A new criterion based on the distribution of cluster onsets for interpreting acoustic emission data signals: Three case studies in structural and process monitoring

Emmanuel Ramasso[1*], Martin Mbarga Nkogo[1†],
Neha Chandarana[2†], Gilles Bourbon[3†], Patrice Le Moal[3†],
Quentin Lefèbvre[4†], Martial Personeni[3†], Constantinos Soutis[5†],
Matthieu Gresil[6†], Sébastien Thibaud[1†]

[1]SUPMICROTECH, Université de Franche-Comté, CNRS, institut FEMTO-ST, Besançon, France.
[2]Department of Aerospace Engineering, Bristol University, United Kingdom.
[3]Université de Franche-Comté, CNRS, institut FEMTO-ST, Besançon, France.
[4]Université de Franche-Comté, Institut FEMTO-ST, F-25000, Besançon, France.
[5]Department of Materials, The University of Manchester, Manchester, United Kingdom.
[6]i-Composites Lab, Department of Materials Science and Engineering & Department of Mechanical and Aerospace Engineering, Monash University, Street, Clayton, VIC, Australia.

*Corresponding author(s). E-mail(s): emmanuel.ramasso@femto-st.fr;
Contributing authors: martin.mbarga@femto-st.fr;
neha.chandarana@bristol.ac.uk; gilles.bourbon@femto-st.fr;
patrice.lemoal@femto-st.fr; quentin.lefebvre@femto-st.fr;
martial.personeni@femto-st.fr; constantinos.soutis@manchester.ac.uk;
matthieu.gresil@monash.edu; sebastien.thibaud@femto-st.fr;
[†]These authors contributed equally to this work.

Abstract

Structural health monitoring (SHM) relies on non-destructive techniques such as acoustic emission (AE) that generate large amounts of data over the lifespan of systems. Clustering methods are used to interpret these data and gain insights into damage progression and mechanisms. Conventional methods for evaluating clustering results utilise clustering validity indices (CVI) that prioritise compact and separable clusters. This paper introduces a novel approach based on the temporal sequence of cluster onsets, indicating the initial appearance of potential damage and allowing for early detection of defect initiation. The proposed CVI is based on the Kullback-Leibler divergence and can incorporate prior information about damage onsets when available.

Three experiments on real-world datasets validate the effectiveness of the proposed method. The first benchmark focuses on detecting the loosening of bolted plates under vibration, where the onset-based CVI outperforms the conventional approach in both cluster quality and the accuracy of bolt loosening detection. The results demonstrate not only superior cluster quality but also unmatched precision in identifying cluster onsets, whether during uniform or accelerated damage growth. The two additional applications stem from industrial contexts. The first focuses on micro-drilling of hard materials using electrical discharge machining, demonstrating, for the first time, that the proposed criterion can effectively retrieve electrode progression to the reference depth, thus validating the setting of the machine to ensure structural integrity. The final application involves damage understanding in a composite/metal hybrid joint structure, where the cluster timeline is used to establish a scenario leading to critical failure due to slippage.

Keywords: Clustering validity index, Acoustic emission, Clusters timeline, Structural health monitoring

1 Introduction

Ensuring the availability of equipment and the integrity of structures is a critical concern within the aeronautical and civil engineering industries, both economically and for user safety. This necessitates the monitoring of structures as well as understanding potential damage mechanisms that can indicate the integrity of the structure. Structural health monitoring (SHM) is a research and development field aimed at implementing and deploying solutions to monitor engineering systems such as equipment and structures [1, 2].

During operation, systems and structures are subjected to mechanical, electrical, or chemical stresses, which generate localised micro-deformations. These micro-deformations may be accompanied by the release of energy in the form of elastic waves, known as acoustic emission (AE). When these changes are irreversible and likely to degrade performance, they are referred to as damage. Accumulated micro-damages eventually lead to material failure, rendering the structure unusable, and potentially leading to catastrophic failure [3, 4]. Piezoelectric transducers can be placed onto or embedded in the host structure to record and convert these micro-vibrations into

usable electrical signals [2, 5]. The recorded signals contain the signatures of the types of damage from which they have been generated [6].

The AE technique is particularly well suited to localise damage [7–9]. It is a passive technique that exploits the transient signals emitted during damage occurrence. Therefore, sensors must be continuously monitored or triggered by significant damage. Although the sampling frequency is relatively high (typically 5 to 10 MHz, depending on the sensors), it is possible to detect damage in real time from AE data streams, even during continuous monitoring, thanks to modern computing capabilities. A key challenge in real-time processing is the acoustic emissivity of the monitored material and/or structure; the more emissive the material and/or structure, the greater the number of AE sources it has, leading to more signals that need to be processed, which increases computational costs. To manage this, an amplitude threshold is typically applied to limit the number of transients to be analysed. After distinguishing between noise and meaningful data, various methodologies can be applied to interpret AE data streams. Generally, these methodologies share a common focus on analysing certain statistical properties, although the methods for deriving these properties vary significantly across publications.

One approach is *sample-based analysis* [10–12]. For instance, Lindley *et al.* [12] apply several thresholds in a probabilistic manner. Depending on how many times these thresholds are crossed, groups are formed using a non-parametric Bayesian approach based on the Dirichlet process. Since the algorithm can generate many clusters, interpretation is done by observing the frequency of cluster assignment. This method was demonstrated on a landing gear subjected to fatigue testing.

Samples can also be analyzed in a *windowed format*, as proposed by Martin-Del-Campo *et al.* [13, 14] in the context of bearing monitoring with particle-contaminated lubricant at varying rotational speeds. Acoustic emissions are analysed online using an unsupervised feature learning method based on orthogonal matching pursuit and dictionary learning. The dictionary consists of waveforms matched to a windowed portion of the stream, optimised to produce a sparse and accurate code, reducing the measurement data rate by an order of magnitude or more. This method was developed for anomaly detection, where part of the data is labelled as "normal," and the algorithm detects deviations, providing insights into structural integrity.

Data analysis can also be performed *at the transient level*. Transient signals must first be detected using wave selection methods from the literature [7, 15–18]. Interpretation follows and aims to address challenges such as damage localisation or clustering for damage classification. For example, Chen *et al.* [19] applied dynamic time warping to measure the similarity of AE signals for source localisation when multiple AE sources are present. Data from thermal-cracking experiments on granite were used to validate the approach. In [20], randomly selected signals are represented by auto-regressive hidden Markov models, subsequently applied to all transients of a dataset to obtain scores indicating goodness-of-fit, which are then used as inputs for clustering algorithms. This method was validated on datasets from carbon fibre composites.

One of the most frequently applied methods encountered in the literature relies *on three main steps* (Figure 1): 1) hit detection - where transient signals are detected based on signal processing tools - and feature extraction where the transient signals are

3

compressed and represented in a common high dimensional space for further analysis, 2) clustering or classification, which either groups features in an unsupervised manner or uses prior labels in a supervised approach to predict signal classes, and 3) evaluation and validation of the groups or classes of signals.
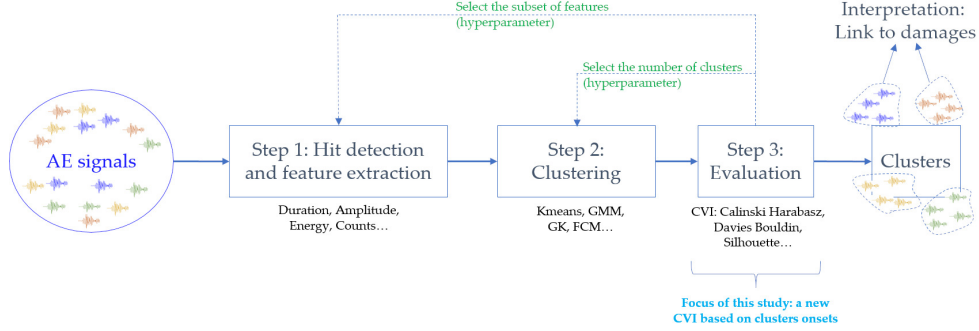
Fig. 1: Feature-based clustering of AE data.

This methodology has been applied, for example, to carbon fibre composite panels for aeronautics [8], high-performance carbon fibre-reinforced thermoset composites in extreme environments [21], bridge and tunnel health assessment [22, 23], railway systems [24], tribology [25, 26], and fatigue crack monitoring [27]. Deep learning approaches, which are supervised methods, follow similar steps but typically automate feature extraction through layers during training [28–30].

The final step, *validation, which is central to this work*, is more challenging for clustering than for supervised learning due to its subjectivity [31], especially given the lack of precise ground truth in the interpretation of AE data. Various criteria, known as *clustering validity indices* (CVI), have been proposed in the pattern recognition community [32–37] to determine whether clusters are meaningful.

For a given application, the challenge for unsupervised interpretation of AE data is the extraction of relevant features that can link the identified clusters to damage mechanisms. This involves answering the following key questions: *What are the most relevant features? How many clusters are needed?*

An answer to these questions is provided by an algorithm used in the literature detailed in appendix A and proposed by [38]. It uses several CVIs evaluated for different combinations of $k \cdot C_n^p$ (subsets of p features from n, considering k different cluster counts). A majority voting scheme, weighted by the user, is used to select the best number of clusters and subset of features.

The CVI used for AE data interpretation, such as the Davies-Bouldin index [39] or Silhouette index [40], *are exclusively focused on clusters' shape, characterised by their compactness and separability in the feature space.* One problem with these CVIs is that the chosen metric (such as Euclidean distance) may differ from those used in the clustering method's objective function (such as likelihood), potentially leading to misleading interpretations [21, 41]. Sawan *et al.* [41] showed that acoustic emission data

should be divided into more clusters than are expected from the number of active damage mechanisms to prevent misrepresentative grouping. Similarly, the results of several clustering methods were combined in [21] to more accurately map the feature space, resulting in a timeline that reflects a plausible damage scenario. Another issue with classical CVIs is that the timeline provided by compact and separated clusters is generally not related to the expected damage scenario, hence justifying novel approaches to achieve more relevant clustering performance [5, 12, 41–46]. These works show that damage can be described by a set of clusters, which may or may not correspond to true mechanisms, but the common factor is that one cluster should act as a precursor to others, following a progression.

In previous literature, clustering results are often presented as sequences after applying shape-based CVIs. However, current methods do not assure users that the results will be relevant in terms of the timeline. Therefore, the challenge remains: how can we select features and determine the number of clusters to guarantee a relevant timeline?

