An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer-composite materials

D.D. Doan[a], E. Ramasso[a,b,*], V. Placet[b], S. Zhang[b], L. Boubakar[b], N. Zerhouni[a]

[a]Department of Automation Science and Micro-Mechatronic Systems
[b]Department of Applied Mechanics
FEMTO-ST, 32 avenue de l'Observatoire, 25000 Besançon, France

Abstract

This work investigates acoustic emission generated during tension fatigue tests carried out on a carbon fiber reinforced polymer (CFRP) composite specimen. Since fatigue data processing, especially noise reduction remains an important challenge in AE data analysis, a Mahalanobis distance-based noise modeling has been proposed in the present work to tackle this problem. A Davies-Bouldin-index-based sequential feature selection has been implemented for fast dimensionality reduction. A classifier offline-learned from quasi-static data is then used to classify the processed data to different AE sources with the possibility to dynamically accommodate with unseen ones. With an efficient proposed noise removal and automatic separation of AE events, this pattern discovery procedure provides an insight into fatigue damage development in composites in presence of millions of AE events.

Keywords: organic-matrix composites, acoustic emission clustering, fatigue datasets, noise reduction, sequential feature selection.

*Corresponding author. Tel.: +33 81 66 69 49.
Email address: emmanuel.ramasso@univ-fcomte.fr (E. Ramasso)

Introduction

AE testing has become a recognized nondestructive test (NDT) method, commonly used to detect and locate defects in mechanically loaded structures and components. AE can provide comprehensive information on the origination of a discontinuity (flaw) in a stressed component and also pertaining to the development of this flaw as the component is subjected to continuous or repetitive load [1]. Moreover, the method has been developed and applied in numerous structural components, such as steam pipes and pressure vessels, and in the research areas of rocks, composite materials and metals [2].

Acoustic emissions (AE) are stress waves produced by the sudden internal stress redistribution of the materials caused by the changes within the structure [3]. For polymer-composite materials, these changes are mainly due to crack initiation and growth, crack opening and closure, fiber breakage and fiber-matrix debonding. The use of AE for structural health monitoring has been investigated several decades ago with the objective to predict material failure [4, 5, 6].

With a huge noisy amount of data originating from fatigue loading tests, a major challenge in the use of AE technique is to associate each signal to a specific AE source related to noise or to a damage mechanism. This analysis is a non-trivial task for two main reasons. First, AE signals are complex so that it has to be characterized by multiple relevant features. Second, there is generally no *a priori* knowledge of the acoustic signatures of damage events which are generally scattered due to the high variability of the properties of composite materials [7].

In the literature, dealing with the challenge of massive data due to high sensitivity of AE sensors and to long-term fatigue loading experiments, several processing approaches have been proposed [8, 9, 10]. In [8], it is considered that only signals with amplitude higher than 70 dB or recorded above 80% of peak load contain information related to damage mechanisms. In [9], “friction emission” tests in which the maximum cyclic load was decreased to a level that was insufficient to generate crack growth were performed to understand the AE signal characteristics arising from hydraulics, machine start/stop and slippage. All of the AE events at this lower peak load were therefore assumed to be due to friction emission. Emission having the characteristics of friction emission was then filtered. A more complex denoising process developed by [10] that combines Principal Component Analysis (PCA) and K-means and several validation techniques was presented to be able to classify more than 60% of the detected signals as noise during long time corrosion monitoring of a pre-damaged post tensioned concrete beam.

High dimensional feature space reduction is a remaining challenge to statistical processing and classification of AE data. In the literature, many approaches for AE data processing [1, 11] rely on the Principal Component Analysis (PCA). The PCA takes a set of features calculated from AE signals, such time-frequency features, and generates a set of articifial variables made of a linear combination of the input features depicting the largest variance. Other approaches [12, 13, 14] rely on a specific subset of features such as energy, rise time, duration, amplitude [12] or have reduced the dimension of the feature space by using complete link hierarchical clustering in order to merge the correlated features into groups [13]. Those apply a greedy approach that generates all possible feature combinations and then selects the one which optimizes a given criterion [14, 15]. The goal of the criterion is generally to evaluate the quality of the partition provided by the clustering. It can be noticed that the PCA and the K-means clustering method are theoretically related to each other as shown in [16]. An alternative approach to Euclidean distance-based clustering methods was proposed [17] and based on the Gustafson-Kessel algorithm (GK) [18]. It makes use of a modified Mahalanobis distance for each cluster which is iteratively adapted to fit ellipse-shaped clusters. The use of hyper-ellipses instead of hyper-spheres is more appropriate for AE clustering in presence of low density and high scattering. In the GK algorithm, the covariance between each pair of features is estimated so that possible redundancy or complementarity between features can be taken into account. The Mahalanobis distance has also been shown to be robust to outliers in statistical analysis [19].

The processing of large AE datasets, in particular originating from fatigue, requires to develop efficient methods in terms of memory and time consumption. Some approaches have been proposed which are able to work online (or real-time), that means that clusters parameters are updated without iterative procedure but as new data arrive. As underlined in the GK-based method proposed in [17] and in the Kmeans-based method developed in [20], external AE sources (corresponding to noise) may have an important influence on the clusters' updating. In this paper we propose a methodology to estimate efficiently the partition of AE data obtained in fatigue loading in presence of noise sources. The methodology also includes an automated sequential feature selection based on the GK algorithm and relying on quasi-static (QS) tests. The clusters obtained are then adapted to be applied on large fatigue tests. The next section is dedicated to presentation of the proposed methodology.

1. Unsupervised pattern recognition

The flow chart of the methodology is shown on Fig. 1.

[Figure 1 about here.]

1.1. AE fatigue data pre-processing

All acoustic emissions even originating from outside the area of interest bounded by the sensors were taken into account (no spatial filtering). Thus a pre-processing step of such AE data is highly important and requires adapted filtering methods [21].

1.1.1. Signal screening

Continuous background noise due to hydraulic flows is essentially eliminated from the AE signal by a floating signal threshold, which is adjusted at a 40 dB level. This threshold makes it possible to loose signals originated from friction. Optimal denoising, for instance using wavelets [22], would be necessary if those signals are important for the monitoring.

1.1.2. Noise model-based filtering

Typical field and environmental noise such as electromagnetic interference (EMI), fretting, mechanical or hydraulic vibration encountered in real applications generate extraneous noise detected by the broadband and high sensitive AE sensors. Assuming that this AE activity is not due to damages, a noise model is built using a multivariate statistical test based on the Mahalanobis distance as used in novelty detection [23, 24]. For that, the AE hits recorded before the loading phase are considered as representative of the AE hits corresponding to external AE sources (such as noise). The statistical mean (center of the noise model) and covariance of those samples define an ellipsoid in the feature space, and its boundary is estimated as the average of the Mahalanobis distances between each sample and the center. An AE hit recorded during loading is then considered as noise if it falls within the boundary of the ellipsoid.

1.2. Sequential selection algorithm of AE features

An automated technique is presented to detect relevant feature subsets for clustering of AE events. In contrast to feature reduction procedures (for example based on correlation dendrogram [1]) or exhaustive search of global optimal feature combinations [14], the principle of the approach is to combine gradually each feature from an available feature space with an initial feature subset [25]. The feature selection is achieved by minimizing the value of Davies and Bouldin (DB) index [26] defined by:

$$DB = \frac{1}{c}\sum_{i=1}^{c} \max_{i \neq j} \left\{ \frac{d_i + d_j}{D_{ij}} \right\} \tag{1}$$

where c is number of clusters, d_i and d_j are the average within-class distances of clusters i and j respectively, and D_{ij} denotes the distance between the two clusters i and j. This clustering validity index has been used by several authors in order to select optimal cluster number [13] or to evaluate feature subset partition [14]. The lower is its value, the better is the compactness and the separability within the partition. Figure 2 shows the diagram of the proposed algorithm based on a feature filtering approach [27].

[Figure 2 about here.]

