



**HAL**  
open science

# Aspects temporels dans l'interprétation des données d'émission acoustique : importance de critères de validation adaptés et pertinence de méthodes variationnelles bayésiennes avec et sans a priori

Martin Mbarga Nkogo

► **To cite this version:**

Martin Mbarga Nkogo. Aspects temporels dans l'interprétation des données d'émission acoustique : importance de critères de validation adaptés et pertinence de méthodes variationnelles bayésiennes avec et sans a priori. Acoustique [physics.class-ph]. Université Bourgogne Franche-Comté, 2022. Français. NNT : 2022UBFCD015 . tel-03881284

**HAL Id: tel-03881284**

**<https://theses.hal.science/tel-03881284v1>**

Submitted on 1 Dec 2022

**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.

**Thèse de doctorat de l'établissement Université Bourgogne Franche-Comté**

**Préparée à l'Université de Franche-Comté**

Ecole doctorale n°37

Ecole Doctorale Sciences Pour l'Ingénieur et Microtechniques

Doctorat de Sciences pour l'Ingénieur

Par

Martin Fabrice MBARGA NKOGO

Aspects temporels dans l'interprétation des données d'émission  
acoustique : importance de critères de validation adaptés et pertinence de  
méthodes variationnelles bayésiennes avec et sans a priori

Soutenue le 7 Juillet 2022 à Besançon

Composition du Jury :

Michèle Rombaut	Professeur des Universités, Université Grenoble Alpes	Présidente
Valérie Kaftandjian	Professeur des Universités, INSA Lyon	Rapporteur
Frédéric Pichon	Professeur des Universités, Université d'Artois	Rapporteur
Emmanuel Ramasso	Maître de Conférences HDR, SUPMICROTECH-ENSMM	Directeur de thèse
Patrice Le Moal	Chargé de recherche CNRS	Co-encadrant de thèse
Gilles Bourbon	Ingénieur de recherche CNRS	Co-encadrant de thèse



# Remerciements

En premier lieu, je remercie mon directeur de thèse Emmanuel Ramasso pour son apport scientifique et ses conseils, pendant la thèse et les enseignements à SUPMICROTECH-ENSMM. Tu avais toujours de nouvelles idées à tester, quelques unes restent malheureusement dans le tiroir . . .

À Patrice Le Moal et Gilles Bourbon, mes co-encadrants : cela a été un vrai plaisir de travailler avec vous. Patrice le « sage » de l'équipe et Gilles très attentif au moindre détail, vous avez été acteurs majeurs jusqu'au bout.

Je tiens aussi à remercier Michèle Rombaut, Frédéric Pichon et Valérie Kaftandjian pour l'évaluation de mes travaux de thèse.

Au bureau 40H du département Mécanique Appliquée : Le meilleur bureau du monde comme j'aime le dire, Camille, Robin, Quentin (s) L. et G., j'ai apprécié toutes ces années. Stéphanie, tu as été ma meilleure rencontre dans ce cadre, merci d'avoir toujours été là pour moi.

Alex, Jason, Svenja, Kevin, Chrisale, Justine, Ludo, Aflah, Anouk, Solène, Fatim, Tiguida, Dakmak, Rafael (s), Amina, Chahrazed et les autres membres de la « Team », ça a été un plaisir de partager des moments avec vous.

Un grand merci à vous Christine Froidevaux, Isabelle Maillotte, Christelle Tisseran, Delphine Travaglini et Alexandra pour votre réactivité quand vous êtes sollicitées. Sans vous, la vie au laboratoire serait morose.

Monique Ntolo, Hugo Abeng, Élisabeth Abeng, Milena Makita, Jeanne Owona, Maan, Emma, Ismaelah, Linda, Ivan, Michel et Geneviève B., vous m'avez toujours encouragé, merci à vous. Vous faites partir de la famille.

Je terminerai par ma famille, vous avez toujours cru en moi, parfois même plus que moi. Papa (Nkogo Simon H.) et T. Jean Paul, sans vos sacrifices, rien de ceci n'aurait été possible, merci !

Je tiens à m'excuser si j'ai omis certains qui ont pu m'assister, il est difficile d'être exhaustif.

---

Je vous dédie ces travaux : Tantine, Brice, partis très tôt. Souriez et trinquez avec nous de là où vous êtes : JE SUIS DOCTEUR!

# Table des matières

Introduction générale	v
<b>1 Interprétation de l'émission acoustique : état de l'art</b>	<b>1</b>
1.1 Généralités sur l'émission acoustique	2
1.1.1 Domaines d'application	2
1.1.2 Sources des signaux d'émission acoustique	4
1.1.3 Acquisition des signaux d'émission acoustique	5
1.1.4 Prétraitement des données d'émission acoustique : cas ORION-EA	7
1.1.5 Interprétation des données d'émission acoustique	16
1.2 Méthodes non supervisées pour l'analyse des émissions acoustiques	19
1.2.1 Approches courantes pour l'interprétation non supervisée	19
1.2.2 Méthodes de partitionnement usuelles pour l'émission acoustique	20
1.2.3 Validation du partitionnement	29
1.2.4 Illustration du partitionnement sur ORION-EA	32
1.3 Problématique et contributions	42
<b>2 Validation du partitionnement par un critère séquentiel</b>	<b>45</b>
2.1 Approches basées sur la séquence d'apparition des clusters	47
2.1.1 Limites des critères de forme	49
2.1.2 Définition du critère séquentiel	51
2.2 Validation séquentielle avec plusieurs partitions	56
2.2.1 Principe de la validation sans estimation préalable de $K^*$	57
2.2.2 Algorithme proposé pour une validation séquentielle	58
2.2.3 Ajustement du critère pour l'estimation du nombre de clusters	60
2.3 Illustration sur ORION-EA	61
2.3.1 Lien entre les partitions et les serrages	62
2.3.2 Détection des niveaux de serrage	63
2.4 Discussions et conclusions	67

<b>3</b>	<b>Approche bayésienne pour l'interprétation des données d'EA</b>	<b>71</b>
3.1	Concepts clés et outils nécessaires . . . . .	73
3.1.1	Entre non supervisé et supervisé . . . . .	73
3.1.2	Labellisation douce et fonctions de masse . . . . .	75
3.1.3	Combinaison des fonctions de croyance . . . . .	77
3.2	Méthode bayésienne variationnelle pour les mélanges gaussiens . . . . .	79
3.2.1	Lois de distribution des variables . . . . .	79
3.2.2	Définition de la fonction objectif . . . . .	81
3.2.3	Inférence approchée et algorithme VB-EM . . . . .	82
3.2.4	Étape VB-E avec a priori sur les clusters . . . . .	86
3.3	Usage du VB-GMM-SOFT pour l'analyse de l'EA . . . . .	88
3.3.1	Détermination du nombre de clusters . . . . .	88
3.3.2	Comment générer les <i>a priori</i> . . . . .	89
3.4	Résultats sur ORION-EA . . . . .	90
3.4.1	Influence de la probabilité d'erreur $\rho$ . . . . .	91
3.4.2	Influence de la corruption des étiquettes et de la taille des données . . . . .	100
3.4.3	Influence des autres hyper-paramètres . . . . .	100
3.5	Discussions et conclusions . . . . .	103
	<b>Conclusion</b>	<b>111</b>
	<b>Publications</b>	<b>115</b>
3.6	Articles de conférence . . . . .	115
3.7	Articles de revue . . . . .	116
	<b>Bibliographie</b>	<b>117</b>
	<b>Annexes</b>	<b>129</b>
A.1	Validation conventionnelle à l'aide du critère séquentielle avec pénalité . . . . .	129
A.2	Validation consensuelle avec des instants de démarrage séquentiel logarithmiques . . . . .	130
A.3	Réduction de dimension . . . . .	130
A.3.1	L'analyse en composantes principales . . . . .	132
A.3.2	Illustration sur ORION-EA . . . . .	136
A.3.3	L'auto-encodeur . . . . .	136
A.4	Estimation de la log-vraisemblance pour le VB-GMM . . . . .	141
A.5	Fiche technique du capteur F50 $\alpha$ . . . . .	144

# Introduction générale

Assurer la disponibilité des équipements et l'intégrité des structures est un enjeu primordial pour les industriels tant sur l'aspect économique que pour la sécurité des usagers, notamment dans les domaines de l'aéronautique et le génie civil. Cela nécessite la surveillance des structures et/ou la connaissance des mécanismes d'endommagement potentiels. Le contrôle-santé des structures (traduction généralement utilisée pour *SHM : Structural Health Monitoring*) est un domaine de recherche et développement qui vise à implémenter et déployer des solutions pour la surveillance de ces équipements et structures. Apparu il y a quelques décennies, notamment aux Etats-Unis sous l'impulsion du Los Alamos National Laboratory (Etats-Unis) et du Department of Mechanical Engineering de l'Université de Sheffield (Royaume-Uni), le *SHM* s'est peu à peu structuré en France [1].

Lors de leur utilisation en service, ces structures et équipements sont soumis à des sollicitations mécaniques, électriques ou chimiques, celles-ci génèrent des micro-déformations locales. Ces micro-déformations peuvent être accompagnées par une libération d'énergie sous forme d'onde élastique connue sous le nom d'*émission acoustique* (EA). Si ces changements sont irréversibles et susceptibles de détériorer leurs performances, on parle d'endommagements (fissures, inclusions, corrosion, délaminage). Cumulés, ces micro-endommagements conduisent à la rupture du matériau rendant la structure hors d'usage. Une partie de l'énergie libérée par une EA se propage à travers la structure suivant ses propriétés physiques et géométriques. Des transducteurs piézoélectriques installés ou incorporés à la structure, permettent de transformer ces micro-vibrations perçues en un signal électrique exploitable. La technique d'EA consiste en l'acquisition de ces ondes élastiques transformées en signaux électriques, puis en leur traitement et leur interprétation. Les signaux enregistrés portent *a priori* les signatures des familles d'endommagements qui les génèrent [2].

La technique d'EA est passive car si aucun endommagement n'a lieu, rien n'est collecté sur les capteurs. D'où la nécessité d'interroger les capteurs en continu ou sur seuil c'est-à-dire à l'apparition d'un « endommagement significatif ». Bien que la fréquence d'échantillonnage soit relativement élevée (en fonction des capteurs utilisés, elle est généralement de 5 à 10 MHz), il est possible de détecter les endommagements en temps réel même en écoute permanente grâce à la capacité des ordinateurs actuels. L'émissivité du matériau surveillé est le principal point dur d'un traitement en temps réel. C'est le cas des matériaux innovants, qualifiés de « tolérants aux dommages », par exemple les composites laminaires et à matrice organique : dans ces matériaux, les endommagements sont très diffus et débutent généralement au niveau de la matrice qui va présenter de nombreuses fissures microscopiques qui vont

ensuite se combiner (phénomène de coalescence) pour se propager le long des fibres dans un même pli puis au travers des plis. A ces phénomènes d'endommagement s'ajoutent des frottements, par exemple entre les plis ou lors de l'ouverture et fermeture de fissures (lorsque le matériau est sollicité en traction et compression par exemple), représentant autant de sources d'EA. Ainsi il est fréquent de se retrouver avec plusieurs milliers de transitoires liés à ces phénomènes qui vont se répéter pendant des essais en laboratoire ou en utilisation en service, et ce, jusqu'à la rupture du matériau. Plus un matériau est émissif, plus il présentera de sources d'EA et plus il y aura de signaux à traiter. Par conséquent les coûts de calcul seront plus importants. C'est pourquoi, dans de nombreuses publications, un seuil sur l'amplitude (des données collectées en sortie de capteurs) est généralement appliqué pour limiter le nombre de transitoires à analyser. Une fois les transitoires d'intérêt détectés, une phase d'extraction de descripteurs intervient. L'objectif est de représenter les transitoires dans un même espace de représentation afin d'appliquer des algorithmes de classification supervisée ou non supervisée. Elle permet par ailleurs de grandement diminuer la quantité de données à stocker.

L'apprentissage non supervisé est généralement de rigueur lorsqu'il s'agit d'interpréter les EA. En effet, dans la majorité des cas d'application de cette technique (excepté dans des essais très spécifiques), il n'est pas possible de connaître avec certitude la source de tous les signaux d'émissions acoustiques. Les milliers de transitoires collectés sur des essais de caractérisation sont donc généralement traités par des algorithmes de partitionnement afin d'associer une classe (ou *cluster*) à chaque vecteur de descripteurs. L'élément clé de cette chaîne d'interprétation des EA est le lien entre les *clusters* et les endommagements car au final c'est ce qui intéresse les spécialistes de la Science des matériaux et de la Mécanique des Structures. Dans la « communauté reconnaissance de forme », cette étape de validation des clusters est réalisée à l'aide de critères. Historiquement, les critères de « forme » sont utilisés dans la « communauté émission acoustique ». Ces critères valident une partition si les *clusters* sont « naturels », c'est à dire s'ils sont **compacts** et **séparés** les uns des autres. Dans cette thèse, nous remettons en question cette manière de valider le partitionnement en nous intéressant plutôt à la manière dont les *clusters* sont répartis dans le temps ou en fonction du chargement.

L'interprétation des clusters est par ailleurs sujette aux incertitudes liées aux étapes précédentes. Par exemple, M. Sause [3] énumère ces incertitudes :

1. *l'incertitude de détection* : Elle est généralement liée au seuil fixé pour la détection et les émissions dites continues (plusieurs transitoires se déclenchent avant l'atténuation des précédentes, la détection est donc difficile). Elle a un impact direct sur l'extraction des descripteurs et par conséquent sur le partitionnement et la validation.
2. *l'incertitude du partitionnement* : Elle est liée à la méthode de partitionnement choisie ainsi que les *a priori* qui en découlent, le nombre et le type de descripteurs choisis, le nombre de clusters  $K$  qui impactent tous la partition obtenue et son interprétation.

3. *l'incertitude sur la validation* : Elle prend en compte essentiellement la méthode de validation (mono/multi critère), le nombre de partitions prises en compte et d'autres hyper-paramètres que nous verrons au premier chapitre.

Au cours de cette thèse, nous présentons et illustrons à l'aide du benchmark « ORION-EA » [4] des méthodes de partitionnement et de validation des données d'EA. Ce benchmark est constitué de données collectées sur un assemblage boulonné de plaques métalliques sous sollicitations vibratoires. Il est utilisé comme une référence pour l'évaluation des algorithmes existants et ceux proposés au cours de cette thèse.

## Organisation du manuscrit

Les contributions de ce manuscrit portent essentiellement sur le partitionnement et sa validation. Nous proposons dans cette thèse : 1) une étude bibliographique sur les méthodes de partitionnement et de validation des EA ; 2) une méthode de validation et de représentation des résultats du partitionnement qui prend en compte la répartition des *clusters* dans le temps ; 3) la prise en compte des incertitudes dans l'interprétation des données d'EAs précédemment mentionnées ; et 4) une méthode de partitionnement partiellement supervisée pouvant prendre en compte des *a priori* sur les *clusters* fournis par un expert.

### Chapitre 1

Une étude bibliographique est réalisée dans ce premier chapitre. Cette étude est centrée sur la manière dont la communauté interprète les émissions acoustiques. Elle présente succinctement les généralités sur la technique d'EA, les défis rencontrés, les solutions apportées par la communauté. Cet état de l'art permet de définir plusieurs verrous levés dans le cadre des autres chapitres. En second une étude comparative des méthodes de partitionnement usuelles est réalisée et illustrée sur le benchmark ORION-EA. Partant des limites liées au partitionnement et à sa validation soulevées dans l'étude bibliographique et illustrées avec le benchmark ORION-EA, nous dégageons des verrous pour lesquels nous proposons des solutions dans les chapitres suivants.