Our work aims to address this question by introducing a new validation criterion based on cluster onsets. The main features and advantages of this criterion are as follows:

- The proposed criterion is the first to formalise the ranking of cluster partitions based on how clusters emerge over time or in relation to a physical parameter. This approach provides a dynamic perspective on clustering behaviour, offering a new way to assess the evolution of clusters beyond static measures. The criterion is innovative and lacks a direct equivalent in the existing literature for selecting the number of clusters and the feature subsets that best explain the data in terms of a timeline.
- The proposed criterion is valuable for onset-based interpretation of AE signals particularly in cases where materials degrade progressively under stress. Three challenging experiments on real-world datasets demonstrate the relevance of the proposed approach. One of the applications is a benchmark on the detection of the loosening of bolted plates under vibration where the proposed criterion outperforms existing methods. It provides superior results, both quantitatively (in terms of accuracy) and qualitatively (in terms of the positioning of estimated onsets and cumulative occurrences).

To illustrate the positioning of our contribution, Figure 2 highlights the fundamental *difference between shape-based and onset-based approaches* for evaluating the results of AE data clustering. In the shaped-based approach (Figure 2a), clusters may appear compact and well-separated in a low-dimensional feature space. However, this does not guarantee that they will correspond to relevant onsets. For example several of them can start very early and almost simultaneously when considering the horizontal axis. In contrast, the latter approach focuses on identifying feature subsets that highlight onsets distributed along the horizontal axis (such as load or time). This is more helpful for interpreting the dynamic behaviour of the material or structure under test, as illustrated in Figure 2b.

5
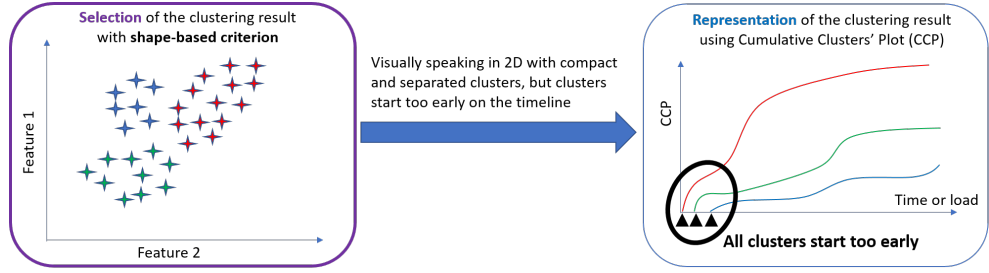
(a) "Selection" (left) is based on clusters shape in the feature space in low-dimensional space for visualization purpose. "Representation" (right) of clusters of the timeline provides onsets which start too early.

(b) "Selection" (right) is based on clusters onsets. The timeline is relevant for monitoring. The "representation" (left) in low-dimension (2D in general) of the clusters can be useless because the clusters are compact and separated only in higher dimensions (used in clustering). The compactness and separation of clusters is ensured during clustering.

**Fig. 2**: Selection and representation of clusters results.

Section 2 introduces the proposed criterion and presents an algorithm for the implemented method, incorporating the proposed criterion for selecting the best parametrisation and the resulting optimal cluster partition. The relevance of the criterion is evaluated in Section 3 using a benchmark obtained from a real structure.

2 Proposed criterion: Sorting subsets of features with respect to onsets

The AE data consist of N d-dimensional feature vectors $\boldsymbol{x}_n \in \Re^d$. Given a subset of features $S \subset \Re^d$ and a number of clusters K, a partition P is obtained by applying a clustering method. P is represented by a vector of length N, where each element $P(n) \in \{1, 2 \ldots K\}$ represents the cluster number assigned to each AE signal. For fuzzy clustering methods, $P(n)$ is the cluster with the largest membership degree.

A cluster's onset is defined as the first occurrence of the cluster. For example, for cluster k, the onset is the first index in P where the k-th cluster appears. Onsets can be represented on an axis related to time, load, or other physical quantities related to

the degradation of the material. Illustrations can be found on the right-hand side of Figure 2 and on Figure 3.
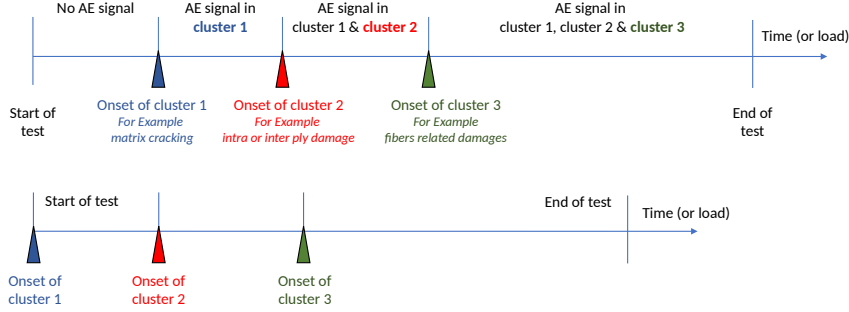
Fig. 3: Illustration used in examples in Section 2: Onsets (triangles) are defined as the first occurrence of each cluster. The horizontal axis can be any monotonically increasing information regarding the evolution of the test like, for example, time or a damage variable. In-between onsets, several clusters, that previously occurred, can keep growing. Reference to composite damages is only for illustration. Top figure: No signal appears from the beginning to the first onset so this part is removed and the starting time shifted accordingly, which leads to the bottom figure.

Let us define $t_k \in [0, T]$ as the first occurrence of cluster k in a dataset. In order to compute the criterion, the onsets are delayed so that the first one starts at $t_1 = 0$ (Figure 3). Then, we compute the normalised difference between two consecutive onsets as

$$\Delta_k = \frac{t_{k+1} - t_k}{T} \quad t_{K+1} = T \tag{1}$$

The set $\mathbf{p}_{\text{est}} = \{\Delta_k\}_{k=1}^{K}$, estimated by clustering, can be interpreted as a probability distribution made of K elements since we have

$$\sum_{k=1}^{K} \Delta_k = 1 \tag{2}$$

Example 1. *As an example, consider Figure 3. We have three clusters, with onsets positioned at $t_1 = 27, t_2 = 57, t_3 = 82$, and with $T = 127$ (horizontal axis is with arbitrary unit; in practice it can be related to time, load, or a physical quantity). We can shift the onsets to $t_1 = 0, t_2 = 30, t_3 = 55$ and $T = 100$. Then $\Delta_1 = 30/T$, $\Delta_2 = 25/T$, $\Delta_3 = 45/T$, satisfying Eq. 2. In that case, $\mathbf{p}_{est} = [0.30, 0.25, 0.45]$.*

2.1 Case 1: A prior on onsets is available

Let's first assume that the end-user can provide the expected onsets (composed of K values). These prior onsets can, for example, be provided by a physics-based model able to predict the onset of each damage or cluster. Prior onsets can be represented as

a distribution, say $\mathbf{p}_{\text{true}}$, following the same reasoning as before. The problem is now to compare $\mathbf{p}_{\text{true}}$ (the prior distribution) and $\mathbf{p}_{\text{est}}$ (the estimated distribution).

The degree of agreement between estimated onsets and the ground truth as quantified by $\mathbf{p}_{\text{est}}$ and $\mathbf{p}_{\text{true}}$ respectively, can be evaluated by different means, for example using the L_1 norm, the Hellinger distance, the Bhattacharyya distance, the total variation distance between probabilities, the Kullback-Leibler (KL) or Renyi divergences, among others. The KL divergence is used in the present work. KL is *not symmetric* and *never negative* and was used in many machine learning algorithms based on probability theory and on artificial intelligence. In information theory, KL is the amount of information lost when one distribution (for example, depending on parameters) is used to approximate another (for example, related to a ground truth). We compute the KL divergence as follows:

$$KL(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{true}}) = \sum_{k=1}^{K} \mathbf{p}_{\text{est}}(k) \log_2 \left(\frac{\mathbf{p}_{\text{est}}(k)}{\mathbf{p}_{\text{true}}(k)} \right) \tag{3}$$

In this expression, $\mathbf{p}_{\text{est}}$ is dependent on S, the subset of features used for clustering, and K the number of clusters. The KL thus comprises two terms. First, a cross entropy loss

$$CE(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{true}}) = -\sum_{k=1}^{K} \mathbf{p}_{\text{est}}(k) \log_2 \left(\mathbf{p}_{\text{true}}(k) \right) \tag{4}$$

and second, Shannon's entropy

$$E(\mathbf{p}_{\text{est}}) = -\sum_{k=1}^{K} \mathbf{p}_{\text{est}}(k) \log_2 \left(\mathbf{p}_{\text{est}}(k) \right) \tag{5}$$

with $KL = CE - E$.

2.2 Case 2: Without a prior on onsets

Even though some initial onsets could be obtained numerically (by simulations) or experimentally (for example using infrared thermography or strain sensors), it is more frequent, in practice, to be in the situation where there is no prior on onsets of clusters. Therefore, one approach to set $\mathbf{p}_{\text{true}}$ without prior is to make use of the *principle of maximum entropy*, since $\mathbf{p}_{\text{true}}$ is a distribution. This ensures that unconscious arbitrary assumptions are not made, and the introduction of biases extrinsic to the data is avoided [47, 48]. Therefore, *in the absence of a prior on onsets, among all possible distributions of onsets for $\mathbf{p}_{true}$, we should select the one with the largest entropy that corresponds to the uniform distribution.*

When using the uniform distribution of onsets for $\mathbf{p}_{\text{true}}$ in CE (Eq. 4), we obtain

$$\mathbf{p}_{\text{true}} = \text{Uniform} \Rightarrow CE(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{true}}) = -\log_2 \frac{1}{K} = \log_2 K \tag{6}$$

which means that KL tends to:

$$\mathbf{p}_{\text{true}} = \text{Uniform} \Rightarrow KL(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{true}}) = \log_2 K - E(\mathbf{p}_{\text{est}}) \tag{7}$$

We can simplify Eq. 7 by dividing by $\log_2 K$, leading to

$$\frac{1}{\log_2 K} KL(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{true}}) = 1 - E_{\text{norm}}(\mathbf{p}_{\text{est}}) \tag{8}$$

where $E_{\text{norm}} \in [0,1]$ is the normalised entropy of $\mathbf{p}_{\text{est}}$. Therefore, for a given K, minimising the initial criterion in Eq. 3 divided by $\log_2 K$ is equivalent to maximising the entropy of onsets. A possible criterion to sort subsets is thus:

$$Q_{\text{proposed}}^{\text{ONSETS}}(S, K): \quad \text{Given } K, \text{ find } \arg\max_S \left[E_{\text{norm}}\left(\mathbf{p}_{\text{est}}\left(\cdot | S, K\right)\right) \right] \tag{9}$$

where it is made explicit that E depends on K (the number of clusters) and S (subsets of features). Therefore, for a given K, the criterion is used to sort the subsets according to the distribution of the onsets.

