



HAL
open science

Emission acoustique et endommagement des composites : Intérêts et limites des techniques de reconnaissance de forme

Nathalie Godin, Mohamed R'Mili, Pascal Reynaud, J. Lamon, Gilbert Fantozzi

► **To cite this version:**

Nathalie Godin, Mohamed R'Mili, Pascal Reynaud, J. Lamon, Gilbert Fantozzi. Emission acoustique et endommagement des composites : Intérêts et limites des techniques de reconnaissance de forme. 17èmes Journées Nationales sur les Composites (JNC17), Jun 2011, Poitiers-Futuroscope, France. pp.35. <hal-00597898>

HAL Id: hal-00597898

<https://hal.science/hal-00597898v1>

Submitted on 2 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Emission acoustique et endommagement des composites : Intérêts et limites des techniques de reconnaissance de forme.

Acoustic emission and composite: Interest and limitation of pattern recognition technique.

N. Godin, M. R'Mili, P. Reynaud, J. Lamon, G. Fantozzi

Université de Lyon, MATEIS, INSA-Lyon, 7 Avenue Jean Capelle, 69621 Villeurbanne, France
Nathalie.godin@insa-lyon.fr, mohamed.rmili@insa-lyon.fr, pascal.reynaud@insa-lyon.fr, jacques.lamon@insa-lyon.fr,
gilbert.fantozzi@insa-lyon.fr

Résumé

La segmentation des données d'émission acoustique – ou partition des données – est un des enjeux actuels permettant d'accéder à des analyses plus approfondies de l'endommagement des matériaux. Parmi les méthodes de classification non supervisées (pour lesquelles on ne connaît pas à priori le nombre de classes présentes dans le jeu de données), une des plus utilisées reste la méthode des K-moyenne. Afin d'améliorer la robustesse de la méthode, un algorithme génétique a été inséré dans l'algorithme de classification. Les résultats du traitement de jeux de données « modèles » issus de données provenant d'essais sur matériaux permettent de mettre en valeur l'intérêt de cette technique et d'en préciser les limites. Les algorithmes de segmentation sont mis à l'épreuve de différents jeux de données modèles visant à apprécier leurs aptitudes à : distinguer une classe minoritaire (faible population de signaux) au sein d'un jeu de données ; isoler une classe dont les signaux, pour un ou plusieurs de leurs descripteurs, présentent des valeurs extrêmes, segmenter un jeu de données contenant des classes se recouvrant dans plusieurs plans de l'espace.

Abstract

The segmentation of acoustic emission data - or data clustering - is one of the current challenges to improve analysis of damage mechanisms in materials. Among the unsupervised classification methods (methods where the number of classes is not known previously), one of the most commonly used is the K-means method. In this study, the robustness of this method has been enhanced by a genetic algorithm-based optimisation. Theoretical data sets derived from experimental AE data obtained on several materials, have been processed to highlight the advantages of this optimised algorithm. For example, this segmentation algorithm tested for different data sets points out its ability to: distinguish a minor class (population with small number of signals) within a dataset; isolate a class where signals for one or some of their descriptors, have extreme values behaving as a group of outliers, segment a dataset containing clusters overlapping for several descriptors.

Mots Clés : Emission acoustique, Technique de reconnaissance de forme, Endommagement, Prévission de la durée de vie.

Keywords : Acoustic emission, Pattern Recognition, Damage, Durability.

1. Introduction

L'endommagement des matériaux composites est un point clé pour la maîtrise de la durabilité et de la fiabilité des pièces en service. Il est nécessaire non seulement de quantifier

l'endommagement mais aussi d'identifier les différents mécanismes qui en sont responsables et d'établir leur chronologie. En effet, dans les composites, il y a plusieurs modes d'endommagement (fissuration matricielle, rupture de fibres, décohésion...) liés à leur structure et aux caractéristiques des constituants. L'émission acoustique (EA) permettant de détecter et de suivre en temps réel l'endommagement au sein d'un matériau composite ou d'une structure apparaît alors comme un outil complémentaire aux investigations microstructurales et à la modélisation mécanique. L'objectif de ce travail est d'identifier en temps réel le mode d'endommagement à partir des données d'EA et d'estimer sa sévérité. Ce travail est basé sur l'utilisation de classificateurs. Si l'on dispose d'une bibliothèque de signaux déjà identifiés, la classification de nouveaux signaux se fait alors de manière supervisée, par comparaison aux signaux de la bibliothèque. Dans le cas contraire, l'approche est non supervisée. Le but de cette analyse statistique est de dégager les significations de données obtenues au cours du suivi par émission acoustique de l'endommagement de matériaux sollicités mécaniquement. Ces outils de classification vont simplifier les données d'EA en rendant l'interprétation plus aisée. Pour cela une méthodologie a été développée au laboratoire [1-9]. En résumé dans cette approche, l'émission acoustique n'est qu'un outil supplémentaire à une meilleure description des mécanismes d'endommagement. Cependant, l'émission acoustique, une fois qu'elle est bien caractérisée et maîtrisée, peut devenir un outil prédictif [10].