Considering an initial subset of features S (empty by default), the algorithm takes each of the available features from F to update S. This subset is then partitioned by the GK clustering algorithm. At the k^{th} iteration, a feature $f_l \in F$ is added to the current subset of features S_k, and the corresponding DB index DB_l of the partition obtained by the GK algorithm is computed. The computation of the DB index makes use of the Mahalanobis-like distance defined in the GK algorithm [18] to estimate the distance between AE hits and cluster centers and finally obtain the estimate of the average within-class distances used in Eq. 1 (d_i and d_j).

The subset of features S_{k+1} for the next iteration is given by $S_k \cup f_{l^*}$ with $l^* = \arg\min_l DB_l$ and the partition is then evaluated by the DB criterion. The feature that minimizes the value of DB index is selected and transfered from F to S. At each iteration, the procedure generates $|F|$ new subsets since each new subset contains the features from S plus a new one taken from the remaining features in F. The algorithm stops when no new subsets can improve the DB criterion.

For each iteration k, an improvement rate $IR(k)$ is calculated as follows:

$$IR(k) = \frac{DB(S_k) - DB(S_{k-1})}{DB(S_{k-1})} \tag{2}$$

where $DB(S_k)$ and $DB(S_{k-1})$ represent the value of the DB-index of the best feature selection for the k^{th} and $(k-1)^{th}$ iteration respectively. The sign of IR indicates if the DB criterion is improved (negative) or not (positive). For the last iteration k^{last} (for which $IR(k^{\text{last}}) > 0$), if $IR(k^{\text{last}}) < min_{k<k^{\text{last}}}|IR(k)|$ then the feature with the best DB-index is added to S to establish the final selected feature set.

1.3. AE source clustering

Quasi-static (QS) tests are first applied to obtain a relatively low amount of data compared to fatigue and by supposing that damage sources in QS tests are mostly similar to fatigue. The GK algorithm is thus applied to estimate the parameters of a given set of k clusters on AE originated from QS tests. To cope with possibly additional AE sources that can occur during fatigue [28], an additional $k+1^{th}$ cluster is estimated based on fatigue data to include all feature vectors located "far" from the previous k clusters. For that, the boundary of each cluster characterized on QS tests is estimated by the average of the Mahalanobis-like distance (used in GK) [24]. A feature vector obtained during fatigue belongs to the $k+1^{th}$ cluster if its distance to nearest cluster is above the corresponding radius. This adaptation of clusters is supposed to take into account one (or more) AE sources that is (or are) not present in quasi-static tests (e.g. noise due to repeated tensile loading, acoustic waves related to cumulated damage ...).

2. Experiments

Composite split disks were considered subjected to cyclic fatigue loading up to failure determined when a complete break of the specimen was observed in the hoop direction. The specimens were cyclically tested under a tensile/tensile sinusoidal loading with constant amplitude and frequency of 5 Hz and under constant stress ratio R = 0.1 at room temperature. Quasi-static tests were preliminarily conducted on five different specimens with a constant loading rate of 0.3 kN.s^{-1}. The static failure stress was equal to 1520 $\pm$ 165 MPa. The tests were performed according to ASTM D2290 "Apparent hoop tensile strength of plastic or reinforced plastic pipe by split disk method". Rings were produced by cutting and machining filament-wound carbon fiber reinforced epoxy tubular structures intended for the manufacturing of flywheel rotors with a $(90°)_6$ lay-up configuration. The transient elastic waves were recorded during test at the material surface using a multi-channels data acquisition system from EPA (Euro Physical Acoustics) corporation (MISTRAS Group). The system is made up of miniature piezoelectric sensors (micro-80) with a range of resonance of 250 - 325 kHz, preamplifiers with a gain of 40dB and a 20 - 1000 kHz filter, a PCI card with a sampling rate of 1MHz and the AEWin software. Two AE sensors were coupled on the specimen faces using silicon grease. The experimental set-up is shown in Fig. 3.

[Figure 3 about here.]

The calibration of the system was performed after installation of the transducers on the specimen and before each test using a pencil lead break procedure. A part of the ambient noise was filtered using a threshold of 40dB. The acquisition parameters: PDT (Peak Definition Time) = 60 μsec; HDT (Hit Definition Time) = 120 μsec and HLT (Hit Lock Time) = 300 μsec were optimized for this specific experimental configuration to extract transient signals. The optimization of these time-driven parameters was performed using the standard pencil-lead breakage proposed by Hsu and Nielsen [29]. Many features such as absolute energy, counts, hits, amplitude, duration, frequency centroid were calculated from recorded waves.

3. Results and discussion

According to different percentages of the ultimate tensile stress determined in the tensile test (1520 MPa), the S-N curve was obtained as illustrated in Figure 4.

[Figure 4 about here.]

Nine samples were used to generate the S-N curve. This was a good compromise between the six specimens recommended by ASTM D-3479 for preliminary and exploratory test campaign and the twelve specimens required for research and development on testing of components and structures. The results presented in this work are part of a wider study including the generation of S-N curves of different types of composites (with different carbon fibers) and with different lay-up configurations. The main goal is to select a composite of choice for the application concerned, namely rotors of flywheels.

Four datasets were considered denoted as A1 (quasi-static test) and A2, A3 and A4 (fatigue tests for 90%, 80% and 70% of the ultimate tensile strength respectively). A brief description of the obtained datasets is summarised in Table 1.

[Table 1 about here.]

3.1. Noise reduction

According to the scenario of the quasi-static test A1 (Fig. 5(a)), around the time-instant t1, the actuator was pressurized and the stress was applied only at t2. The noise modeling phase (Section 1.1.2) has been made from AE data recorded before t1, i.e. while the specimen was let in its environment without any mechanical loading. Noise during loading is then filtered by this model.

[Figure 5 about here.]

Figure 5(b) and 5(c) represent dataset A1 (made of 52,832 AE hits) in the duration-amplitude space segmented into three populations: noise before and during loading in Figure 5(b), and denoised data after application of noise model in Figure 5(c). The two first populations (noise) possess the same characteristics, the same location and the same scattering. This observation is justified by the graphic of AE cumulated energy in Fig. 6(a). Indeed, the level of AE cumulated energy of noise before and during loading is negligible and the total energy is conserved within denoised data while the latter occupies only 12% of the whole dataset in terms of quantity (Fig. 6(b)).

[Figure 6 about here.]

The application of the noise model to fatigue dataset A3 made of 1,682,434 AE hits led to a similar separation between noise and denoised data (Fig. 7(a)). In spite of 93% of AE hits recorded associated to "noise" (Fig. 7(c)), this highest population represents negligible AE cumulated energy level in comparison with that of denoised data (Fig. 7(b)).

[Figure 7 about here.]

3.2. Feature selection

Many energy-based approaches of damage characterization or identification have been studied since AE energy provides a good correlation with damage mechanisms. Thus, in this work, absolute energy (Fig. 8) is used to initialize the subset of relevant features. As the number of clusters is unknown, 3 cases were addressed to check the stability of the selection algorithm by considering 4, 5 and 6 clusters.

The selection algorithm was applied on the quasi-static dataset A1 with 4 clusters. At the first iteration, given the absolute energy feature, the optimal DB index is given by the combination with the amplitude feature (Fig. 8(a)). At the second iteration, the best score was obtained by the combination with the MARSE energy (Fig. 8(b)). No more improvement of the DB index is made at the next iteration, so the algorithm is stopped by selecting the subset made of absolute energy, amplitude and MARSE energy. The same selection result was obtained with 5 and 6 clusters. In what follows, 4 clusters are used as initial number of AE sources.

[Figure 8 about here.]

3.3. AE source classification

3.3.1. Sequence of AE hits in the quasi-static case

The denoised and selected feature subset obtained previously is now used to identify the clusters in quasi-static dataset A1 using the GK clustering algorithm. Four well-separated clusters with different sizes and shapes have been obtained in the duration-amplitude space (Fig. 9(a)). After projection onto the amplitude dimension, four distinct distributions can be obtained, among which three are located above 75 dB. These distributions have been often used to identify AE sources [30, 31].