Example 2. *Let us continue the previous example, with $T = 100$ and $K = 3$ we can define three onsets distributed uniformly in order to define $\mathbf{p}_{true}$, located at $t_1^{true} = 0$, $t_2^{true} = T/3 = 33.33$, $t_3^{true} = 2T/3 = 66.66$, with all $\Delta_k = 33.33/T$. Therefore $\mathbf{p}_{true} = \mathbf{p}_{uniform} = [0.33, 0.33, 0.33]$ (at two digits of precision, the sum must be one). Using $\mathbf{p}_{est}$ of the previous example, we get $KL = 0.0454714$ (Eq. 3) with $CE = 1.585$ and $E = 1.539$. The maximum of E for 3 clusters is $\log_2(3) = 1.585$ therefore $E_{norm} = 1.539/1.585 = 0.971$.*

Remark 1. *One important point to observe is that the proposed CVI only requires the partition, not the features, conversely to a shape-based CVI. Therefore, according to [34], the proposed criterion belongs to the family of "external" CVIs whereas shape-based CVIs are referred to as "internal". However, the proposed criterion is a special case because external CVIs actually require a complete reference partition to be compared with, meaning that the ground truth must be known for all feature vectors, whereas the proposed CVI only requires the onsets of clusters.*

Remark 2. *Our criterion can be used in an online setting for SHM applications, provided T is known. T represents, for example, the maximum number of cycles for fatigue testing, or the maximum crack size. Since in an online setting, K is assumed to evolve, we need to iteratively store the onsets. The list of onsets will be updated with a new value (appending a new element in a vector) when a new cluster is created. Then, Eq. 1 and 2 can be computed, followed by Eq. 9 (for example, without prior).*

2.3 Proposed algorithm

The criterion presented in the previous section is implemented within a broader procedure that incorporates the steps of the feature-based clustering method outlined in Figure 1. This procedure proceeds as follows.

First, a standard clustering method (in this case, K-means) is applied to the dataset, assuming a range for the number of clusters and possible feature subsets. Unlike other methods, for each partition an additional characteristic, called the cluster onset, is defined as the first occurrence of that cluster in the dataset. The criterion, given by Eq. 5 or Eq. 9, depending on whether prior information on the onsets is available, is then calculated for each partition. The best partitions are selected based on the maximisation of this criterion (*e.g.*, the top percentile of the values of $Q_{\text{proposed}}^{\text{ONSETS}}(S, K)$).

Finally, the onsets of each cluster in the selected partitions are extracted and compiled into a histogram representing the cumulative distribution of onsets across all selected partitions. Similar to interactive clustering [49], the end-user can analyse this histogram, and based on the positions and amplitudes of peaks, they can choose to either accept, reject, or further refine the results—potentially triggering a re-clustering process by using different features or alternative clustering methods.

An algorithm detailing this procedure is provided in Alg. 1.

The interpretation of the histograms can vary across applications and specificity of the dataset. The location of the peaks can be helpful for the end-user to conclude that some damages potentially start or grow from these locations. It can also lead to the conclusion that the extracted features or the clustering method are meaningful or not. Examples are provided in the following sections.

3 Experiments

Three applications have been considered. Table 1 presents the details of their parameterization, including the number of features per subset, the possible number of clusters, and the clustering method. For each application, the results of the proposed onset-based CVI (Alg.1) and the shape-based CVI [38] (Alg. 2) are presented.

3.1 Dataset #1: Monitoring of a bolted metallic structure

In this section, three state-of-the-art shape-based CVIs used in the AE community, namely the Silhouette, Davies-Bouldin, and Calinski-Harabasz CVIs, are used to sort the partitions with the voting scheme [38] and compared to the proposed onset-based CVI to provide a benchmark. It is important to note that these CVIs are adapted for hard clustering methods like K-means. Soft clustering techniques such as Gaussian mixture models (GMM) or fuzzy C-means (FCM), which assign probabilistic memberships, may not align well with these metrics. Therefore, K-means is used in both cases as well as because it is a widely adopted method in AE literature.

3.1.1 Setup and data description

The benchmark dataset ORION-AE is presented in [50]. In this dataset, the true labels are provided, which allowed us to evaluate the performance of the proposed criterion and compare it with the standard solution. The data were collected on a setup designed to reproduce the loosening phenomenon observed in aeronautics, automotive, or civil engineering structures where parts are assembled together by means of bolted joints (Figure 4).

Inputs:
    Matrix of features $\mathbf{Y}$ with size $N \times d$
    Range of number of clusters $[K_{\min}, K_{\max}]$
    Subsets of features $\mathcal{S}$
Outputs:
    Criterion
    Histogram of onsets

1 **for** $K \in [K_{min}, K_{max}]$ **do**
2   **for** $S \in \mathcal{S}$ **do**
3   *// Apply the clustering method and store the partition and parameters;*
4 $[P(S,K), \theta(S,K)] \leftarrow \mathsf{Clustering}(\mathbf{Y}(:,S),\ K)$;
5   *// Compute quality according to how onsets are distributed;*
6 $Q_{\mathrm{proposed}}^{\mathrm{ONSETS}}(S,K) \leftarrow \mathsf{Onsets\_Clusters\_Quality}(P)$;
7   **end**
8   *// Sort and select a few subsets of subsets of features for each K using Eq. 9. The threshold can be adapted according to the number of subsets considered;*
9 $q_{0.99} = \mathsf{quantile}$ with probability 0.99 of $Q_{\mathrm{proposed}}^{\mathrm{ONSETS}}(S,K)$;
10 $S_{\mathrm{sub}}(K) = \{S : Q_{\mathrm{proposed}}^{\mathrm{ONSETS}}(S,K) >= q_{0.99}\}$;
11   *// Find onsets of the selected partitions;*
12 $\mathrm{Onsets}(K) = \mathrm{Find\_onsets}(F(K))$;
13 **end**
14 *// Accumulate evidence from various numbers of clusters using histogram;*
15 *// The number of intervals in the histogram depends on the application;*
16 $\mathrm{Acc} = \mathrm{Histogram}\,(\mathrm{Onsets}(\forall K))$;
17 *// Present the histogram to the end-user who can then either reject, accept, or be ignorant of the result, triggering a re-clustering of the data [49].;*

Algorithm 1: Onsets-based algorithm using Eq. 9 by evidence accumulation.

The benchmark is composed of different campaigns of measurements. In the following, the set of measurements called "*measurementSeries_B*" (campaign #B) was used for illustration. The structure was submitted to a 120 Hz harmonic excitation force. Seven tightening levels were applied on the upper bolt, starting from 60 cNm (set with a torque screwdriver) and held during a 10-seconds vibration test. The shaker was stopped and this vibration test was repeated after a torque modification at 50 cNm. Torque modifications at 40, 30, 20, 10 and 5 cNm were then applied in this order. In each campaign, three sensors are available and we used the AE sensor named "F50A" for illustration. The raw data are illustrated in Figure 5. From the raw data (green), *the goal is to infer the tightening levels by clustering and to compare the results with the ground truth* (blue stairs).

The features are freely available at the following link [50]: https://drive.google.com/drive/folders/1H413RxYu4ya7YMEgF_lTh_fHr7flvvOO?usp=sharing. They are

Table 1: Applications used to illustrate the proposed criterion. GMM stands for Gaussian mixture model, GK for Gustafson-Kessel, EDM for electrical discharge machining.

Application	# sensors used	Total # features	# features per subset	# clusters	Ground truth	Clustering method
Bolted metallic structure	1	19	4	$\{2, 3, 4 \ldots 10\}$	Yes (7 classes)	K-means
EDM µ-drilling hard material	1	24	3	$\{4, 6, 8, 10\}$	No	GK
Glued hybrid tubular structure	4	19	3	$\{4, 6, 8, 10\}$	No	GK
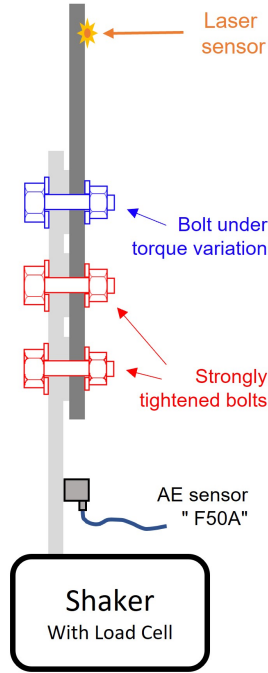
Fig. 4: Setup configuration for ORION-AE application.

common features used in the AE community [51]: Rise time, counts, counts to peak, duration, MARSE and absolute energies, amplitude, average signal level, signal strength, RMS, reverberation frequency, initiation frequency, average frequency, partial power in [0,200], [200,400], [400,600], [600,1000] kHz, peak frequency, and frequency centroid. The set of feature vectors was post-filtered using a 31-sample moving median applied to each dimension in order to ensure temporal coherence, and one point in three were kept in order to reduce the amount of points (there are about 1200 cycles per tightening level in the original dataset).
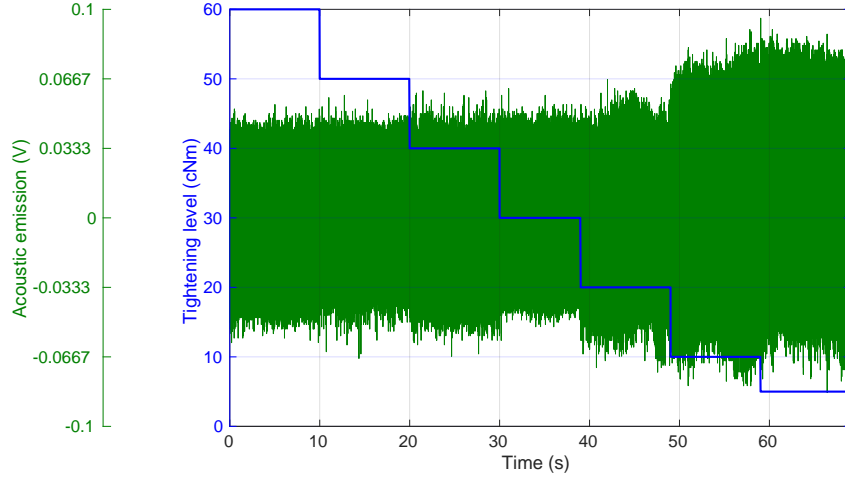
Fig. 5: Raw data (green) and tightening levels (in blue represented by stairs).

