. Shen, P.D. Wilcox, Rail steel health analysis based on a novel genetic density-based clustering technique and manifold representation of acoustic emission signals, Appl. Artif. Intell. 36 (2022), https://doi.org/10.1080/08839514.2021.2004346.
[30] J. Hillenbrand, J. Detroy, J. Fleischer, Investigation of defects in roll contacts of machine elements with acoustic emission and unsupervised machine learning, IOP Conf. Ser. Mater. Sci. Eng. 1193 (2021), 012085, https://doi.org/10.1088/1757-899X/1193/1/012085.
[31] H. Bi, Z. Li, D. Hu, I. Toku-Gyamerah, Y. Cheng, Cluster analysis of acoustic emission signals in pitting corrosion of low carbon steel, Materwiss. Werksttech. 46 (2015) 736–746, https://doi.org/10.1002/mawe.201500347.
[32] M. Knapek, P. Minárik, J. Čapek, R. Král, J. Kubásek, F. Chmelík, Corrosion of pure magnesium and a WE43 magnesium alloy studied by advanced acoustic emission analysis, Corros. Sci. 145 (2018) 10–15, https://doi.org/10.1016/J.CORSCI.2018.09.006.
[33] G. Du, J. Li, W.K. Wang, C. Jiang, S.Z. Song, Detection and characterization of stress-corrosion cracking on 304 stainless steel by electrochemical noise and acoustic emission techniques, Corros. Sci. 53 (2011) 2918–2926, https://doi.org/10.1016/J.CORSCI.2011.05.030.
[34] P. Pallu, R. Suryawnashi, A. Dubey, A. Abha Choubey, A. Systematic, Review on K-means clustering techniques related papers a novel approach for dat a clust ering using improved K-means algorit hm A systematic review on K-means clustering techniques, Int. J. Sci. Res. Eng. Technol. 6 (2017).
[35] K. Khan, S.U. Rehman, K. Aziz, S. Fong, S. Sarasvady, A. Vishwa, DBSCAN: Past, present and future, 5th Int. Conf. Appl. Digit. Inf. Web Technol. ICADIWT 2014. (2014) 232–238. doi:10.1109/ICADIWT.2014.6814687.
[36] A. Ghosal, A. Nandy, A.K. Das, S. Goswami, M. Panday, A. Short, Review on different clustering techniques and their applications, Adv. Intell. Syst. Comput. 937 (2020) 69–83, https://doi.org/10.1007/978-981-13-7403-6_9/COVER.
[37] ISO 4623–2:2016 Paints and varnishes - Determination of resistance to filiform corrosion - Part 2: Aluminium substrates, (2016).
[38] ASTM B368–21 Standard Test Method for Copper-Accelerated Acetic Acid-Salt Spray (Fog) Testing (CASS Test), (2021).
[39] S.J. Vahaviolos, Acoustic emission: standards and technology update, ASTM International, 19428-2959. 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA, 1999, https://doi.org/10.1520/STP1353-EB.
[40] L. Ni, S. Jinhang, The analysis and research of clustering algorithm based on PCA, ICEMI 2017 - Proc. IEEE 13th Int. Conf. Electron. Meas. Instruments. 2018-Janua (2017) 361–365. doi:10.1109/ICEMI.2017.8265817.
[41] A. Ben-Hur, I. Guyon, Detecting stable clusters using principal component analysis, Funct. Genom. (2003) 159–182, https://doi.org/10.1385/1-59259-364-X:159.
[42] D. Xu, Y. Tian, A comprehensive survey of clustering algorithms, Ann. Data Sci. 22 (2) (2015) 165–193, https://doi.org/10.1007/S40745-015-0040-1.
[43] R.D.A. Jain, Algorithms for clustering data, 1988.
[44] D. Steinley, K-means clustering: a half-century synthesis, Br. J. Math. Stat. Psychol. 59 (2006) 1–34, https://doi.org/10.1348/000711005×48266.
[45] K. Sawant, Adaptive methods for determining DBSCAN parameters, IJISET-Int. J. Innov. Sci. Eng. Technol. 1 (2014).
[46] P. Makwana, T.M. Kodinariya, P.R. Makwana, Review on determining of cluster in k-means clustering review on determining number of cluster in K-means clustering, Int. J. Adv. Res. Comput. Sci. Manag. Stud. 1 (2013).
[47] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, J. Liu, A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm, Eurasip J. Wirel. Commun. Netw. 2021 (2021) 31, https://doi.org/10.1186/s13638-021-01910-w.
[48] K.Y. Yeung, D.R. Haynor, W.L. Ruzzo, Validating clustering for gene expression data, Bioinformatics 17 (2001) 309–318, https://doi.org/10.1093/BIOINFORMATICS/17.4.309.
[49] L. Hubert, P. Arabie, Comparing partitions, J. Cl. 2 (1985) 193–218, https://doi.org/10.1007/BF01908075.