Ono and Gallego [2] recently underlined a misconception that fiber fracture always produces high-energy event, and that still persists to this day. For the considered material, the damage process involves fiber *tow* breakage. If the breakage of an elementary fiber ($7\mu m$ diameter) can cause the release of low energy transient, the breakage of fiber tows including hundreds or thousands of elementary fibers (up to 12,000 in the considered material) are likely to induce highly energetic signals.

As a complementarity view, the temporal evolution of the logarithm of the Cumulated Sum of Cluster Appearance (logCSCA) [17] has been depicted in Fig. 9(c) (for each cluster) together with the cumulated energy and the load. When an AE hit (emitted after the activation of an AE source) is associated to a given cluster at a given time, the corresponding logCSCA curve depicts a *step*. When several consecutive steps appear in a short time period, this visualisation allows to point out that the activity of the corresponding AE source is particularly sustained which may be related to propagations of cracks [32].

In the sequence shown in Fig. 9(c), the first cluster is activated at the very beginning before applying the load. Despite the number of AE hits in this cluster is important (73% at the end), the cumulated energy of AE hits in this cluster is the lowest one among all clusters (Fig. 9(e)). These observations are coherent with the activation of an AE source related to mechanical and hydraulic emission such as vibration and friction between the specimen and the half-cylinders.

Both cluster 2 and 3 start early when the actuator has been pressurized. The main activity of cluster 2 occurs after a certain level of load (Fig. 9(c)) and the cumulated energy of AE hits in this cluster (Fig. 9(e)) as well as the amplitudes ($>$ 95 dB, Fig. 9(a)) are the highest ones compared to all other clusters. The number of AE hits in this cluster is particularly important at the end of the test, as expected with the ruine of the specimen induced by a cascade of fiber tow breakage. This cluster is thus related to the activity of highly energetic sources, in particular carbon fiber tow breakage.

The low cumulated energy in cluster 3 as well as the amplitudes around 75 and 90 dB make this cluster related to minor damage (probably matrix micro-cracks).

The partition also emphasizes an important cascade of AE hits at $t \approx 352$ s during which the activity of cluster 3 increases importantly and this increase is synchronised with both the appearance of cluster 4 and a high activity of cluster 3. The load level at this time, the mean value of amplitudes in cluster 4 (around 95 dB, Fig. 9(a)) and the level of the cumulated energy in this cluster (around 13% of the total cumulated energy, Fig. 9(e)) make this cluster related to macro-cracking and interface failures starting around the specimen's notches and propagating gradually in the hoop direction.

3.3.2. Sequence of AE hits in a fatigue test

Afterwards, the model estimated on A1's AE hits is used to infer the partition on the fatigue dataset A3. Direct application of the model generates overlapping zones between clusters in the duration-amplitude space of A3 (Fig. 9(b)). We can observe a similar distribution of clusters in this feature space compared to A1 (Fig. 9(a)). However, we can also observe clusters overlap, particularly important between clusters 2 and 3. As a consequence, the projection onto the amplitude axis would not give distinct distributions as for the quasi-static test. This phenomenon finds its origins in the fact that, compared to quasi-static tests, additional mechanisms can play a role during fatigue such as the temperature [33] or the cycling which implies crack opening/closing initially not observed during QS tests [28]. Therefore, it was expected to find out that a pattern recognition model learned from a quasi-static test and simply applied on a fatigue test may present a limited generalization capability. Based on the assumption that a new AE source is activated during fatigue and which has not been observed in quasi-static tests, the proposed methodology (Section 1.3) includes the creation of new cluster to cope with this problem. The result is a new segmentation with less overlapping between clusters as shown in Fig. 9(d).

[Figure 9 about here.]

The comparison of partitions with the previous quasi-static test yield similar conclusions concerning the possible damage scenario. The main difference holds in the position of the new cluster, which has been automatically found from AE hits. Indeed, the cluster 3 identified as the friction and possibly micro-cracking in the quasi-static test (Fig. 9(a)) was split (AE sources 3a and 3b, Fig. 9(b)). The signatures of AE hits in both clusters in terms of amplitudes, durations and energies (Fig. 9(f)) are quite different despite the fact that the clusters are pretty close in the duration-amplitude space. The evolution of AE cumulated energy of each source (Fig. 9(f)) brings useful interpretations about the damaging process during fatigue. Despite its smallest population, AE source 2 is dominant in term of energy at the end of test as for the quasi-static test and is associated to severe damage mechanisms related to carbon fibers. AE source 1 is the most scattered and populated but represents negligible contribution compared to the total energy. As for the quasi-static case, this cluster may represent the activation of an AE source related to mechanical and hydraulic systems [34]. AE source 4 generates AE hits with the longest duration and the highest energy that may be related to macro-cracking and interface failures.

[Figure 10 about here.]

[Figure 11 about here.]

Figure 11 represents the positioning of clusters onto the load level for the fatigue dataset A3. This figure enables one to visualise the load level when the AE sources are activated. On dataset A3, it can be observed in Fig. 10(a) to 10(f) that during 20%-fatigue-life of the specimen, many AE hits appear, related to all AE sources. This phenomenon is well known as the accommodation phase [35] which generally appears at the first stage of materials undergoing fatigue testing and may lead to partial fractures. Indeed, AE hits with high energy (from AE sources 2 and 4) are activated during this phase (and during failure). After this stage, the clusters' activities globally slow down for a while (stabilization phase). Beyond 65% of the fatigue life, an important number of highly energetic AE hits occur up to the ruine of the specimen (from AE sources 2 and 4). It is interesting to notice a repetitive phenomena that takes place all along the test and represented by the activation of AE sources 3a and 3b: The latter is mainly activated in loading phases while the former occurs in unloading phase (Fig. 11(b)). As for the quasi-static test, the latter may correspond to internal frictions and interfaces fretting as observed in previous papers [36]. It can also be observed that the AE hits originated from these clusters occur between 5-7 kN at the beginning of the tests and between 3-4 kN at the end. AE source 3a is much more activated than AE source 3b between 20% and 50%, then the activity of 3b substantially increases until the ruine. This increasing is followed by the activation of AE source 5 that is particularly active between 70% and 90%, just before the ruine. Therefore, as expected, the fatigue plays a role on the loading level required to activate some sources and the chronology of activation may give insights to the understanding of damage mechanisms.

3.3.3. Sequence of AE hits in two other fatigue tests

The complexity of damage mechanisms involved during fatigue is illustrated in this section. For that, two other specimens denoted as A2 and A4, corresponding to 90% and 70% of the tensile strength respectively, are considered. The behavior of A2 is similar to the previous specimen A3 as depicted in Fig. 12(a). The activity of the AE hits including high energy and high duration signals is rather high (relatively to the remaining AE hits) at the very beginning of loading and increases again at about 60% of the specimen life, as for A3. Although AE hits generated by AE source 3a are more scattered than in the previous test, overlaps between clusters related to this source and to AE source 3b have also been detected by the proposed algorithm. Table 2 summarises the clusters assigned to each AE source according to the previous observations.

[Table 2 about here.]

Rather different than the previous tests, the partition obtained on dataset A4 at 70% of the ultimate static strength is depicted in Fig. 12(b). The initial (accommodation) phase occurs within the first cycles as for the previous loading levels, but it is then followed by a silence of most of AE sources. Only AE source 1 is activated (representing possible external sources which has been filtered out) and a few highly energetic AE hits occur (such as fiber tow breakage). Then, at 20% of the fatigue life, a progressive activation of all AE sources can be observed. In the load band $2-10$ kN, only cluster 1 is activated but this band is gradually reduced with respect to the number of cycles to reach $4-7$ kN when approaching the end-of-life. The progressive and continuous reduction of the band beyond which clusters are activated can be of interest for predicting the remaining lifetime of the composite if confirmed on other specimens and lower loading levels. It can also be noticed that more AE hits related to AE sources 2 and 4 (i.e. with the highest energy) can be found compared to the two previous specimens. Therefore, the failure process of specimen A4 tested at 70% of the ultimate tensile strength is more gradual and more related to the progressive weakness of the material during the repeated stress until the ruine.

[Figure 12 about here.]

