

Application of an Unsupervised Pattern Recognition Approach for AE Data Originating from Fatigue Tests on CFRP

Dinh Dong DOAN *, Emmanuel RAMASSO, Vincent PLACET, Lamine BOUBAKAR,
Noureddine ZERHOUNI
FEMTO-ST Institute, Besançon, France

Abstract. Acoustic Emission (AE) technique is gaining more and more interest for structural health monitoring (SHM) in polymer-composite materials. Recent literature has shown that using appropriate pattern recognition techniques (PRT), the identification of the natural clusters of acoustic emission data can be obtained. Despite these recent and valuable advances and to achieve health assessment of composite materials, the scientific community faces two major challenges: (i) develop real-time approaches and (ii) propose clustering approaches able to process in in-service-like situation, i.e. in case of high AE activity generated simultaneously from many damage sources in material, from damage progression and cumulated damage and from noise.

This work investigates acoustic emission generated during tension fatigue tests carried out on a carbon fiber reinforced polymer (CFRP) composite specimen. The acoustic emission signals detected during testing were analyzed by feature based pattern recognition techniques. In previous studies, it was demonstrated that the presented approach for detection of AE sources related to noise or damage is suitable. In the present paper, AE data originating from different stress amplitudes of cyclic loading tests have been used to reveal the effectiveness and the capacity of generalization of the proposed methodology including noise removal, feature selection and automatic separation of AE events.

1 Introduction

AE testing has become a recognized nondestructive test (NDT) method, commonly used to detect and locate faults in mechanically loaded structures and components. Acoustic emissions (AE) are the stress waves produced by the sudden internal stress redistribution of the materials caused by the changes in the internal structure. Possible causes of the internal-structure changes are crack initiation and growth, crack opening and closure, dislocation movement, twinning, and phase transformation in monolithic materials and fiber breakage and fiber-matrix debonding in composites. Most of the sources of AE are damage-related; thus, the detection and monitoring of these emissions are commonly used to predict material failure.

With a huge amount of noisy 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 a damage mechanism. This analysis is a non-trivial task for two main reasons. First, AE signals are complex objects that must be characterized by multiple



relevant features. Second, there is no *a priori* knowledge of the acoustic signatures of damage events and these are assumed rather scattered.

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 [1], [2], [3]. In [1], [2], 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 [3], "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 and stop, slippage, grating between fracture surfaces (also referred to as "fretting"), and abrasion of load train. 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.

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 [4][5][6] are conditioned by Principal Component Analysis (PCA). The latter provides a feature space reduction as well as extraction of relevant components subset from the original features set. This algorithm assumes that the linear combination of features improves the relevancy of the principal components and that a large variance implies meaningfulness. Other approaches [7], [8], [9] rely on a specific subset of features such as energy, rise time, duration, amplitude [7] or reduce feature dimension space by using complete link hierarchical clustering in order to merge the correlated features into groups [8]. Those apply a greedy approach that generates all possible feature combinations and then selects the one which optimizes a given criterion [9], [10]. The goal of the criterion is generally to evaluate the quality of the partition provided by the clustering. Most of criteria are based on the Euclidean distance to assess the membership of an AE hit to a given cluster. Thus the applicability of this approach is limited to clustering algorithms which are based on the Euclidean distance. The PCA and K-means are theoretically related to each other as shown in [11]. The main reason to account for the performance of this couple is actually due to the link between both tools. An alternative approach based on the Gustafson-Kessel algorithm (GK) [12] was proposed in [13] which used 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.