[50] G. Camps-Valls, J.L. Rojo-Álvarez, M. Martínez-Ramón, Kernel methods in bioengineering, signal and image processing, Kernel Methods Bioeng. Signal Image Process. (2006) 1–415, https://doi.org/10.4018/978-1-59904-042-4.
[51] M. Shutaywi, N.N. Kachouie, Silhouette analysis for performance evaluation in machine learning with applications to clustering, Entropy 23 (2021) 759, https://doi.org/10.3390/e23060759.
[52] M. Hazewinkel Ed., Encyclopaedia of Mathematics. Dordrecht, in: Springer, Netherlands, 1997, https://doi.org/10.1007/978-94-015-1288-6.
[53] C. Radhakrishna Rao, Advanced Statistical Methods in Biometric Research, Wiley, 1952, https://doi.org/10.1002/ajpa.1330120224.
[54] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Stat. 23 (1952) 493–507, https://doi.org/10.1214/aoms/1177729330.
[55] Solomon Kullback, Information Theory and Statistics, Wiley, 1959.
[56] A.N. Kolmogorov, On the approximation of distributions of sums of independent summands by infinitely divisible distributions, Contrib. Stat. (1965) 159–174, https://doi.org/10.1016/B978-1-4832-3160-0.50016-6.
[57] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, J. R. Stat. Soc. Ser. B 28 (1966) 131–142, https://doi.org/10.1111/J.2517-6161.1966.TB00626.X.
[58] R. Kala, M. Krzyśko, W. Wołyński, Two preliminary tests for discriminant analysis, Commun. Stat. - Simul. Comput. 34 (2005) 179–189, https://doi.org/10.1081/SAC-200047074.
[59] H.N. McMurray, G. Williams, S. O'Driscoll, Chromate inhibition of filiform corrosion on organic coated AA2024-T3 studied using the scanning Kelvin probe, J. Electrochem. Soc. 151 (2004) B406, https://doi.org/10.1149/1.1757460.
[60] H.N. McMurray, G. Williams, Inhibition of filiform corrosion on organic-coated aluminum alloy by hydrotalcite-like anion-exchange pigments, Corrosion 60 (2004) 219–228, https://doi.org/10.5006/1.3287724.
[61] G. Williams, H.N. McMurray, Inhibition of filiform corrosion on organic-coated AA2024-T3 by smart-release cation and anion-exchange pigments, Electrochim. Acta 69 (2012) 287–294, https://doi.org/10.1016/J.ELECTACTA.2012.03.002.
[62] L. Breiman Random For. 45 2001 5 32.
[63] L. Bbeiman Bagging, Predict., 24, 1996, pp. 123–140.
[64] K. Wu, J.W. Byeon, Morphological estimation of pitting corrosion on vertically positioned 304 stainless steel using acoustic-emission duration parameter, Corros. Sci. 148 (2019) 331–337, https://doi.org/10.1016/J.CORSCI.2018.12.031.
[65] L.J. Graham, G.A. Alers, Acoustic emission in the frequency domain, Monit. Struct. Integr. Acoust. Emiss. (1975) 11–39.
[66] L.M. Callow, J.D. Scantlebury, Using Acoustic Emission to Investigate Disbonding at the Polymer-Metal Interface, in: 1986: pp. 115–122. doi:10.1021/bk-1986–0322.ch011.
[67] H. Hansmann, Application of acoustic emission analysis on adhesion and structural problems of organic and metallic coatings, Ind. Eng. Chem. Prod. Res. Dev. 24 (1985) 252–257.
[68] R.D. Rawlings, Acoustic emission testing of coatings, in: A.D. Wilson, J. W. Nicholson, H.J. Prosser (Eds.), Surf. Coatings—2, Springer Netherlands, Dordrecht, 1988, pp. 71–105, https://doi.org/10.1007/978-94-009-1351-6_3.
[69] T.W. Rettig, M.J. Felsen, Acoustic emission method for monitoring corrosion reactions, Corrosion 32 (1976) 121–126, https://doi.org/10.5006/0010-9312-32.4.121.
[70] T. Tsuru, A. Sagara, S. Haruyama, Acoustic emission measurements to evaluate the degradation of coating films, Corrosion 43 (1987) 703–707, https://doi.org/10.5006/1.3583853.
[71] W.B. Yao, C.Y. Dai, W.G. Mao, C. Lu, L. Yang, Y.C. Zhou, Acoustic emission analysis on tensile failure of air plasma-sprayed thermal barrier coatings, Surf. Coat. Technol. 206 (2012) 3803–3807, https://doi.org/10.1016/J.SURFCOAT.2012.03.050.
[72] L. Yang, H.S. Kang, Y.C. Zhou, W. Zhu, C.Y. Cai, C. Lu, Frequency as a key parameter in discriminating the failure types of thermal barrier coatings: cluster analysis of acoustic emission signals, Surf. Coat. Technol. 264 (2015) 97–104, https://doi.org/10.1016/j.surfcoat.2015.01.014.