Conclusion

An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer-composite materials has been presented to tackle different existing challenges of AE analysis and damage detection: 1) data pre-processing, especially noise reduction; 2) automatic and fast feature selection; 3) clustering of massive data from fatigue tests with cluster adaptation. The methodology relies on the estimation of clusters during static tests. Its application to big fatigue data based on the adaptation phase allows to add a new cluster to cope with new AE sources. The assignment of a cluster to a AE hit is not iterative and only requires to find the closest cluster by using a Mahalanobis-like distance that allows to cope with data scattering. The processing of a fatigue dataset is made faster than iterative procedures which requires to load a dataset and to perform interative optimization on large matrices.

The first results on three real fatigue tests of thermoset ring-shaped CFRP involving until 10 millions AE hits demonstrate that the proposed methodology allows to identify some relevant clusters. Of particular interest:

- Four main phases have been identified: Accommodation with many AE hits with the highest energy and amplitude (0-20% of the lifetime), a slowdown of AE activity (20-50%), a resumption of the AE activity (50-85%) and a failure progress up to the final failure (85-100%). The fatigue at 70% of the ultimate strength depicts a particular pattern during the degradation involving an envelop which gradually reduces until the ruine.

- Two clusters detected by the adaptation phase occur at similar loading levels. A modification of their kinetics with report to the cumulated loading lets suppose that those two clusters can be due to damage. It is also interesting to emphasize that the level required to activate the

AE sources related to those two clusters depicts a slight and progressive decreasing together with the degradation of the material until the ruine.

The visualization of clusters in the amplitude-duration feature, the logarithm of the cumulated AE hits and energy in each cluster as well as the the positioning of clusters onto the loading level have allowed to connect some clusters to possible AE sources. In order to validate the identification of AE sources observed, complementary non-destructive techniques and in-situ measurements is under study on more specimens. The application of the proposed methodology is currently investigated on thermoplastic CFRP composites and compared to finite element models [37]. Finally, the proposed methodology is under improvement for robust AE-based prognostics of composite structures.

Acknowledgement

This work has been carried out in the framework of the Laboratory of Excellence ACTION through the program "Investments for the future" managed by the National Agency for Research (references ANR-11-LABX-01-01). The authors thank the reviewers for the relevant and fruitful comments.

[1] A. A. Anastassopoulos, T. P. Philippidis, Clustering methodology for the evaluation of ae from composites, J. of Acoust. Emission 13 (1–2) (1995) 11–22.

[2] K. Ono, A. Gallego, Research and applications of ae on advanced composites, Journal of Acoustic Emission.

[3] M. Huang, L. Jiang, P. Liaw, C. Brooks, R. Seeley, D. Klarstrom, Using acoustic emission in fatigue and fracture materials research, The Journal of The Minerals, Metals and Materials Society 50 (1998) 1–12.

[4] L. J. Graham, Characterization of acoustic emission signals and application to composite structures monitoring, in: Proceedings of the ARPA/AFML Review of Progress in Quantitative NDE, 1975.

[5] C. Murthy, B. Dattaguru, A. Rao, Application of pattern recognition concepts to acoustic emission signals analysis, Journal of Acoustic Emission 6 (1987) 19–28.

[6] M. Otsu, K. Ono, Pattern recognition analysis of acoustic emission from unidirectional carbon fiber-epoxy composites by using autoregressive modeling, Journal of Acoustic Emission 6 (1) (1987) 61–70.

[7] P. Pineau, F. Dau, Subsampling and homogenization to investigate variability of composite material mechanical properties, Computer Methods in Applied Mechanics and Engineering 241244 (2012) 238 – 245.

[8] J. Henry, Z. Aboura, K. Khellil, S. Otin, Suivi de l'endommagement en fatigue d'un composite a renfort interlock carbone/epoxy par emission acoustique, Mater. & Tech. 100 (6-7) (2012) 643–652.

[9] J. Yu, P. Ziehl, B. Zarate, J. Caicedo, L. Yu, V. Giurgiutiu, B. Metrovich, F. Matta, Quantification of fatigue cracking in CT specimens with passive and active piezoelectric sensing, in: Proc. SPIE, Vol. 7649, 2010, pp. 76490R–76490R–12.

[10] 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.

[11] N. Godin, S. Huguet, R. Gaertner, L. Salmon, Clustering of ae signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers, Ndt&e Int 37 (4) (2004) 253–264.

[12] R. Gutkin, C. Green, S. Vangrattanachai, S. Pinho, P. Robinson, P. Curtis, On ae for failure investigation in CFRP: pattern recognition and peak frequency analyses, Mech. Syst. Signal Pr. 25 (4) (2011) 1393–1407.

[13] M. Moevus, N. Godin, M. RMili, D. Rouby, P. Reynaud, G. Fantozzi, G. Farizy, Analysis of damage mechanisms and associated ae in two SiCf/[SiBC] composites exhibiting different tensile behaviours. Part II: unsupervised ae data clustering, Compos. Sci. Tech. 68 (6) (2008) 1258–1265.

[14] M. Sause, A. Gribov, A. Unwin, S. Horn, Pattern recognition approach to identify natural clusters of ae signals, Pattern Recogn. Lett. 33 (1) (2012) 17–23.

[15] M. Halkidi, Y. Batistakis, M. Vazirgiannis, On clustering validation techniques, J. Intell. Inf. Syst. 17 (2-3) (2001) 107–145.

[16] C. Ding, X. He, K-means clustering via principal component analysis, in: Proc. ICML, Banff, Canada, 2004.

[17] V. Placet, E. Ramasso, L. Boubakar, N. Zerhouni, Online segmentation of ae data streams for detection of damages in composite structures in unconstrained environments, in: Int. Conf. Struct. Saf. Reliab., New York, 2013.

[18] D. Gustafson, W. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: IEEE Int. Conf. Decis. Control, Vol. 17, 1978, pp. 761–766.

[19] D. Peel, G. J. Mclachlan, Robust mixture modelling using the t distribution, Statistics and Computing 10 (2000) 339–348.

[20] E. Pomponi, A. Vinogradov, A real-time approach to acoustic emission clustering, Mechanical Systems and Signal Processing 40 (2) (2013) 791 – 804.

[21] V. Barat, Y. Borodin, A. Kuzmin, Detection of AE signals against background friction, J. of Acoust. Emission 29.

[22] D. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory 41 (3) (1995) 613–627.

[23] C. Farrar, K. Worden, Unsupervised learning novelty detection, in: Struct. Health Monit., 2012, pp. 321–360.

[24] L. Serir, E. Ramasso, P. Nectoux, N. Zerhouni, E2GKpro: An evidential evolving multi-modeling approach for system behavior prediction with applications, Mechanical Systems and Signal Processing 37 (1-2) (2013) 213–228.

[25] S. Guerif, Unsupervised variable selection: when random rankings sound as irrelevancy, JMLR Proc. 4 (2008) 163–177.

[26] D. Davies, D. Bouldin, A cluster separation measure, IEEE T. Pattern Anal. PAMI-1 (2) (1979) 224 –227.

[27] G. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, Mach. Learn.: Proc. Int. Conf. (1994) 121–129.

[28] I. D. Baere, W. V. Paepegem, J. Degrieck, Assessment of the interlaminar behaviour of a carbon fabric reinforced thermoplastic lap shear specimen under quasi-static and tension-tension fatigue loading, Polymer Testing 32 (7) (2013) 1273 – 1282.

[29] N. Hsu, F. Breckenridge, Characterization and calibration of acoustic emission sensors, Materials Evaluation 39 (1) (1981) 60–68.

[30] L. J. Graham, R. K. Elsley, Characteristics of ae signals from composites, in: ARPA/AFML Rev. Prog. Quant. NDE, 1977, pp. 219–225.

[31] D. Short, J. Summerscales, Amplitude distribution ae signatures of unidirectional fibre composite hybrid materials, Composites 15 (3) (1984) 200–206.

[32] D. Sornette, Statistical physics of rupture in heterogeneous media, in: Handb. Mater. Model., 2005, pp. 1313–1331.