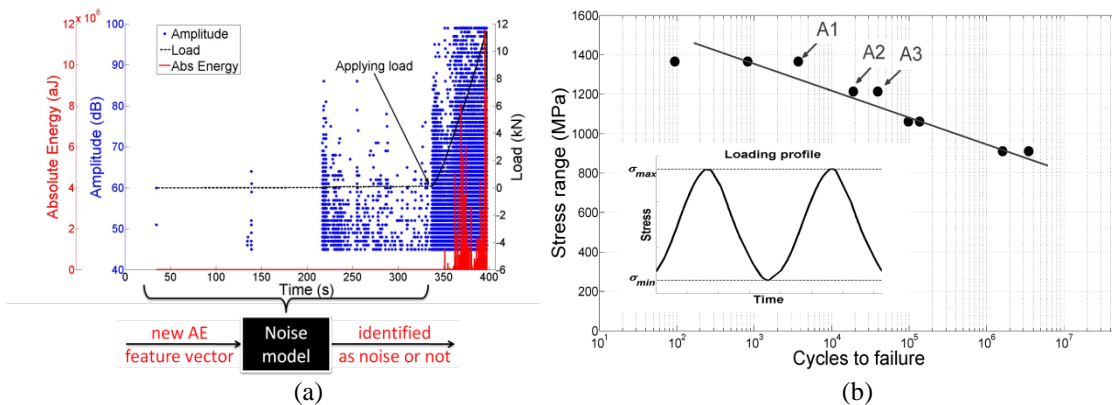
In previous work [14], a methodology to estimate the partition of AE data obtained in fatigue loading in presence of noise sources was developed. The methodology includes an automated filtering step and a sequential feature selection. The algorithm proposed in [13] for quasi-static tests was adapted to be applied on fatigue tests. The next sections are dedicated to present the experimentation and then remind briefly different steps of the proposed methodology as well as their illustrative results.

2 Experimentation

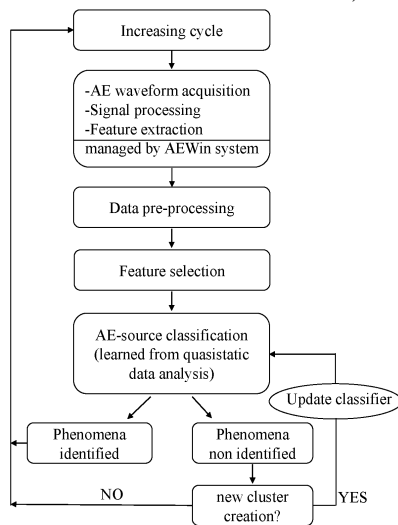
In this paper, health was assessed on carbon-epoxy composite split disks (6 layers $[(90^\circ)_2/\pm 45^\circ/(90^\circ)_2]$) when subjected to cyclic fatigue loading up to failure. Composite specimens were cyclically tested under a tensile/tensile sinusoidal loading with constant amplitude and frequency of 5 Hz and under constant stress ratio $R = 0.1$ at room temperature. Previously, quasi-static test (Pic. 1a) with a constant loading rate of $0.3 \text{ kN}\cdot\text{s}^{-1}$ was conducted to determine the static ultimate strength of the material. The tests were

performed according to ASTM D2290 "Apparent hoop tensile strength of plastic or reinforced plastic pipe by split disk method". 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. The sensors were coupled on the specimen faces using silicon grease. 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 μ s; HDT (Hit Definition Time) = 120 μ s and HLT (Hit Lock Time) = 300 μ s were identified using preliminary measurements. Many features such as absolute energy, counts, hits, amplitude, duration, frequency centroide were calculated from recorded waves. In this experimental configuration, AE sources are many. AE waves can be generated by damage in composite, by the cumulated damage, by the friction between the specimen and the clamps or by the hydraulic and EM systems used to perform the test.

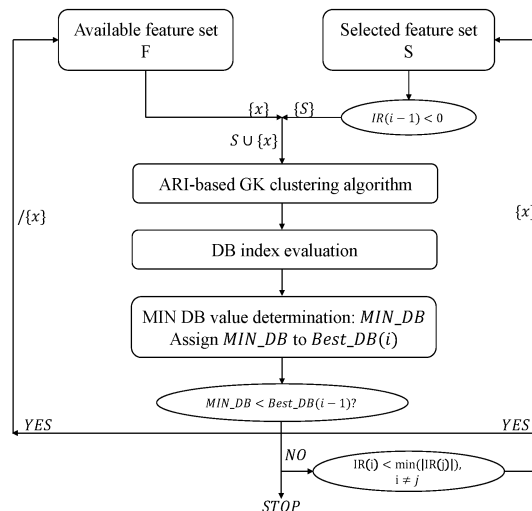
Three datasets A1, A2, A3 originating respectively from fatigue loading tests at 90% and 80% of the ultimate tensile strength (Pic. 1b and Table 1) as well as additional quasi-static dataset, denoted A0 are used to present the proposed methodology in this paper.



