Title: **Clustering acoustic emission time-series using unsupervised-shapelets** for Proceedings of the 12-th International Workshop on Structural Health Monitoring 2019

Authors :   Emmanuel Ramasso
            Pauline Butaud
            Vincent Placet
            Thomas Jeannin
            Fabrizio Sarasini

ABSTRACT

The sensitivity of AE sensors makes the AE technique very interesting for detecting damages at various scales as encountered in complex heterogeneous materials made of various constituents with different damage kinetics. However, there are unknowns behind the generation process of AE signals, as well as behind the modification of those signals along the propagation path until the sensors. Those make the interpretation of AE signals a difficult task in terms of pattern recognition. Even though some damage families are expected to occur for a given material under specific loading, there is a lack of knowledge to bridge the gap between AE signals collected on sensors with the related AE source. This accounts for the use of unsupervised learning when one is interested to discover relationships between AE data.

When using clustering for unsupervised learning, the output is a set of clusters representing a set of labels assigned to each datum and forming a partition. In AE work, clusters are estimated using features, computed from AE signals. Feature extraction has become a must-to-do step due to the use of particular clustering methods and to the fact that AE streaming has not been particularly analysed for damage sources identification.

The contribution of this paper is a new methodology for AE signal clustering which does not require feature extraction as usually done. It works directly on raw AE signals derives from several tools of the literature, namely unsupervised shapelets, anomaly detection and statistical modelling by Autoregressive Weakly Hidden Markov Models. When used in inference on unknown signals or streaming, those models allow to generate a set of novelty scores which are then processed by a chronology-based clustering algorithm to get a partition, accompanied by the uncertainty around clusters and a quantification of the robustness of the results obtained.

INTRODUCTION

When collecting acoustic emission (AE) streaming during mechanical tests, two

E. Ramasso, P. Butaud, V. Placet and T. Jeannin, FEMTO-ST institute, Univ. Bourgogne Franche-Comté, CNRS, 15B avenue des Montboucons, 25030 Besançon, cedex, France.

F. Sarasini, Sapienza-Università di Roma, Dept. of Chemical Engineering Materials Environment, Via Eudossiana 18, 00184 Rome, Italy.
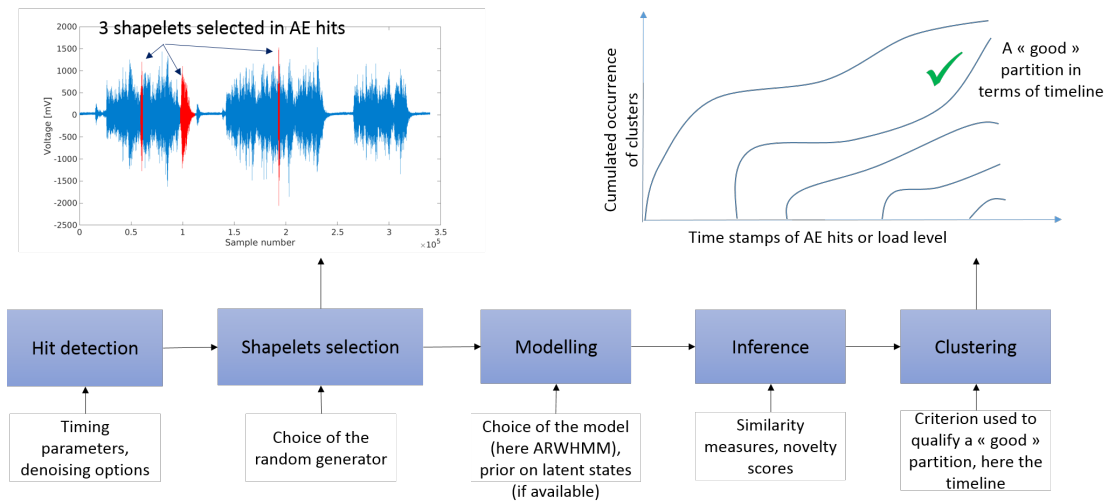
Figure 1. AE streaming (blue), 3 shapelets selected (red), and "good" clustering (after application of the methodology proposed) with respect to the timeline obtained.

main questions arise: How to manage the massive flow of data? And how to discover knowledge from it?