[33] J. R. Gregory, S. M. Spearing, Constituent and composite quasi-static and fatigue fracture experiments, Composites Part A: Applied Science and Manufacturing 36 (5) (2005) 665 – 674.

[34] M. Wevers, G. V. dijck, W. Desadeleer, M. Windelmans, K. V. D. Abeele, Acoustic emission for on-line monitoring of damage in various application fields, Journal of Acoustic emission 22.

[35] G. Degallaix, A. Seddouki, S. Degallaix, Low cycle fatigue of a duplex stainless steel alloyed with nitrogen, in: Proc. 3rd Int. Conf. Low Cycle Fatigue Elasto-Plast. Behav. Mater., Vol. 3, Berlin, 1992.

[36] J. Cheng, H.-J. Li, S.-Y. Zhang, L.-Z. Xue, W.-F. Luo, W. Li, Internal friction behavior of unidirectional carbon/carbon composites after different fatigue cycles, Mater. Sci. Eng.: A 600 (2014) 129–134.

[37] M. G. Sause, S. Richler, Finite element modelling of cracks as acoustic emission sources, Journal of Nondestructive Evaluation 34 (1).

List of Figures

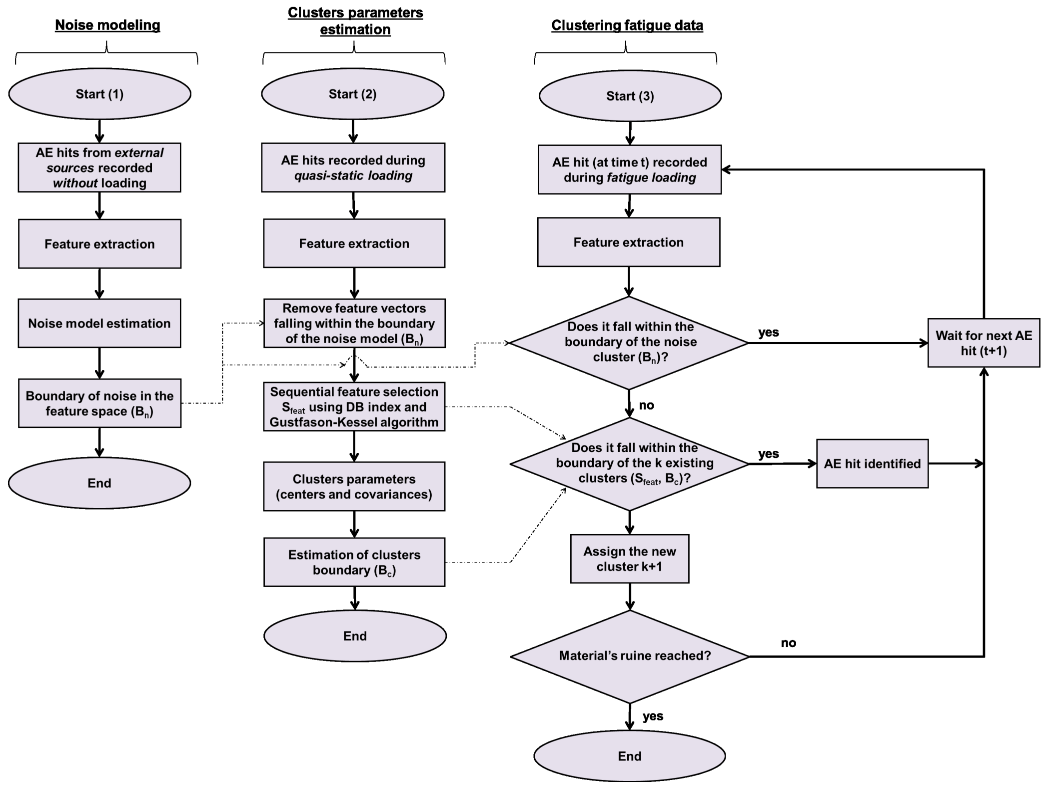


Figure 1: Unsupervised damage detection methodology

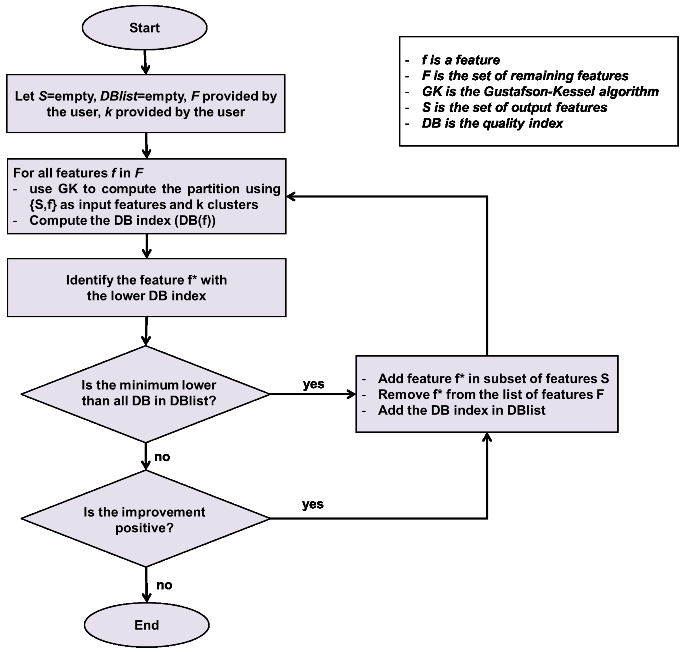
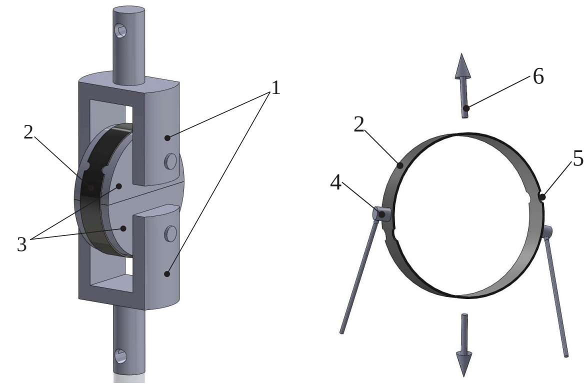


Figure 2: Sequential feature selection diagram



(a) (b)

Figure 3: Experimental set-up for tensile test on split-disk specimen. (1) fixture, (2) notched ring specimen, (3) half-cylinder, (4) AE sensor, (5) notched region, (6) load.

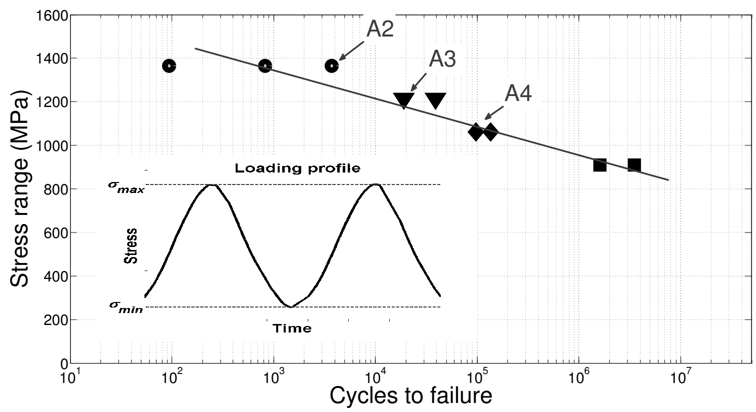
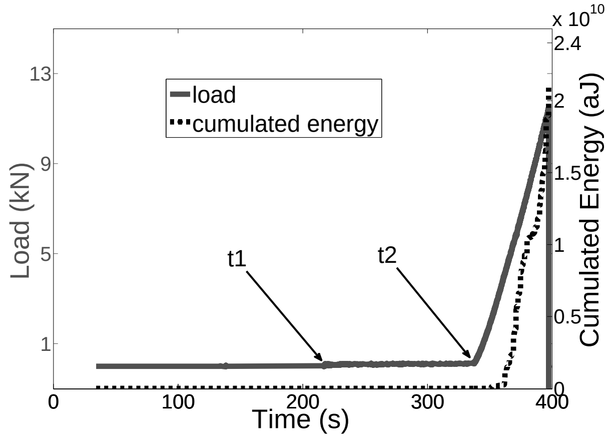
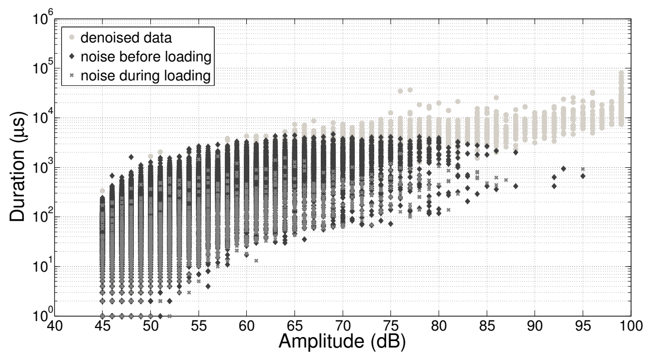