### 3.1.2 Parametrisation of algorithms

Algorithms 1 and 2 were used on every combination of 4 features within the set of 19 mentioned above, resulting in 3876 combinations for each $K = 3, 4 \ldots 10$ value with a total of 31008 combinations. K-means clustering was run 5 times for every combination and the result leading to the minimum sum of squared Euclidean distances over all points and clusters was selected. If an empty cluster was obtained, the corresponding subset of features was not used. The set of partitions obtained was used in both criteria. For Algorithm 2 (based on shape), the voting scheme proposed in [38] was applied with three internal CVIs implemented in Matlab's Statistical and Machine Learning Toolbox: Silhouette, Calinski-Harabasz, and Davies-Bouldin. Algorithm 2 was applied for each K independently in order to sort the subsets of features and select the best one. For Algorithm 1 (onset-based), the 20 best partitions were selected for each number of clusters using the onset-based CVI.

### 3.1.3 Link between clusters onsets and change in tightening levels

We evaluated whether the clusters' onsets coincide with the changes of tightening levels. For that, we represented the distribution of the estimated onsets using histograms. This was done in all partitions found by the previous algorithms and for all numbers of clusters ($K = 3 \ldots 10$). Histograms depict the ground truth with blue stairs and the vertical dotted lines represent the instant of the true changes in the tightening level. This ground truth is available in [50]. The stairs show the duration of each level, equal to $10.0, 10.0, 10.0, 9.0170, 10.0, 10.0,$ and 10.0 seconds for $60, 50, 40, 30, 20, 10,$ and 5 cNm, respectively.

Histograms of onsets for shape-based and onset-based CVIs are presented in Figure 6 with a bin size of 0.5 s. *These results only differ by the criterion since the same partitions obtained by the K-means method were used to evaluate both criteria in the same conditions.* With the shape-based CVI (Fig. 6a), we can observe three missing

onsets (no vertical bar at 10, 30, and 60 s corresponding to 50, 30, and 5 cNm), one delayed onset (20 s / 40 cNm) and one clear false positive (at around 44 s). *Based on this histogram, an end-user is likely to select 4 clusters, with one wrong and 3 missed tightening levels.*
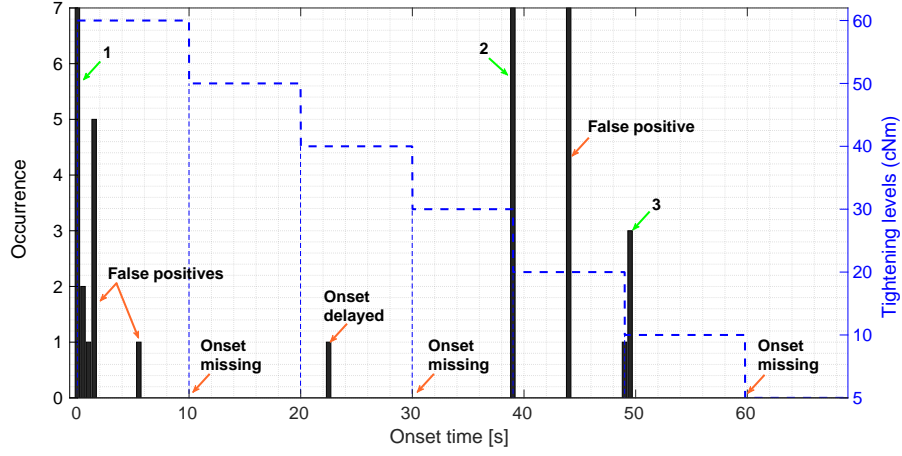
(a) Shape-based CVI with the best vote
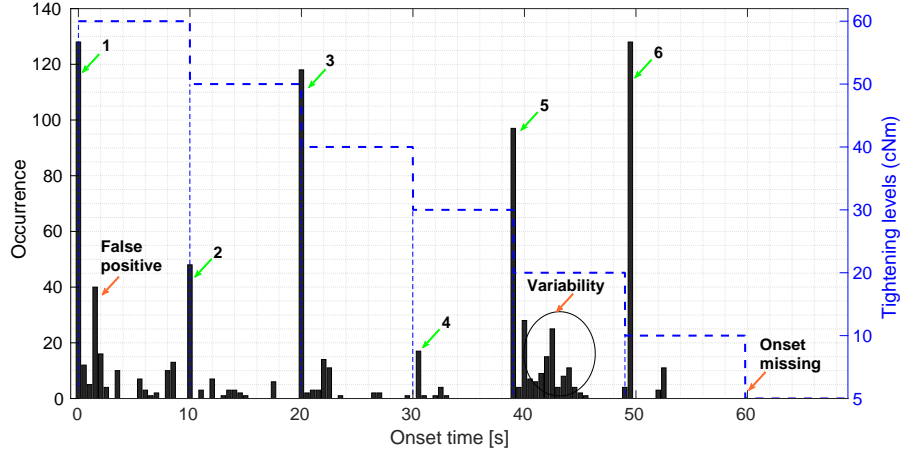
(b) Proposed onset-based CVI with 20 best partitions

Fig. 6: Validation of the clustering provided by a K-means using the voting scheme and onset-based criteria.

Figure 6b shows that all changes in tightening levels, except the last one, are detected by the onset-based CVI. Four primary peaks distinctly appear, located at expected times corresponding to changes in tightening levels at 60 cNm (0 s), 50 cNm (10 s), 40 cNm (20 s), 20 cNm (40 s), 10 cNm (50 s). *The same partitions were used*

14

in both cases which means that only the CVI makes the difference. The peak at 30 s is clearly distinguishable with 17 occurrences. The secondary peaks (lower occurrence) are globally distributed around the primary peaks and represent the variability on onsets in the 20 partitions.

High peaks mean that changing the number of clusters and features does not modify the location of these onsets. However, as demonstrated by Sawan *et al.* [41], we expect that increasing the number of clusters gives different insights, which is clearly not the case with the shape-based CVI. Moreover, we observe that the number of small peaks for the onset-based CVI is larger, explained by the fact that 20 times more partitions were used, adding variability to the results. We observe that the average level of these peaks is below 10, representing less than 8% of values of highest peaks. This variability highlights the difficulty in detecting the tightening level at 40 cNm, which may be represented by more than one cluster. The high number of small peaks at the highest tightening levels (left-hand side) can be attributed to the challenge in accurately quantifying bolts loosening from vibration data [52–54].

3.1.4 Link between clusters onsets and clusters overall quality

We have shown above that the onset-based CVI provides more relevant onsets. To further evaluate its performance, we assess whether the clusters generated coincide with the ground truth. For this, the 20 best partitions provided by each criterion were evaluated using the Rand index (RI) [55]. The RI is a value in the range $[0, 1]$, where "1" indicates a perfect match between the clustering assignments and the ground truth. The number of points in each partition is 4684, based on which the RI is computed. The results are presented in Figure 7 using boxplots whereby the red line represents the median value, the boxes indicate variability based on the 25th and 75th percentiles, and the outliers are marked by red crosses.

We observe that the median RI values are generally close, indicating that, on average, both criteria provide similar cluster assignments across 4684 points. Note that the onsets are obscured in these results: only the histograms shown earlier allow for an assessment of their precise positions.

Concerning the variability across partitions, it differs between the two criteria. When considering values corresponding to the ground truth ($K_{true} = 7$), the onset-based CVI demonstrates better consistency, with less variability in results compared to the shape-based CVI. This observation is confirmed in Table 2, which summarizes the RI values for $K = 6, 7, 8$. The minimum RI values show a 4% to 7% improvement for the onset-based CVI, while maximum values are comparable for both criteria. This suggests that the worst partitions identified by the onset-based criterion are more relevant than those produced by the shape-based CVI. Additionally, four outlier values below the minimum were found for the shape-based CVI (11 across all K values), representing 10% of the results. In contrast, no outliers were detected for the onset-based criterion at $K = 6, 7, 8$, and only three for all K values. These results demonstrate its superior robustness in response to feature changes.

(a) Partitions selected by shape-based CVI   (b) Partitions selected by onset-based CVI
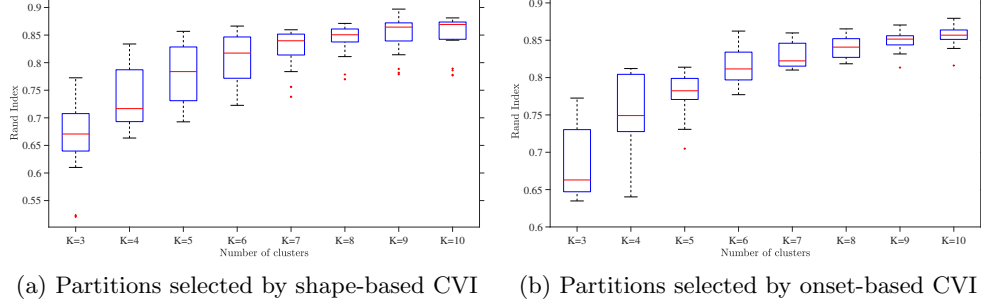
Fig. 7: Performance comparison evaluated by the Rand index

# cluster/perf	min		max		median		# outliers	
CVI	onset	shape	onset	shape	onset	shape	onset	shape
K=6	**77.7**	72.3	86.2	**86.6**	81.2	81.7	**0**	0
K=7	**81.0**	73.8	86.0	86.0	82.2	**83.9**	**0**	2
K=8	**81.8**	77.0	86.5	87.1	84.1	**85.9**	**0**	2

Table 2: Rand index for onset-based and shape-based CVI with K=6,7,8 (±1 around the ground truth).