### Chapitre 2

La validation (choix du sous-ensemble de descripteurs pertinents) du partitionnement dans la communauté de l'émission acoustique a longtemps reposé sur des méthodes assez simples et finalement assez peu adaptées comme argumenté dans le chapitre 1. Ce point est crucial car il conditionne grandement l'interprétation finale. Il représente notamment une source d'incertitudes sur le nombre de *clusters* et l'analogie qu'on en fait avec des phénomènes physiques. Nous proposons un critère de validation externe prenant en compte la répartition des déclenchements de *clusters* dans le temps. La méthodologie proposée repose sur une recherche exhaustive des sous-ensembles les plus pertinents par l'intermédiaire de ce critère. Une méthode de représentation basée sur des histogrammes de déclenchement des *clusters* est développée pour combiner les différentes partitions en se focalisant sur la manière dont les

*clusters* apparaissent graduellement dans le temps. La validation par des critères séquentiels répond à la limite des critères de formes (internes) à trouver des clusters séparés dans le temps. La prise en compte d'un ensemble de partitions permet par ailleurs d'estimer d'une manière robuste ces instants de déclenchement, qui *a priori* correspondent à des changements de conditions expérimentales, ce qui peut être utile pour les tâches de monitoring.

### Chapitre 3

La détermination du nombre de *clusters* de façon objective, est un des verrous identifiés dans le chapitre 1. Il est fonction du critère utilisé, donc incertain. Pour réduire cette incertitude nous proposons d'exploiter le formalisme *Bayésien variationnel* pour déterminer le nombre de *clusters* sans l'intermédiaire d'un critère de validation. Les illustrations sur le benchmark ORION-EA montrent que ce formalisme fournit des *clusters* pertinents, ainsi le partitionnement bayésien est une option sérieuse pour l'interprétation des données d'EA. La possibilité d'introduire des *a priori* pour améliorer le partitionnement lorsqu'on ne dispose que d'informations partielles est un avantage qu'offre également l'algorithme proposé. Des *a priori* sont introduits sur les *clusters* en utilisant la théorie des fonctions de croyance. Les résultats obtenus avec le benchmark ORION-EA montrent que peu d'*a priori* suffit à améliorer le partitionnement même lorsque l'information est incomplète.

### Mots clés

SHM/PHM, inférence, méthodes non supervisées, incertitude, vibrations, partitionnement, émission acoustique

# Chapitre 1

## Interprétation de l'émission acoustique : état de l'art

### Sommaire

---

<b>1.1 Généralités sur l'émission acoustique</b> . . . . .	<b>2</b>
1.1.1 Domaines d'application . . . . .	2
1.1.2 Sources des signaux d'émission acoustique . . . . .	4
1.1.3 Acquisition des signaux d'émission acoustique . . . . .	5
1.1.4 Prétraitement des données d'émission acoustique : cas ORION-EA . . . . .	7
1.1.5 Interprétation des données d'émission acoustique . . . . .	16
<b>1.2 Méthodes non supervisées pour l'analyse des émissions acoustiques</b> . . . . .	<b>19</b>
1.2.1 Approches courantes pour l'interprétation non supervisée . . . . .	19
1.2.2 Méthodes de partitionnement usuelles pour l'émission acoustique . . . . .	20
1.2.3 Validation du partitionnement . . . . .	29
1.2.4 Illustration du partitionnement sur ORION-EA . . . . .	32
<b>1.3 Problématique et contributions</b> . . . . .	<b>42</b>

---

## Introduction

Dans ce premier chapitre centré sur l'état de l'art, nous définissons ce qu'est une émission acoustique (EA), ses domaines d'application et les différentes interprétations faites à partir de ces signaux. L'interprétation fait intervenir différentes étapes qui seront détaillées. Ces étapes seront illustrées à l'aide d'un benchmark *ORION-EA* [4] qui permettra de comparer les méthodes de la littérature. Nous terminerons par une discussion sur l'analyse des principales méthodes usuelles d'interprétation des EA.

### 1.1 Généralités sur l'émission acoustique

La plupart des mécanismes de défaillance des structures découlent du choix des matériaux retenus à leur conception, de l'environnement (potentiellement agressif), du cyclage en fatigue, de la nature et de l'intensité des sollicitations mécaniques. Comprendre et caractériser ces mécanismes est une étape nécessaire pour améliorer la prédiction des scénarios d'endommagement et augmenter la durée de vie utile de ces structures [5]. Pour cela des techniques de contrôles non destructifs sont utilisées pour assurer leur surveillance.

Les techniques de contrôle sont actives ou passives. Dans le premier cas, une onde est transmise à la structure via un émetteur "artificiel", tel un transducteur piézoélectrique. L'onde se propage dans le matériau jusqu'à un récepteur qui enregistre le signal déformé par le milieu de propagation. La mesure de la distorsion entre le signal émis et le signal perçu permet d'évaluer, par exemple, la présence voire la taille des endommagements. Dans le cas des techniques passives, il n'y a pas de source externe, seul un récepteur est utilisé pour collecter les ondes issues des endommagements. Dans cette thèse, nous nous intéressons à une technique passive appelée émission acoustique (EA) [2,3,6,7]. La norme AFNOR NF EN 1330-9 relative à la terminologie des essais non destructifs, définit l'EA comme : *un phénomène de libération d'énergie sous forme d'ondes élastiques transitoires résultant des micro-déplacements locaux internes dans un matériau (fissures, inclusions, corrosion, délaminage ...)*. La figure 1.1 présente le principe de la technique d'EA. Sur cette figure, un changement dans l'intégrité du matériau engendre des micro-vibrations (en pratique, elles sont entre le pico et le nanomètre) perçues généralement par un capteur piézoélectrique permettant d'obtenir un signal exploitable.

#### 1.1.1 Domaines d'application

En sismologie, il y a environ un siècle, les ondes élastiques émises lors d'un séisme ont été utilisées pour localiser et caractériser les failles en fonction de leurs profondeurs et de l'énergie libérée. En 1923, Hodgson proposa d'analyser des bruits audibles pour prédire la rupture mécanique des roches et les tremblements de terre dans les mines. Plus tard, des sons audibles perçus juste avant la rupture de matériaux métalliques et ceux liés à leur traitement thermo-chimique furent qualifiés d'EA [1,9].

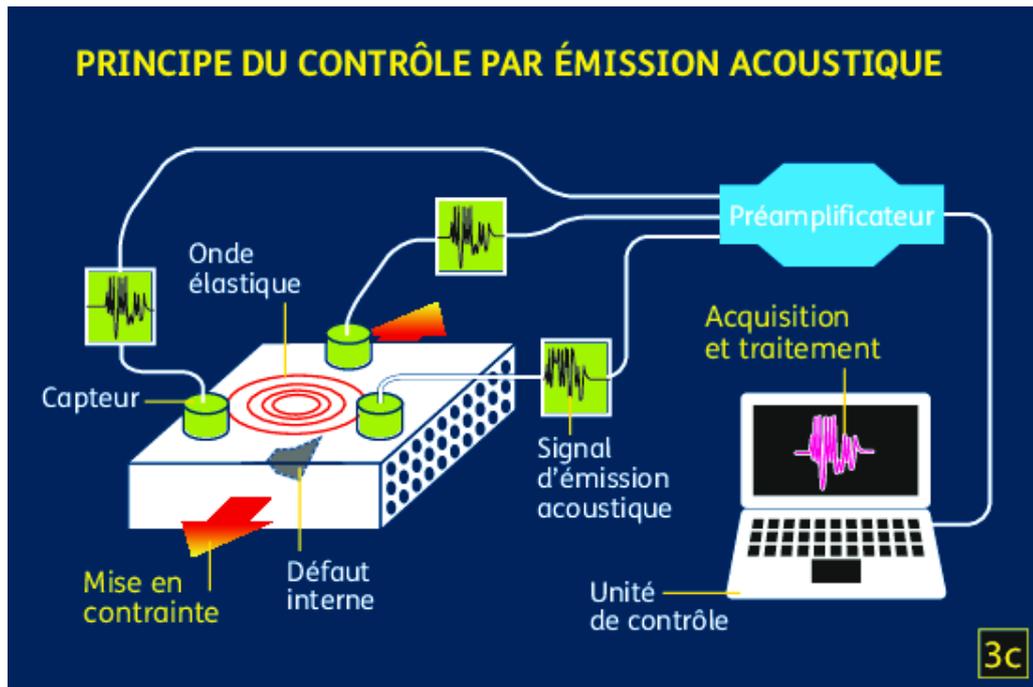


Figure 1.1 – Technique de l'émission acoustique en *SHM* [8]

Toujours dans le cadre des matériaux métalliques, Mason et al. [9] ont suggéré que les EA sont des vibrations hautes fréquences provoquées par des dislocations [1].

Joseph Kaiser (1953) et Bradford Schofield (1955) sont reconnus comme étant les précurseurs de l'utilisation de la technique de l'EA moderne. Les travaux de Kaiser ont permis de mettre en évidence les EA lors d'essais de caractérisation de matériaux polycristallins (zinc, acier, aluminium, argent et plomb). Schofield a quant à lui montré que le phénomène d'EA est aussi présent dans les matériaux monocristallins. Il conclut également que la fréquence et l'amplitude des signaux d'EA ne sont pas des descripteurs suffisants pour distinguer des familles d'endommagement. Ces travaux ont fait grandir l'intérêt de la technique de l'EA pour la caractérisation et le monitoring [1].

En 1969, Dunegan réalise une des premières études quantitatives de l'EA sur une structure en aluminium [10]. Il conclut que les EA ne sont pas proportionnelles à la taille d'un défaut ou au niveau de contrainte, mais corrélées au champ de contrainte (qui est une fonction des deux) autour d'un défaut. Il note aussi que les émissions sont liées aux déformations plastiques et ruptures et que l'absence de celles-ci est une preuve de l'absence de défauts actifs dans la structure. Enfin l'EA est dépendante du nombre de cycles (ou tout autre variable d'endommagement) et il la recommande pour une écoute en continu. Dunegan justifie ainsi l'utilisation de la technique de l'EA pour le monitoring des structures (*SHM : Structural Health Monitoring*).

Les premières applications industrielles de la technique de l'EA dès les années 1970 ont concerné des structures métalliques telles que les cuves de stockage dans l'industrie pétrochimique, les conduites d'acheminement de matières, les soudures dans l'industrie navale ou encore l'intégrité des systèmes rotatifs. Dans chacun de ces cas, des réseaux de capteurs d'EA sont principalement installés pour

détecter des anomalies et les localiser.

Depuis lors, la technique de l'EA a été utilisée dans de nombreux domaines, par exemple pour la détection des signes précurseurs de rupture de roches [11] et des ciments, ainsi que pour la surveillance de l'intégrité des structures industrielles telles que les réacteurs nucléaires et ponts [12]. Assez récemment, la société Mistras Group a été en charge de l'installation de près de 700 capteurs raccordés par 55 km de câbles sur le pont de Normandie [13]. La technique de l'EA est également utilisée pour le monitoring des équipements en fonctionnement (pales d'éoliennes, ailes d'avion, turbines, engrenages) [14].

De nos jours les domaines de l'aéronautique [15], du génie civil [16] et la recherche académique et industrielle sur les matériaux sont les principaux utilisateurs de la technique d'EA. Des applications comme la recherche sur les matériaux composites [2,6,7,17] et le contrôle de procédés industriels [18,19] participent aussi à l'avancée de la technique de l'EA. Dans ces domaines, l'EA sert généralement à détecter ou caractériser des anomalies ou les phases d'un procédé.

Parmi les principaux acteurs contribuant à l'interprétation de l'EA, citons par exemple : Lyon (MA-TEIS : Matériaux, Ingénierie, Science) [2,20,21], Le Mans (LAUM : Laboratoire d'Acoustique) [22], Grenoble (CEA :Centre d'Études Nucléaires de Grenoble) [23,24], Besançon (FEMTO-ST) [25,26].

L'émission acoustique fait également partie des techniques utilisées en *SHM*, un domaine qui s'est récemment structuré en France avec deux réseaux de compétence : la COFREND (Confédération Française pour les Essais Non Destructifs) et la PRECEND (Plateforme Régionale de Contrôles et Essais non Destructifs).

### 1.1.2 Sources des signaux d'émission acoustique

On distingue les sources d'EA naturelles des sources artificielles. Les sources naturelles sont dépendantes du milieu environnant, du type de structure et des matériaux qui la constituent. Parmi ces sources naturelles, on peut distinguer des sources mécaniques, par exemple, dues aux contraintes appliquées, et des sources externes, généralement thermo-chimiques, causées par un milieu agressif [5]. Les sources artificielles sont quant à elles des stimulations externes, indépendantes de la structure, du milieu environnant ou des conditions en fonctionnement normal.

Dans [5,21,24,27] les principales sources d'EA citées sont :

1. « **Dans un matériau** » : Dans ce cas, le matériau est la source et le milieu de propagation de l'EA. Les EA sont en général provoquées par des déformations plastiques (apparition et propagation des fissures) et la rupture (clivage, glissement..). Pour un composite, il peut s'agir de la rupture et arrachement des fibres, la fissure de la matrice, le délaminage et le décollement fibre/matrice. La Figure 1.2 issue de [5] illustre chacun de ces endommagements, sources potentielles d'EA. En plus de ces sources mécaniques, on distingue des sources thermo-chimiques dues à l'interaction avec un environnement agressif. Il s'agit principalement des

phénomènes de corrosion (locale ou généralisée) et des traitements thermiques (transformation martensitique) [5, 21].

2. « **À l'interface des structures** » : dans ce cas, la source et le milieu de propagation de l'EA sont possiblement différents. En plus des sources citées dans le matériau, on a la relaxation de contraintes dans les soudures, les fuites (de liquides ou de gaz) et les frottements. Les contacts entre pièces mécaniques assemblées génèrent des frottements lorsqu'elles sont en mouvement l'une par rapport à l'autre. Ces mécanismes s'accompagnent d'une montée de température et d'un arrachement de matière localisés aux zones de contact. Ces changements locaux sont à l'origine des EA, ce qui fait de cette technique une option pour l'analyse tribologique des surfaces en contact lors des essais vibratoires [4] ou de frottement et rupture [28].
3. **Les sources artificielles** : On retrouve le cassé de mine, la rupture d'un capillaire en verre, le jet de gaz, le lâché de bille et les générateurs d'onde [25]. Ces signaux artificiels sont essentiellement utilisés pour calibrer les capteurs et vérifier les montages en laboratoire. Le cassé de mine et le lâché de bille excitent le matériau sur un large spectre, ce qui permet d'évaluer la sensibilité du capteur. Le générateur d'onde, généralement mono-fréquentiel sert à évaluer la réponse du capteur à des bandes de fréquence spécifiques. Ces essais servent à la construction d'un modèle régressif pour la localisation des ondes.

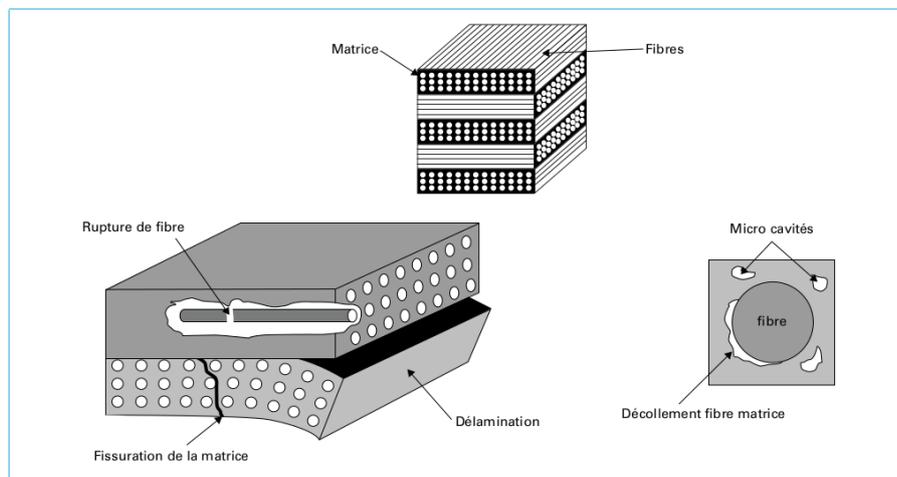


Figure 1.2 – **Endommagements dans un matériau composite** [5] : la rupture et l'arrachement des fibres, la fissure matricielle, le délaminage et le décollement fibres/matrice.

### 1.1.3 Acquisition des signaux d'émission acoustique

Une chaîne d'acquisition conventionnelle est constituée d'un capteur, un préamplificateur/amplificateur et une unité de traitement, reliés entre eux par un câblage (voir la figure 1.1). Le capteur joue un rôle de transducteur et le préamplificateur celui de conditionneur. Le principe de transduction majeur mis en œuvre par les capteurs d'EA est l'effet piézo-électrique direct (apparition de charges électriques sous sollicitation mécanique) [3, 21]. D'autres principes de transduction sont investigués au stade de

la recherche comme les capteurs électrostatiques [29] et optiques [3, 21].