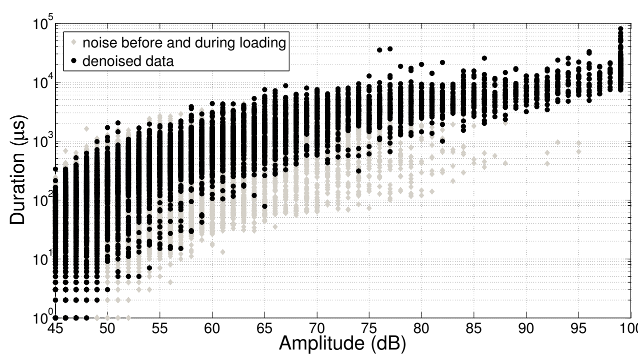
Figure 4: S-N curve of all tested specimens with ●: 90% of the ultimate tensile strength, ▼: 80%, ◆: 70% and ■: 60%.



(a)

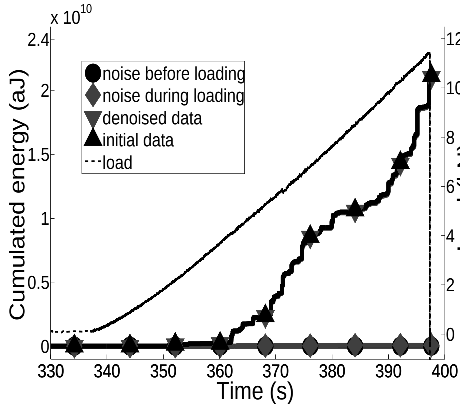


(b)

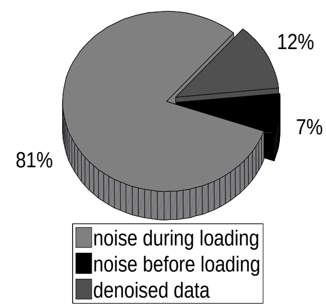


(c)

Figure 5: Quasi-static dataset A1: (a) Loading profile; (b) and (c) Duration vs. Amplitude for AE hits detected as noise (remaining data surimposed in light gray) and for for denoised data respectively

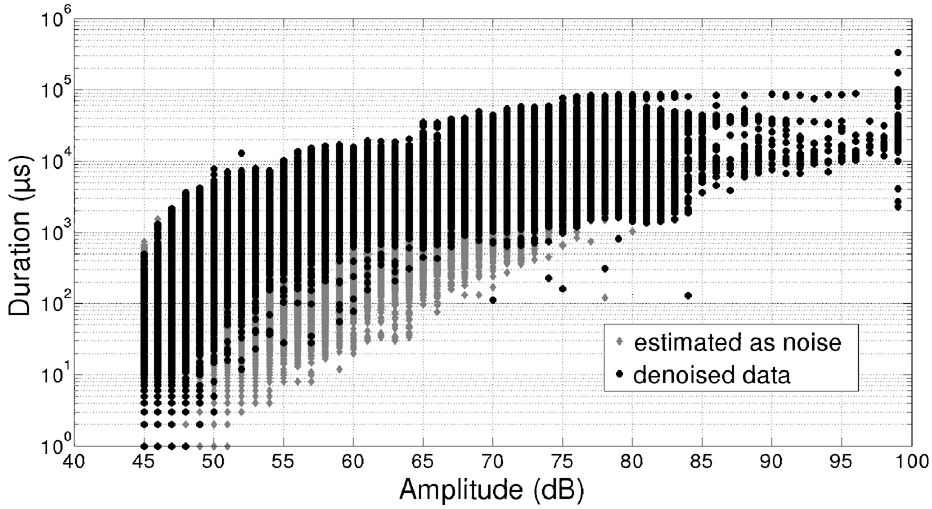


(a)

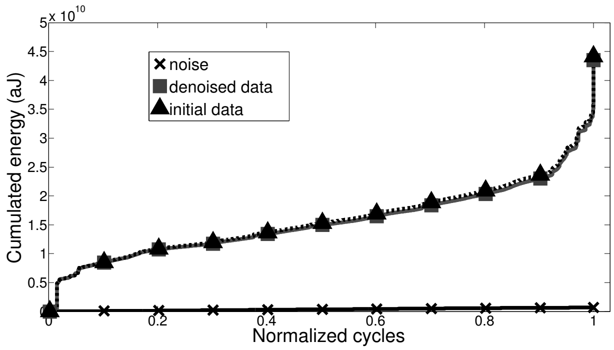


(b)

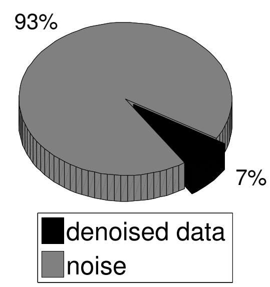
Figure 6: Quasi-static dataset A1: (a) AE cumulated energy; (b) Percentage in terms of population



(a)



(b)



(c)

Figure 7: Fatigue dataset A3: (a) Duration vs. Amplitude; (b) AE cumulated energy; (c) Percentage in terms of population