Pic. 1. Tensile testing: (a) quasistatic loading test; (b) fatigue cyclic loading at 90% (A1) and 80% (A2 and A3) of static ultimate strength



Pic. 2. Unsupervised damage detection methodology



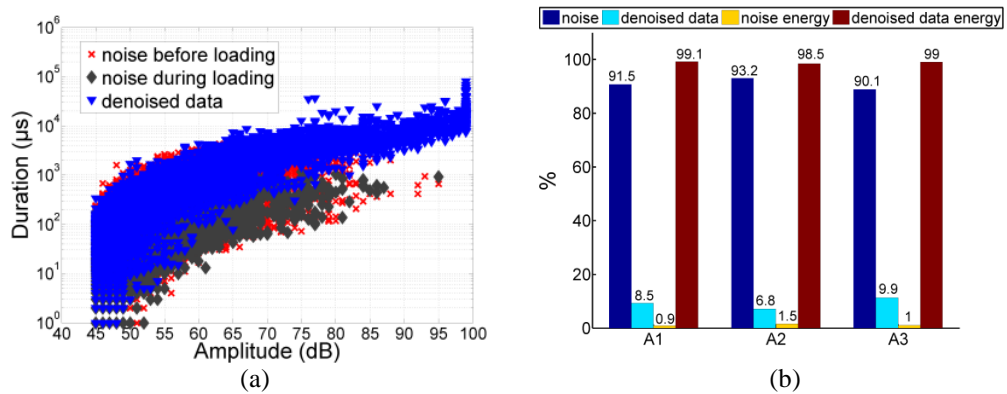
Pic. 3. Sequential feature selection diagram

3 Unsupervised pattern recognition

The flow chart of the methodology is shown on Pic.2.

3.1 AE fatigue data pre-processing

Despite the presence of an amplitude threshold of 40 dB detecting the arrival time of an AE signal, the acquisition system records a big amount of noisy data due to high sensitivity of AE sensors. Based on the assumption that there is no damage during the setting-in-place time, a noise model is built using multivariate statistical test based on the Mahalanobis distance as used in novelty detection [15]. This model is then used to filter out AE events during the test which have the same characteristics as the modeled noise. Pic.1a represents amplitude of AE hits originating from a quasistatic test and their energy. All AE signals detected before applying load having negligible energy are so associated to noise. AE hits during loading identified as noise possesses the same location and the same scattering as those before loading (Pic. 4a). Pic. 4b and Table 1 summarizes result of the denoising method on different AE datasets from fatigue loading tests. It is seen that in all cases, AE hits corresponding to noise represents the majority of recorded AE data but insubstantial phenomenon in terms of energy.



Pic. 4. Denoising method: (a) separation of denoised data from noise in case of a quasistatic test; (b) Percentage of noise vs. denoised data in terms of quantity (AE hits) and in terms of energy (aJ)

Table 1. Description of different AE datasets

Dataset / loading type	% of the ultimate strength	AE hits	AE hits filtered as noise	Denoised data	Time-to-failure (s)
A0 / quasistatic	X	52,832	45,936	6,896	0.40E+3
A1 / fatigue	90	481,595	440,762	40,833	0.74E+3 (3.7E+3 cycles)
A2 / fatigue	80	1,682,434	1,568,905	113,529	4.11E+3 (2.0E+4 cycles)
A3 / fatigue	80	3,502,311	3,155,142	347,169	7.86E+3 (3.9E+4 cycles)

3.2 Sequential selection algorithm of AE features

The goal of this section is to propose an automated technique to detect relevant feature subsets for clustering of AE events. In contrast to feature reduction procedures (e.g. based on correlation dendrogram in [4]) or exhaustive search of global optimal feature combinations in [9], the principle of the presented approach is to combine gradually each feature from an available feature space with an initial feature subset. The feature selection

is achieved by minimizing the value of Davies and Bouldin (DB) index [16] presented on Equation (1):