3.1.5 Case of a dataset with non-uniformly distributed onsets

In the ORION-AE dataset, the tightening levels have approximately the same duration, which is a special case. In this section, we propose to evaluate the proposed criterion when the duration in each tightening level differs. The dataset becomes imbalanced. For that, we selected only a part of the data in each level so that the duration of the levels follows a non linear evolution. Practically, it means that the metallic plates remain tightened for a longer period for a given level compared to the subsequent levels. For 60, 50, 40, 30, 20, 10 and 5 cNm, the first seconds of data were kept as follows: $10\,s$, $8.5\,s$, $7\,s$, $5\,s$, $3\,s$, $2\,s$ and $1.5\,s$, and the remaining data were discarded.

The results are reported in Figure 8. We can observe that the proposed onset-based CVI outperforms the standard approach by selecting relevant partitions for which onsets are located as expected. As in the previous case, only the last onset is not discovered by the onset-based CVI. Conversely, the shape-based CVI only found three clear onsets. For 50 cNm (10 s) and 40 cNm (20 s) only one occurrence is found, which is not enough to distinguish these onsets from the false positives.

The difference in behaviour of both criteria between this test (irregular timing of damage mechanisms) and the previous one (regular timing of damage mechanisms) is the size of the clusters. Since we shortened the duration of each tightening level, there are fewer points in the classes. Consequently, the distribution of the features within the classes has changed, making the clustering task more difficult. Some classes becomes smaller relatively to others due to this change of distribution which affects both the intra-cluster compactness and inter-cluster separability, therefore degrading the values of the shape-based CVIs.
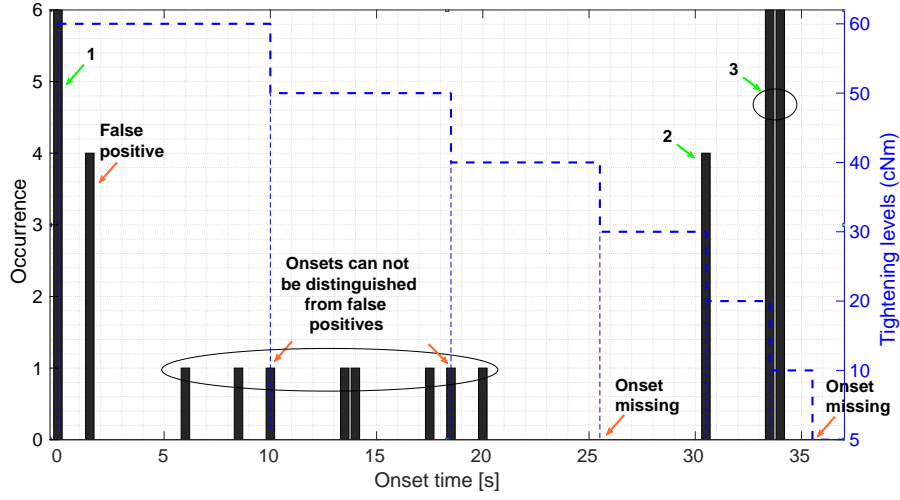
(a) Shape-based CVI with the best vote
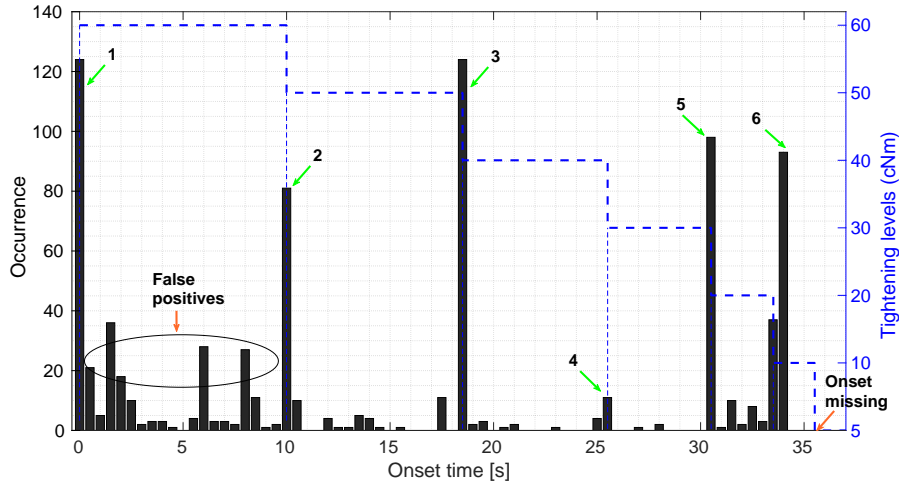
(b) Proposed onset-based CVI with 20 best partitions

Fig. 8: Case of non-uniform onsets: Validation of the clustering provided by a K-means using the voting scheme and onset-based criteria.

3.1.6 Summary

When averaged over 20 partitions of ˜4600 points, we observed that the quality of clusters found by the onset-based criteria is improved by 4% to 7%. Using histograms, a distinct improvement was observed on the positioning of onsets with the proposed CVI, which makes this criterion suitable for structural health monitoring.

3.2 Dataset #2: Micro-drilling by electrical discharge machining

This application is about monitoring of a manufacturing process known as micro-electrical discharge machining (micro-EDM) used for drilling micro-diameter and deep holes. The goal of AE clustering is to gain insights about the process, and to evaluate if it is possible to detect when the electrode has reached the end of drilling and if the process proceeds similarly between two drilling events.

3.2.1 Materials and method

Micro-EDM relies on electrical discharges emitted by a tool-electrode to remove material from a work-piece immersed in a dielectric fluid (Fig. 9a). An electric arc forms between the conducting part and the electrode, which are not in physical contact, enabling the machining of very hard conductive materials with a high degree of flexibility in terms of shape [56]. Both the feed rate of the tool-electrode and the sequence of discharges, which are necessary to remove material and achieve a specified depth, are controlled by proprietary software. During this process, both the work-piece and the tool-electrode experience wear, which is challenging and time-consuming to evaluate precisely and in real-time. Therefore, we used AE to collect data during drilling to gain insights into the process. Specifically, the goal is to assess whether the chronology of AE signals can aid in monitoring the progression of the electrode to the reference depth.

Experiments were conducted on a SARIX EDM machine, which features auto-regulation of input parameters for spark generation. A 0.3 mm diameter tubular electrode was used to drill a 2 mm hole into a 0.3 mm thick cavity. The material being drilled is a Ni-based superalloy. Figure 9b illustrates the setup on the EDM machine. The camera, synchronised with the start of the process, records the moment when the electrode begins drilling. It is then used to visually assess when a spark appears in the cavity, corresponding to the hole opening (third image in Figure 9a). This event is marked in the figures below as we anticipate specific behaviour in the AE signals and clusters. AE data streams were acquired at a sampling frequency of 5 MHz using a "micro-80" sensor made by Mistras-EuroPhysical Acoustics, with an operating frequency range of 200-900 kHz. Data acquisition was performed using a Picoscope from Picotech. The AE hits were detected using the same method as in application #1 (ORION-AE) and similar features, except for the partial powers, which were considered in the following intervals: [0 20], [20 100], [100 200], [200 300], [300 400], [400 500], [500 600], [600 800], and [800 1000] kHz, resulting in a total of 24 features.

3.2.2 Results

The tests described below were conducted using a brand new electrode and after cleaning the machine. Two tests were performed successively. Algorithm 1, which is based on onsets, was applied to every combination of 3 features from a set of 24 (totalling 2024 combinations). The 20 best partitions for each number of clusters $K \in \{4, 6, 8, 10\}$ were used to generate the histograms.

(a) Principle of EDM drilling.

(b) Illustration of the AE acquisition set-up.
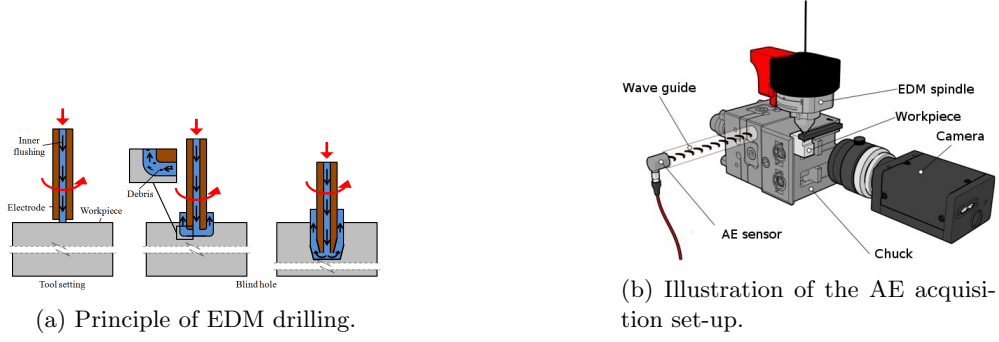
Fig. 9: EDM principle and instrumentation [56].

Figure 10 presents histograms for Test #1 (with a new electrode) and Test #2 (after one use), showing peaks with more than 10 occurrences. Due to the continuous wear of the electrode during machining, its position was estimated using a camera to determine when the first spark appeared at the hole's opening. The time instances identified are: 1) Test #1: First spark at 14-15 s; drilling from 17-18 s; Test #2: First spark at 18-19 s; drilling from 22-23 s. Cumulated energy is superimposed on the histograms, as this feature traditionally indicates damage evolution trends.

For Test #1 (Fig. 10a), four main stages are identified by the dashed lines, based on cumulated energy and onset positions. *Stage I (0-6 s)* is characterised by three primary peaks with low energy, related to initial spark appearances. Drilling progresses with a few sparkles until a hole is made. *Stage II (6-15 s)* shows significant increase in cumulated energy in two phases (6-12 s, 12-15 s), with a primary onset around 12 s indicating the machine's adaptation to drilling. *Stage III (15-17.6 s)* depicts two primary onsets (5 and 6) and corresponds to the first spark observed by the camera, marking the approach to the first part of the work-piece. *Stage IV (18 s onwards)* shows one more onset (7) indicating the start of drilling the second part, with similar characteristics to the first part. For Test #2 (Fig. 10b), Stage IV is shorter, and the process is more stable, starting at about 21.5 s. The cumulated energy increases more regularly, indicating a more stable drilling process. The rectangle highlights the first spark (around 18.2 s), with onsets before (7) and after (8) the observed spark. The end of this phase (just before starting drilling the next part) is marked by onset (9).

The innovative aspects of these results are supported by patented technology from [57], which underscores the novelty and significance of this approach for industrial applications.