Les capteurs commercialisés en EA sont essentiellement piézo-électriques et produits par la société Mistras. Ces capteurs sont constitués d'une céramique piézoélectrique (*PZT* : zirconate titanate de plomb, *PM* : méta-niobate de plomb...) qui est packagé dans un cylindre métallique. Ce packaging présente l'avantage d'être optimisé par un backing adapté et un préamplificateur qui réalise l'adaptation d'impédance et amplifie la tension de sortie du capteur éventuellement faible. Ceci confère aux capteurs Mistras un bon rapport signal sur bruit compatible avec la problématique EA [3, Chap.4]. Ils sont opérationnels dans des conditions industrielles difficiles (haute pression, bruits...) et couvrent des bandes fréquentielles relativement grandes. Ils sont plus sensibles autour d'une fréquence caractéristique dite de *résonance*. Parmi les capteurs utilisés pour des applications standards, on retrouve le Micro80 (200 - 900 kHz, Figure 1.3a), le micro200HF (500 – 4500 kHz, Figure 1.3b) et le F50 $\alpha$  (utilisé dans ce manuscrit, sa fiche technique est en annexe 5).



Figure 1.3 – Exemples de capteurs pour des applications standards

Ces capteurs sont caractérisés par des dimensions centimétriques généralement cylindriques avec une hauteur de plus grande dimension. Par exemple, le capteur  $\mu 80$  dit miniature a un diamètre de l'ordre de 9mm et une hauteur autour de 11mm. Ainsi, ces capteurs sont généralement disposés en surface des structures. Des travaux de recherche actuels visent à intégrer de tels capteurs au sein même des structures (par exemple entre les plis d'un matériau composite stratifié) pour soit être au plus près des sources d'émission acoustique, soit créer des matériaux qui intégreraient dans leur procédé de fabrication des capteurs de suivi de santé. Dans cette optique, d'autres solutions technologiques sont envisagées. Les pastilles piézoélectriques (*PWAS* : *Piezoelectric wafer active sensors*) font parties des solutions pour la miniaturisation et l'exploitation d'autres technologies. Elles peuvent être embarquées, utilisées en mode passif pour la détection des EA et en mode actif en tant qu'émetteur [30, 31]. Elles

présentent aussi l'avantage de résister à des conditions extrêmes (hautes températures et radiations) [32].

Des techniques de micro-usinage du silicium contribuent également à cette recherche, il s'agit des PMUT et CMUT (*Piezoelectric et Capacitive Micromachined Ultrasonic Transducers*). Les PMUT reposent sur le même principe piézoélectrique que les capteurs MISTRAS par exemple, la céramique piézoélectrique dite "bulk" étant remplacée par une couche mince (de quelques dixièmes à quelques microns) d'un matériau piézoélectrique (PZT, AlN, ZnO, ...). Les CMUT sont généralement des réseaux de membranes vibrantes constituant des condensateurs dont les capacités varient et génèrent ainsi un signal électrique mesurable. La mise en œuvre de membranes de dimensions différentes confèrent aux réseaux de CMUT la capacité de couvrir une large bande fréquentielle [29].

Indépendamment de la technologie utilisée et du principe de transduction pour l'acquisition, le capteur utilisé doit être caractérisé par une sensibilité, un rapport signal sur bruit et une largeur de bande optimaux permettant de percevoir des signaux transitoires de faible amplitude, éventuellement bruités dans une bande fréquentielle autour de 100kHz - 1MHz. Le capteur est généralement collé ou couplé à la surface de la structure à l'aide d'un gel spécifique dit « couplant » qui favorise un bon couplage acoustique entre le capteur et la structure. La fonction du gel est de « remplacer » (ou réduire) une éventuelle couche d'air qui induit usuellement une désadaptation d'impédance acoustique à l'origine de pertes acoustiques importantes. Le signal issu du capteur est conditionné et pré-amplifié. Le préamplificateur, installé en amont de la chaîne d'acquisition, améliore le rapport signal sur bruit en réalisant une adaptation d'impédance lorsque le signal issu du capteur est très faible (en dessous du mV), ce qui permet d'amplifier le signal du capteur et améliorer sa sensibilité. Les fréquences d'échantillonnage des EA sont relativement grandes (entre 1 MHz et 10 MHz). Le flux brut obtenu en sortie est appelé *stream* (nom anglo-saxon), ce dernier est constitué de l'ensemble des signaux transitoires d'EA et du bruit lié à l'acquisition. Ce flux subira un prétraitement avant l'utilisation des méthodes de reconnaissance de formes pour relier chaque signal d'EA à une source potentielle d'EA.

#### 1.1.4 Prétraitement des données d'émission acoustique : cas ORION-EA

L'interprétation des EA suit généralement une procédure standard (Figure 1.4) constituée des étapes suivantes : 1) Détection des signaux transitoires d'intérêt, 2) extraction de descripteurs, 3) partitionnement et 4) validation. Ces étapes seront illustrées par la suite sur un benchmark décrit au paragraphe suivant.

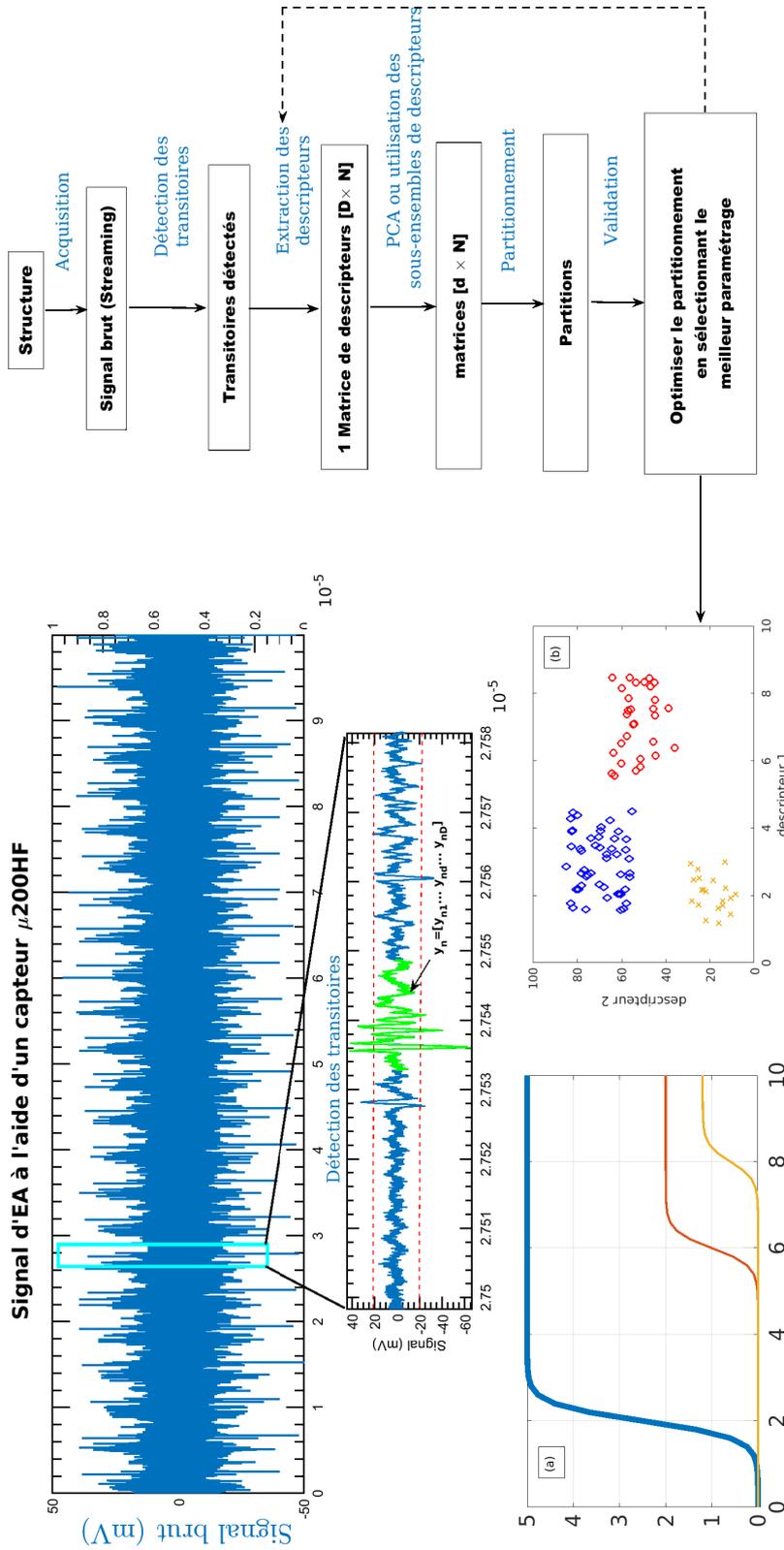


Figure 1.4 – **Chaîne de reconnaissance de formes** | Un capteur couplé en surface d'une structure enregistre un signal constitué des signaux d'E/A et du bruit d'acquisition. Les transitoires détectés après prétraitement sont représentés par des descripteurs. Des méthodes de reconnaissance de formes permettent de faire le lien entre signaux et phénomènes physiques. Si l'interprétation est difficile, une nouvelle extraction de descripteurs est possible. (PCA : Analyse en composantes principales) ; (a) Représentation de la chronologie des événements ; (b) Représentation dans l'espace des descripteurs. Modifiée à partir de [33].

### a) Présentation des données utilisées

L'ensemble de données de référence appelé « **ORION-EA** » est utilisé dans ce manuscrit pour illustrer les méthodes de la littérature ainsi que celles proposées dans les chapitres suivants. L'objectif est de comparer les méthodes de la littérature et celles développées au cours de cette thèse à l'aide d'un benchmark. Le protocole de génération de cet ensemble de données est présenté dans [4, 34] (figure 1.5). Les expériences ont été conçues pour reproduire le phénomène de desserrage observés dans les structures aéronautiques, automobiles ou de génie civil où les pièces sont assemblées entre elles au moyen d'assemblages boulonnés. Dans ces applications, l'enjeu est d'assurer la disponibilité des équipements et la sécurité des personnes. Pour cela, il est nécessaire de détecter de manière précoce le desserrage. Au cours de leur utilisation en service, ces structures sont soumises à différentes conditions de fonctionnement, notamment des vibrations et chargements pouvant modifier le couple de serrage. La surveillance de l'état du desserrage dans ces structures pendant le fonctionnement est un défi car les contacts et les frottements dans les assemblages induisent un comportement stochastique non linéaire.

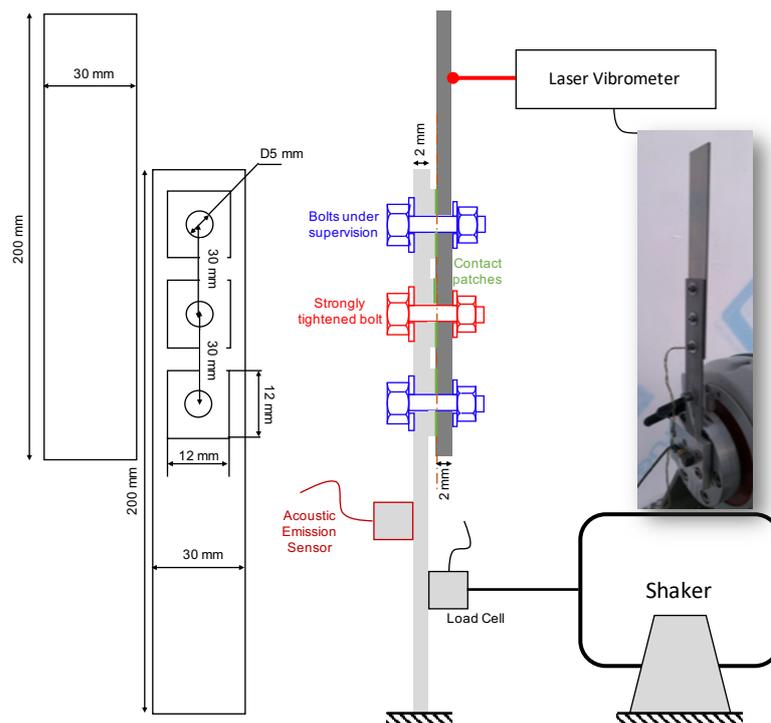


Figure 1.5 – Structure étudiée : ORION-EA [4]

L'assemblage considéré est constitué de deux plaques d'aluminium reliées par trois boulons (Figure 1.5). Le contact entre les plaques est réalisé par des patches usinés sur les surfaces. La structure est soumise à un test de vibration harmonique de 120 Hz pendant environ 10 secondes pour 7 niveaux de serrage différents sur le boulon supérieur. Le serrage est initialement réglé à 60 cNm avec un tournevis dynamométrique. Après 10 secondes de vibration, le pot vibrant est arrêté et le couple modifié à 50 cNm pendant 10 secondes, et ceci est refait pour des couples de serrage de plus en plus faibles 40, 30,

20, 10 et 5 cNm. Ce test est répété cinq fois menant à autant de jeux de données. Les durées exactes pour les cinq essais de chaque niveau de serrage sont répertoriées dans le tableau 1.1. On y constate que pour le deuxième essai, il n'y a pas d'acquisition pour le couple de serrage 20 cNm.

Tableau 1.1 – Récapitulatif des conditions expérimentales des essais « ORION-EA »

Essais	Échantillonnage (MHz)	Couple de Serrage (cNm)						
		60	50	40	30	20	10	5
B	5	10.000	10.000	10.000	9.0170	10.000	10.7725	9.225
C	5	10.000	10.000	10.000	10.000	×	10.000	10.000
D	5	10.000	10.000	10.8119	9.1881	10.000	10.000	10.000
E	5	10.000	10.000	10.000	10.000	10.000	10.000	10.000
F	5	10.000	10.000	10.000	10.000	10.000	10.000	10.000

Pour chacune des campagnes, quatre capteurs ont été utilisés : trois capteurs d'EA différents ( le « micro-200HF », le « micro-80 » et le « F50 $\alpha$  » avec différentes bandes de fréquences déjà présentés à la section 1.1.3) fixés sur la plaque inférieure. Les capteurs d'EA sont prévus pour détecter les frottements ou les chocs dans l'interface qui représentent les sources d'EA lors des essais vibratoires [35]. La mesure du déplacement de la plaque supérieure (graphe rouge, Figure 1.6) donnée par le vibromètre est constante indépendamment du niveau de serrage. C'est une caractéristique très importante du jeu de données ORION-EA. Pour ne pas induire de « biais » lié à l'amplitude intrinsèque de la vibration, l'excitation de la plaque est pilotée de manière à ce que l'amplitude de vibration mesurée par le vibromètre laser soit constante quelque soit le niveau de serrage appliqué. Le flux brut d'EA a été représenté (en vert), avec les sept niveaux de serrage (en bleu) au cours du temps. On y observe également qu'une approche mono-paramétrique basée sur l'amplitude n'est pas suffisante pour discriminer les niveaux de serrage. Il est donc nécessaire d'utiliser d'autres descripteurs et faire appel à des méthodes de traitement du signal et reconnaissance de formes plus sophistiquées.

Le benchmark peut être utilisé de différentes manières, par exemple, pour tenter d'identifier les sources d'EA (par exemple, contacts et/ou frottements) et leurs signatures à partir des signaux ; ou alors pour évaluer le niveau de serrage. Dans cette thèse nous nous sommes focalisés sur le deuxième point avec l'objectif d'évaluer quantitativement les méthodes d'apprentissage non supervisées utilisées dans la littérature. Cela permettra de situer les performances des algorithmes proposés dans les chapitres suivants.

Un intérêt du jeu ORION-EA est qu'il peut servir de référence pour l'évaluation des méthodes de la littérature. On dispose des niveaux de serrages, ceux-ci peuvent servir de référence pour l'évaluation des méthodes de partitionnement et/ou pour la mise en œuvre des méthodes d'apprentissage supervisé.