Un des objectifs de ce travail est de valider cette approche basée sur les techniques de reconnaissance de forme [11] mais aussi d'en montrer les limites en travaillant dans un premier temps sur des jeux de données d'EA qualifiées de « modèles ». L'algorithme développé au laboratoire [1-9] a été optimisé par l'ajout d'un algorithme génétique [12].

Ces jeux de données vont nous permettre de nous intéresser en particulier à :

- la performance de l'algorithme optimisé par une stratégie génétique par rapport à l'algorithme initial ;
- l'influence du choix des descripteurs des signaux d'EA sur les résultats de la segmentation ;
- l'influence du choix du critère de validation. Deux critères de validation de la segmentation sont comparés, le critère de Davies et Bouldin [13] et le critère des silhouettes [14]. Ces derniers permettent de définir la segmentation optimale soit le nombre de classes à considérer.

Avec les jeux de données modèles, nous pouvons aussi répondre aux questions suivantes :

- Quelle est l'influence d'un mécanisme source minoritaire, est-il possible de discriminer les signaux associés à un mécanisme minoritaire par exemple quelques ruptures de fibres ?
- Quelle est l'influence de la présence de signaux qualifiés « d'extrêmes », par exemple des signaux associés à du délaminage dans des CMO ?

2. Méthodologie

2.1 Analyse statistique multivariante

- choix des descripteurs

Chaque signal d'émission acoustique peut être décrit par plusieurs paramètres (amplitude, énergie, durée, temps de montée, fréquence moyenne ...), dont certains peuvent être corrélés et donc redondants. Il est donc nécessaire de choisir judicieusement ces descripteurs. Ils doivent être peu corrélés et en nombre suffisant pour ne pas perdre d'information sur les signaux mais pas trop nombreux pour ne pas alourdir les calculs ou fausser l'importance relative de chaque paramètre. Pour chaque signal d'émission acoustique, 18 descripteurs ont été définis. Ces descripteurs sont normalisés afin de ramener tous les paramètres dans l'intervalle [-1; 1]. Une analyse en composante principale (ACP) est ensuite réalisée afin de trouver une nouvelle base de vecteurs propres non corrélés par combinaisons linéaires des 8 descripteurs sélectionnés permettant ainsi de se ramener à un nouvel espace orthogonal. On peut alors réduire cet espace par exemple à 4 dimensions en conservant plus de 95% de l'information.

- *Algorithme des k-moyennes*

Cet algorithme de classification non supervisée est un processus itératif qui se déroule en 7 étapes :

- 1) Choix du nombre de classes k ,
- 2) Choix d'une mesure de similarité (distance euclidienne pondérée),
- 3) Initialisation des centres des k classes (C_k) de façon aléatoire,
- 4) Calcul de la distance entre chaque vecteur et chaque centre : le calcul des distances est établi à partir de la distance euclidienne pondérée par les valeurs propres λ_i issues de l'analyse en composantes principales :

$$d^2(X, Y) = \sum_{i=1}^4 \lambda_i (X_i - Y_i)^2 \quad (1)$$

- 5) Affectation de chaque vecteur à la classe la plus proche (distance minimale entre le signal et le centre de la classe),
- 6) Calcul des nouveaux centres des classes ainsi définis à la fin de cette étape d'itération,
- 7) Si les centres des classes sont stables, alors l'algorithme a convergé et la procédure est terminée sinon il faut répéter les étapes 4 à 6.