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left\{ \frac{d_i + d_j}{D_{ij}} \right\} \quad (1)$$

where 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 [8] or to evaluate feature subset partition [9]. Due to the way it is defined, as a function of the ratio of the within cluster scatter and the between-cluster separation, a lower value of this criterion means a good compactness and a good separation of dataset partition. Pic. 3 shows the diagram of the proposed algorithm based on a feature filtering approach [17]. Considering an initial selected feature set denoted S (empty by default), the algorithm will take each of available features from F to create a new subset with S . This subset is then partitioned by the clustering algorithm proposed in [13]. At the k^{th} iteration, a feature f_i in F is added to the current subset of features S_k , and the DB index DB_i of the partition obtained by the GK algorithm is computed. The subset of features S_{k+1} for the next iteration is given by $S_k \cup f_{i^*}$ with $i^* = \operatorname{argmin}_i DB_i$ and the partition is then evaluated by the DB criterion. The additional feature of subset that minimizes the value of DB index is selected as the relevant one. Thus, this feature will be removed from F to S . At each iteration, the procedure generates k new subsets if the number of features remaining in F is k , because each new subset contains the features from S plus a new one taken from the remaining ones in F . The algorithm stops when no new subsets can improve the DB criterion. For each iteration i , an improvement rate is calculated by Equation (2):

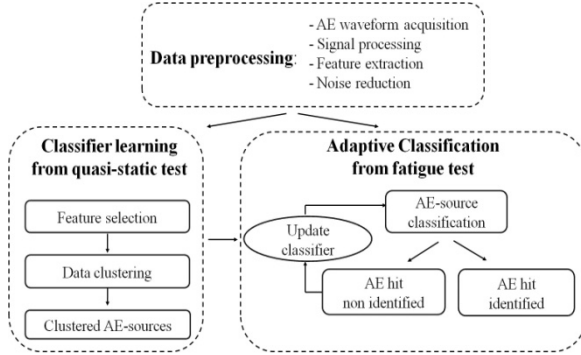
$$IR(i) = \frac{DB(S_i) - DB(S_{i-1})}{DB(S_{i-1})} \quad (2)$$

where $IR(i)$ is the improvement rate in the i^{th} iteration, $DB(S_i)$ and $DB(S_{i-1})$ are value of DB index of the best feature selection for the i^{th} and $(i-1)^{th}$ iteration. The sign of IR indicates if the DB criterion is improved (negative) or not (positive). In the last iteration j , *i.e.* $IR(j) > 0$, if $IR(j) < \min_{i \neq j} |IR(i)|$ then the best-DB-index feature can be added to S to establish the final selected feature set.

As depicted in the previous work [14], feature subset composed of absolute energy, MARSE and amplitude giving the best score of selection has been used as parameter of the proposed clustering algorithm in [13].

3.3 AE source detection

Quasi-static tests are first applied to obtain a relatively low amount of data compared to fatigue by assuming that damage sources are similar in quasi-static and fatigue tests. The GK algorithm as proposed in [13] is applied to estimate the parameters of a given set of k clusters. An additional $(k+1)^{th}$ cluster is estimated during fatigue to include all feature vectors located far from the previous k clusters. The average Mahalanobis-like distance (used in GK) in each cluster representing its radius is estimated after the quasi-static data partitioning. A feature vector obtained during fatigue belongs to the $(k+1)^{th}$ cluster if its distance to nearest cluster is above the corresponding radius. Pic. 5 resumes the developed procedure used for AE data analysis, showing its main steps.

