

Contribution of functional approach to the classification and the identification of acoustic emission source mechanisms

O.I. Traore¹, P. Cristini¹, N. Favretto-Cristini¹, L. Pantera², P. Vieu³, S. Viguier-Pla^{4&3}

Abstract In a context of nuclear Reactivity Initiated Accident, we describe acoustic emission signals, for which a problem of classification is open. As classical approaches with a reduced number of variables do not give satisfactory discrimination, we propose to use the envelopes of the received signals. We perform a kmeans clustering and discuss the first results of this approach.

Introduction

Several non-destructive methods are used for the monitoring of nuclear safety experiment reactors. Among them the acoustic emission (AE) technique is of major interest. It has the advantage of being simple to adapt to nuclear-oriented purposes and allows a quasi-real-time monitoring of experiments. One goal of the AE testing is to process the AE signals recorder from the reactor in such a way that they can be associated with specific physical source mechanisms occurring in the tested structure or material. In general, this characterization is done by computing classical AE variables from the received signal like the energy, the rise time, the duration... When the number of recorded signals to process becomes very important, classical data-mining methods based on these variables are used to classify them and then the physical source mechanism associated with each class is identified. However, depending on the diversity of source mechanisms (cracks, fractures, delaminations ...) and the type material (nuclear fuel, zircaloy, inox...) very different type of variable constructed from the AE signals can be discriminant [4, 3, 2, 9]. In the case of

¹Aix-Marseille Univ., CNRS, Centrale Marseille, L.M.A., France, e-mail: toumarissika@gmail.com · ²CEA, DEN, DER/SRES, Cadarache, F13108 Saint-Paul-Lez-Durance, France, e-mail: laurent.pantera@cea.fr · ³Equipe de Stat. et Proba., Institut de Mathématiques, UMR5219, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France, e-mail: vieu@math.univ-toulouse.fr · ⁴ Université de Perpignan via Domitia, LAMPS, 52 av. Paul Alduy, 66860 Perpignan Cedex 9, France, e-mail: viguier@univ-perp.fr

nuclear safety experiment which is of interest in this article, the test device is composed of several type of materials and interact with a very complex environment leading to a difficulty to get enough discriminant variables for a very heterogeneous sample of source mechanisms.

Unsupervised classification based on functional approaches has been, in literature, widely of interest, as the problem on the overall curves shape, their origins and regularity. Some authors use a basis expansion, and process a clustering based on the coefficients of expansions (see for example Abraham et al. [1] for the use of a B-spline basis or Giacomini et al.[6], for the use of wavelet basis). Other use the scores of the functional principal components analysis (as for example, Peng and Muller [11]).

In this work, we present a functional approach based on the envelopes of the received signals. This choice rather than the use of the raw received signals has been motivated by their more effectiveness and robustness to estimate the time-delay between signals recorded from the same source mechanism at two different sensors, a crucial variable for source identification.

1 Context of the study and Raw data processing

Reactivity Initiated Accident (RIA) is a nuclear safety experiment which involves an unexpected and very fast increase in fission rate and reactor power due to the ejection of a control rod. The power increases may damage the fuel clad and the fuel pellets of the reactor. The French Alternative Energies and Atomic Energy Commission (CEA) operates a pool-type reactor dedicated to fuel behavior study in RIA conditions. During these RIA experiments, the test device is equipped with two AE sensors (microphones) allowing to record information about the fuel behavior.

Since 1993, fourteen RIA experiments have been operated by the laboratory in charge of the preparation and the realization of the experiments, after each of them, the raw microphone signals are processed in order to give to the experimentalists the first estimations about the fuel behavior. This experiment result analysis process is composed of several steps. In the first one, the physical measurements performed by the two sensors are converted into numerical signals (fig. 1).

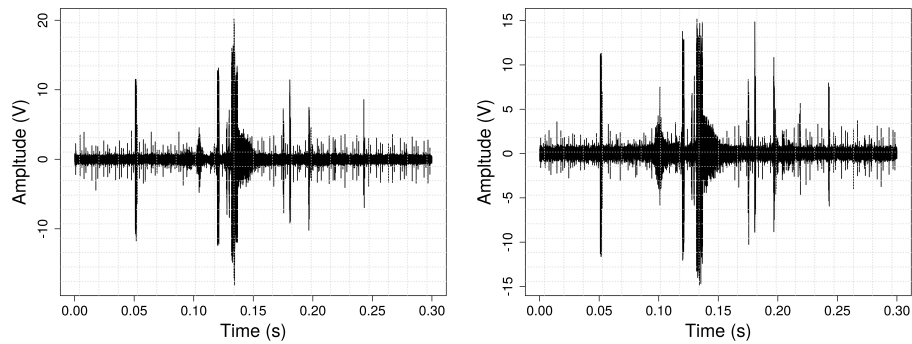


Fig. 1: Microphone numerical signals recorded during an experiment realized in 1994. Microphone 1 (left) and microphone 2 (right)