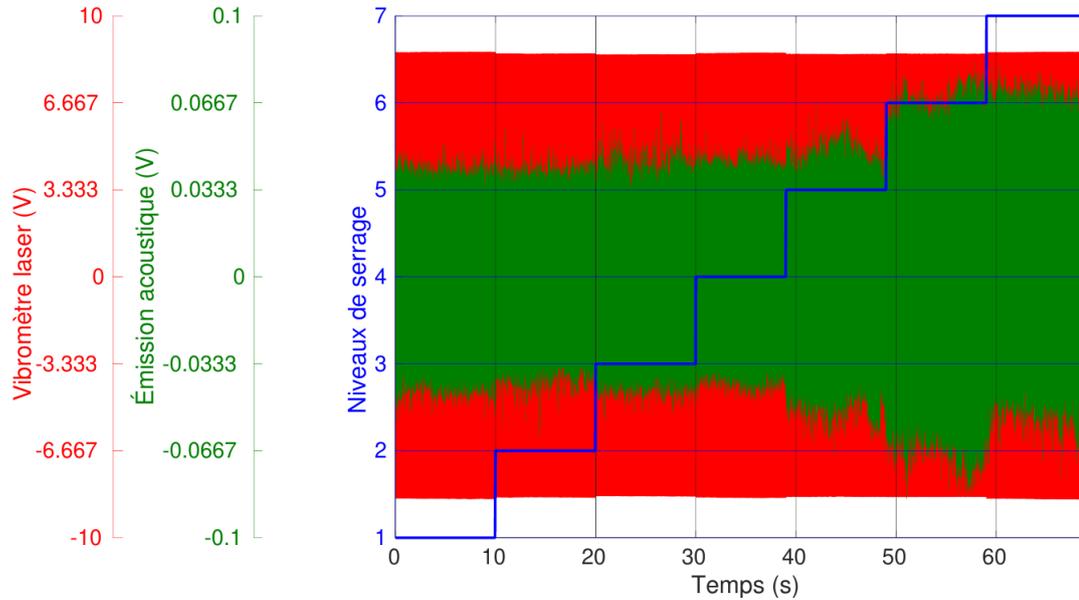


Figure 1.6 – Niveaux de serrage, émission acoustique et signal du vibromètre| L'axe du temps a été concaténé pour différents niveaux de serrage. De 0 à 10s, le serrage est 60 cNm (niveau 1), de 10 à 20s, le serrage est à 50 cNm (niveau 2), ainsi jusqu'à la fin de l'essai. La courbe rouge représente un signal sinusoïdal à 120 Hz pour lequel un zoom est illustré à la figure 1.7a.

## b) Détection des transitoires

La détection des transitoires est la première étape dans la méthodologie d'interprétation des EA. La détection est une tâche capitale, car elle se fait en amont de l'extraction des descripteurs, de la caractérisation et/ou de la localisation.

La première méthode, aussi la plus utilisée est *l'activation sur seuil*. C'est également celle commercialisée par l'entreprise MISTRAS (figure 1.9). Elle consiste à fixer un seuil en dessous duquel les données sont considérées comme du bruit et donc non traitées. Si ce seuil est atteint, une EA est automatiquement enregistrée. Le signal étant transitoire, on a plusieurs passages en dessous de ce seuil. À chacun de ces passages, un compteur mesurant le temps en dessous du seuil est déclenché. Si le temps passé en dessous du seuil est égal à la valeur *HDT* (*hit definition time*) on considère qu'il s'agit de la fin du signal. On fixe également une durée minimale pour le passage d'un signal à l'autre (*HLT* : *hit lockout time*) (voir [3, chap.4] et [36, chap.5]). La méthode repose essentiellement sur la détermination des valeurs de ces seuils. Si le seuil de détection est trop bas ou la valeur du HDT trop grande, deux ou plusieurs signaux compteront pour un seul. Inversement si le seuil est haut et le HDT trop bas une EA sera divisée en deux ou plusieurs transitoires de petite taille. Enfin si la valeur du HLT est trop grande on perdra une partie du signal suivant [37].

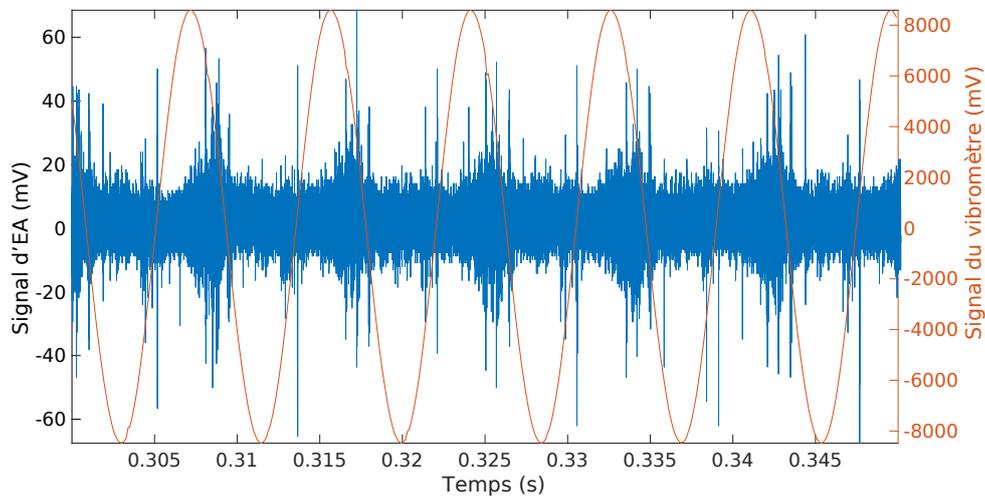
La deuxième méthode [25] repose sur une procédure de détection de transitoires basée sur le débruitage par ondelettes. L'algorithme précédent est précédé d'un débruitage par ondelettes pour

trouver de nouveaux instants de démarrage et de fin des signaux d'EA ayant la quasi-totalité de l'énergie des signaux initialement détectés. Une compensation du retard introduit par le filtrage est faite en ajoutant un nombre d'échantillons égal au nombre de coefficient dans l'ondelette.

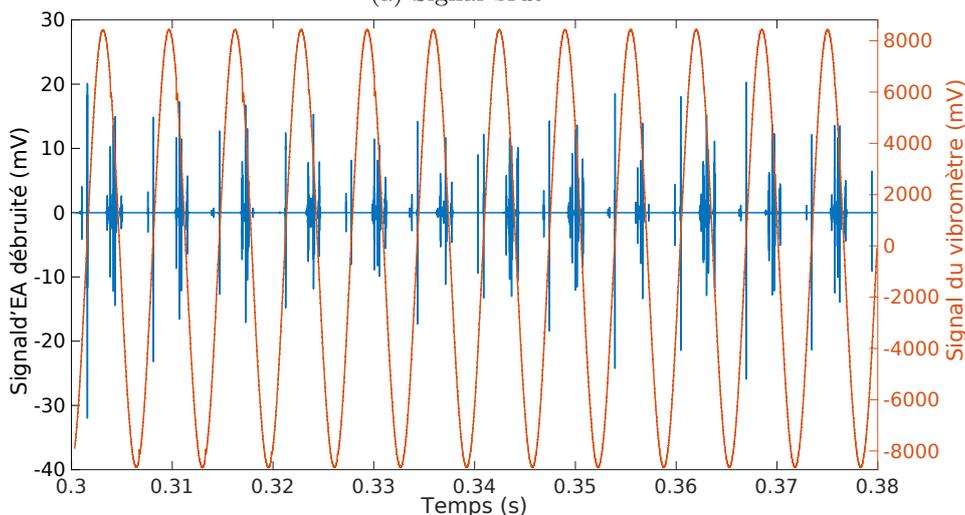
Les paramètres de la méthode proposée dans [4, 25] sont :

- Le *HDT* est réglé à  $1100\mu\text{s}$
- Le *HLT* est fixé à  $80\mu\text{s}$

La figure 1.7a montre le signal brut autour de 0.3s, les transitoires (en bleu) ne sont pas mis en évidence. Des pics d'amplitudes sont cependant observés une à 2 fois par période. Après débruitage, les signaux d'EA sont bien détectés avec un début et une fin.



(a) Signal brut



(b) Signal après débruitage

Figure 1.7 – Débruitage des signaux | zoom autour de 0.3s pour l'essai #B pour un couple de serrage de 60 cNm. La détection du seuil est faite à partir du transitoire filtré (b) et l'extraction des descripteurs à partir du signal brut (a).

Lorsqu'un matériau est très émissif, ce qui est notre cas avec ORION-EA, il arrive que les émissions surviennent en cascade, le flux prend alors la forme d'un bruit dont il est difficile de séparer les signaux d'EA : on parle alors d'*émission continue*. C'est un cas particulier pris en compte par la seconde méthode de détection. Dans le cas d'une *émission discrète*, les signaux d'EA observés sont des transitoires ayant un début et une fin. La figure 1.8 (issue de [25]) montre à gauche une émission « continue » et à droite un signal transitoire isolé observé avec le benchmark ORION-EA.

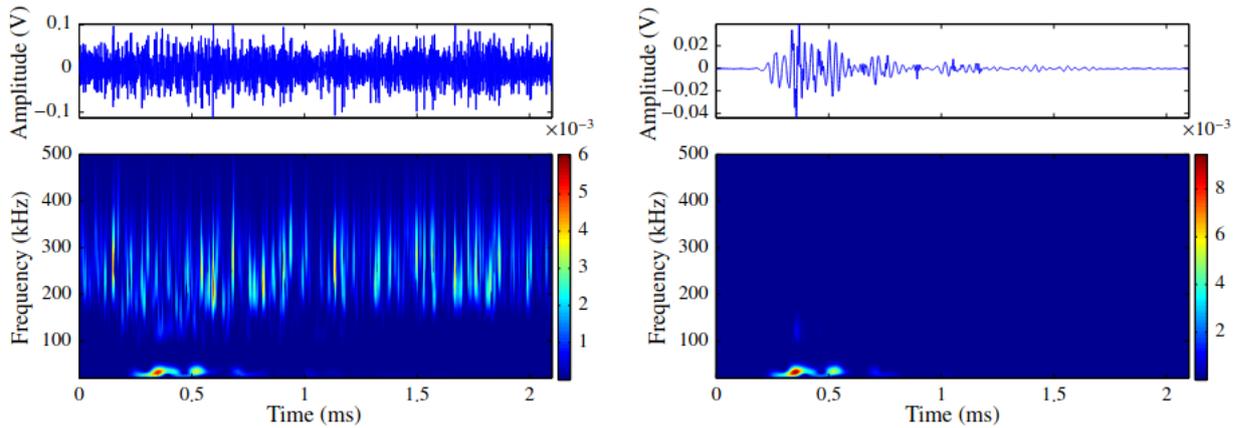


Figure 1.8 – Émission acoustique continue (à gauche) et émission discrète (à droite) [25]

Le bon déroulement de l'étape de détection est une condition nécessaire pour l'optimisation des futures étapes. Deux approches permettent d'identifier et caractériser les signatures des sources d'EA, la première est basée sur les signaux bruts (transitoires) et la seconde à partir de descripteurs extraits à partir de ces derniers.

La première approche consiste à utiliser directement les « *signaux bruts* » dans le domaine temporel [26, 38]. Elle nécessite la construction d'une base de signaux de référence dont chacun représente la signature d'un *cluster* (famille d'endommagement). Les signaux peuvent aussi être choisis aléatoirement dans [26]. Puis, une mesure de similarité permet d'attribuer les nouveaux signaux au *cluster* le plus proche, un nouveau cluster peut également être créé si le transitoire a une signature très différente des autres *clusters*.

L'approche la plus courante consiste à représenter les signaux dans un espace de descripteurs.

### c) Extraction des descripteurs

L'étape d'extraction des descripteurs survient après cette phase. En effet les signaux extraits du flux brut sont de tailles variables, l'extraction de descripteurs permet de les représenter dans un espace commun tout en réduisant la dimension des données à traiter. Cette étape permet également un gain en temps de calcul.

Parmi les représentations basées sur l'extraction des descripteurs, on retrouve principalement :

1. **La représentation en descripteurs statistiques** : dans cette approche standard [3, 39],

les signaux d'EA sont transformés en  $D$  caractéristiques statistiques appelées *descripteurs*. Le choix de descripteurs pertinents nécessite une expertise dans le cas non supervisé, car il dépend du domaine et du type de signaux à analyser.

2. **Représentation à partir de la transformée de Fourier rapide (*FFT : Fast Fourier Transform*)** : Elle consiste à utiliser le spectre fréquentiel ou les enveloppes des signaux [40–42]. De la même manière qu'avec les descripteurs statistiques, les signaux sont représentés dans un espace commun.

Dans cette étude bibliographique et dans les prochains chapitres nous nous intéressons uniquement à l'approche standard basée sur des descripteurs statistiques.

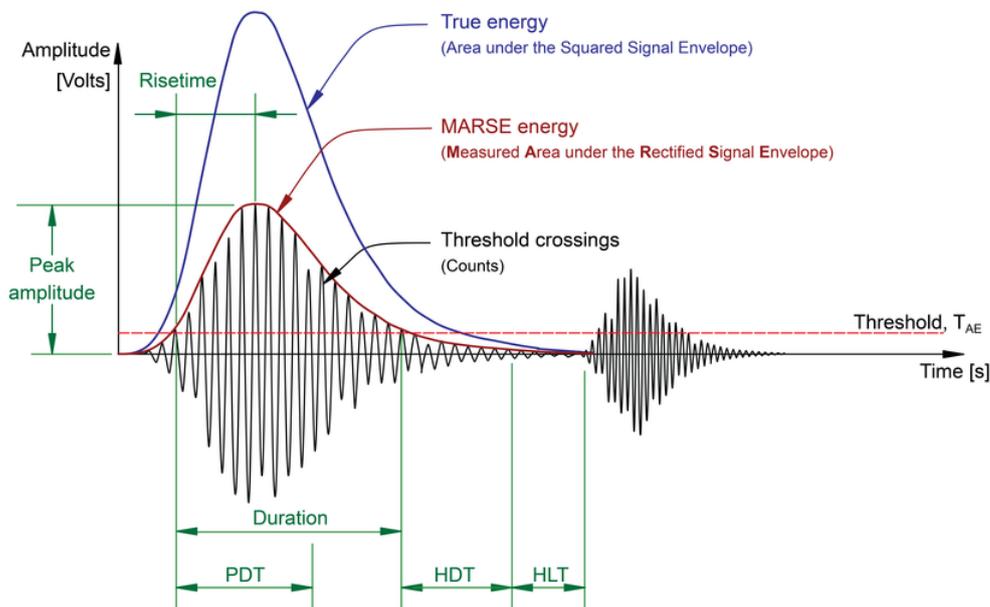


Figure 1.9 – Image d'un transitoire muni de quelques descripteurs [37]

La société *MISTRAS Group*, qui fournit des solutions en termes d'instrumentation (capteurs spécifiques) et d'analyses des données (logiciels dédiés) et qui constitue une référence industrielle dans le domaine de l'EA, propose par exemple une liste de descripteurs communément repris dans la littérature [43, 44]. La figure 1.9 présente un transitoire abstrait pour lequel sont illustrés le seuil (rouge en pointillés), le début du transitoire, ainsi que quelques uns des descripteurs proposés dans [43].

Les données ont été prétraitées par un algorithme présenté dans [25], et un ensemble de 31 descripteurs ont été extraits. La liste complète et un résumé de l'algorithme est détaillée dans [33]. La matrice de descripteurs obtenue est disponible sur le lien [https://drive.google.com/drive/folders/1H413RxYu4ya7YMEgF\\_1Th\\_fHr7flvv00?usp=sharing](https://drive.google.com/drive/folders/1H413RxYu4ya7YMEgF_1Th_fHr7flvv00?usp=sharing). Tous les descripteurs n'ont pas été retenus car certains avaient des corrélations importantes. Un seuil de 0.9 a été fixé sur la corrélation, et un seul des deux descripteurs ayant une corrélation supérieure a été conservé. Après cette étape, 13 descripteurs ont été conservés. Cette étape permet de réduire le temps de calcul et évite

les problèmes numériques lors de l'implémentation de certains algorithmes de partitionnement (ceux incluant, par exemple, le calcul d'une inverse de matrice de covariance).