The main line of works consists in considering only features extracted from AE streaming (after hit detection), thus without considering the raw data. Behind this choice, one important assumption is made: Features are supposed to be representative enough of the content of AE signals and relevant enough to distinguish between damage families. This choice also makes the data interpretation task easier since it allows to use most of standard clustering methods available in the literature and in common softwares. Indeed, such methods mostly work with feature vectors organised as a matrix.

Additionally, the standard clustering methods used in the literature (used to assign a label to each AE signal in an unsupervised way) rely on assumptions which are counterintuitive when considering the specifities of AE data. The clustering of AE data is all the more complex than there is no ground truth (the true cluster for each AE signal is unknown) in order to validate the clustering results. This lack of prior knowledge prevents from finding the right optimal set of parameters of a clustering method. The set of parameters is called a parameterisation and is made, for example, of the number of clusters, the subsets of features and all other parameters used in a clustering method.

This paper suggests a breakthrough approach for interpreting AE data that works directly on raw signals, without requiring the standard feature extraction step. The method is based on the idea that only a global information can be drawn from AE streaming, without looking for a precise relationship with damages. In place, the global information is about how the clusters are organised along time. For that, the method aims at discovering a set of parameterisations that reflects a good timeline of AE events.

After the presentation of AE data specifities (Section 1), the feature-free method for AE data interpretation is presented in Section 2. The clustering phase is also enhanced by considering a modification of a recent criterion to select a set of good parameterisations. Section 3 is dedicated to an illustration on impacted composites.

SPECIFICITIES OF AE DATA

Specificity 1 *AE signals constitute time-series. AE transients are thus not iid and could either be indexed by a time-stamp or by the value of a variable representing damage accumulation.*

Using Kmeans or Gaussian Mixture Models on AE features imply that AE transients are supposed independent and identically distributed (iid). This differs from the Markov assumption which supposes that the probability of occurrence of a given set of transients is a function of the previous transients in the sequence.

Specificity 2 *AE signals are generally unevenly-spaced in time events.*

The principle of the AE technique is to passively detect the propagation of a stress wave, without knowing when a damage would occur. The time-stamps (or index) between AE transients are likely to be irregularly spaced (or only in very special cases).

Specificity 3 *The statistical process behind AE data generation is non-stationary.*

AE signals characteristics are modified by damage accumulation and AE generation and propagation are theoretically dependent on material properties and geometries which are likely to change in practice due to loadings.

Specificity 4 *The characteristics of AE sources are generally unknown.*

The numerical generation of a full AE streaming that would correspond to a *twin* of a streaming obtained experimentally, has never been done due to both computational burden and to lack of knowledge.

FEATURE-FREE RAW WAVEFORM-BASED AE DATA CLUSTERING

U-shapelets and adaptation to AE signals

Shapelets have been proposed for supervised classification [1] and was extended to unsupervised learning in [2]. The idea of unsupervised-shapelets (u-shapelets) is to build a distance map which contains the distance between each of the u-shapelets and all the time series in the dataset. With m u-shapelets for a dataset of N time series, the size of the distance map is $N \times m$, each column being a distance vector of a u-shapelet.

Definition 1 (Adaptation to AE data interpretation) *An AE streaming is decomposed into N AE signals using a wave picking algorithm forming a dataset D. Each AE signal correspond to a time-series. An "AE-shapelet" M is a mathematical model of an AE signal selected in the set (randomly in the case of unsupervised learning). The distance map then contains the reconstruction errors of AE signals in D using model M. If there are m models (using various AE-shapelets), then the distance map is a $N \times m$ matrix.*

Algorithm 1 Principle of the method.

Require: A dataset D (full streaming data)
Ensure: A partition of the dataset, uncertainty around clusters, robustness of the results

    % Hyper-parameters
1: θ_{wav} *% Parameters of the hit detection method*
2: θ_{clus} *% Parameters of the clustering method*
3: $R \leftarrow 30$ *% number of repetitions for bootstrapped ensembles*

    *% **Step 1: Hit detection**, use the method in [3]*
4: $AE\_hits \leftarrow \text{hit\_detection}(D, \theta_{wav})$

    *% **Step 2: Model estimation***
    % The following for loop can be done using parallel processing
5: **for** $i \in [1, R]$ **do**

6:     $ae\_signal_i \leftarrow$ Pick an AE hit at random in AE_hits
      % The random number generator can be set according to the application

7:     $model_i \leftarrow$ Estimate the parameters of a model for ae_signal_i
      % The model is here an Autoregressive Weakly Hidden Markov Model (AR-WHMM) [4]. The number of regressors are set using partial autocorrelation.