Then, a hits (events) detection function is used in order to isolate segments of signal which are potentially associated with physical source mechanism of interest. This hits detection strategy based on a moving variance consists, for each microphone, on :

1. defining a learning sample of signal corresponding to the background noise of the reactor and on fixing a threshold equal to the variance of this noise sample
2. computing the moving variance vector associated with the microphone signal
3. identifying the segments of the microphone signal corresponding to a threshold violation.

After the realization of the hits detection for the two microphones, their results are merged in order to get the same hits starts and ends times. At the end of this first, a certain number of hits, so of potential physical source mechanisms of interest, are associated with the experiment (fig. 2).

In order to implement a statistical classification algorithm able to perform an automatic source mechanism identification for the future experiments, the results of the hits detection process for the fourteen first experiments have been gathered to form a sample of 161 hits.

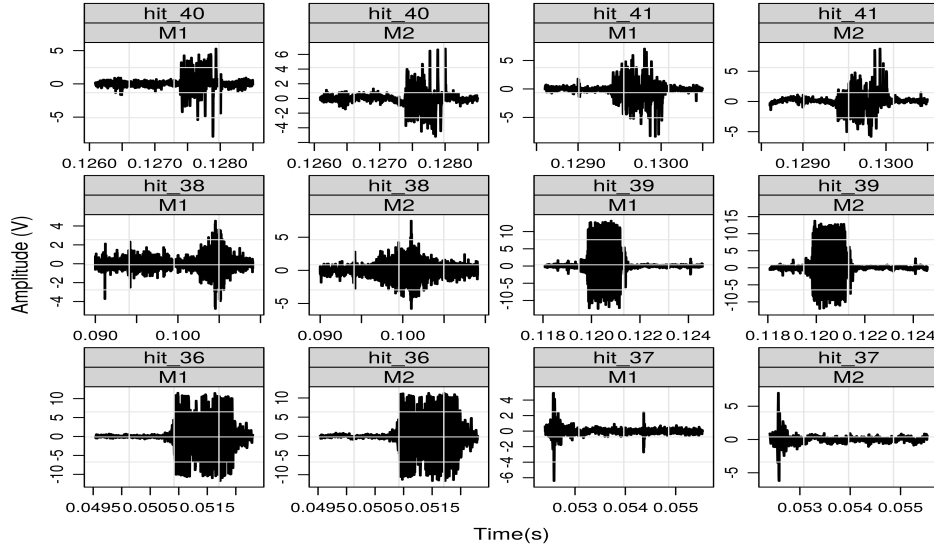


Fig. 2: Representation of the first six hits detected after the application of the moving variance strategy on the experiment signal presented in figure 1

Functional dataset creation

The 161 hits resulting from the application of the hits detection strategy on the fourteen experiments constituted a dataset of signal of very different duration (length). In order to get a dataset with discretized curves of the same length and to cope with some limitations dues to changes on signal sampling ratio and differences of period between the two microphone signals, we have chosen to perform the classification on the re-sampled envelopes of the hit signals. The envelope of a signal has got the advantage of being characteristic of a signal and smoother than the raw data. Here is the method for the computation of this envelope [10].

Let $x(t)$ be the real-valued signal associated with a given hit. The analytic signal $z(t)$ of $x(t)$ is defined as follows:

$$z(t) = x(t) + iy(t), \quad (1)$$

where $y(t)$ is the Hilbert transform of $x(t)$, that is

$$y(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (2)$$

Then the signal $z(t)$ can be written as follows:

$$z(t) = A(t)e^{i\phi(t)} \quad (3)$$

The function $A(t)$ is named the envelope of the signal $x(t)$. It can also be written $A(t) = \sqrt{x^2(t) + y^2(t)}$.

Then, the envelope based on the Hilbert transform of each hit is computed and its length is coerced to 500 observation points by resampling (fig. 3).