Parmi les descripteurs retenus on retrouve :

- (1) *Temps de montée* : le temps entre le début du transitoire et l'amplitude crête [ $\mu$ s].
- (2) *Nombre de pics* : le nombre de dépassements de seuil pendant l'EA
- (3) *L'énergie* : l'aire entre la partie du signal supérieure au seuil et l'axe horizontal fixé au seuil [ $\mu$ Vs].
- (4) *Amplitude* ou *amplitude crête* : l'amplitude maximale atteinte par le signal [dB].
- (5) *RMS* (root mean square) : la valeur efficace en  $\mu$ V.
- (6) *Nombre de pics à la montée* : le nombre de dépassements de seuil du début du transitoire à l'amplitude crête.
- (7) *La fréquence de réverbération* : le rapport entre les différences (nombre de pics - nombre de pics à la montée) et (durée- temps de montée)
- (8) *La fréquence d'initiation* : le rapport entre le nombre de pics jusqu'au seuil et le temps de montée.
- (9) *La puissance partielle* dans l'intervalle [0 200] kHz
- (10) *La fréquence du pic* : la fréquence qui correspond à l'amplitude maximale dans le domaine de Fourier [kHz].

Les derniers descripteurs sont obtenus à partir du scalogramme.

- (11) *L'énergie (spectre) moyenne sur le scalogramme*
- (12) *L'énergie maximale du signal sur le scalogramme*
- (13) *La fréquence centroïde* du signal du scalogramme .

Tableau 1.2 – Descripteurs obtenus pour les essais "ORION-EA"

Essais	B	C	D	E	F
Nombre de niveaux serrage	7	6	7	7	7
Nombre de transitoires	14089	13863	13107	10146	11190

Une médiane mobile de 100 échantillons a été appliquée à chaque dimension de la matrice de descripteurs résultante afin d'assurer une cohérence temporelle des données. Les valeurs obtenues sont ensuite normalisées (centrées et réduites) suivant chaque dimension pour réduire les effets des ordres de grandeurs lors du calcul des distances utilisées par les méthodes d'apprentissage.

Le tableau 1.2 présente pour chacun des essais, le nombre de niveaux de serrage et le nombre de transitoires détectés.

Le nombre  $D$  de descripteurs est aussi important que la nature de ces derniers, car peu de descripteurs favorise la perte d'information et un nombre très grand peut créer de la redondance et augmenter

exponentiellement le temps de calcul.

### 1.1.5 Interprétation des données d'émission acoustique

L'interprétation des émissions acoustiques est généralement effectuée soit par apprentissage, soit par localisation (voire le couplage des deux). L'apprentissage se fait à l'aide des méthodes supervisées et non supervisées. En pratique l'interprétation non supervisée consiste à *trouver le nombre de clusters* (données ayant des caractéristiques similaires) et les *identifier*. Les méthodes supervisées quant à elles, déterminent une *fonction* qui caractérise chaque *classe* et la différencie des autres. Les méthodes supervisées exigent que les données soit annotées (figure 1.10) .

#### a) Par apprentissage

Comme mentionné précédemment, bien qu'il soit avéré que les endommagements génèrent des EA, il est très difficile d'annoter manuellement les signaux transitoires collectés sur un système. En effet, une source d'EA est caractérisée par différents paramètres comme sa position dans la matière, son orientation, l'énergie libérée etc. Ces paramètres dépendent de la nature de la source, et vont changer au fur et à mesure que le système (ou localement, le matériau) est utilisé et s'endommage. Il peut aussi exister des sources méconnues, ce qui ajoute de l'incertitude quant à l'origine précise des signaux. Lorsque nous collectons des milliers d'EA pendant un essai (par exemple entre 10000 et 20000 pour un essai et un capteur donné dans le benchmark ORION-EA), il est donc en général de mise d'utiliser des méthodes d'apprentissage non supervisées. Pour le cas particulier d'ORION-EA, les méthodes supervisées pourraient être explorées en utilisant l'information sur le niveau de serrage comme une « classe ». Ces classes seront utilisées par la suite pour évaluer la qualité des méthodes de partitionnement et de validation utilisées ou développées dans cette thèse.

Dans [45] par exemple, les auteurs mettent en œuvre une méthode supervisée pour discriminer la rupture de fibres de la fissure matricielle pour un composite unidirectionnel. Des tests sont préalablement faits sur les différents composants (fibres et matrice) afin d'identifier leurs signatures. En supposant que les éléments gardent les mêmes propriétés après l'assemblage, de nouveaux signaux acquis sont classifiés.

Les méthodes non supervisées permettent quant à elles une analyse exploratoire basée sur des données non annotées. Il s'agit essentiellement du partitionnement traduit de l'anglais *clustering* ou *cluster analysis* (figure 1.10). Cette approche nécessite de choisir un nombre de *clusters* et un sous-ensemble de descripteurs. Différentes partitions sont donc obtenues en variant ces paramètres. La validation permet ensuite de choisir la meilleure partition. Les algorithmes de partitionnement et de validation seront présentés dans le détail dans la section suivante [46, chap.4].

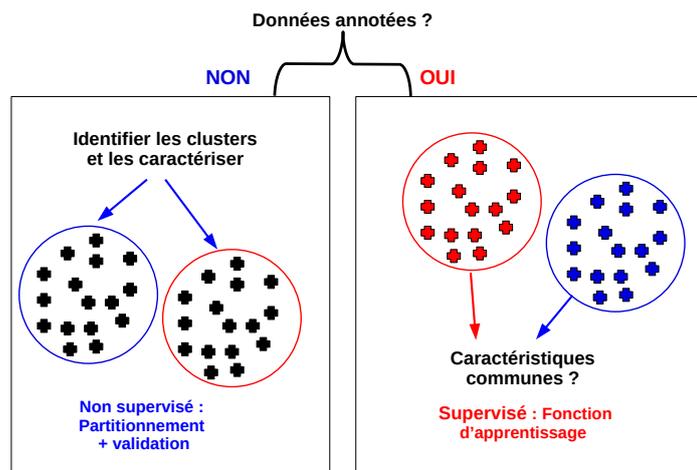


Figure 1.10 – Méthodes d'apprentissage supervisée et non supervisée

### b) Localisation d'une émission acoustique

En utilisant plusieurs capteurs, il est théoriquement possible de localiser la source d'une émission acoustique. Des algorithmes ont été proposés dans le cas linéaire (poutre), 2D (plaques) et 3D (volumes) [47, 48]. Le principe, simplifié mais utilisé largement dans le cas linéaire sur une poutre, consiste à mesurer les temps d'arrivée sur deux capteurs. Soit  $S_1, S_2$  la position des capteurs puis  $t_1, t_2$  les temps respectifs d'arrivée du signal sur ces capteurs. Soit  $c$  la vitesse de propagation des ondes dans le matériau dans la direction donnée par la droite  $(S_1, S_2)$ , Après avoir vérifié que la différence entre  $t_1$  et  $t_2$  est inférieure au temps limite d'arrivée du signal (dicté par la vitesse et la distance entre capteurs), nous déterminons par le calcul la position de la source par :

$$OP = \frac{1}{2}(OS_1 + OS_2 - (t_2 - t_1) \times c) \quad (1.1)$$

O est l'origine du repère. Dans ce calcul il est nécessaire de faire les hypothèses suivantes : la vitesse est connue, un évènement arrive sur les deux capteurs, avant qu'un deuxième évènement ne survienne, les capteurs ont un rayon d'action suivant  $(S_1, S_2)$ .

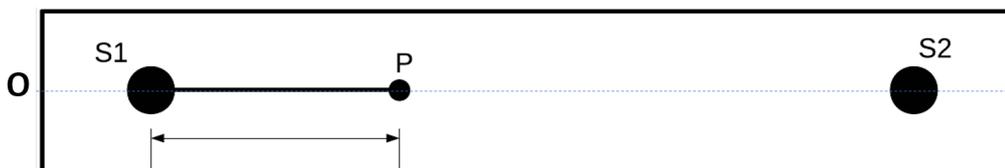


Figure 1.11 – Localisation d'un signal d'EA sur une poutre isotrope

Concernant la première hypothèse, il est connu que la vitesse peut changer et donc doit être ré-estimée au cours du temps (en général on interrompt l'essai). La deuxième hypothèse est très restrictive

sur certaines familles de matériaux, comme les composites. Ces matériaux émettent en effet beaucoup d'évènements en même temps et de manière très diffuse dans la matière rendant impossible la localisation précise. La troisième hypothèse nécessite de bien positionner les capteurs car l'amplitude des EA diminue avec la distance parcourue, rendant difficile leur détection si les capteurs sont trop éloignés. En complément sur la vitesse, il est nécessaire de prendre en compte les variations en fonction de la direction de propagation, dans le cas 2D et 3D. Cela rend bien plus difficile la localisation et nécessite une connaissance fine sur le milieu de propagation qui évolue au cours des essais. Enfin, l'algorithme de détection des transitoires est critique pour la localisation puisque le temps d'apparition des signaux sur les capteurs intervient dans les formules de localisation [48].

Dans [49], une approche différente est proposée en construisant un modèle purement numérique. Ce modèle de prédiction est basé sur des réseaux de neurones qui apprennent les positions à partir d'une base de signaux dont les positions des sources sont connues.

### c) Défis liés à l'interprétation des signaux d'émission acoustique

Les défis concernant l'utilisation de la technique des EA peuvent être groupés en différentes catégories chacune liée à une phase du processus d'interprétation. Dans [3, chap.4], M. Sause définit une analyse des EA performante si elle peut permettre de manière précise et concise :

1. **La détection des émissions** : le premier défi porte sur le choix des capteurs car ils sont caractérisés par une sensibilité qui dépend de la fréquence des ondes libérées dans le matériau surveillé. L'expérience guide l'utilisateur dans le choix des capteurs, mais parfois des essais sont nécessaires avec une panoplie de capteurs pour déterminer les plus adaptés. En général, un seul type de capteur est utilisé. La chaîne d'acquisition est aussi très importante pour collecter des signaux de qualité (alimentation des capteurs, préamplificateur, oscilloscope numérique). La détection nécessite parfois un débruitage préalable (dont il faut régler les paramètres) et le choix des paramètres (HLT et HDT, figure 1.9) pour définir la longueur des signaux.
2. **La localisation des signaux détectés** : Afin d'élaborer un bon diagnostic et une intervention, la connaissance de la position d'une source d'émission est parfois nécessaire. Comme vu précédemment, la localisation repose sur de nombreuses hypothèses, en partie liées aux propriétés physiques du matériau surveillé (dont les lois de propagation des ondes dans la structure). Des groupes tels que MISTRAS ont acquis une grande expertise dans la localisation spatiale sur des structures industrielles, comme les gigantesques unités de stockage utilisées dans l'industrie pétrochimique.
3. **L'identification des classes ou clusters** : Pour interpréter les EA, les méthodes d'apprentissage supervisées sont plus aisées à mettre en oeuvre mais nécessitent de disposer d'*a priori* sur les signaux. Dans la grande majorité des applications, l'apprentissage non supervisé s'im-

pose. L'inconvénient des méthodes non supervisées concerne la validation des résultats. Ces derniers varient en fonction de plusieurs hyper-paramètres utilisés dans ces méthodes ajoutant de l'incertitude à l'interprétation [3, chap.4].

M. Sause [3, chap.4] définit pour ces trois défis des incertitudes (PoD : Probabilité de détection, UoL : Incertitude de localisation, UoC : Incertitude de *classification*). La réduction de ces incertitudes conduit à l'amélioration du diagnostic.

Par la suite nous nous intéressons uniquement aux défis liés au troisième point. Les méthodes et les algorithmes proposés sont utilisés hors ligne, il s'agit d'une fouille de données expérimentales.

## 1.2 Méthodes non supervisées pour l'analyse des émissions acoustiques

Nous nous intéressons dans cette section aux méthodes de partitionnement usuelles basées sur l'utilisation des descripteurs.

### 1.2.1 Approches courantes pour l'interprétation non supervisée

L'idée d'utiliser les descripteurs date des toutes premières caractérisations des endommagements par la technique de l'EA. La première approche consiste à faire du partitionnement avec un seul descripteur. Chaque source est alors caractérisée par un intervalle ou par une distribution gaussienne du descripteur [50–54]. Pour la plupart de ces études, les descripteurs utilisés sont l'amplitude ou la fréquence moyenne du signal. Des représentations 2D (avec deux descripteurs) sont également utilisées pour la caractérisation des signaux. Les signaux d'EA sont représentés par des points dans un plan, les sources d'EA correspondent alors à des distributions localisées aux zones ayant une grande densité de points [55, 56]. Ces approches ont montré des limites pour des structures et équipements complexes, dans ces cas un ou deux descripteurs se révèlent insuffisants pour décrire les *clusters* [57, 58].

Par la suite des méthodes multi-dimensionnelles ont vu le jour. Quelques descripteurs choisis par des experts permettent de segmenter les signaux à partir des algorithmes de partitionnement (les k-moyennes, les regroupements hiérarchiques, les mélanges gaussiens, les cartes auto-organisatrices, le partitionnement flou...) [59].

Une autre manière de procéder consiste à se servir de  $D$  (assez grand) descripteurs et de faire du partitionnement avec des sous-ensembles de  $D_1$  (avec  $D_1 < D$ ) descripteurs. L'idée est que parmi les descripteurs initialement extraits, il existe un sous-ensemble permettant d'avoir des *clusters* qui optimisent une fonction objectif [3, 17]. L'étape de validation permet ainsi de trouver ce sous-ensemble

optimal en plus du nombre de *clusters*. La plupart des méthodes de partitionnement font intervenir une mesure de distance. Lorsque la dimension augmente, le volume de l'espace croît également, les données se retrouvent alors isolées. Ce problème est connu sous le nom de « fléau ou malédiction de la dimension » (*curse of dimensionality*) [60, 61]. Utiliser différents sous-ensembles permet de contourner ce problème et aussi de déterminer les descripteurs les plus discriminants.

La réduction de dimension à l'aide d'une analyse en composantes principales (ACP) répond également au problème sus-mentionné. L'ACP permet de générer des descripteurs décorrélés [20, 62]. Étant des combinaisons linéaires des descripteurs extraits, ces nouveaux descripteurs n'ont pas nécessairement un sens physique. Cette méthode est détaillée à l'annexe A.3.

### 1.2.2 Méthodes de partitionnement usuelles pour l'émission acoustique

Nous nous proposons de définir les termes et expressions mathématiques que nous utiliserons par la suite. On dispose de  $N$  signaux d'EA représentés par un ensemble de  $D$  descripteurs, notés  $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ . Chaque observation  $\mathbf{y}_n$  est constituée de  $D$  composantes  $[y_{n1}, y_{n2}, \dots, y_{nD}]$ ,  $\mathbf{y}$  est donc une matrice de dimension  $[N \times D]$ . Dans l'espace des descripteurs, chaque transitoire est représentée graphiquement par un point.

Le nombre de *clusters*,  $K$  est un paramètre fixé lors du partitionnement. Il est considéré comme le nombre des sources d'EA ayant *a priori* généré les signaux observés. Pour chaque donnée  $\mathbf{y}_n$ , une variable  $\mathbf{x}_n$  est définie pour représenter le *cluster*.  $\mathbf{x}_n$  est un vecteur de  $K$  valeurs binaires :  $[x_{n1}, x_{n2}, \dots, x_{nK}]$  données par l'équation 1.2.

$$x_{nk} = \begin{cases} 1, & \text{si } y_n \in C_k \\ 0 & \text{sinon} \end{cases} \quad (1.2)$$

où  $C_k$  est le  $k^{\text{ème}}$  cluster.