- *Validation de la segmentation, choix de la solution*

Le nombre de classes de signaux étant inconnu a priori, on fait varier ce nombre entre 2 et 10 avec 30 itérations pour avoir le meilleur résultat possible. La meilleure classification est celle qui permet de minimiser (Fig.1a) le coefficient de Davies et Bouldin DB [13] qui est défini comme suit :

$$DB = \frac{1}{k} \sum_{j=1}^n \max_i \left(\frac{d_i + d_j}{d_{ij}} \right) \quad (2)$$

Avec k : le nombre de classes, d_i : la distance moyenne à l'intérieur de la classe i , d_j : la distance moyenne à l'intérieur de la classe j , d_{ij} : la distance moyenne entre les classes i et j

Par ailleurs, un autre critère a également été implémenté, il s'agit des silhouettes [14]. Elles sont déterminées pour chaque signal par :

$$s_{ind} = \frac{b_i - a_i}{\max(a_i; b_i)} \quad (3)$$

où a_i est la distance signal i au centre sa classe de rattachement et b_i est la distance entre ce signal et la classe la plus proche. Ce critère est à considérer comme un indice de confiance de l'appartenance du signal à sa classe, une valeur de 1 signifiant une parfaite appartenance tandis qu'une valeur négative indique une mauvaise affectation. Ce critère, individuel, permet d'évaluer de manière plus précise la qualité de la classification (Fig. 1b). Il peut également être utilisé comme critère global par le calcul de S_i , moyenne des silhouettes de tous les signaux.

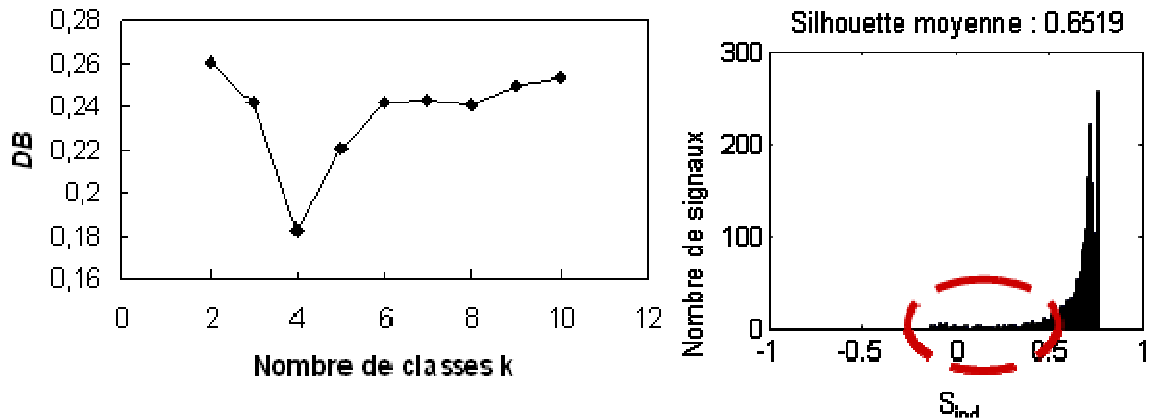


Fig.1 : a) Evolution du DB en fonction du nombre de classes, b) Histogramme des silhouettes individuelles pour une classe.

- optimisation avec une stratégie génétique [7-8]

Une optimisation de l'algorithme précédent a été réalisée en insérant un algorithme génétique dans l'algorithme des K-moyennes. L'intérêt de cet algorithme est de permettre un balayage de l'ensemble de l'espace des solutions. Dans ce cadre, ce n'est plus un jeu de centres de classes qui est initialisé mais un ensemble (population) de jeux de centres de classes qui sont créés, chaque jeu étant dénommé individu. Le choix du nombre d'individus doit se faire de manière raisonnée afin de ne pas allonger inutilement les temps de traitement. Dans ce travail, une valeur de 100 individus a été retenue. Les générations suivantes sont construites par des opérations qui reprennent la terminologie de la génétique. Les individus les plus performants (pour lesquels le critère d'évaluation présente une valeur élevée) sont conservés d'une génération à la suivante ; ils sont appelés les élites. Leur nombre a été fixé à deux. Les autres « enfants » sont créés par différents processus : sélection, mutation et croisement. La sélection des individus qui vont permettre de construire la génération suivante se fait avec une probabilité proportionnelle à la valeur des coefficients d'évaluation attachés à chaque individu. Cette opération est commune à tous les futurs individus. En fait, une partie de la population est issue de croisement (80%), l'autre partie de mutation (20%). Les croisements s'effectuent entre deux individus sélectionnés, les parents, en mode multipoints : pour chaque caractéristique de l'enfant (gène), les probabilités de transmission du « père » et de la « mère » sont égales. La mutation quant à elle frappe les gènes des autres individus avec une faible probabilité (1%) : la valeur du gène à muter est déterminée de manière aléatoire dans la gamme de valeur occupée par le jeu de données. Contrairement à l'algorithme précédent, l'arrêt de la classification se fait par convergence du critère d'évaluation, la valeur du DB. Ici, à partir d'une variation de moins de 1E-7 par génération, on considère que l'algorithme a convergé.