## 3.3 Dataset #3: Hybrid composite/metal joint structure

This application is about monitoring of damages in a hybrid composite/metal joint structure. The proposed criterion is used to evaluate when damage mechanisms occur and how they grow during loading until the failure of the composite/metal joint.

3.3.1 Materials and method

The specimen studied is a composite tube manufactured by Collins Aerospace, featuring a joint between a composite part and a metal part. The composite section consists
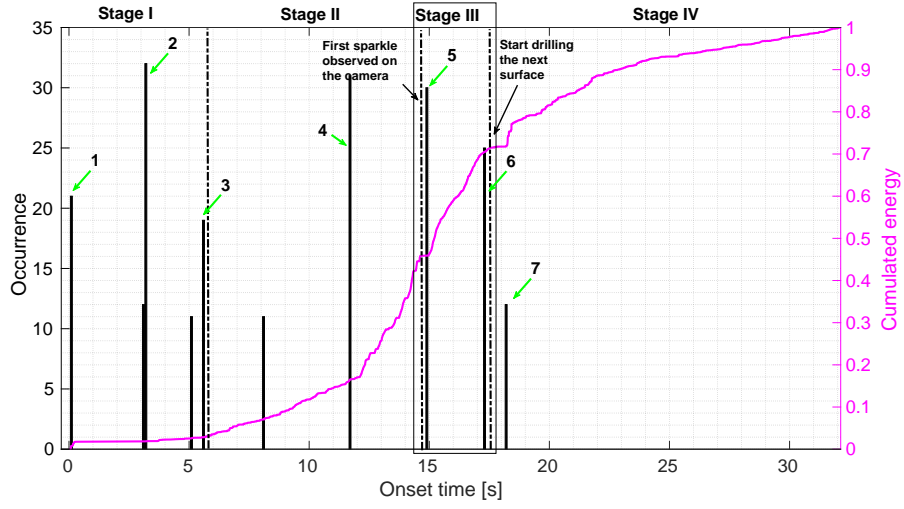
(a) Test #1.
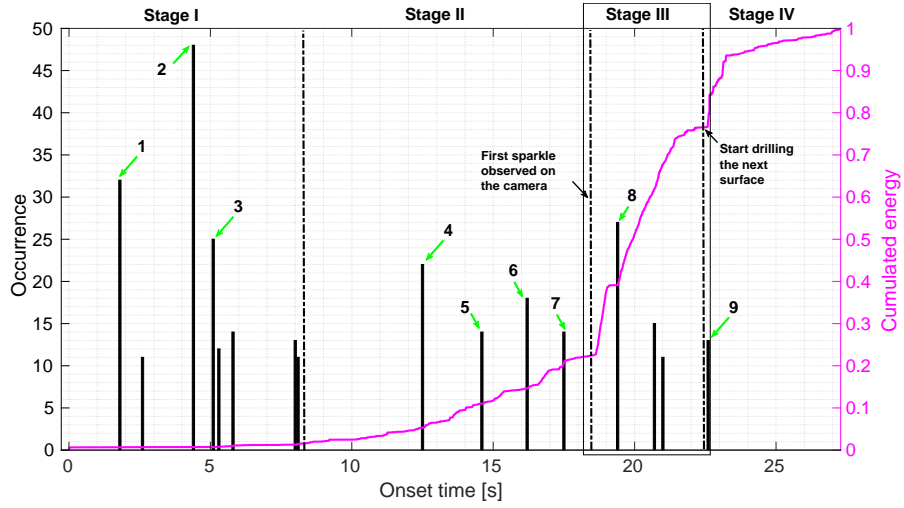
(b) Test #2.

Fig. 10: Histogram of onsets for the two tests.

of eight plies of filament-wound carbon fibre tows, forming a tube with a nominal internal diameter of 45.2 mm and a wall thickness of 4 mm. This geometry was chosen to handle loads typical of aerospace applications, where certifying bonded joints can be challenging.

The joint technology integrates a bespoke helical screw thread, where an epoxy-based adhesive is applied to the metal thread for lubrication during assembly. Preloading the joint applies through-thickness contact pressure to the composite, and

20

loads are transmitted in two ways: (i) through reaction forces at multiple teeth and (ii) through friction at flat regions within the helical thread. If the frictional forces are exceeded, relative movement between the composite and metal parts can occur. Therefore, it is crucial to consider both slip and matrix damage in the design. The test was designed to ensure failure occurs at the interface regions rather than within the composite tube or metal end fitting.

The specimen was tested in a quasi-static tensile loading configuration on an Instron 5989 electro-mechanical testing machine fitted with a 600 kN load cell. It comprises a composite tube and two metallic end fittings that form two joints, connected to the loading rig via rod ends with spherical bearings, which are screwed into the metallic end fittings. These rod ends are pinned to clevises that are screwed into the rig. The described test set-up is shown in Figure 11-a).
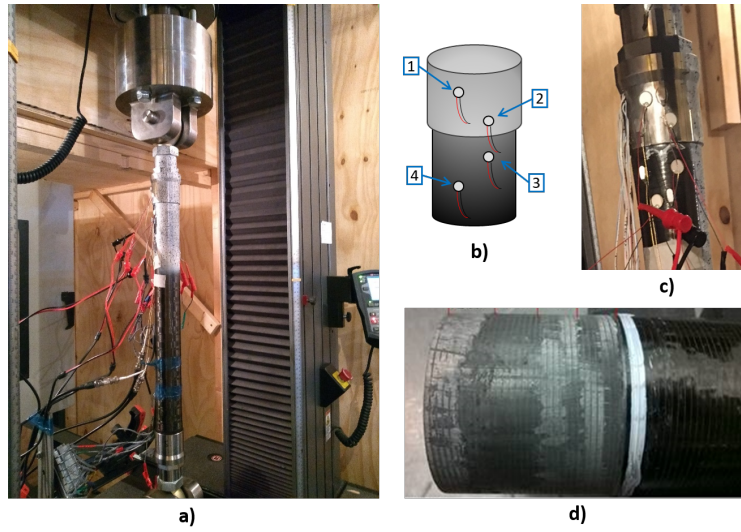
Fig. 11: a) Quasi-static tensile loading of the hybrid composite/metal specimens. The specimen is connected to the loading rig at either end via a rod end with spherical bearings that are screwed into the metallic end fittings of the joint. b) and c) Schematic and photograph showing the arrangement of PWAS on both specimens. d) Photograph illustrating final failure of a hybrid composite/metal joint in the form of resin cracking and extensive slippage due to thread crushing. The joint was subjected to quasi-static tensile loading.

The specimen was loaded at 0.5 mm/min crosshead speed in a single cycle until failure. The specimen failed at the instrumented joint. The breaking load was 268 kN. The final failure consisted of significant damage caused by initial matrix cracking and followed by slippage at the composite/metal interface of one of the two joints (Figure 11-d) where the screw threads have been almost erased during failure.

Piezoelectric wafer active sensors (PWAS) [2] were bonded to the metal and composite parts of the joint for AE monitoring when the specimen is subjected to quasi-static tensile loading. PWAS of type PIC255, supplied by PI Ceramic-Germany, with 10 mm diameter and 0.5 mm thickness – were used. The positioning of PWAS is shown in Figure 11-b) and c).

AE data were recorded by a PCI-2 based system, and captured in the software "AEWin" (Mistras, USA), with a sampling rate of 10 MHz and 20 dB of pre-amplification per sensor. Hit detection was performed by the software using a threshold of 55 dB and HDT=800 µs, PDT=200 µs and HDT=1 ms. In the following, only the data from sensors on the composite part (number 3, 4 on Fig. 11b)) were used for the analysis. The same 19 features as in the first application were used.

3.3.2 Results

The objective of applying the onset-based CVI in this context is to assess its capability to reveal phenomena in a complex engineering structure and to detect incipient damages within the joint. Data-driven methods like clustering can uncover patterns in large datasets, which is especially useful when prior knowledge is uncertain, as with the metal/composite joint scenario. This approach is beneficial when assigning labels is challenging and supervised algorithms are not feasible.

The Gustafson-Kessel clustering method [58] was employed in Algorithm 1, considering all combinations of 3 features out of 19 (totalling 969 combinations) and varying the number of clusters $K \in \{4, 6, 8, 10\}$, resulting in 3876 partitions. The 20 best partitions for each K were used to construct the histograms shown in Figure 12a. Due to the lack of ground truth for this dataset, we compared the results with the cumulated energy from the previous application and provided the distribution of amplitudes in clusters for $K = 10$ (Fig. 12b).

The analysis reveals three distinct stages before failure. Each stage shows a unique evolution of cumulated energy and amplitude distributions. In the first two stages (0-275 s and 265-465 s), the cumulated energy exhibits two plateaus, indicating lower energy AE signals compared to the later stage III. This is confirmed by Fig. 12b, where clusters in stage I have the lowest amplitudes (55-75 dB), stage II shows middle-range amplitudes (75-100 dB), and stage III has the highest amplitudes (>100 dB).

The histogram of onsets (Fig. 12a) shows a primary peak at 270 s, closely aligning with the initial increase in cumulated energy, marking a significant release of AE signals. This peak corresponds to primary peaks #4 and #5 at 265 s and 305 s, indicating incipient damage. During the plateau phase (275-465 s), fewer onsets are observed, with a significant peak (#7) at 465 s, aligning with the start of increased cumulated energy and subsequent damage cascade.