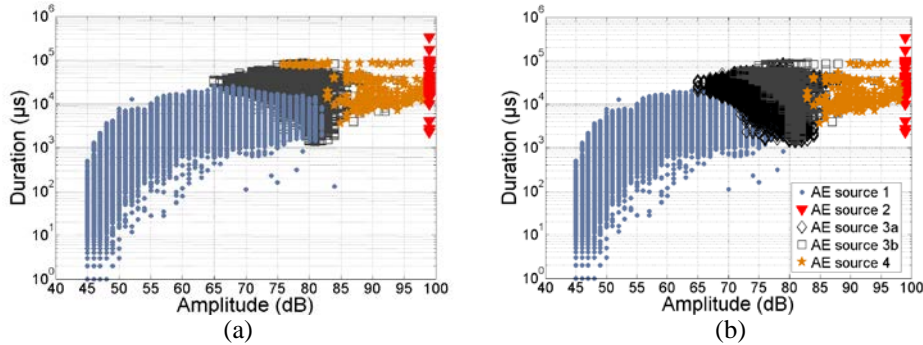


Pic. 5. AE data analysis flow chart

Table 2. Identification of AE sources

AE source	Possible origins
1	Extraneous noise (external friction, hydraulic vibration, EMI)
2	Fiber-related damage (rupture of single or bundle of fibers, pull-out)
3a	Internal friction due to crack closure
3b	Matrix-related damage (micro/macro cracking, splitting)
4	Interface-related damage (fiber/matrix debonding)

The denoised and selected feature subset in previous sections is now used to estimate the parameters of the clusters in quasi-static dataset A0 using the GK clustering algorithm proposed in [13]. Afterwards, testing directly this classifier with fatigue dataset may lead to an unsatisfactory separation (Pic. 6a). In fact, overlapping zones between clusters can be observed due to a new AE source that did not exist in the case of quasi-static testing. Consequently, according to this hypothesis, the proposed method offers an adaptive classification by creating a new class to obtain better segmentation.



Pic. 6: Testing phase: (a) direct classification without adaptation; (b) adaptive classification giving better separation

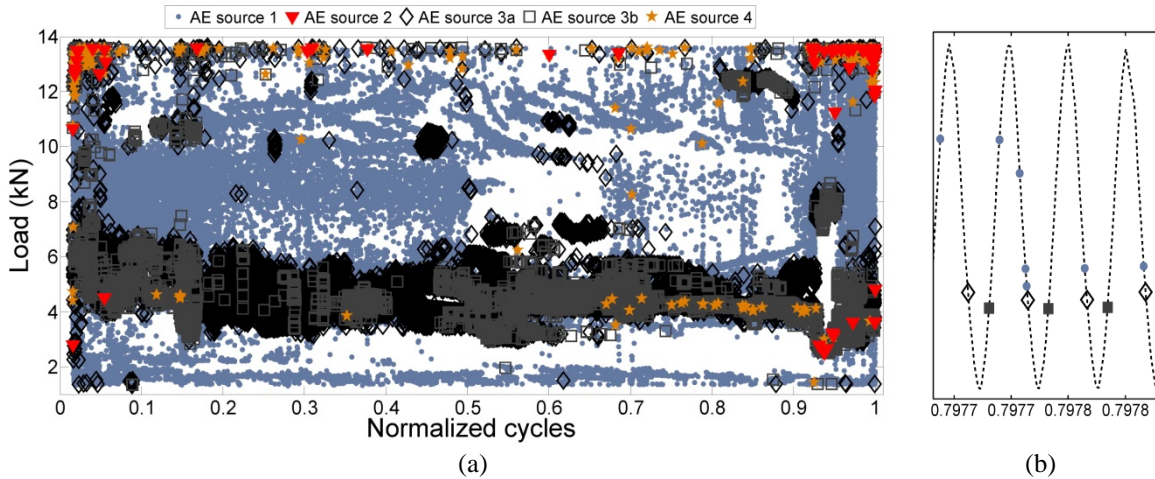
Pic. 6 shows the results of this adaptive method on a fatigue dataset and the separation between clusters is satisfying. The amplitude range of 5 detected AE-sources is clearly different. In the previous work [14], the AE sources were labeled using *in situ* observations, non-destructive testing and *post-mortem* observations. The identified sources are listed in Table 2.