On définit également la distance  $d(y_1, y_2)$ , elle est une mesure de la similarité entre deux signaux  $y_1$  et  $y_2$ . Tout au long de ce manuscrit  $d$  représentera la distance euclidienne définie par :

$$d(y_1, y_2) = \sqrt{\sum_{d=1}^D (y_{1d} - y_{2d})^2} \quad (1.3)$$

Un *cluster*  $C_k$  est représenté par son centre  $\mu_k$  et sa dispersion  $\Sigma_k$  qui mesure la similarité entre les signaux du *cluster*. Plus elle est petite, plus le *cluster* est compact.  $N_k$  est défini comme le nombre de signaux appartenant au cluster  $C_k$ .

$$\begin{aligned}
 \mu_k &= \frac{1}{N_k} \sum_{n=1}^N x_{nk} y_n \\
 \Sigma_k &= \frac{1}{N_k} \sum_{n=1}^N x_{nk} (y_n - \mu_k)(y_n - \mu_k)^T \\
 N_k &= \sum_{n=1}^N x_{nk}
 \end{aligned} \tag{1.4}$$

On note aussi  $\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})$  (défini pour  $\boldsymbol{\mu} = \{\mu_k\}$  l'ensemble des centres des clusters), l'inertie de la partition, elle caractérise la dispersion au sein des différents clusters. Elle est définie par :

$$\mathcal{J}(\mathbf{y}, \boldsymbol{\mu}) = \sum_{n=1}^N \sum_{k=1}^K x_{nk} d^2(y_n, \mu_k) \tag{1.5}$$

La séparabilité des *clusters* est donnée par :

$$\mathcal{B}(\mathbf{y}, \bar{\mathbf{y}}) = \frac{1}{K} \sum_{k=1}^K d^2(\mu_k, \bar{\mathbf{y}}) \tag{1.6}$$

$\bar{\mathbf{y}}$  est le barycentre de l'ensemble des signaux, ses composantes sont les valeurs moyennes des descripteurs des  $N$  signaux. La mesure de  $\mathcal{B}(\mathbf{y}, \bar{\mathbf{y}})$  permet d'évaluer la dispersion des centres des *clusters*. Plus elle est grande, plus les *clusters* sont distincts.

Les cinq principaux algorithmes de partitionnement utilisés dans la communauté EA sont décrits ci-après : les *k-moyennes* [3, 35, 39], les *regroupements hiérarchiques* [63, 64], les *modèles de mélanges gaussiens* [65, 66], le *partitionnement flou* [67, 68] et la *méthode consensuelle* [69].

### a) K-moyennes

Traduit de l'anglais *K-means*, la première version des K-moyennes a été proposée en 1957 par S. Lloyd [70]. Cet algorithme est une méthode heuristique visant à construire  $K$  clusters qui minimisent la dispersion  $\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})$  intra-clusters. L'intuition de l'algorithme est qu'une concentration d'observations dans une zone de l'espace des descripteurs est caractéristique d'un *cluster*.

Les centres  $\mu_k$  des clusters sont initialisés puis mis à jour à chaque itération en associant au cluster  $C_k$  l'observation  $y_n$  si elle est plus proche de  $\mu_k$  que des autres centres. L'algorithme des K-moyennes se décompose en ces quatre étapes suivantes :

1. **Initialiser les centres de clusters** : On définit aléatoirement  $K$  parmi les  $N$  observations comme les centres des clusters.
2. **Estimer les distances  $d(y_n, \mu_k)$  par rapport à chaque centre et estimer  $x_{nk}$**  de l'équation

(1.2).

$$x_{nk} = \begin{cases} 1, & \text{si } k = \operatorname{argmin} (d(y_n, \mu_k)) \\ 0 & \text{sinon} \end{cases} \quad (1.7)$$

$x_{nk}$  prendra donc la valeur 1 si et seulement si le centre  $\mu_k$  du cluster  $C_k$  est le centre le plus proche de  $y_n$ .

### 3. Réestimer les centres des clusters :

$$\mu_k = \frac{1}{N_k} \sum_n x_{nk} y_n \quad (1.8)$$

$$N_k = \sum_n x_{nk} \quad (1.9)$$

$N_k$  est le nombre de signaux appartenant au cluster  $C_k$ .  $x_{nk}$  est nul pour toute observation n'appartenant pas à  $C_k$ ,  $\mu_k$  représente le centre du *cluster*.

4. **Calculer la dispersion  $\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})$  et vérifier sa convergence.** On estime que celle-ci est atteinte si  $\mathcal{J}^{(i)} - \mathcal{J}^{(i-1)} < tol$ , une tolérance fixée ou si le nombre d'itération maximal est atteint.  $i$  et  $i - 1$  sont respectivement l'itération en cours et la précédente. Si l'une de ces deux conditions est vérifiée, l'algorithme s'arrête. Dans le cas contraire, on retourne à la deuxième étape.

La convergence de  $\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})$  a été démontrée dans [71], un nombre d'itérations maximal est quand-même fixé pour optimiser le temps de calcul pour les cas où la convergence est lente.

L'algorithme des k-moyennes présente l'avantage d'être rapide et facile à implémenter, il est plus adapté aux cas où le nombre de *clusters* est déjà connu et les observations regroupées en amas. Dans [71], MacQueen cite aussi des difficultés possibles avec des observations dont les composantes sont linéairement dépendantes, une analyse en composantes principales préalable est alors requise.

Le résultat final dépend de l'initialisation. Il existe des méthodes telles que le `kmean++` pour réduire cette dépendance [72]. Le choix de la distance a également une grande importance, l'utilisation de la distance euclidienne par exemple impose des clusters en forme de disque (2D), sphère (3D) ou hypersphère ( $D > 3$ ). Dans ce cas, il est supposé que toutes les composantes de l'observation ont la même importance, ce qui n'est pas toujours vrai. Supposons par exemple une observation  $\mathbf{y}$  de dimension 2D, dont une dimension est d'un ordre grand avec très peu de variance et la seconde d'un ordre faible mais très fluctuante. En utilisant la distance euclidienne nous arriverons à des clusters prenant en compte uniquement  $\mathbf{y}^1$  comme s'il s'agissait d'une variable 1D. Il est recommandé d'utiliser les variables centrées réduites ou la distance de Mahalanobis (qui est fonction de la matrice de covariance) pour pallier la situation [71].

## b) Méthodes hiérarchiques

Le partitionnement hiérarchique peut être ascendant ou descendant. Il repose sur des dissimilarités par paires, ce qui permet d'évaluer la similarité entre les données d'un même *cluster* et entre *clusters*.

## Hierarchique ascendante

Le but de cet algorithme est de réunir les points présentant le plus de similarité dans le même *cluster*. Au départ chaque point est considéré comme un *cluster*, la dispersion intra-*cluster*  $\mathcal{J}(\boldsymbol{\mu}, \mathbf{y})$  est alors nulle. À chaque itération on fusionne les deux *clusters* dont les centres sont les plus proches, la dispersion  $\mathcal{J}(\boldsymbol{\mu}, \mathbf{y})$  croît au fur et à mesure des itérations. Si  $K$  est fixé on a  $N - K$  itérations, dans le cas contraire on s'arrête lorsqu'il y a une forte variation de  $\mathcal{J}(\boldsymbol{\mu}, \mathbf{y})$  entre deux valeurs de  $K$  consécutives. Ce critère d'arrêt est connu sous le nom de *critère de Ward*, l'intuition est que l'association de deux *clusters* "non naturels" crée un saut de  $\mathcal{J}(\boldsymbol{\mu}, \mathbf{y})$  [73]. Un dendrogramme permet de visualiser l'évolution des regroupements de *clusters*. L'algorithme est fait en quatre étapes :

1. **L'initialisation** : on considère chaque point comme un cluster.
2. **Le calcul des distances entre les différents centres de clusters**
3. **La fusion des deux clusters les plus proches** (figure 1.12 (flèche verte (1))). Le centre du nouveau *cluster* fusionné est ré-estimé.
4. **La vérification de la condition d'arrêt** s'en suit, si cette dernière n'est pas atteinte on retourne à la deuxième étape.

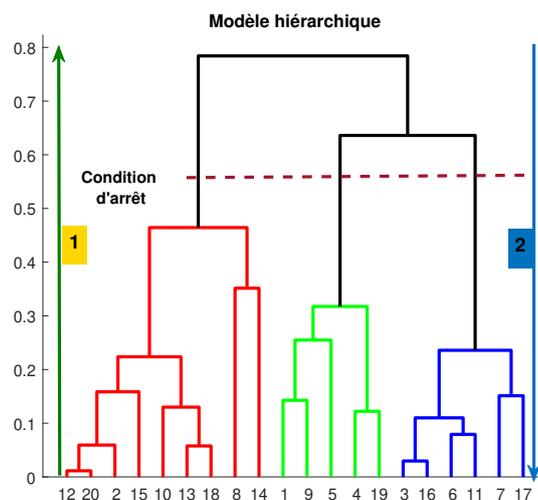


Figure 1.12 – Dendrogramme pour un jeu de 20 données générées aléatoirement | Le nombre de cluster est fixé à 3.

La méthode présente les mêmes avantages que le k-moyennes. En outre elle peut être utilisée pour des cas où  $K$  est inconnu en observant  $\mathcal{J}(\boldsymbol{\mu}, \mathbf{y})$ . Elle permet également la visualisation des données et du partitionnement. Le partitionnement hiérarchique ascendant n'a pas d'*a priori* sur la forme des clusters [74, chap.3].

Le temps de calcul peut devenir long lorsque la taille des données augmente car le nombre d'itération ( $N - K$ ) augmente.

## Hierarchique descendante

Illustrée à la FIGURE 1.12 (flèche bleu 2), la méthode hiérarchique descendante suit la logique contraire à la méthode ascendante. Au départ on a un seul *cluster*, puis à chaque itération on génère deux nouveaux *clusters* plus compacts en segmentant le *cluster* le plus dispersé [75]. L'algorithme fonctionne en 4 étapes :

1. *Initialiser l'algorithme en considérant les  $N$  points comme un seul cluster.*
2. *On calcule la dispersion  $\mathcal{J}_k$  au sein de chaque cluster.*

$$\mathcal{J}_k(\mathbf{y}, \mu_k) = \frac{1}{N_k} \sum_{n=1}^N x_{nk} d^2(y_n, \mu_k) \quad (1.10)$$

avec  $N_k = \sum_{n=1}^N x_{nk}$  pour une valeur de  $k$  fixée.

3. *Segmenter en deux le cluster ayant la plus grande valeur d'inertie ( $k = \arg_k \max(\mathcal{J}_k)$ ).*  
La division se fait à l'aide du k-moyennes ( $K_i = 2$ )
4. *Vérifier la condition d'arrêt*, si elle n'est pas atteinte on retourne à la deuxième étape.

Cet algorithme présente l'avantage de produire des résultats plus stables et des *clusters* avec des bonnes performances suivant les critères fixés (compact et distincts). Il est à noter que les *clusters* estimés ont souvent des tailles assez similaires. Il est donc rare d'avoir des *clusters* presque vides comme avec le k-moyennes ou la méthode hiérarchique ascendante, cependant il est plus difficile à implémenter [72]. Il est recommandé de travailler avec les valeurs centrées réduites des descripteurs afin d'homogénéiser les dimensions.

### c) Modèle de mélanges gaussiens

Décrits dans [76, 77] et [78, chap.9] et appliqués sur les données d'EA dans [65, 66], les modèles de mélanges gaussiens (*GMM : Gaussian Mixture Model*) sont des modèles probabilistes reposant sur un paradigme différent des algorithmes précédemment présentés. L'hypothèse principale est que chaque *cluster* peut être représenté par une distribution. Il s'agit d'un modèle de mélange. Un point peut donc appartenir aux *clusters* avec un certain degré d'appartenance (*soft clustering*).

#### Intuition des mélanges gaussiens :

Un GMM est un modèle statistique pour lequel on considère que la distribution des points dans chaque *cluster* suit une gaussienne. La probabilité d'observer  $y_n$  issue des observations  $\mathbf{y}$ , générées par  $K$  distributions gaussiennes  $\mathcal{N}_k(y_n | \mu_k, \Sigma_k)$  s'écrit :

$$p(y_n) = \sum_{k=1}^K \pi_k \mathcal{N}_k(y_n | \theta_k) \quad (1.11)$$

Cette loi est une somme pondérée des distributions gaussiennes  $\mathcal{N}_k(y_n|\theta_k)$  de chaque *cluster*  $C_k$  dont  $\theta_k=(\mu_k, \Sigma_k)$  correspondant à la moyenne et la dispersion de chaque *cluster*. Utiliser un GMM pour partitionner les signaux d'EA équivaut à considérer chaque mécanisme (ou famille) d'endommagement comme une gaussienne.

Dans l'équation 1.11 les  $\pi_k$  sont les proportions du mélange. Chaque  $\pi_k$  correspond à la probabilité qu'une observation appartienne au cluster  $C_k$ .

$$p(y_n \in C_k) = p(x_{nk} = 1) = \pi_k \quad (1.12)$$

$\mathcal{N}_k$  correspond à la probabilité de  $y_n$  sachant que  $\mathbf{y}_n \in C_k$ . Son expression est donnée par 1.13 :

$$p(y_n|x_{nk} = 1) = \mathcal{N}_k(y_n|\theta_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ \frac{-1}{2} (y_n - \mu)^T \Sigma^{-1} (y_n - \mu) \right\} \quad (1.13)$$

où  $(\mu, \Sigma)$  représentent respectivement la moyenne et la matrice de covariance de la variable ;  $^T$  est la transposée ;  $D$  est la dimension des données (nombre de descripteurs) et  $N$  la taille de  $\mathbf{y}$ .

### L'algorithme Espérance Maximisation (EM) :

Dans le cadre du partitionnement, ce qui nous intéresse c'est d'estimer  $p(x_{nk} = 1|y_n)$ , la probabilité *a posteriori* d'avoir le *cluster*  $C_k$  pour chaque donnée  $\mathbf{y}_n$ . Cette probabilité est obtenue en utilisant le théorème d'inversion de Bayes, en supposant  $\pi_k$  et  $\mathcal{N}_k(y_n|\theta_k)$  connus. On obtient ainsi :

$$p(x_{nk} = 1|y_n) = \frac{p(y_n, x_{nk} = 1)}{p(y_n)} \quad (1.14)$$

En gardant la même définition de  $x_{nk}$ , on définit  $E(x_{nk}) = p(x_{nk} = 1|y_n)$  l'espérance de  $x_{nk}$  :

$$E(x_{nk}) = \frac{\pi_k \mathcal{N}_k(y_n|\theta_k)}{\sum_{j=1}^K \pi_j \mathcal{N}_j(y_n|\theta_j)} \quad (1.15)$$

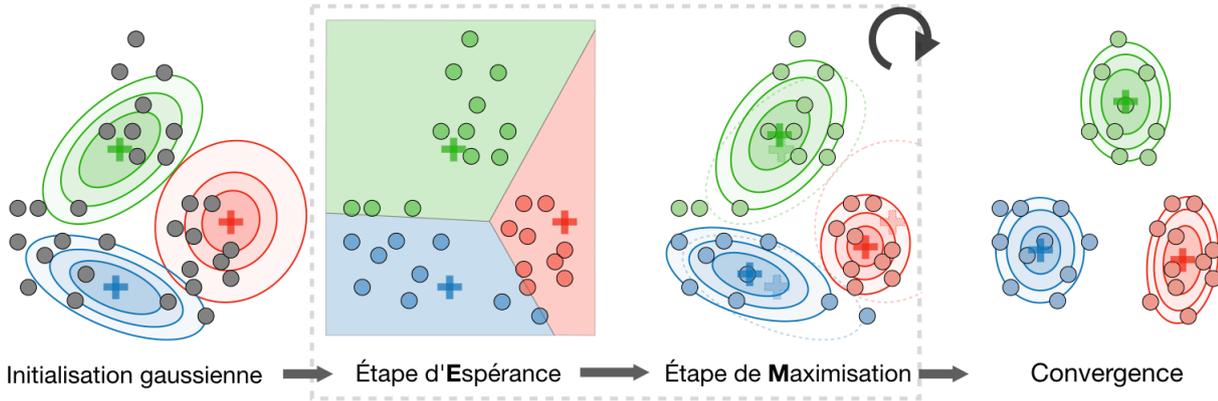
L'algorithme Espérance Maximisation (EM) permet d'estimer  $\boldsymbol{\pi} = \{\pi_1, \pi_2 \dots \pi_K\}$  et  $\boldsymbol{\theta} = \{\theta_1, \theta_2 \dots \theta_K\}$ , les paramètres du modèle qui optimisent la log-vraisemblance des données observées :

$$\mathcal{L}(\mathbf{y}) = \sum_{n=1}^N \log \sum_{k=1}^K \pi_k \mathcal{N}_k(y_n|\theta_k) \quad (1.16)$$

L'algorithme est illustré dans la Figure 1.13 pour un cas avec trois *clusters*. Les trois *clusters* sont initialisés avec des centres aléatoires puis ajustés jusqu'à la convergence comme pour l'algorithme des k-moyennes.