2.2 Données traitées

Des jeux de données ont été générés numériquement par création et combinaison de plusieurs classes de signaux. Les signaux de chaque classe sont décrits par un ensemble de paramètres dénommés descripteurs dont les valeurs ont été choisies proches de celles issues de données expérimentales (monitoring sur acier, céramique, fibre de verre, composite,...). Cinq classes ont ainsi été créées. Plusieurs jeux de données sont testés successivement. Toutes les données ont été exprimées dans la base des composantes principales en conservant 95% de l'information.

Données écoles

Jeu de données N ° 1, contient 4 classes de 2000 signaux chacune (classes 1 à 4) (fig. 2a). Les descripteurs pertinents définis par l'utilisateur sont le nombre de coups, la durée du signal et de l'énergie du signal. Cet ensemble de données nous permettra de comparer l'efficacité de deux algorithmes. Est-ce qu'ils produisent des résultats différents lorsqu'ils sont appliqués aux mêmes données?

Jeu de données N ° 2, « classe minoritaire », contient 4 classes identiques au jeu de données précédent (classes 1 à 4) avec une classe pour laquelle le nombre de signaux varie de 2000 à 100. Les descripteurs pertinents définis par l'utilisateur sont le nombre de coups, la durée du signal et de l'énergie du signal. Cet ensemble de données permettra de répondre à la question: est-il possible de discriminer un mécanisme source si ce mécanisme source est largement minoritaire?

Jeu de données N ° 3, « données extrêmes », contient 4 classes (cluster de 1 à 3 même ensemble de données que 1 & 2), mais avec une classe (cluster 5) présentant des valeurs extrêmes pour ses descripteurs et contenant 1937 signaux (fig. 2b). Ces analyses sont très sensibles aux « outliers ». Cet ensemble de données permettra de répondre à cette autre question: que se passe-t-il si un mécanisme source est très différent des autres ayant comme des valeurs « aberrantes » (mouvement de dislocations, signaux provenant de délaminage)?

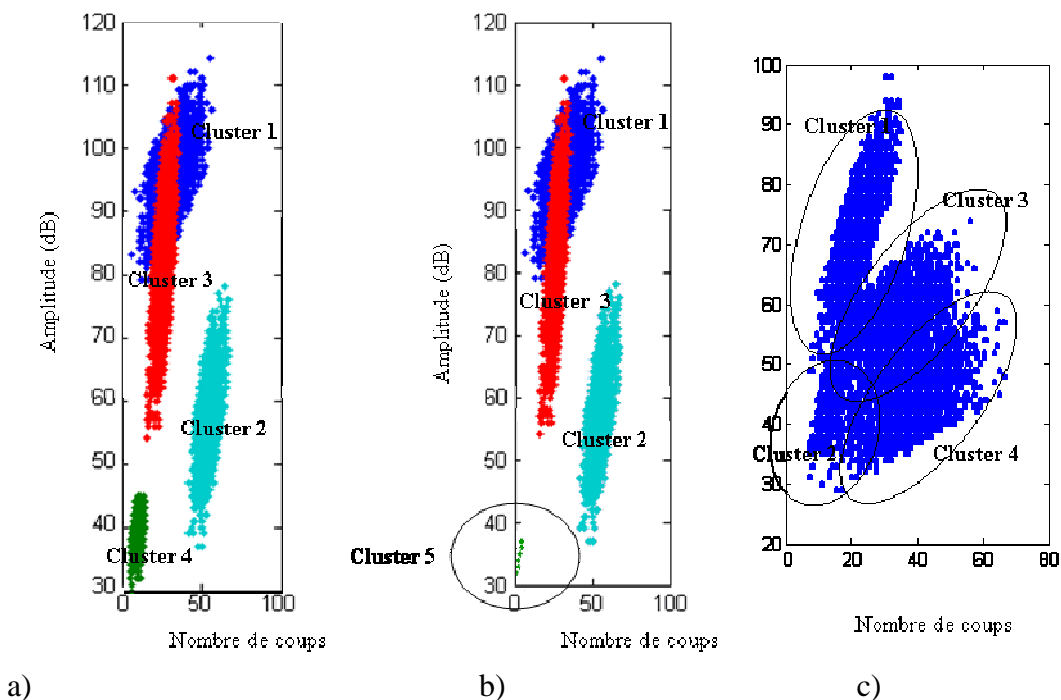


Fig. 2. Visualisation des jeux de données, projection dans le plan (amplitude, nombre de coups) (a) jeux de données N°1 (b) jeux de données N°3, (c) jeux de données N°4

Données artificielles proches de données réelles.