4 Results and discussion

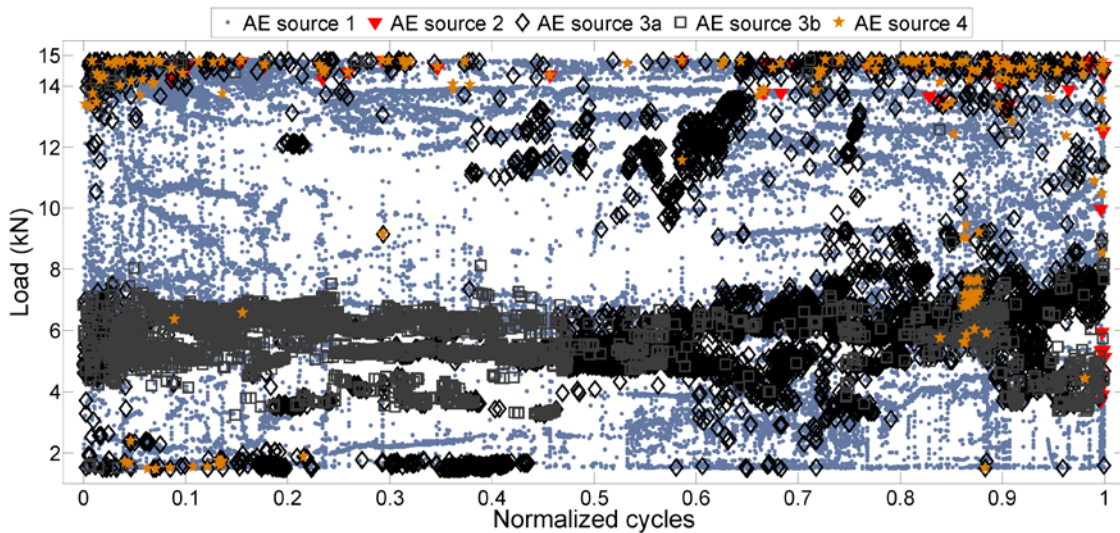
In this section, cyclic representation using the two dataset A1 and A2 are presented to perform the fatigue behavior of the material by analyzing the apparition of each AE source following loading/unloading cycles. Considering the dataset A2, it has been observed in Pic. 7a that during the first 10% of the fatigue life appears a high concentration of AE hits including all modes, followed by a steady state where all activities due to damage slow down. This stage is attributed to accommodation phase of composite material subjected to fatigue loading. Again, until 90% of the fatigue life, an acceleration and accumulation of high energy AE hits occurs up to the ruine of the specimen. Concerning AE sources 3a and 3b, they have an interesting interpretation: according to their apparition following loading and unloading, a repetitive phenomenon takes place all along of the test that AE source 3b locates mainly in loading phase and AE source 3a in unloading phase (Pic. 7b). The latter could correspond to internal friction or fretting between the faces of the previously

developed matrix cracks that could be associated to AE source 3b. Table 2 summarizes the AE sources assigned to possible origins related to damage mechanisms or not.

Again, testing with another cyclic fatigue dataset denoted A1, a similar behavior is obtained along the cyclic loading depicted in Pic. 8. Indeed, the temporal density of the AE events including high energy and high duration signals is rather high in the very beginning of loading process and increases again at about 60% of the specimen life. Although AE hits generated by acoustic source 3 is more scatter than the previous test, overlap phenomenon between this one and AE source 4 has also been detected by the proposed algorithm. This dissimilarity justifies the good applicability of cluster detection approach presented in this paper.



Pic. 7: Dataset A2 - Visualization of classified AE hits during cyclic loading: (a) whole test; (b) snapshot of cyclic loading



Pic. 8: Dataset A1 - Visualization of classified AE hits during cyclic loading

5 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 big data from fatigue tests with cluster adaptation. The proposed methodology permits to overcome problems related to computing approaches involving time consuming, computational cost and accuracy gain. The first results on real fatigue tests demonstrate that the proposed methodology allows identifying

some relevant clusters in loading and unloading phases, some clusters of different levels of energy which seems important for structural health monitoring.

Acknowledgment

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).