8:     **for all** AE_hits (indexed by j) **do**

9:        $inference_j \leftarrow$ Apply $model_i$ to reconstruct ae_signal_j

10:      $C_{ij} \leftarrow$ Compare the reconstruction and the real signal
        % The comparison makes use of multiple similarity measures

11:      $M_{ij} =$ Update distance map with C_{ij}

12:     **end for**

13: **end for**

    *% **Step 3: Clustering***
    % Perform clustering using [5,6] incorporating the modification of [7]. This method evaluates the uncertainty of clusters and the robustness.
    $clusters, uncertainty, robustness = \text{Clustering}(M_{ij}, \theta_{clus})$

14: **return** $clusters, uncertainty, robustness$

Is the distance map similar to a feature matrix? By seeing the big picture of the method, the distance map is actually computed from raw signals. Therefore, it is a different form of knowledge extracted from raw data. With this method, this is no more necessary to question the use of particular features: energy-based? time-based? amplitude-based? or frequency-based? ... Indeed, the question is now moved to identify which model one may use to extract the best content of AE signals in order to compute this distance map: statistical? neural-based? ... This allows to move the problem

from low-level data (features) to high level knowledge (model). The expansion of machine learning through the recent years make it possible to exploit many models not yet explored for AE interpretation.

General algorithm and some choices made in the proposed method

The proposed method is decomposed as in Algorithm 1. The main steps are the following. Firstly, after performing hit detection using a wavelet-based method proposed in [3], a set of AE signals are drawn randomly. Here the random number generator can be set according to the application. For example, one can give preference to AE signals occurring at the beginning of the tests (similarly to novelty detection [8]), or to some given levels of loading during fatigue tests. Secondly, for every AE signal sampled, a model is built. After several tests made on various datasets, we selected a statistical model called Autoregressive Weakly Hidden Markov Model (ARWHMM) proposed in [4]. The parameters of this model are set using partial autocorrelation (for regressors) [9] and using AIC citerion (for mixture components). Thirdly, the models are then applied on all AE signals to build the aforementionned distance map. We suggested to use several novelty scores proposed in [10]. The distance map is finally taken as input of a clustering method based on [5, 6] and incorporating a modification proposed in [7].

Related work

Clustering AE signals using features is a widely coverered topic [11]. However, questioning the use of features for unsupervised AE interpretation is more recent. For instance, [12, 13] were interested in improving source location using AE signals. [13] used dynamic time warping (DTW) to evaluate the similarity between AE signals for multi-source events localisation. Some decades ago, [14] suggested to use a paired-sample statistical analysis using spectral densities of AE signals to distinguish between different fracture modes. [15] suggested to evaluate change detection [16] methods to detect transitions in a material behavior from AE.

EXPERIMENTS

The method has been applied on various datasets. We here show a result on impacted carbon fibre reinforced laminates. The material and method are presented in [10]. After impact, some specimens were scanned using a nano-micro-tomograph to evidence the initial state of damage (Fig. 2-a)). The cross-section depicts a typical cone-shaped profile of defects initiated by the impact. It shows a network of delaminations interlinked by intralaminar matrix cracks occurring parallel to the direction of the fibres and forming a "spiral staircase" of delaminations around the impact site [17].

This complex initial state of damage is the source of stress concentrations under loading. Both static and fatigue tests were performed using a four point bending device and in accordance with ASTM D 7264. Only results on quasi-static test are shown below.