Les différentes étapes de l'algorithme EM pour générer un modèle sont :

1. *L'initialisation des valeurs  $\pi_k, \theta_k$  du modèle*


 Figure 1.13 – **Illustration 2D de l'algorithme Espérance Maximisation** [79]

2. **L'étape E** : on estime les espérances des états  $x_{nk}$  données par l'équation 1.15.

$$E(x_{nk}) = \frac{\pi_k \mathcal{N}_k(y_n | \theta_k)}{\sum_{j=1}^K \pi_j \mathcal{N}_j(y_n | \theta_j)} \quad (1.15)$$

Cette étape permet d'attribuer les points à différents *clusters*. Elle prend en compte les paramètres du modèle estimés à l'itération précédente.

3. **L'étape M** : on estime  $\theta^{(i+i)}$  qui maximise 1.16 :

$$\theta^{(i+i)} = \operatorname{argmax}_{\theta} \mathcal{L}(\mathbf{y}) \quad (1.17)$$

Les résultats obtenus dans la littérature correspondent aux expressions déjà définies précédemment :

$$\begin{aligned} \mu_k &= \frac{1}{N_k} \sum_{n=1}^N E(x_{nk}) \mathbf{y}_n, \text{ les centres des clusters} \\ \Sigma_k &= \frac{1}{N_k} \sum_{n=1}^N E(x_{nk}) (\mathbf{y}_n - \mu_k) (\mathbf{y}_n - \mu_k)^T, \text{ la dispersion au sein des clusters} \\ \pi_k &= \frac{N_k}{N}, \text{ les proportions de chaque cluster dans le mélange} \\ N_k &= \sum_{n=1}^N x_{nk}, \text{ le nombre de points par cluster} \end{aligned} \quad (1.18)$$

4. *L'évaluation de la convergence de  $\mathcal{L}(\mathbf{y})$ .*

L'algorithme EM n'est pas déterministe, comme les K-moyennes il est fonction de l'initialisation. Le GMM est similaire à un modèle de type k-moyennes où la distance a été remplacée par des probabilités pour mesurer les similarités [76, 77], [78, chap.9].

Nous avons mis plus de détails à la description des GMM que pour les autres méthodes à cause de sa complexité et surtout la nécessité de sa compréhension qui est incontournable pour le second chapitre.

#### d) Partitionnement flou

On garde les mêmes notations que lors des précédents paragraphes. Proposé dans [80], il est supposé que  $x_{nk}$  précédemment défini comme binaire est désormais une probabilité

$$\begin{cases} 0 \leq x_{nk} \leq 1 & \text{et} \\ \sum_{k=1}^K x_{nk} = 1 & ; x_{nk} \end{cases}$$

est le degré d'appartenance de  $y_n$  au cluster  $C_k$ .

L'algorithme cherche à minimiser la fonction objectif :

$$\begin{aligned} J(x, \theta) &= \sum_{n=1}^N \sum_{k=1}^K x_{nk}^\alpha D_{nk} ; \alpha \geq 1 \\ D_{nk} &= D_n(\theta_k) = (y_n - \mu_k)^T M_k^{-1} (y_n - \mu_k) \end{aligned} \quad (1.19)$$

avec  $1 \leq k \leq K$  est une norme métrique. La méthode vise à trouver les valeurs  $\theta_k = (\mu_k, M_k)$  qui minimise la fonction objectif.  $J(x, \theta)$  représente une mesure de similarité au sein des clusters.

L'algorithme est le suivant :

1. *Initialiser les valeurs  $\theta_k$ .*
2. **Étape E** : on estime les valeurs de  $x_{nk}$

$$x_{nk} = \frac{1}{\sum_{j=1}^K \left(\frac{D_{nk}}{D_{nj}}\right)^{1/(1-\alpha)}}, \text{ le degré d'appartenance à un cluster.} \quad (1.20)$$

3. **Étape M** : On estime  $\theta^{(i+1)}$  qui minimise  $J(\mathbf{y}, \theta^{(i)})$  :

$$\theta^{(i+1)} = \underset{\theta}{\operatorname{argmin}} \left( J(\mathbf{y}, \theta^{(i)}) \right) \quad (1.21)$$

Les résultats obtenus dans la littérature après dérivation de  $J$  sont donnés par :

$$\begin{aligned} \mu_k &= \frac{\sum_{n=1}^N x_{nk}^\alpha y_n}{\sum_{n=1}^N x_{nk}^\alpha}, \text{ les centres des clusters} \\ P_k &= \frac{\sum_{n=1}^N x_{nk}^\alpha (x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^N x_{nk}^\alpha} \\ M_k &= \left[ \frac{1}{\rho_k |P_k|} \right]^{1/n} P_k \\ D_{nk} &= D_n(\theta_k) = (y_n - \mu_k)^T M_k^{-1} (y_n - \mu_k), \text{ la dispersion au sein des clusters} \end{aligned} \quad (1.22)$$

4. *Evaluer  $J(x, \theta)$  et retourner à l'étape 2. s'il n'y a pas convergence.*

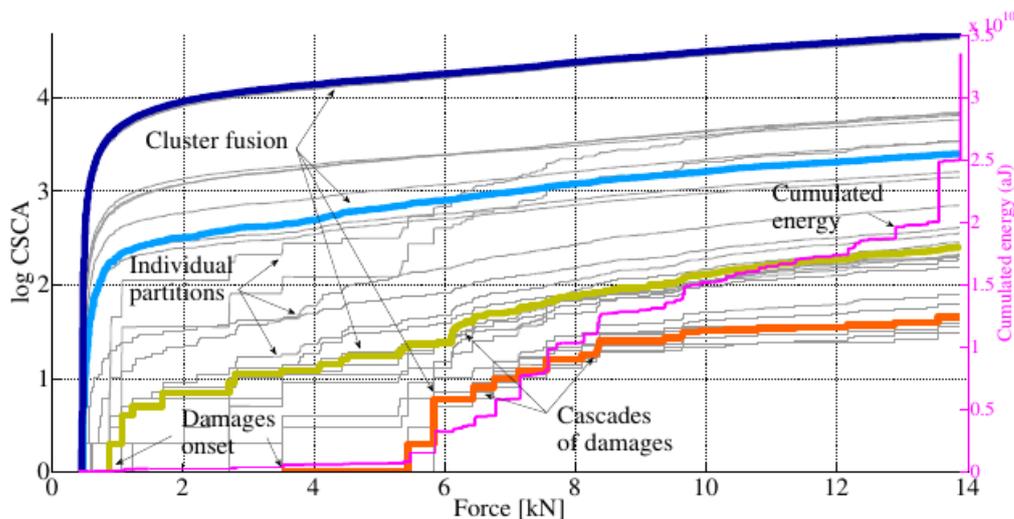
Le tableau 1.3 récapitule ce qu'il y a d'essentiel à retenir sur cette section pour faire le choix d'un algorithme de partitionnement [72, 74, 81]. Le choix peut ainsi être fait en fonction du type de logique (flou ou dur) ou d'autres paramètres. Il n'y a pas à notre connaissance une étude comparative des algorithmes de partitionnement appliqués en fonction du type de données d'EA.

Tableau 1.3 – Récapitulatif des méthodes de partitionnement classiques

Caractéristiques	K-moyennes	Hiérarchique	GMM	Partitionnement flou
<i>a priori</i>	$K$ , critère d'arrêt	$K$	$K$ , critère d'arrêt	$K$ , critère d'arrêt, $\alpha$
Forme du cluster	imposée	aucune	imposée	imposée
Logique	dure	dure	floue	floue
Déterministe	non	oui	non	non
Taille des clusters	peu variable	très variable	peu variable	peu variable
partitionnement	dur	dur	dur / soft	soft
Complexité	$O(D \times N \times K)$	$O(D \times N^2)$	$O(D^2 \times N \times K)$	$O(D^2 \times N \times K)$

### e) Méthode consensuelle

Si les données sont regroupées en amas bien distincts, les différents paramétrages (algorithme, initialisation, type de distance, prétraitement sur les descripteurs...) convergent vers la même partition. Cependant, pour des données réelles, il est fréquent que différents paramétrages mènent à des partitions également différentes. Il devient alors difficile de choisir objectivement la meilleure partition. Illustrée dans [82, 83], la méthode consensuelle cherche à fusionner des partitions et à construire une partition de consensus. L'approche est basée sur l'idée que chacune des partitions possède une fraction de la vérité.


 Figure 1.14 – Illustration du partitionnement consensuel pour  $K = 4$  clusters [69].

L'approche appliquée pour les données d'EA dans [69] est illustrée à la figure 1.14. Elle consiste à fusionner  $P$  différentes partitions pour un nombre de clusters  $K$  fixé. Pour cette approche, l'ordre

des données est fondamental, même si les auteurs utilisent des algorithmes de partitionnement qui ne prennent pas en compte l'ordre des données. Un partitionnement est fait au préalable avec différents sous-ensembles de descripteurs. Puis le consensus est effectué sur les meilleures partitions (sélectionnées à partir d'un critère) en suivant ces étapes :

1. Labelliser les clusters en fonction de leur ordre d'occurrence dans le temps. Ainsi le *cluster* apparaissant en premier est  $C_1$ , le cluster apparaissant en second est  $C_2$ , etc.
2. Trier les partitions suivant un critère en lien avec les proportions cumulées dans chaque *cluster*.
3. Représenter les 20 meilleures partitions (en gris figure 1.14) après tri sur un même graphe ce qui permet d'évaluer l'incertitude sous forme d'enveloppe sur la cumulée d'apparition.
4. Fusionner les partitions pour obtenir la partition « médiane » (qui correspond à la partition consensuelle, courbes en couleurs à la figure 1.14).

### 1.2.3 Validation du partitionnement

Valider un apprentissage non supervisé revient à choisir dans l'ensemble des partitions celle qui optimise un ou plusieurs critères de qualité. Ce choix doit répondre à deux interrogations :

- Pour un sous-ensemble de descripteurs donné, quel est le nombre de *clusters* qui optimise les critères de validation ? Ici on essaie de déterminer  $K_s$  le nombre de clusters "optimal" pour chaque sous-ensemble  $S^*$ .
- Parmi les différents sous-ensembles de descripteurs, lequel optimise le partitionnement (clusters bien distincts et compacts) ?

Pour y répondre la notion de **qualité d'un partitionnement** doit être définie. Les principales approches sont basées sur la forme des *clusters*, l'information mutuelle et la validation par plusieurs critères.

#### a) Basée sur la forme des clusters

Les critères basés sur la forme des *clusters* sont aussi souvent appelés critères internes car ils utilisent uniquement l'information contenue dans la partition à évaluer. Ils mesurent spécifiquement la densité des *clusters*, la partition sera bien notée si l'inertie intra-*clusters*  $\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})$  est minimale et l'inertie inter-*clusters* maximale. Cette approche est courante dans le cadre du SHM pour caractériser les endommagements et les modes de fonctionnement. Dans [84–87], les auteurs recherchent des sous-ensembles de caractéristiques qui distinguent le mieux les principaux mécanismes d'endommagement des matériaux composites que sont les ruptures de fibres, la fissure de la matrice et la décohésion. Dans les articles [88,89], on retrouve une liste exhaustive de critères de validation basée sur la forme des *clusters*. Nous avons choisi de ne présenter que les trois qui seront par la suite illustrés sur l'ensemble

des partitions issues des essais ORION-EA. Il s'agit des critères : *Silhouette*, *Calinski-Harabasz* et *Davies-Bouldin* déjà implémentés dans la *toolbox* présentée dans [90] afin d'assurer la reproductibilité des résultats. On retrouve :

1. **Le critère *Silhouette*** : Proposé pour la première fois dans [91], le *coefficient Silhouette* d'un point estime la différence entre la distance moyenne avec les points du **cluster le plus proche** et la distance moyenne entre ce point et ceux du **même cluster**. Pour un point  $y_n$  appartenant au  $k^{ieme}$  cluster : ( $y_n \in C_k$ ) :

$$C_{SIL}(n) = \frac{\beta(n) - \alpha(n)}{\max(\beta(n), \alpha(n))} \quad (1.23)$$

avec  $\beta(n) = \min_{k' \neq k} (\frac{1}{n_{k'}} \sum_{y_{n'} \in C_{k'}} d(y_n, y_{n'}))$  la distance moyenne du point avec ceux du cluster le plus proche, tandis que  $\alpha(n) = \frac{1}{n_k} \sum_{y_{n'} \in C_k} d(y_n, y_{n'})$  est la distance moyenne entre ce point et ceux du même cluster. La division par le maximum des deux valeurs permet la normalisation dans l'intervalle [-1 1], ce qui permet la comparaison avec les autres points. Si le partitionnement est bon,  $\alpha(n) \ll \beta(n)$ , le coefficient silhouette tendra alors vers 1, à l'inverse si le point n'est pas alloué au bon cluster cette valeur tendra vers -1. On peut étendre cette notion à un cluster  $C_k$  en moyennant les coefficients des points. En suivant la même analogie on tendra vers -1 pour les clusters avec une forte dispersion et proche de 1 pour les clusters compacts.

$$C_{SIL}[k] = \frac{1}{N_k} \sum_{y_n \in C_k} C_{SIL}(n) \quad (1.24)$$

Le critère silhouette de la partition est la moyenne des valeurs des différents clusters. Ce critère varie entre les valeurs -1 pour le pire partitionnement et 1 pour des clusters compacts et bien séparés.

2. **Le critère *Calinski-Harabasz*** : Le critère *Calinski-Harabasz* prend en compte la dispersion au sein de chaque cluster et la séparabilité entre les différents clusters. Il est défini par :

$$C_{CH}(\mathbf{y}, \boldsymbol{\mu}) = \frac{(N - K)\mathcal{B}(\boldsymbol{\mu}, \bar{\mathbf{y}})}{(K - 1)\mathcal{J}(\mathbf{y}, \boldsymbol{\mu})} \quad (1.25)$$

avec  $\mathcal{B}(\boldsymbol{\mu}, \bar{\mathbf{y}}) = \frac{1}{K} \sum_k d^2(\boldsymbol{\mu}_k - \bar{\mathbf{y}})$  la dispersion inter-classe, il estime la séparabilité entre les centres de clusters,  $\mathcal{J}$  garde le même sens qu'à l'introduction de cette section. Le partitionnement est bon lorsque la dispersion  $\mathcal{J}$  dans les clusters est faible et  $\mathcal{B}$  grand, on aura une valeur de  $C_{CH}$  maximale. Par contre si le partitionnement est mauvais il tendra vers 0. Le rapport produit  $\frac{N-K}{K-1}$  joue un rôle de régulateur pour défavoriser les modèles très complexes ( $K$  proche  $N$ ) [88, 92].

3. **Le critère *Davies-Bouldin*** : Tout comme le critère Calinski-Harabasz, le critère *Davies-Bouldin* prend en compte la séparabilité des clusters.

$$C_{DB}(\mathbf{y}, \boldsymbol{\mu}) = \frac{1}{K} \sum_{k' \neq k} \max\left(\frac{\sigma_k + \sigma_{k'}}{d(\boldsymbol{\mu}_k, \boldsymbol{\mu}_{k'})}\right) \quad (1.26)$$

La valeur  $\sigma_k$  représente la distance moyenne par rapport au centre  $\mu_k$  au sein du cluster  $C_k$ . Le partitionnement sera donc optimal si la valeur de  $C_{DB}$  est minimale et tend vers 0, et médiocre si le critère tend vers  $+\infty$  [93].

L'utilisation la plus simple de ces critères consiste à choisir la meilleure partition suivant un critère. Dans [94] où les critères Davies-Bouldin et Silhouette sont utilisés séparément pour déterminer les endommagements sur des composites renforcés par des fibres de verre. Le même processus est réitéré pour le tri des sous-ensembles de descripteurs. Il existe néanmoins deux approches plus complexes.