References

- [1] J. Henry, Z. Aboura, K. Khellil, et S. Otin, « Suivi de l'endommagement en fatigue d'un composite à renfort interlock carbone/époxy par émission acoustique », *Matér. Tech.*, vol. 100, n° 6-7, p. 643-652, août 2012.
- [2] J. P. Yu, P. Ziehl, et A. Pollock, « Signal identification in acoustic emission monitoring of fatigue cracking in steel bridges », p. 83471Z-83471Z, avr. 2012.
- [3] J. Yu, P. Ziehl, B. Zarate, J. Caicedo, L. Yu, V. Giurgiutiu, B. Metrovich, et F. Matta, « Quantification of fatigue cracking in CT specimens with passive and active piezoelectric sensing », 2010, vol. 7649, p. 76490R-76490R-12.
- [4] A. A. Anastassopoulos et T. P. Philippidis, « Clustering methodology for the evaluation of acoustic emission from composites », in *Journal of acoustic emission*, 1995, vol. 13, p. 11-22.
- [5] M. Johnson, « Waveform based clustering and classification of AE transients in composite laminates using principal component analysis », *NDT E Int.*, vol. 35, n° 6, p. 367-376, sept. 2002.
- [6] N. Godin, S. Huguet, R. Gaertner, et L. Salmon, « Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers », *NDT E Int.*, vol. 37, n° 4, p. 253-264, juin 2004.
- [7] R. Gutkin, C. J. Green, S. Vangrattanachai, S. T. Pinho, P. Robinson, et P. T. Curtis, « On acoustic emission for failure investigation in CFRP: Pattern recognition and peak frequency analyses », *Mech. Syst. Signal Process.*, vol. 25, n° 4, p. 1393-1407, mai 2011.
- [8] M. Moevus, N. Godin, M. R'Mili, D. Rouby, P. Reynaud, G. Fantozzi, et G. Farizy, « Analysis of damage mechanisms and associated acoustic emission in two SiCf/[Si-B-C] composites exhibiting different tensile behaviours. Part II: Unsupervised acoustic emission data clustering », *Compos. Sci. Technol.*, vol. 68, n° 6, p. 1258-1265, mai 2008.
- [9] M. G. R. Sause, A. Gribov, A. R. Unwin, et S. Horn, « Pattern recognition approach to identify natural clusters of acoustic emission signals », *Pattern Recognit. Lett.*, vol. 33, n° 1, p. 17-23, janv. 2012.
- [10] M. Halkidi, Y. Batistakis, et M. Vazirgiannis, « On Clustering Validation Techniques », *J Intell Inf Syst.*, vol. 17, n° 2-3, p. 107-145, déc. 2001.
- [11] C. Ding et X. He, « K-means Clustering via Principal Component Analysis », in *Proceedings of the Twenty-first International Conference on Machine Learning*, New York, NY, USA, 2004, p. 29-.
- [12] D. E. Gustafson et W. C. Kessel, « Fuzzy clustering with a fuzzy covariance matrix », in *IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, 1978, vol. 17, p. 761-766.
- [13] V. Placet, E. Ramasso, L. Boubakar, et N. Zerhouni, « Online segmentation of acoustic emission data streams for detection of damages in composite structures in unconstrained environments », in *Int. Conf. on Structural Safety and Reliability*, New York, NY, 2013.
- [14] D. D. Doan, E. Ramasso, V. Placet, M. L. Boubakar, et N. Zerhouni, « An Unsupervised Pattern Recognition Approach for AE Data Originating from Fatigue Tests on Polymer-Composite Materials », 7th European Workshop on Structural Health Monitoring, Nantes, France.
- [15] C. R. Farrar et K. Worden, « Unsupervised Learning – Novelty Detection », in *Structural Health Monitoring*, John Wiley & Sons, Ltd, 2012, p. 321-360.
- [16] D. L. Davies et D. W. Bouldin, « A Cluster Separation Measure », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, n° 2, p. 224-227, avr. 1979.
- [17] G. H. John, R. Kohavi, et K. Pfleger, « Irrelevant Features and the Subset Selection Problem », in *Machine Learning: Proceedings Of The Eleventh International*, 1994, p. 121-129.