Figure 3 summarizes the results of clustering by representing the sequence of clusters. Black curves represent the logarithm of the cumulated occurrence of clusters against time. The stress is represented (red) together with the cumulated MARSE (blue)
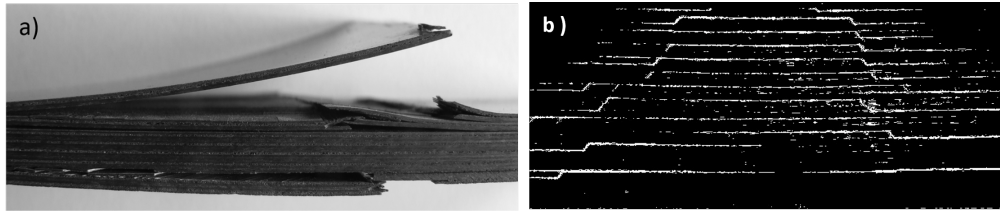
Figure 2. a) Post-mortem profile after a fatigue test of an impacted sample. b) Tomography image of a specimen after impact (obtained through MIFHySTO platform).
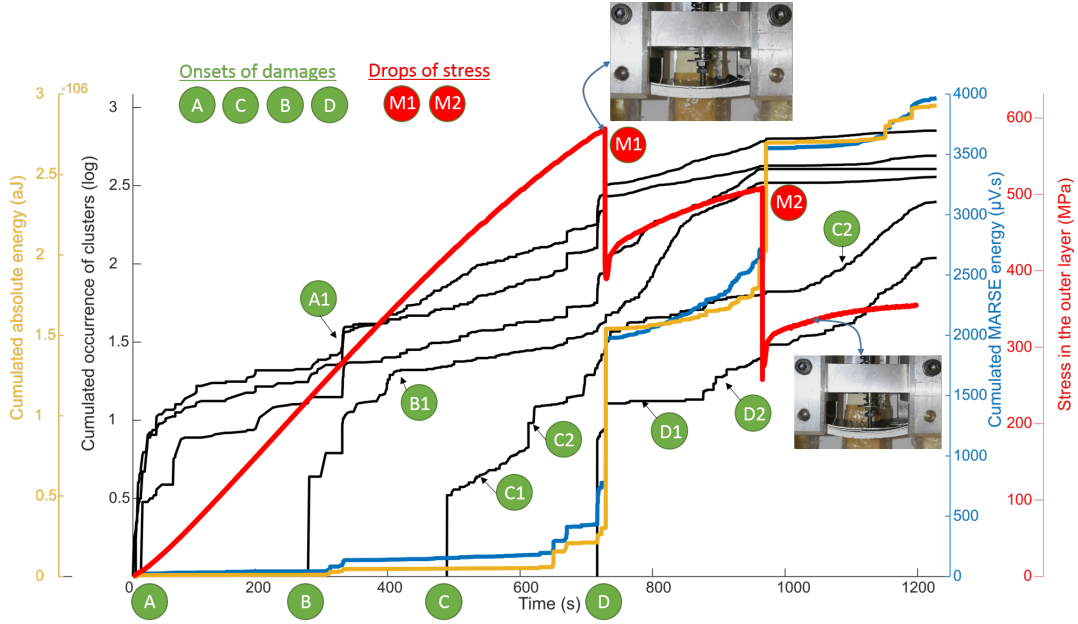

Figure 3. Quasi static test: Damage sequence estimated. Cumulated energy (MARSE and absolute) shown for information but not used in clustering.

and absolute (yellow) energies. Those features are shown since they are related to the delamination process [18] but were not used in clustering. Only the distance map made of novelty scores) were used in clustering, and were computed directly from the application of the statistical models (ARWHMM) using randomly selected AE signals.

From Figure 3, we can observe the first macro-delamination at stage D and ply failure (see image at top, close to M1) when the stress drops suddenly (M1 at 730 s) while AE energies increase importantly. At this stage, the specimen still holds but is highly damaged. The pattern recognition method evidences a scenario to explain the occurrence of stage (or state) D through a sequence:

$$A_0^{200MPa} \rightarrow B_{200}^{450MPa} \rightarrow C_{450}^{600MPa} \rightarrow D_{600}^{ruin}$$

where $S_{\sigma_s}^{\sigma_f}$ is a state starting at stress σ_s and finishing at σ_f. To each state corresponds a set of substates which depict interesting patterns. State A is made of three clusters starting early. State A is "followed" by a new state, B, at $t = 290$ s (corresponding

to 200 MPa). The onset of state B is "followed" by a clearly visible slope change in A (see A1 on the figure). Just after, B shows a slope change too (B1). From B1 until C, we can see that the cumulated occurrence (in log-scale) evolves gradually for all clusters likewise the damage process is relatively linear. A new chapter is opened in this scenario with state C. After the onset of C, an exponential progression is observed (C1) followed by several "steps" (C2) which correspond to sudden releases of AE signals and corresponding to early signs of the occurrence of a new state, D (macro-delamination). D1 is a plateau which represents stress accumulation until a new advance, at D2, of the delamination which is followed by a new drop in the stress curve (M2). After M2, only D and C evolve importantly, the remaining clusters remaining almost steady. Therefore, C and D are likely to be related to delaminations, ply failure and fibre-breakage (see multiple ply delaminations through the thickness as well as ply failures, in the bottom-right image). Since C started clearly earlier than D, the cluster related to C can be used for anomaly detection purpose, allowing to raise an alarm before the occurrence of D.