Jeu de données N°4, « ensemble de données denses », contient 4 classes (2000 signaux chacune) qui sont plus représentatives des données expérimentales réelles (Fig. 2c). Cet ensemble de données « denses », très semblables à des données réelles, nous permettra d'illustrer la notion de descripteur pertinent. En effet, après l'acquisition des données, chaque signal est représenté par 18 descripteurs. Est-il nécessaire ou utile de conserver les 18 descripteurs, ou est-il préférable de choisir des

descripteurs pertinents? Cet ensemble de données a été décrit initialement par 18 descripteurs puis par 3 descripteurs (amplitude, ln [nombre de coups au pic], ln [énergie]). Les paramètres pertinents ont été définis par l'utilisateur en fonction de la structure connue des données.

Jeu de données N ° 5, contient les mêmes classes que la série précédente mais avec des tailles différentes. La classe 1 contient encore 2000 signaux, Classe 2 seulement 500 signaux, classe 3 1000 signaux et la classe 4 a seulement 100 signaux. Dans un essai mécanique, les mécanismes d'endommagement sont activés très rarement dans des proportions égales. Cet ensemble de données nous permettra de voir si une disproportion des mécanismes a une influence sur la segmentation des données. Cet ensemble de données a été décrit par seulement 3 descripteurs pertinents (amplitude, ln [nombre de coups au pic], ln [énergie]).

Pour évaluer la performance des deux classificateurs, le taux d'erreur a été calculé. Il est défini comme le rapport entre le nombre signaux mal classés sur le nombre total de signaux. Il a été calculé que si le nombre correct de clusters a été retrouvé. Afin de vérifier la pertinence des critères d'évaluation et d'évaluer la capacité des algorithmes à distinguer les quatre classes qui composent les jeux, des partitions ont été réalisées 10 fois pour des nombres de classes allant de 2 à 10.

3. Résultats et discussion

3.1 Apports de la stratégie génétique ; traitement d'un jeu de données modèles.

Le jeu de données n°1 a été traité successivement par les deux algorithmes. Pour chacun des algorithmes et pour chaque classification, la valeur du coefficient de Davies et Bouldin (DB) et la moyenne des silhouettes (Si) ont été calculées. Ces résultats permettent de mettre en lumière plusieurs éléments. Dans ce cas, les valeurs de silhouettes calculées sont corrélées avec les valeurs de DB ; leur maximum coïncide avec le minimum de DB marquant pour chacun des algorithmes la classification optimale. L'algorithme génétique parvient à dissocier les quatre classes composant le jeu (10 fois sur 10). Seuls 139 signaux ont été mal classés, le taux d'erreur est de 1,75%. Avec l'algorithme initial, les résultats sont dispersés. La bonne solution a été obtenue que 4 fois sur 10. En fait, l'algorithme initial est fortement dépendant de l'initialisation et donc ne conduit pas systématiquement à la même solution.

Un défi majeur est de distinguer un mécanisme de source minoritaire qui peut être le mécanisme d'endommagement de plus grande sévérité. Pour évaluer les performances des deux algorithmes, nous avons utilisé l'ensemble de données N ° 2. L'usage de l'algorithme initial conduit à des résultats très contrastés. Jusqu' à une population de 500 signaux pour la classe 4, une segmentation optimale composée de quatre classes est retrouvée (seulement 5 fois sur 10), avec des taux d'erreur relativement faibles. Pour une population de 250 signaux, l'algorithme initial ne parvient plus à retrouver la solution. La sensibilité de l'algorithme initial au déséquilibre des populations, conduisant à des résultats aléatoires, est ainsi mise en évidence.

Au contraire, les résultats démontrent la supériorité de la stratégie génétique dans la segmentation de jeux de données aux populations inégales. Quelle que soit la population de la classe 4, jusqu'à une valeur réduite à 250 signaux, la structure du jeu de données, constituée de trois classes de 2000 signaux et d'une classe minoritaire, est retrouvée à chaque essai (10 fois sur 10). Les valeurs des critères de validité, DB et Silhouettes, sont systématiquement optimisées par rapports aux valeurs obtenues par l'algorithme initial (DB minimisé et Si maximisé). Les taux d'erreur consécutifs aux segmentations calculées sont également plus faibles, inférieurs dans tous les cas à 2,3 %. Ces éléments indiquent non seulement la plus grande efficacité de la stratégie génétique mais également

sa plus grande stabilité. La méthode présente cependant une limite puisque pour une population de la classe minoritaire de 100 signaux, ces derniers se retrouvent redistribués au sein de trois classes majoritaires. La stratégie génétique a donc un intérêt significatif pour la classification en effet les mécanismes de sources d'émission acoustique sont rarement activés de manière équilibrée.