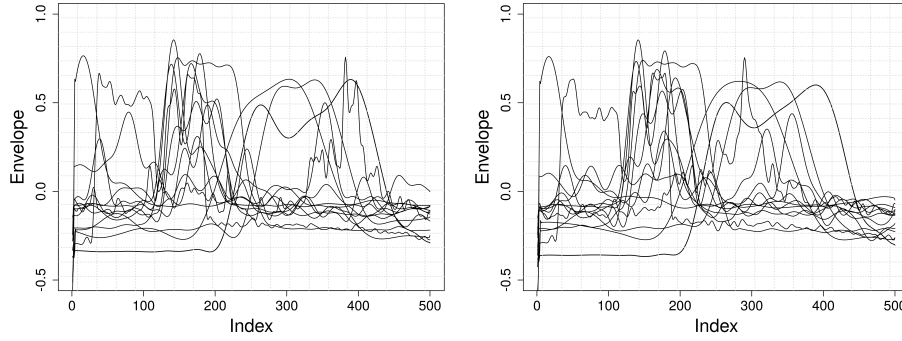


Fig. 3: Illustration of hits envelopes dataset associated with the experiment signals presented in figure 1 and hits sample presented in figure 2. Microphone 1 (left) and microphone 2 (right)

At the end of this raw data processing step, each hit can be considered as the realization of a functional random variable written as:

$$X = \{X(t), t \in \mathcal{T}\} \quad (4)$$

where \mathcal{T} is the interval of hit observation time.

Of course, this formulation is a misnomer as in practice we have a finite set of observation points $\in \{1, \dots, 500\}$. Then, in practice, X_i the i^{th} realization of X should be written as:

$$X_i = \{X_{ij}, j = 1, \dots, 500\} \quad (5)$$

In order to transform the envelopes (discretized curves) into functional data, several methods exist. According to Jacques and Preda [8], the most common solution to this problem is to consider that sample paths belong to a finite dimensional space spanned by some basis of functions. In our case, as the sampling rate of signal were about $2.5\mu s$, the observation grid is very fine, then we can consider each envelope as a continuous curve [5].

2 Unsupervised classification method and parameters settings

Based on the classification of clustering methods for functional data proposed by Jacques and Preda [8], we distinguish four principal approaches. Among them, the non parametric one is a priori very well adapted to the envelopes. Let consider the usual non parametric kmeans clustering method for which a description can be found in several articles (see for example [13, 7]). The key point when applying such algorithm in non parametric functional context is of course the choice of a function to measure the proximity between two curves. As the envelopes are smooth curves, semi-metrics are good candidates to do so [5].

Let X be the functional random variable associated with the envelopes taking its values in an infinite dimensional semi-metric space (E, d) . The semi-metric considered in order to compute the proximity between two envelopes x_i and $x_{i'}$ is defined as follow :

$$d_l(x_i, x_{i'}) = \left(\int_{\mathcal{I}} (x_i^{(l)}(t) - x_{i'}^{(l)}(t))^2 dt \right)^{1/2}, \quad (6)$$

where \mathcal{I} is the interval where X is defined, and $x_i^{(l)}$ is the l derivative of x .

Recalling that in our case the observation grid is enough fine to consider the discretized envelopes as good approximation of functional ones, $d_l(x_i, x_{i'})$ is estimated by

$$\hat{d}_l(x_i, x_{i'}) = \left(\sum_{t=1}^{500} (x_i^{(l)}(t) - x_{i'}^{(l)}(t))^2 \right)^{1/2} \quad (7)$$

i indexing the n realizations of X , and t indexing the $T = 500$ points of observation of each realization.

Then performing a successful clustering is equivalent to choose the best value of the order of derivative l of the curves, this is discussed by Ferraty and Vieu [5]. An over crucial parameter being of course the number of cluster. To do so, some strategies are proposed in the literature among which the use of silhouette values [12].

Four our purpose we have given priority to physical considerations. Indeed, except for non physical source mechanisms taking place out of the test device, the two microphones are supposed to give the same signals. Then, the derivative order have been chosen in order to get the closest possible classification result between the two microphones. Furthermore, as experimentalist have a good a priori about the characteristics many source mechanisms, the number of clusters have been set according to their appreciation and give the priority to physical interpretability of each cluster.

3 Some analysis results

The figure 4 show the result of the clustering in 6 clusters of the 161 envelopes of the dataset by using derivative of order 2 of the curves. It highlight a very important closeness between envelopes belonging to the same cluster and a good concordance between the two microphone results.