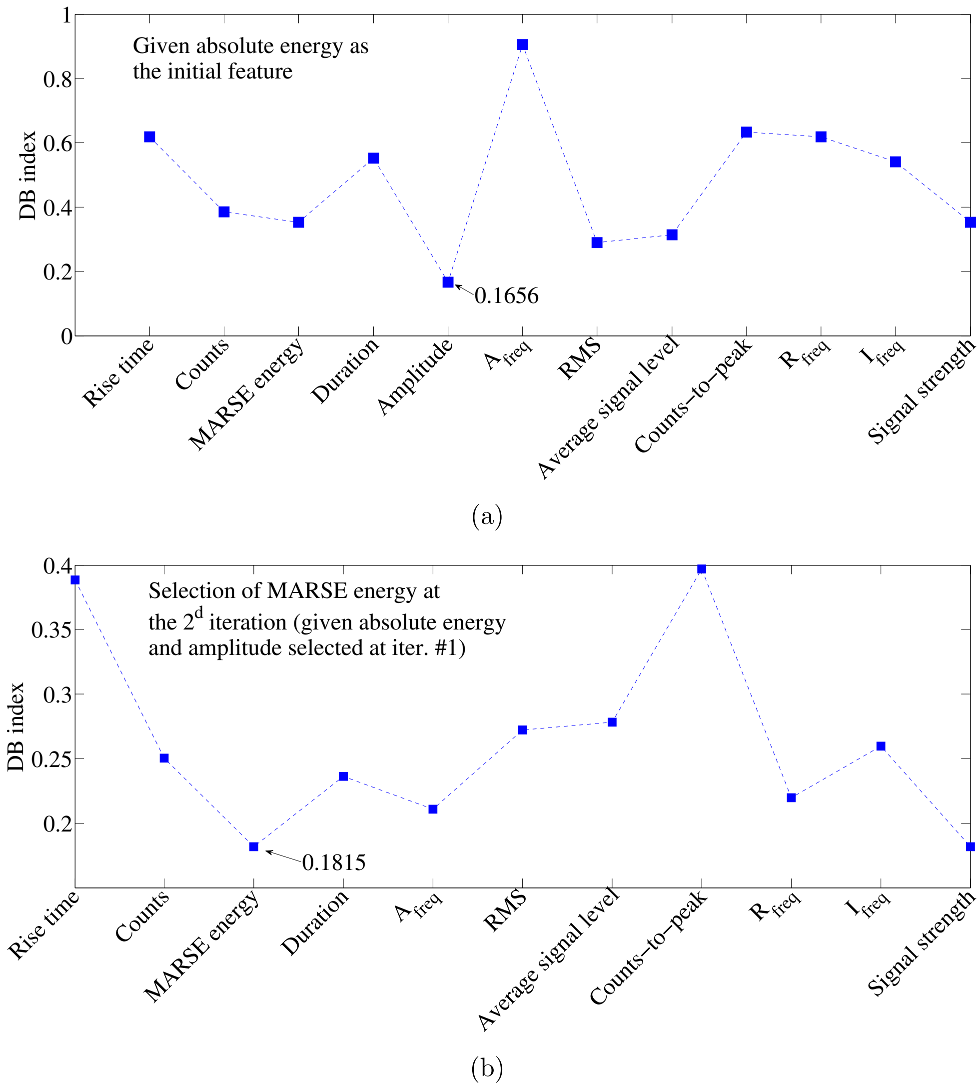


Figure 8: Case of 4 clusters: (a) first selection giving amplitude feature as the best; (b) second selection giving feature MARSE energy as the best.

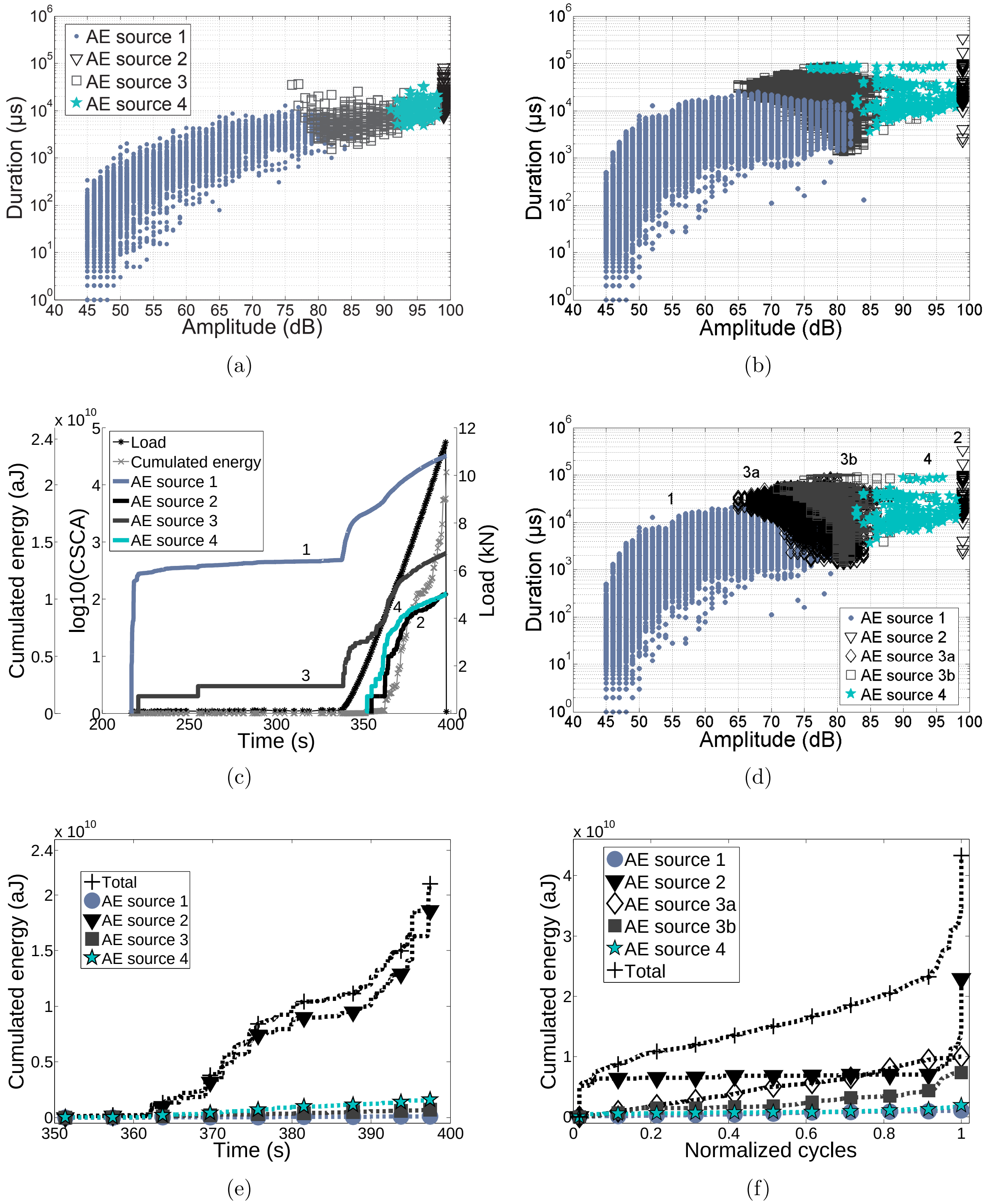


Figure 9: **Left** – Clustering result on dataset A1 (quasi-static): (a) Partition in the Duration vs. Amplitude space; (c) Evolution of the cumulated number of hits in each cluster (log CSCA); (e) Cumulated energy of each source. **Right** – Testing phase on dataset A3 (fatigue): (b) Direct classification without adaptation; (d) Adaptive classification; (f) Cumulated energy of each AE source.