Pour l'ensemble de données N° 3 qui contient un cluster avec des valeurs extrêmes, selon l'indice de DB, les deux algorithmes (tableau 1, Fig. 3a) conduisent à une classification optimale à deux classes (DB = 0,17 et Si = 0,67). La segmentation se compose d'un cluster de 2000 signaux (la valeur moyenne des silhouettes pour cette classe : 0,96) et un ensemble de signaux 5937 (silhouette moyenne pour ce groupe: 0,56). Les deux procédures isolent la classe extrême et regroupent les autres. Seuls quelques signaux de la classe 3 (environ 63) ont été inclus dans le cluster extrême. Selon le critère des silhouettes (tableau 1, Fig. 3b), la meilleure segmentation a été obtenue avec 4 classes (Si= 0,68 et DB = 0,25). Pour l'algorithme génétique, les clusters ont des populations respectives de 1817 (Cluster 1), 1968 (Cluster 3), 1971 (Cluster 5) et 2181 (Cluster 2) des signaux. Seuls 215 signaux ont été mal classés, ce qui correspond à un faible taux d'erreur d'environ 2,7%. Ce résultat est bien meilleur que celui sélectionné à l'aide de l'indice DB qui a combiné trois groupes ensemble. Avec l'algorithme initial et en fonction du critère des silhouettes, la solution à 4 classes a été obtenue seulement 6 fois sur 10. Ce résultat met en évidence l'intérêt d'utiliser plusieurs indices de validité et de définir le plus adapté au problème de partitionnement.

	DB validation		S(i) validation	
Algorithme initial	K=2	100 %	K=2	30 %
			K=4	60 %
			K=5	10 %
Approche génétique	K=2 (DB=0.17)	100%	K=4 (S(i)=0.68)	100%

Tab. 1. Résultats de segmentation pour le jeu de données N°3 en fonction des critères de validation permettant le choix de la solution.

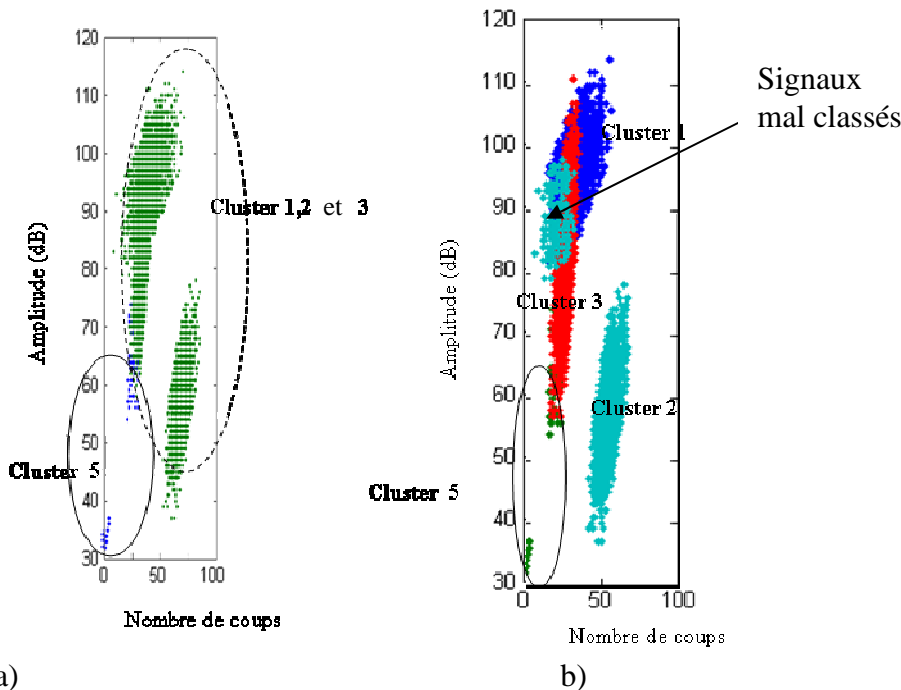


Fig. 3. Résultats de la segmentation pour le jeu de données N°3 par approche génétique, a) Validation par le critère du DB, b) Validation par le critère des silhouettes.