Simulation results using Abaqus [5, Chap. 7] show slip accumulation between 90-180 kN, followed by fatigue until failure at 280 kN (compared to 262 kN experimentally). This correlates with Stage II in the AE data, involving clusters #4-8, suggesting that damages are represented by multiple clusters. The high number of onsets before 265 s (90 kN) and the low cumulated energy during this period may indicate friction, while the critical load accumulation at 465 s (160 kN) aligns with the fatigue period
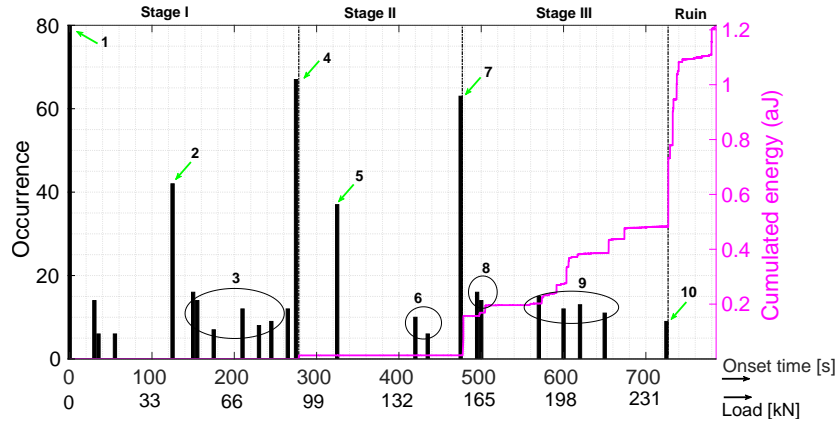
(a) Histogram of onsets over time and load.
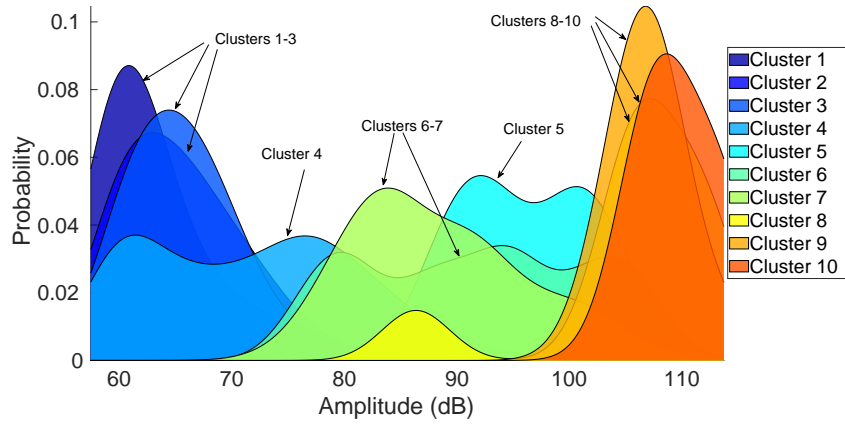
(b) Distribution of amplitudes in clusters.

Fig. 12: Analysis of clustering results for the third application.

found in simulations (Stage III). Clusters around 600 s (78% of test duration, 205 kN) represent incipient critical damages leading to final failure detected with cluster #10.

4 Conclusion and future work

This study introduces a novel approach for interpreting acoustic emission (AE) data through a new clustering validation index (CVI) that emphasises the distribution of cluster onsets over time, cycles, or load scales. Unlike traditional methods that prioritise compact and separable clusters, our approach focuses on onsets, which are crucial

for identifying different defects, such as mechanical loosening, process stages, or composite damage. This shift in focus is particularly valuable in monitoring applications where the initiation of defects is a key characteristic.

The proposed onset-based CVI demonstrated superior performance over conventional methods in a benchmark dataset, accurately detecting levels of loosening in bolted plates and reflecting changes in loosening with high precision. This criterion's ability to select optimal clustering results without prior knowledge, and to incorporate prior knowledge when available, underscores its uniqueness and applicability.

In the application to micro-drilling using electrical discharge machining, the onset-based CVI highlighted distinct process stages, offering potential for enhanced quality control through AE feedback. Additionally, in the analysis of hybrid composite/metal joints, the criterion facilitated the identification of damage progression leading to failure, which is crucial for structural health monitoring.

This work emphasises the need for specialised tools in AE data analysis to better interpret material and system behaviour. The onset-based CVI proves to be a valuable tool, as evidenced by its effectiveness across different applications. Future research will explore integrating this criterion into supervised learning algorithms for Structural Health Monitoring, aiming to detect incipient defects and trigger prognostics algorithms. Additionally, we plan to develop damage models that generate onsets and utilise this prior knowledge to enhance clustering and classification with physically-informed algorithms.

Data and code

ORION-AE dataset used in application #1 is available at https://doi.org/10.7910/DVN/FBRDU0. Codes will be made available at https://emmanuelramasso.github.io/.

References

[1] Farrar, C.R., Worden, K.: Structural Health Monitoring: A Machine Learning Perspective. John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom (2013). https://doi.org/10.1002/9781118443118

[2] Giurgiutiu, V.: Structural Health Monitoring: with Piezoelectric Wafer Active Sensors. Academic Press, Oxford (2015). https://doi.org/10.1016/C2012-0-07213-4

[3] Manuello Bertetto, A., Masera, D., Carpinteri, A.: Acoustic emission monitoring of the turin cathedral bell tower: Foreshock and aftershock discrimination. Applied Sciences **10**(11), 3931 (2020) https://doi.org/10.3390/app10113931

[4] Pinho, R., Scattarreggia, N., Orgnoni, A., Lenzo, S., Grecchi, G., Moratti, M., Calvi, G.M.: Forensic estimation of the residual capacity and imposed demand on a ruptured concrete bridge stay at the time of collapse. In: Structures, vol. 55, pp. 1595–1606 (2023). https://doi.org/10.1016/j.istruc.2023.05.114 . Elsevier

[5] Chandarana, N.: Combining passive and active methods for damage mode diagnosis in tubular composites. PhD thesis, Manchester University, Faculty of Science and Engineering, Departement of Materials (18th November 2019)

[6] Zhang, L., Tang, Y., Zeng, J., Li, H., Liu, Q., Zhang, T.: Acoustic emission signatures for quantifying damage patterns in half grouted sleeve connection under tensile load. Construction and Building Materials **430**, 136452 (2024) https://doi.org/10.1016/j.conbuildmat.2024.136452

[7] Holford, K.M., Eaton, M.J., Hensman, J.J., Pullin, R., Evans, S.L., Dervilis, N., Worden, K.: A new methodology for automating acoustic emission detection of metallic fatigue fractures in highly demanding aerospace environments: An overview. Progress in Aerospace Sciences **90**, 1–11 (2017) https://doi.org/10.1016/j.paerosci.2016.11.003

[8] Al-Jumaili, S.K., Holford, K.M., Eaton, M.J., McCrory, J.P., Pearson, M.R., Pullin, R.: Classification of acoustic emission data from buckling test of carbon fibre panel using unsupervised clustering techniques. Structural Health Monitoring **14**(3), 241–251 (2015) https://doi.org/10.1177/1475921714564640

[9] Melchiorre, J., Manuello Bertetto, A., Rosso, M.M., Marano, G.C.: Acoustic emission and artificial intelligence procedure for crack source localization. Sensors **23**(2), 693 (2023) https://doi.org/10.3390/s23020693

[10] Pomponi, E., Vinogradov, A.: A real-time approach to acoustic emission clustering. Mechanical Systems and Signal Processing **40**(2), 791–804 (2013) https://doi.org/10.1016/j.ymssp.2013.03.017

[11] Placet, V., Ramasso, E., Boubakar, L., Zerhouni, N.: Online segmentation of acoustic emission data streams for detection of damages in composite structures in unconstrained environments. In: Deodatis, G., Ellingwood, B.R., Frangopol, D.M. (eds.) Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, pp. 525–532. CRC Press, London (2014). https://doi.org/10.1201/b16387

[12] Lindley, C., Jones, M., Rogers, T., Cross, E., Dwyer-Joyce, R., Dervilis, N., Worden, K.: A probabilistic approach for acoustic emission based monitoring techniques: With application to structural health monitoring. Mechanical Systems and Signal Processing **208**, 110958 (2024) https://doi.org/10.1016/j.ymssp.2023.110958

[13] Martin-del-Campo, S., Schnabel, S., Sandin, F., Marklund, P.: Detection of particle contaminants in rolling element bearings with unsupervised acoustic emission feature learning. Tribology International **132**, 30–38 (2019)

[14] Martin-del-Campo, S., Sandin, F.: Online feature learning for condition monitoring of rotating machinery. Engineering Applications of Artificial Intelligence **64**, 187–196 (2017)

[15] Kurz, J.H., Grosse, C.U., Reinhardt, H.-W.: Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasound signals in concrete. Ultrasonics **43**(7), 538–546 (2005) https://doi.org/10.1016/j.ultras.2004.12.005

[16] Pomponi, E., Vinogradov, A., Danyuk, A.: Wavelet based approach to signal activity detection and phase picking: Application to acoustic emission. Signal Processing **115**, 110–119 (2015) https://doi.org/10.1016/j.sigpro.2015.03.016

[17] Madarshahian, R., Ziehl, P., Caicedo, J.M.: Acoustic emission bayesian source location: Onset time challenge. Mechanical Systems and Signal Processing **123**, 483–495 (2019) https://doi.org/10.1016/j.ymssp.2019.01.021

[18] Quy, T.B., Kim, J.-M.: Crack detection and localization in a fluid pipeline based on acoustic emission signals. Mechanical Systems and Signal Processing **150**, 107254 (2021) https://doi.org/10.1016/j.ymssp.2020.107254

[19] Chen, S., Yang, C., Wang, G., Liu, W.: Similarity assessment of acoustic emission signals and its application in source localization. Ultrasonics **75**, 36–45 (2017) https://doi.org/10.1016/j.ultras.2016.11.005

[20] Ramasso, E., Butaud, P., Jeannin, T., Sarasini, F., Placet, V., Godin, N., Tirillo, J., Gabrion, X.: Learning the representation of raw acoustic emission signals by direct generative modelling and its use in chronology-based clusters identification. Engineering Applications of Artificial Intelligence **90**, 103478 (2020) https://doi.org/10.1016/j.engappai.2020.103478

[21] Ramasso, E., Placet, V., Boubakar, M.L.: Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites. IEEE Trans. on Instr. and Meas. **64**(12), 3297–3307 (2015) https://doi.org/10.1109/TIM.2015.2450354

[22] Marasco, G., Chiaia, B., Ventura, G., *et al.*: Ai based bridge health assessment. In: 9th International Workshop on Reliable Engineering Computing (REC 2021) Is "Risk and Uncertainty in Engineering Computations (2021)

[23] Li, D., Chen, Q., Wang, H., Shen, P., Li, Z., He, W.: Deep learning-based acoustic emission data clustering for crack evaluation of welded joints in field bridges. Automation in Construction **165**, 105540 (2024) https://doi.org/10.1016/j.autcon.2024.105540

[24] Carboni, M., Crivelli, D.: An acoustic emission based structural health monitoring approach to damage development in solid railway axles. International Journal of Fatigue **139**, 105753 (2020) https://doi.org/10.1016/j.ijfatigue.2020.105753

[25] Olorunlambe, K., Eckold, D., Shepherd, D., Dearn, K.: Bio-tribo-acoustic emissions: Condition monitoring of a simulated joint articulation. Biotribology **32**, 100217 (2022) https://doi.org/10.1016/j.biotri.2022.100217