CONCLUDING REMARKS

Changing our viewpoint is a necessary condition for developing novel and break-through data analysis methods specifically dedicated to AE data interpretation. This paper suggests to consider only raw signals for that task. Bypassing the feature extraction step was possible by using some concepts inherited from data-mining (shapelets) and anomaly detection together with advanced unsupervised learning based on bootstrapped ensembles. The patterns obtained in the application described in this paper (timeline and evolution of states) were also observed for multiple specimens, various loading and different materials.

ACKNOWLEDGEMENT

REFERENCES

1. Ye, L. and E. Keogh. 2009. "Time series shapelets: a new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 947–956.

2. Zakaria, J., A. Mueen, and E. Keogh. 2012. "Clustering time series using unsupervised-shapelets," in *12th Int. Conf. on Data Mining*, IEEE, pp. 785–794.

3. Kharrat, M., E. Ramasso, V. Placet, and M. Boubakar. 2016. "A signal processing approach for enhanced Acoustic Emission data analysis in high activity systems: Application to organic matrix composites," *Mechanical Systems and Signal Processing*, 70:1038–1055.

4. Juesas, P., E. Ramasso, S. Drujont, and V. Placet. 2016. "On partially supervised learning and inference in dynamic Bayesian networks for prognostics with uncertain factual evidence: Illustration with Markov switching models," *International Journal of Prognostics and Health Management*, (7):1–9.

5. Ramasso, E., V. Placet, and M. L. Boubakar. 2015. "Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites," *IEEE Transactions on Instrumentation and Measurement*, 64(12):3297–3307.

6. Ramasso, E., V. Placet, X. Gabrion, and M. Boubakar. 2017. "A robust-to-parameterization clustering fusion approach for monitoring damage onsets and growth using acoustic emission," vol. 2, pp. 2162–2169.

7. Chandarana, N., E. Ramasso, C. Soutis, and M. Gresil. 2019. "Damage identification in composites through acoustic emission monitoring," in *9th Int. Conf. on Acoustic Emission (ICAE-9) & 61st Acoustic Emission Working Group Meeting (AEWG-61)*, Chicago, IL, USA.

8. Farrar, C. and K. Worden. 2013. *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, Ltd.

9. Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time series analysis: forecasting and control*, John Wiley & Sons.

10. Ramasso, E., P. Butaud, T. Jeannin, F. Sarasini, V. Placet, N. Godin, J. Tirillò, and X. Gabrion. 2019. "Learning the representation of raw acoustic emission signals by direct generative modeling and its use in natural clusters identification," *Engineering Applications of Artificial Intelligence*, submitted.

11. Ono, K. 2018. "Review on structural health evaluation with acoustic emission," *Applied Sciences*, 8(6):958.

12. Holford, K. M., M. J. Eaton, J. J. Hensman, R. Pullin, S. L. Evans, N. Dervilis, and K. Worden. 2017. "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, ISSN 0376-0421.

13. Chen, S., C. Yang, G. Wang, and W. Liu. 2017. "Similarity assessment of acoustic emission signals and its application in source localization," *Ultrasonics*, 75:36–45.

14. Egan, D. and J. Williams-Jr. 1978. "Acoustic-emission spectral analysis of fiber composite failure mechanisms," Tech. Rep. 2938, NASA Contractor Report, MIT, Cambridge, Massachusetts, USA.

15. Martin-Del-Campo, S. and F. Sandin. 2017. "Online feature learning for condition monitoring of rotating machinery," *Engineering Applications of Artificial Intelligence*, 64:187 – 196.

16. Kuncheva, L. I. 2008. "Classifier ensembles for detecting concept change in streaming data: Overview and perspectives," in *2nd Workshop SUEMA*.

17. Borstnar, G., F. Gillard, M. Mavrogordato, I. Sinclair, and S. Spearing. 2016. "Three-dimensional deformation mapping of Mode I interlaminar crack extension in particle-toughened interlayers," *Acta Materialia*, 103:63 – 70, ISSN 1359-6454.

18. Silversides, I., A. Maslouhi, and G. LaPlante. 2013. "Acoustic emission monitoring of interlaminar delamination onset in carbon fibre composites," *Structural Health Monitoring*, 12(2):126–140.