3.2 Segmentation d'un jeu de données artificielles proches de données réelles

L'influence du choix des descripteurs est examiné avec ensemble de données N ° 4, proches de données réelles. Les données expérimentales sont composées de neuf descripteurs enregistrés en temps réel et 9 autres calculés en post-traitement. Lorsque les 18 descripteurs sont considérés, la meilleure segmentation obtenue est toujours une solution à 3 classes même avec la stratégie de génétique (tableau 2, Fig. 4a). Seulement un Cluster a été correctement identifié avec 1985 signaux. La Silhouette moyenne pour ce groupe est supérieure à 0,6, tandis que pour les deux autres groupes, la Silhouette moyenne est inférieure à 0,5. La structure réelle des données n'a pas été révélée quelque soit l'algorithme de segmentation utilisé.

Considérant que trois descripteurs pertinents (amplitude, Ln [nombre de chiffres à pic], Ln [énergie]), la solution à 4 classes correspondant à la structure réelles des données a été obtenue. Avec l'algorithme initial, la solution correcte a été obtenue 8 fois sur 10, la stratégie génétique a obtenue la bonne solution 10 fois sur 10 (tableau 2, Fig. 4b). Les quatre classes ont une silhouette moyenne supérieure à 0,6 et il y a seulement 120 signaux mal classés. Ce résultat montre que le choix des descripteurs d'entrée est essentiel. Dans la classification non supervisée, il est impossible pour l'utilisateur de déterminer visuellement les paramètres pertinents. Ainsi, il sera utile de tester différents ensembles de descripteurs. Une autre solution serait d'utiliser la matrice de corrélation des 18 descripteurs et de sélectionner les descripteurs non corrélés avec un regroupement hiérarchique. Mais cette méthode permet seulement de sélectionner les descripteurs non corrélés mais ne donne pas d'indication sur leur pertinence.

Pour l'ensemble de données N ° 5, les données ont été décrites par l'amplitude, Ln [nombre de coups au pic] et Ln [énergie]. Avec l'algorithme initial, la structure de données n'a pas été révélée, en effet les deux solutions obtenues avec $k = 4$ étaient incorrects, les 4 classes qui ont été créés ne sont pas les bonnes. La classe majoritaire a été divisée en deux groupes et il est apparu que l'algorithme a tendance à créer des classes avec des tailles équivalentes. Avec la stratégie de génétique (9 fois sur 10), les critères DB et S (i) conduisent à une solution à 4 classes correspondant à la structure réelle des données. La valeur du DB est de 0,23 et celle de S_i est de 0,65. La stratégie génétique est moins sensible aux différentes tailles de cluster à la différence de l'algorithme initial. Cluster 1 contient 1969 (au lieu de 2000), cluster 2, 551 signaux (au lieu de 500), le cluster 3, 979 signaux au lieu de 1000 signaux et le cluster 4, 151 signaux au lieu de 100 signaux. Le taux d'erreur n'est que de 3%.

	18 descripteurs		3 descripteurs Amplitude, Ln[nombre de coups au pic], Ln[énergie]	
	Algorithme initial	K=3	50%	K=3
	K=4 (DB=0.36) (SI=0.48)	40 %	K=4 (DB=0.24. S(k)=0.62)	80 %
	K=5	10 %	K=5	20 %
Stratégie génétique	K=3 (DB=0.33) (SI=0.5)	100%	K=4 (DB=0.26) (S(k)=0.62)	100%

Tab. 2. Résultats de la segmentation pour le jeu de données N°4 pour les deux algorithmes en fonction du choix des descripteurs

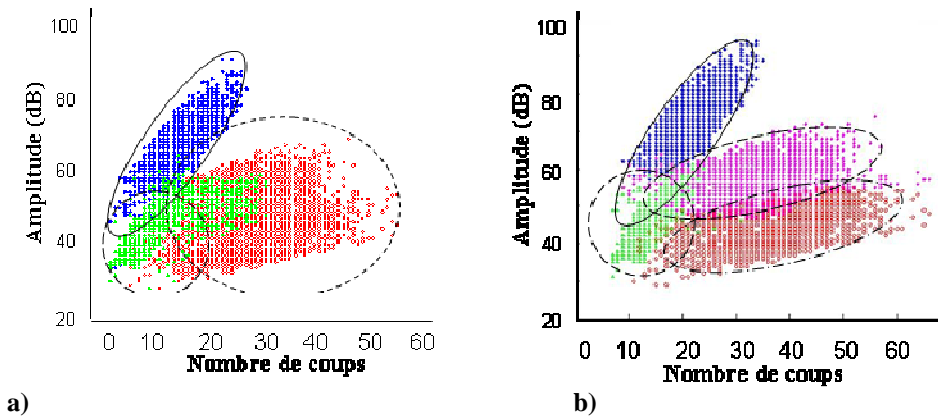


Fig. 4. Résultats de la segmentation pour le jeu de données N°4 par stratégie génétique, problème décrit par a) 18 descripteurs b) 3 descripteurs pertinents.