[26] Tervo, J., Junttila, J., Lämsä, V., Savolainen, M., Ronkainen, H.: Hybrid methodology development for lubrication regimes identification based on measurements, simulation, and data clustering. Tribology International **195**, 109631 (2024) https://doi.org/10.1016/j.triboint.2024.109631

[27] Karimian, S.F., Modarres, M., Bruck, H.A.: A new method for detecting fatigue crack initiation in aluminum alloy using acoustic emission waveform information entropy. Engineering fracture mechanics **223**, 106771 (2020) https://doi.org/10.1016/j.engfracmech.2019.106771

[28] Du, J., Zeng, J., Wang, H., Ding, H., Wang, H., Bi, Y.: Using acoustic emission technique for structural health monitoring of laminate composite: A novel cnn-lstm framework. Engineering Fracture Mechanics, 110447 (2024) https://doi.org/10.1016/j.engfracmech.2024.110447

[29] Hesser, D.F., Mostafavi, S., Kocur, G.K., Markert, B.: Identification of acoustic emission sources for structural health monitoring applications based on convolutional neural networks and deep transfer learning. Neurocomputing **453**, 1–12 (2021) https://doi.org/10.1016/j.neucom.2021.04.108

[30] Ramasso, E., Teloli, R.d.O., Marcel, R.: On the condition monitoring of bolted joints through acoustic emission and deep transfer learning: generalization, ordinal loss, and super-convergence. Structural Health Monitoring, 1–23 (2024) https://doi.org/10.1177/14759217241259668

[31] Carmichael, J., Julius, R.: Finding natural clusters. Systematic Biology **17**(2), 144–150 (1968) https://doi.org/10.1093/sysbio/17.2.144

[32] Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. Pattern recognition **46**(1), 243–256 (2013) https://doi.org/10.1016/j.patcog.2012.07.021

[33] Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on pattern analysis and machine intelligence **24**(12), 1650–1654 (2002) https://doi.org/10.1109/TPAMI.2002.1114856

[34] Vendramin, L., Campello, R.J.C.B., Hruschka, E.R.: Relative clustering validity criteria: A comparative overview. Statistical Analysis and Data Mining **3**(4), 209–235 (2010) https://doi.org/10.1002/sam.10080

[35] Vendramin, L., Jaskowiak, P.A., Campello, R.J.: On the combination of relative clustering validity criteria. In: Proceedings of the 25th Int. Conf. on Scientific and Statistical Database Management, p. 4 (2013). ACM

[36] Bezdek, J.C., Pal, N.R.: Some new indexes of cluster validity. IEEE Trans. on Syst., Man, and Cybernetics, Part B (Cybernetics) **28**(3), 301–315 (1998) https://doi.org/10.1109/3477.678624

[37] Bayá, A.E., Larese, M.G.: Dstab: estimating clustering quality by distance stability. Pattern Analysis and Applications, 1–17 (2023) https://doi.org/10.1007/s10044-023-01175-7

[38] Sause, M., Gribov, A., Unwin, A., Horn, S.: Pattern recognition approach to identify natural clusters of acoustic emission signals. Pattern Reco. Lett. **33**, 17–23 (2012) https://doi.org/10.1016/j.patrec.2011.09.018

[39] Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence (2), 224–227 (1979) https://doi.org/10.1109/TPAMI.1979.4766909

[40] Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics **20**, 53–65 (1987) https://doi.org/10.1016/0377-0427(87)90125-7

[41] Sawan, H.A., Walter, M.E., Marquette, B.: Unsupervised learning for classification of acoustic emission events from tensile and bending experiments with open-hole carbon fiber composite samples. Composites Science and Technology **107**, 89–97 (2015) https://doi.org/10.1016/j.compscitech.2014.12.003

[42] Ramasso, E., Doan, D.D., Kharrat, M., Placet, V., Boubakar, L.: Reconnaissance des sources acoustiques dans les composites à matrice organique: quel (s)

critère (s) utiliser pour une classification non-supervisée des signaux? In: Congrès Français d'Acoustique (CFA 2014), Poitiers, France (2014)

[43] Chandarana, N., Ramasso, E., Wu, Z., Bernard, J., Pethick, J., Chatzi, P., Soutis, C., Gresil, M.: Damage identification in a tubular composite/metal joint through chronology-based robust clustering of acoustic emissions. In: European Workshop on Structural Health Monitoring Series, Manchester, UK (2018)

[44] Hohl, A., Griffith, A.D., Eppes, M.C., Delmelle, E.: Computationally enabled 4d visualizations facilitate the detection of rock fracture patterns from acoustic emissions. Rock Mechanics and Rock Engineering **51**(9), 2733–2746 (2018) https://doi.org/10.1007/s00603-018-1488-z

[45] Rastegaev, I., Merson, D., Danyuk, A., Afanasyev, M., Vinogradov, A.: Using acoustic emission signal categorization for reconstruction of wear development timeline in tribosystems: Case studies and application examples. Wear **410**, 83–92 (2018) https://doi.org/10.1016/j.wear.2018.06.006

[46] Rastegaev, I., Merson, D., Rastegaeva, I., Vinogradov, A.: A time-frequency based approach for acoustic emission assessment of sliding wear. Lubricants **52**(7), 1–24 (2020) https://doi.org/10.3390/lubricants8050052

[47] Jaynes, E.T.: Information theory and statistical mechanics. Physical review **106**(4), 620 (1957) https://doi.org/10.1103/PhysRev.106.620

[48] De Martino, A., De Martino, D.: An introduction to the maximum entropy approach and its application to inference problems in biology. Heliyon **4**(4), 00596 (2018) https://doi.org/10.1016/j.heliyon.2018.e00596

[49] Bae, J., Helldin, T., Riveiro, M., Nowaczyk, S., Bouguelia, M.-R., Falkman, G.: Interactive clustering: a comprehensive review. ACM Computing Surveys (CSUR) **53**(1), 1–39 (2020) https://doi.org/10.1145/3340960

[50] Ramasso, E., Verdin, B., Chevallier, G.: Monitoring a bolted vibrating structure using multiple acoustic emission sensors: A benchmark. MDPI DATA **7**, 31–45 (2022) https://doi.org/10.3390/data7030031

[51] Kattis, S.: Noesis: Advanced data analysis, pattern recognition & neural networks software for acoustic emission applications. In: Kolloquium Schallemission, Statusberichte zur Entwicklung und Anwendung der Schallemissionsanalyse, pp. 1–8. Deutsche Gesellschaft für Zerstörungsfreie Prüfung e.V., Fulda (2017)

[52] Niikura, T., Hosoya, N., Hashimura, S., Kajiwara, I., Giorgio-Serchi, F.: Loosening detection of a bolted joint based on monitoring dynamic characteristics in the ultrasonic frequency region. In: Oberst, S., Halkon, B., Ji, J., Brown, T. (eds.) Vibration Engineering for a Sustainable Future, pp. 191–196. Springer, Cham (2021)

[53] Eraliev, O., Lee, K.-H., Lee, C.-H.: Vibration-based loosening detection of a multi-bolt structure using machine learning algorithms. Sensors **22**(3), 1210 (2022)

[54] Silva, S., Omori Yano, M., Teloli, R.d.O., Chevallier, G., Ritto, T.G.: Domain adaptation of population-based of bolted joint structures for loss detection of tightening torque. ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg **10**(1) (2024)

[55] Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association **66**(336), 846–850 (1971) https://doi.org/10.2307/2284239

[56] Lefebvre, Q., Personeni, M., Ramasso, E., Thibaud, S.: Observation and detection of machining phases in edm microdrilling by acoustic emission. In: Structural Health Monitoring 2019: Enabling Intelligent Life-cycle Health Management for Industry Internet of Things (IIOT), Stanford University, CA, USA, pp. 1–6 (2019). https://doi.org/10.12783/shm2019/32099

[57] Ramasso, E., Ducas, M., Remacha, C., Thibaud, S.: Method for Detecting the End of Drilling by Electroerosion of a Hole in a Part of Electrically Conductive Material. Patent FR3113468B1, August 2020

[58] Gustafson, D.E., Kessel, W.C.: Fuzzy clustering with a fuzzy covariance matrix. In: IEEE Conference on Decision and Control Including the 17th Symposium on Adaptive Processes, pp. 761–766 (1979). https://doi.org/10.1109/CDC.1978.268028

A Voting scheme algorithm

Algorithm 2 was proposed in [38].

1 Input: Matrix of features $\mathbf{Y}$ with size $N \times d$;

2 Input: The range of number of clusters $[K_{\min}, K_{\max}]$;

3 Input: The subsets of features $\mathcal{S}$;

4 Input: The set of weights for two voting schemes W_1, W_2;

5 Output: Voted partition of the data P^*;

6 Output: Clusters parameters θ^*;

7 for $S \in \mathcal{S}$ **do**

8   **for** $K \in [K_{min}, K_{max}]$ **do**

9   *// Apply the clustering method and store the partition and parameters;*

10 $[P(S,K), \theta(S,K)] \leftarrow \mathsf{Clustering}(\mathbf{Y}(:,S), K)$;

11   *// Quality can be evaluated with one or several indices;*

12   *// Indices used for AE are mostly internal and focus on clusters shape;*

13 $Q_{\text{literature}}^{\text{SHAPE}}(S,K) \leftarrow \mathbf{Internal\_Shape\_Indices\_Quality}(P, Y(:,S), \theta)$;

14   **end**

15 end

16 *// For each subset, firstly find the clusters using a voting scheme;*

17 *// The vote uses weights set by the end-user;*

18 for $S \in \mathcal{S}$ **do**

19   nb_clusters$(S) \leftarrow \arg\max_k \text{Vote}(Q(S,k), W_1)$;

20 end

21 *// Then find the subset and the final number of clusters using a second voting scheme;*

22 $S^* \leftarrow \arg\max_S \text{Vote}(Q(S, \text{nb\_clusters}(S)), W_2)$;

23 $K^* \leftarrow \text{nb\_clusters}(S^*)$;

24 *// Return results;*

25 $P^* \leftarrow P(S^*, K^*)$;

26 $\theta^* \leftarrow \theta(S^*, K^*)$;

Algorithm 2: Evaluation of clustering methods: Shape-based *internal* indices with voting scheme used in the literature [38]. The output partition is on selected by a vote from the subset of partitions generated by clustering.