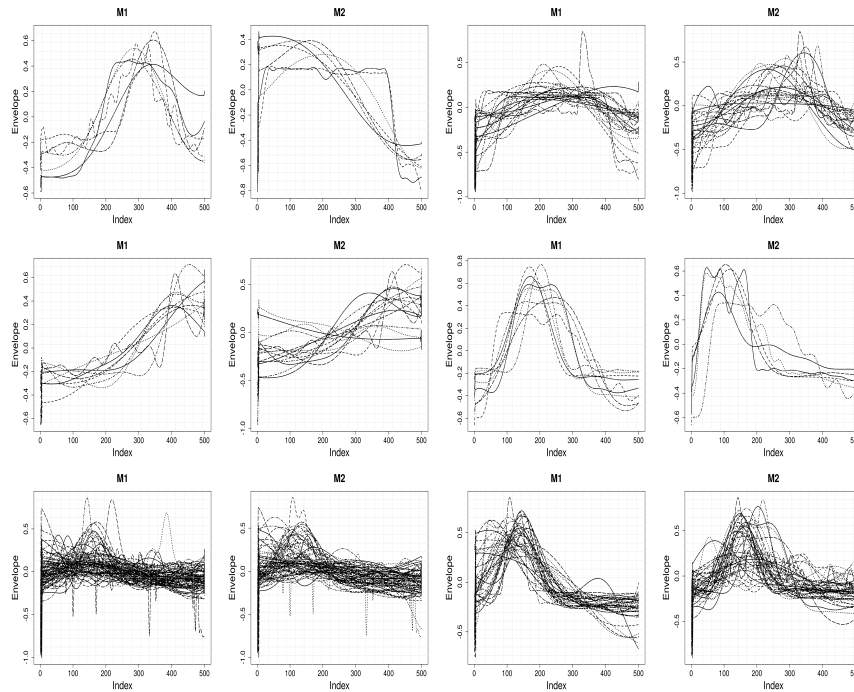


Fig. 4: Results of kmeans clustering of the datasets of the two microphones. Clusters from 1 to 6 to be read from top to bottom and left to right.

Back to physical source mechanisms

The natural continuation of the process is the identification of source mechanisms associated with each cluster. For some clusters, this is very simple after a return to the raw detected signals. For example, hits associated with cluster 4 are supposed to correspond to fuel clad failure source mechanism as those of cluster 5 are correspond to noise (not associated with source mechanism of interest).

4 conclusion

In this work, we have study the opportunity of applying functional data clustering to classify acoustic emission signals recorded from two microphones during nuclear safety experiments. A functional dataset construction process in several steps have been presented. We propose a hit detection strategy based on moving variance. The detected hits for two microphones are then merged and their resampled envelopes computed. A kmeans classification algorithm have been performed, the number of classes being chosen in order to get the best physical interpretability, we have chosen the semi-metric taking into account the non regular character of the envelopes and the necessity to get close classification result between the two microphones. The results confirms the potential of the functional approaches for this kind of data. A multivariate classification with the two microphones at the same time is a natural continuation to this work.

References

1. C. Abraham, P.A. Cornillon, E. Matzner-Løber, and N. Molinari. Unsupervised curve clustering using b-splines. *Scandinavian journal of statistics*, 30(3):581–595, 2003.
2. Q. Ai, C. Liu, X. Chen, P. He, and Y. Wang. Acoustic emission of fatigue crack in pressure pipe under cyclic pressure. *Nuclear Engineering and Design*, 240(10):3616–3620, 2010.
3. A. A. Anastassopoulos and T. P. Philippidis. Clustering methodology for the evaluation of acoustic emission from composites. *Journal of Acoustic Emission*, 13(1-2):11–22, 1995.
4. N. Favretto-Cristini, L. Hégron, and P. Sornay. Identification of the fragmentation of brittle particles during compaction process by the acoustic emission technique. *Ultrasonics*, 67:178–189, 2016.
5. F. Ferraty and P. Vieu. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media, 2006.
6. M. Giacomini, S. Lambert-Lacroix, G. Marot, and F. Picard. Wavelet-based clustering for mixed-effects functional models in high dimension. *Biometrics*, 69(1):31–40, 2013.
7. F. Ieva, A.M. Paganoni, D. Pigoli, and V. Vitelli. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3):401–418, 2013.
8. J. Jacques and C. Preda. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3):231–255, 2014.
9. S. Keyvan and J. Nagaraj. Pattern recognition of acoustic signatures using art2: A neural network. *Journal of Acoustic Emission*, 14(2):97–102, 1996.
10. R.G. Mbu Nyamsi, T. Aubin, and J.C. Bremond. On the extraction of some time dependent parameters of an acoustic signal by means of the analytic signal concept. its application to animal sound study. *Bioacoustics*, 5(3):187–203, 1994.
11. J. Peng and H.G. Müller. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, pages 1056–1077, 2008.
12. A. Struyf, M. Hubert, P. Rousseeuw, et al. Clustering in an object-oriented environment. *Journal of Statistical Software*, 1(4):1–30, 1997.
13. T. Tarpey and K.K.J. Kinader. Clustering functional data. *Journal of classification*, 20(1):093–114, 2003.