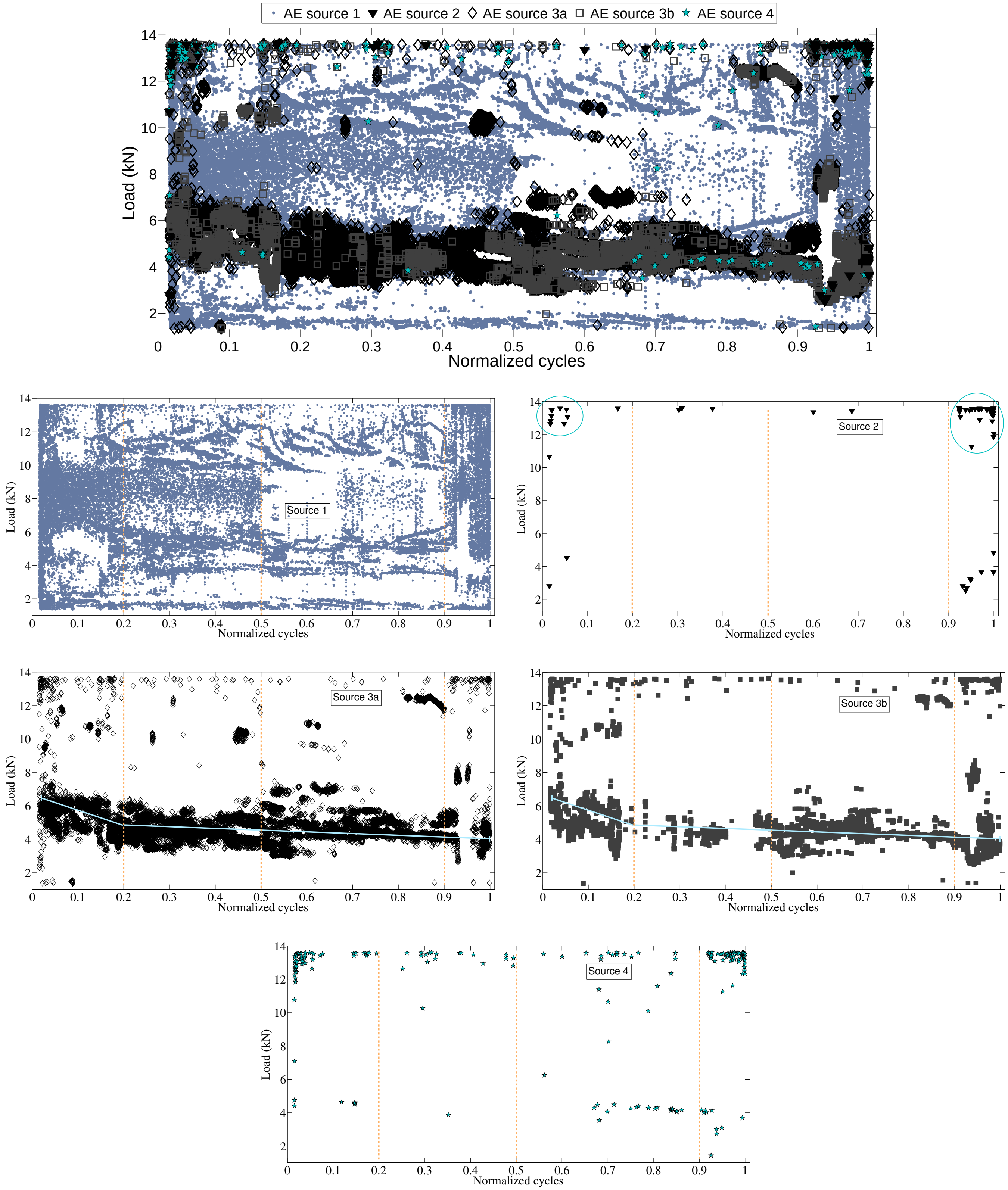


Figure 10: Classified AE events during cyclic loading of specimen A3 (80%): (a) All sources during the whole test; (b)-(f) Individual AE source

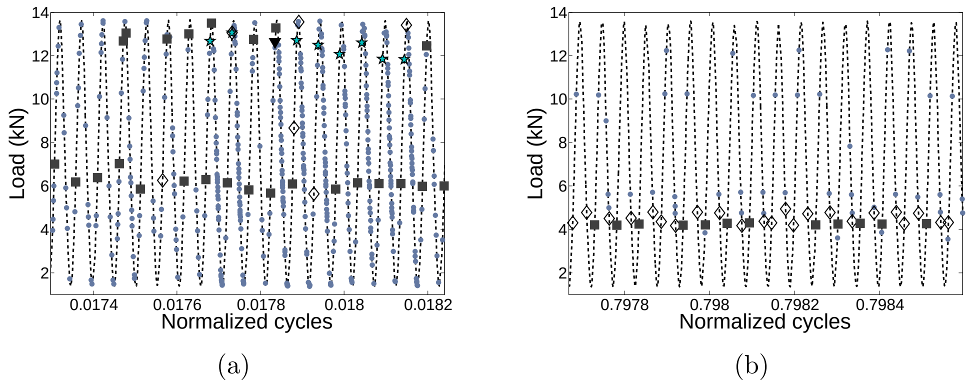
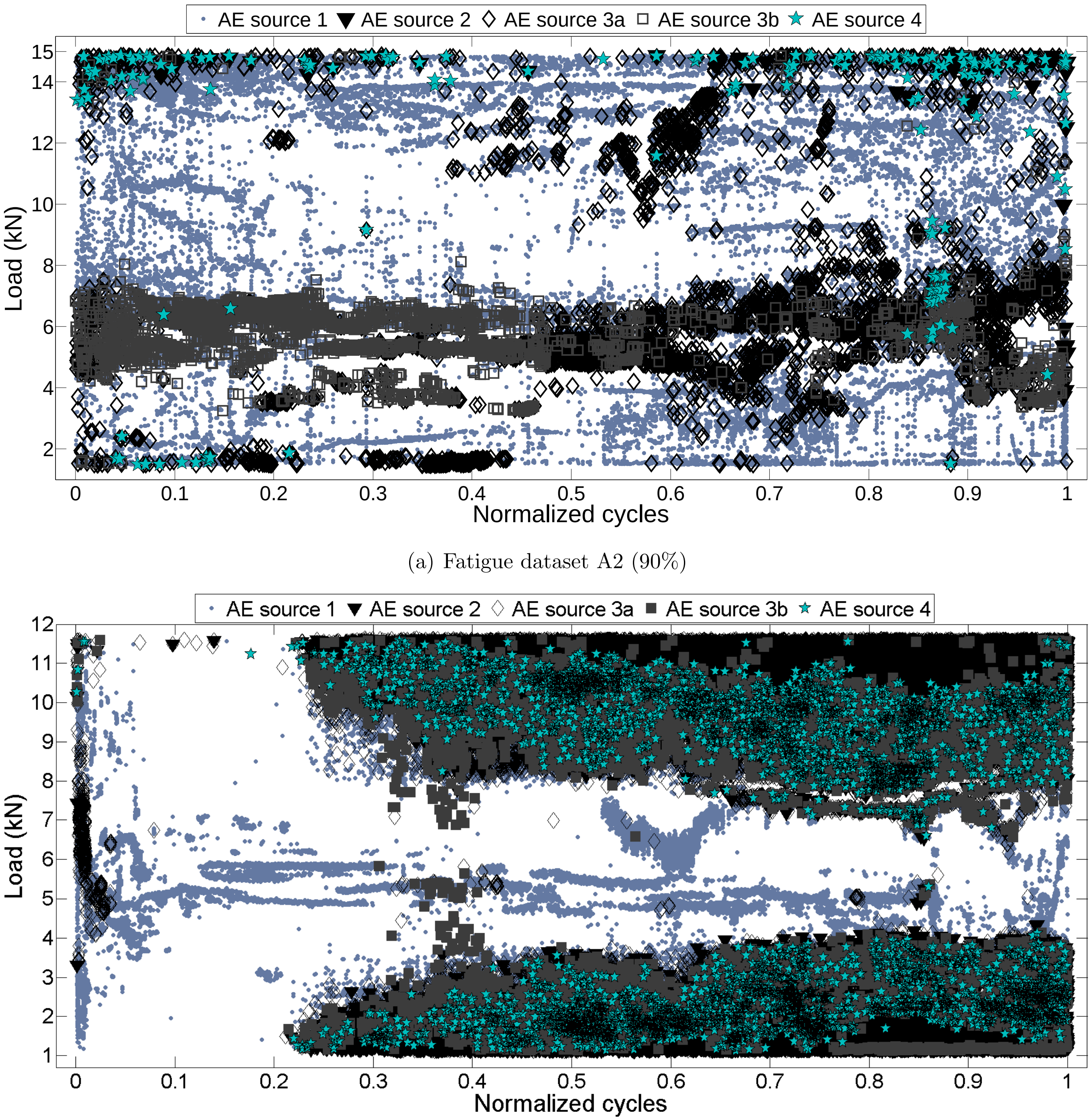


Figure 11: AE events during cyclic loading of specimen A3 (80%): Close-up view (a) at the beginning and (b) at the end of the test.



(a) Fatigue dataset A2 (90%)

(b) Fatigue dataset A4 (70%)

Figure 12: Visualization of classified AE events during cyclic loading A2 (90% of the ultimate strength) and A4 (70%)

List of Tables

Dataset N/ loading type	% of the ultimate strength	AE hits	Time-to-failure (s)
A1 / quasi-static	x	52,832	0.40E+3
A2 / fatigue	90	481,595	0.74E+3 (3.7E+3 cycles)
A3 / fatigue	80	1,682,434	4.11E+3 (2.0E+4 cycles)
A4 / fatigue	70	9,555,227	2.14E+4 (1.0E+5 cycles)

Table 1: Characteristics of AE datasets considered.

Cluster	AE source
1	Extraneous noise (external friction, hydraulic vibration, EMI)
2	Fiber-related damage (rupture of tows, pull-out)
3a	Friction-related source due to fatigue crack closure under cyclic loading
3b	Matrix-related damage (micro/macro cracking, splitting)
4	Interface-related damage (fiber/matrix)

Table 2: Assigned-to-damage clusters