4. Conclusion

L'application des algorithmes de segmentation au jeu de données modèles a permis de mettre en évidence un certain nombre de points. Systématiquement, la stratégie génétique se distingue de l'algorithme initial par l'augmentation de la stabilité du traitement des données et l'obtention de segmentations qui optimisent les valeurs des critères de validité. Cependant, le choix des descripteurs permettant de décrire les données d'EA est un point clé quelque que soit l'algorithme de segmentation utilisé. L'algorithme de traitement, optimisé autour d'un algorithme génétique est plus apte à distinguer une classe minoritaire (faible population de signaux) au sein d'un jeu de données. Le choix du critère de validation permettant de définir la meilleure segmentation doit être bien adapté au problème, le critère des silhouettes est moins sensible à la présence de signaux, dont un ou plusieurs de leurs descripteurs, présentent des valeurs extrêmes, tels les « outliers ».

Références

- [1] Godin, N., Huguet, S., Gaertner, R. Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers NDT & E International 2004; 37 : 253-264.
- [2] Godin N, Huguet S, Gaertner R. Integration of the Kohonen's self-organising map and k-means algorithm for the segmentation of the AE data collected during tensile tests on cross-ply composites. NDT&E Int. 2005; 38(4): 299-309.
- [3] Huguet S., Godin N., Gaertner R.. Use of acoustic emission to identify damage modes in glass fibre reinforced polyester. Composites Science and technology, 62, 2002, pp. 1433-1444
- [4] Moevus M., Rouby D, Godin N., R'Mili M, Reynaud P, Fantozzi G, Fayolle G. Analyse of damage mechanisms and associated acoustic emission in two SiC/[Si-B-C] composites exhibiting different tensile curves. Part I: Damage patterns and acoustic emission activity. Composite Science and Technology, 2008; 68 (6): 1250-1257.
- [5] Moevus M, Godin N., Rouby D., R'Mili M., Reynaud P., Fantozzi G., Farizy G. Analyse of damage mechanisms and associated acoustic emission in two SiC/[Si-B-C] composites exhibiting different tensile curves. Part II: Unsupervised acoustic emission data clustering. Comp Sci Technol 2008; 68 (6): 1258-1265.
- [6] Shaira M., Godin N., Guy P., Vanel L., Courbon J.. Evaluation of the strain-induced martensitic transformation by Acoustic Emission monitoring in 304L austenitic stainless: identification of the AE signature of the Martensitic transformation and power-law statistics. Material Science and Engineering A 492, 2008, pages 392-399.
- [7] Sibil A. Etude de l'endommagement thermo-mécanique de nouveaux réfractaires verriers, Thèse Génie des Matériaux. Soutenance prévue en 2011. Lyon: Institut National des Sciences Appliquées de Lyon

- [8] Sibil A., Godin N., R'Mili M., Maillet E., Fantozzi G. Clustering of Acoustic emission data with a Davies and Bouldin index optimisation by a genetic algorithm. Soumis à Mechanical systems and signal processing.
- [9] Momon S., Godin N., Reynaud P., R'Mili M., Fantozzi G.. Unsupervised and supervised classification of AE data collected during fatigue test on CMC at high temperature, Soumis à Composites Part A.
- [10] Momon S., Moevus M., Godin N., R'Mili M., Reynaud P., Fantozzi G., Fayolle G., Acoustic emission and lifetime prediction during static fatigue tests on ceramic-matrix-composite at high temperature under air, Composites Part A: Applied Science and Manufacturing, Volume 41, Issue 7, July 2010, Pages 913-918
- [11] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Computing Surveys. 1999. Vol. 31. n°3. pp. 264-323.
- [12] David Goldberg, Addison Wesley. Algorithmes génétiques (1994).
- [13] D.L. Davies, D.W. Bouldin. A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intelligence (1979), 1:224–227.
- [14] Rousseeuw. P. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, J. Comput. Appl. Math. 20 (1987), 53-65.