### b) Basée sur l'information mutuelle normalisée

L'information mutuelle normalisée (*NMI : Normalized Mutual Information*) est une mesure de qualité du partitionnement inspirée de la mesure de l'entropie. C'est un critère externe, tout comme l'indice de Rand (*RI : Rand Index*) [95] et la pureté (*purity*) [96]. Ces critères exigent une partition référente (les classes) pour l'évaluation.

L'information mutuelle est définie par :

$$MI(C, \hat{C}) = \sum_k^K \sum_{k'}^K P(C_k, \hat{C}_{k'}) \log \frac{P(C_k, \hat{C}_{k'})}{P(C_k)P(\hat{C}_{k'})} \quad (1.27)$$

où  $P(C_k) = n_k/N$ ,  $P(\hat{C}_{k'}) = n_{k'}/N$ ,  $P(C_k, \hat{C}_{k'}) = n_{kk'}$  sont les proportions normalisées correspondant, respectivement, au nombre de points dans le *cluster*  $C_k$ , dans  $C_{k'}$  et dans  $C_k \cap C_{k'}$ .

L'information mutuelle est majorée par la moyenne des entropies  $H(C)$  du *cluster* et  $H(\hat{C})$  de la classe. Cette valeur est donc normalisée pour pouvoir faire une comparaison des partitions ayant un nombre de *clusters* différent.

$$NMI(C, \hat{C}) = \frac{2MI(C, \hat{C})}{H(C) + H(\hat{C})} \quad (1.28)$$

Cette valeur tendra vers 0 si la partition évaluée et la référence sont indépendantes (mauvais partitionnement) et tendra vers 1 plus la partition est proche de la référence [97, 98].

Dans [69], l'idée est la suivante : la partition consensuelle (section 1.2.2, issue de l'algorithme qui prend  $n$  partitions et qui en génère une) est vue comme une référence, ce qui permet de calculer le NMI. Ceci est fait de manière répétée (bootstrap avec remise) en retirant des partitions parmi les  $n$ , recalculant le consensus, et le NMI.

Dans [97], il est proposé de choisir le nombre de *clusters* qui minimise la variance du NMI après bootstrap. Pour les auteurs de [97], cette manière de choisir  $K$  assure une certaine robustesse dans le choix de  $K$ , robustesse par rapport aux partitions utilisées pour calculer le consensus.

### c) Basée sur plusieurs critères

Pour un même ensemble de partitions, les critères de validation donnent souvent différents *clusters*. Valider avec plusieurs critères permet d'établir un compromis entre différents critères. Le choix de la partition finale semble alors plus objectif. La méthode de vote s'inscrit dans ce registre.

Le vote proposé dans [17] consiste à calculer plusieurs critères de qualité et trier les partitions en fonction de chaque critère, ce pour chaque sous-ensemble de descripteurs. La méthode consiste à faire deux « votes » successifs : le premier pour déterminer le nombre de *clusters* pour chaque sous-ensemble et le second pour le choix du sous-ensemble de descripteurs optimal.

Soit  $Q$  le nombre de critères utilisés,  $N_s$  le nombre de sous-ensembles. Le nombre de clusters  $k$  varie entre  $K_{min}$  et  $K_{max}$  et  $P = K_{max} - K_{min} + 1$  le nombre de valeurs possibles de  $k$ . Le nombre de partitions à trier est  $P \times N_s$ .

La première étape consiste à choisir la meilleure partition parmi les  $P$  partitions de chaque sous-ensemble. Les  $Q$  critères de validation sont calculés pour chacune des partitions, les  $P$  partitions sont triées suivant chaque critère. La première obtient une note  $P$  et la seconde  $P - 1$  et 1 pour la dernière. Le nombre *clusters*  $K_s$  est celui avec la meilleure note cumulée.

Pour le choix du sous-ensemble de descripteurs, l'approche est similaire. Pour chaque critère, on accorde une note 25 à la meilleure partition, 24 à la seconde, ainsi de suite 1 à la 25<sup>ième</sup> et 0 au reste des partitions. Une fois cette première étape faite, les notes sont additionnées pour chacune des partitions. Un compromis est établi entre les critères, seule une partition est sélectionnée, celle présentant le meilleur score cumulé.

La méthode a été proposée dans le cadre de la caractérisation de matériaux composites par la technique de l'EA. Elle peut être implémentée pour tout type de critères de validation (interne, externe ou relatif). Nous l'illustrerons dans la section suivante.

#### 1.2.4 Illustration du partitionnement sur ORION-EA

Nous nous proposons dans cette section d'illustrer une méthodologie de partitionnement avec trois des principales méthodes usuelles (k-moyennes, hiérarchique ascendante et GMM) déjà disponibles dans la boîte à outils du logiciel Matlab. La validation est celle du vote, elle se fait avec trois critères de validation internes (Silhouette, DB, CH) sus-présentés. Une analyse qualitative (à partir des graphes) et quantitative (tableau) est assurée. Les sous-ensembles sont constitués de  $d = 4$  descripteurs parmi  $D = 13$  descripteurs au total, soit 715 sous-ensembles de descripteurs possibles. Aucune connaissance sur le nombre de *clusters*  $K$  n'est supposée,  $K$  varie de 2 à 10 pour couvrir le nombre de serrage qui est de 6 ou 7. Sa valeur « optimale »  $K_{opt}$  sera estimée selon cette méthode de valida-

tion. Le nombre total de partitions à trier lors de la validation est de 6435 pour chaque méthode de partitionnement et pour tous les essais.

Le but est d'évaluer la capacité de la méthode multicritère par vote à retrouver les niveaux de serrage. L'évaluation quantitative est basée sur le nombre de changements de niveaux trouvés par rapport au nombre de *classes*; et le critère *ARI* (Adjusted Rand Index). Le critère *ARI* est une mesure de similarité (taux d'accord) entre 2 partitions qui sont généralement le vecteur des *clusters* estimés  $\mathbf{x}_n$  et le vecteur de *classes* (fournies par un expert). Cette mesure varie en général dans l'intervalle  $[0, 1]$ . Plus on se rapproche de 1, plus la partition est en accord avec la référence [95]. Formellement, cet indice est défini par :

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{N}{2}}{\frac{1}{2} \left[ \sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{N}{2}} \quad (1.29)$$

avec

	$C_{1,est}$	$C_{2,est}$	...	$C_{K',est}$	Somme
$C_{1,ref}$	$n_{11}$	$n_{12}$	...	$n_{1K'}$	$a_1$
$C_{2,ref}$	$n_{21}$	$n_{22}$	$\vdots$	$n_{2K'}$	$a_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
$C_{K,ref}$	$n_{K1}$	$n_{K2}$	...	$n_{KK'}$	$a_K$
Somme	$b_1$	$b_2$	...	$b_{K'}$	

où  $C_{k,ref}$  et  $C_{k',est}$  sont les numéros de clusters dans la partition de référence et dans celle estimée, respectivement. Le nombre de clusters peut être différent dans les deux partitions. L'expression de l'ARI dépend de l'indice de Rand qui calcule, pour deux partitions données (l'une étant la vérité terrain et l'autre le résultat d'un partitionnement), la proportion des paires d'éléments qui sont conjointement groupées ou conjointement séparées (représentée par les  $n_{kk'}$ ).

D'un point de vue qualitatif, l'interprétation de clusters sera également réalisée visuellement en corrélant les résultats du partitionnement avec les niveaux de serrage. De la sorte, l'analyse des résultats prend à la fois en compte la forme des *clusters* et leur chronologie.

## Essai B

Quelle que soit la méthode de partitionnement mise en oeuvre, la partition « optimale » établie par le vote détermine 6 clusters distincts. Le vote n'est donc pas capable de discriminer la partition correspondant aux 7 niveaux de serrage. Nous illustrons pour chaque essai les graphes correspondant à la partition « optimale » établie par le vote. Pour l'algorithme des k-moyennes, les données ont été projetées dans des espaces 2D pour une visualisation (Figure 1.15). Le sous-ensemble sélectionné est constitué du *temps de montée* (1), l'*énergie* (3), la *puissance partielle* [0-200] (9) et la *fréquence centroïde* (13). Le même sous-ensemble est sélectionné en utilisant les mélanges gaussiens au lieu

des k-moyennes comme algorithme de partitionnement. Pour le partitionnement hiérarchique, seul le descripteur *énergie* (3) est remplacé par la *fréquence du pic* (10), les trois autres descripteurs sont communs. En observant les premières partitions du tri, nous avons également constaté que les « meilleurs sous-ensembles » sont généralement les mêmes indépendamment de l'algorithme de partitionnement.

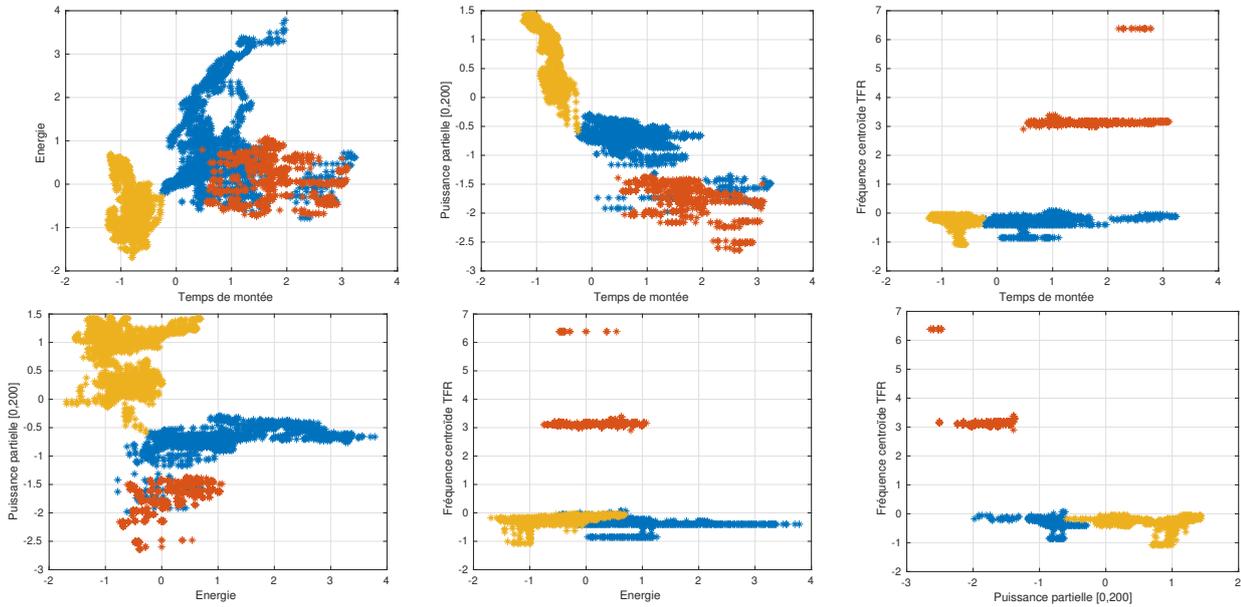


Figure 1.15 – **Partition après validation par la méthode de vote pour les k-moyennes sur l'essai B** | 3 clusters distincts sont obtenus. Le sous-ensemble sélectionné est constitué du *temps de montée* (1), l'*énergie* (3), la *puissance partielle [0-200]* (9) et la *fréquence centroïde TFR* (13).

La projection précédente est de nouveau faite pour les *classes* correspondant aux niveaux de serrage (figure 1.16). Les sept *classes* ne sont pas discriminées par ce sous-ensemble de descripteurs. Ce graphe illustre la complexité d'une tâche de partitionnement pour des données réelles d'EA.

Nous nous intéressons cette fois à la chronologie d'apparition des *clusters*. Pour cela les occurrences cumulées des différents *clusters* sont tracées pour les partitions sélectionnées pour chacun des algorithmes de partitionnement. Pour  $K$  fixé, un compteur est initialisé à 0 pour chaque *cluster*. Les données étant temporelles, on trace les logarithmes des cumulées de  $x_{nk}$  en fonction des instants de démarrage des transitoires. Dans la figure 1.17, nous avons à gauche le graphe d'occurrences cumulées pour les k-moyennes, la méthode hiérarchique au milieu et à droite les GMM. Pour les k-moyennes (figure 1.17a) et les GMM (figure 1.17c), nous avons trois clusters dont deux ont les instants de déclenchement bien distincts et correspondent à des instants de changement de serrage. Le premier *cluster* (bleu) englobe une partie des observations du premier serrage (60 cNm) et l'ensemble des trois autres premiers serrages (50 cNm, 40 cNm et 30 cNm). Le second *cluster* (rouge), actif au début de l'essai a les observations restantes du premier serrage. Au bout des 10 premières secondes, il n'est plus actif jusqu'à la fin de l'essai (70s). Le troisième cluster (jaune) englobe ensuite les trois derniers serrages. Dans les deux cas bien que le partitionnement accompagné d'une validation par vote ne permet pas de

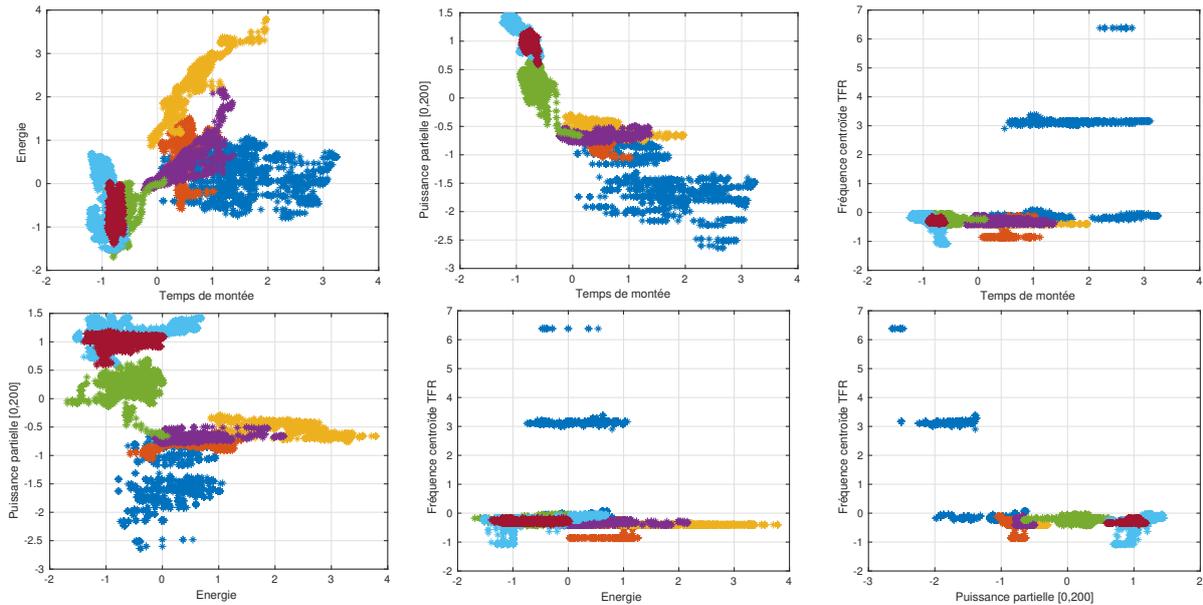


Figure 1.16 – **Classes correspondant aux serrages** | Les 7 serrages ne sont pas distincts dans cet espace de descripteurs. Le sous-ensemble sélectionné est constitué du *temps de montée* (1), l'*énergie* (3), la *puissance partielle [0-200]* (9) et la *fréquence centroïde* (13)

discriminer les 7 niveaux de serrage, la méthode permet de discriminer les serrages forts des serrages faibles.

Le partitionnement hiérarchique fournit six *clusters* dont trois sont actifs dès les premières secondes de l'essai (figure 1.17b). Le premier cluster (bleu) est actif uniquement au début de l'essai. Le second (rouge) est constitué d'une partie du premier serrage, il est commun aux différentes méthodes de partitionnement. Le troisième *cluster* (jaune) est actif pendant les 4 premiers serrages. Les quatrième et sixième *clusters* sont actifs quelques secondes pendant le deuxième et sixième serrages. Le cinquième *cluster* (vert) correspond aux serrages faibles, bien que certaines observations n'y sont pas attribuées. Le partitionnement hiérarchique associé au vote fournit ainsi un plus grand nombre de *clusters*, mais qui sont difficilement interprétables et n'ont pas de lien direct avec les niveaux de serrage.

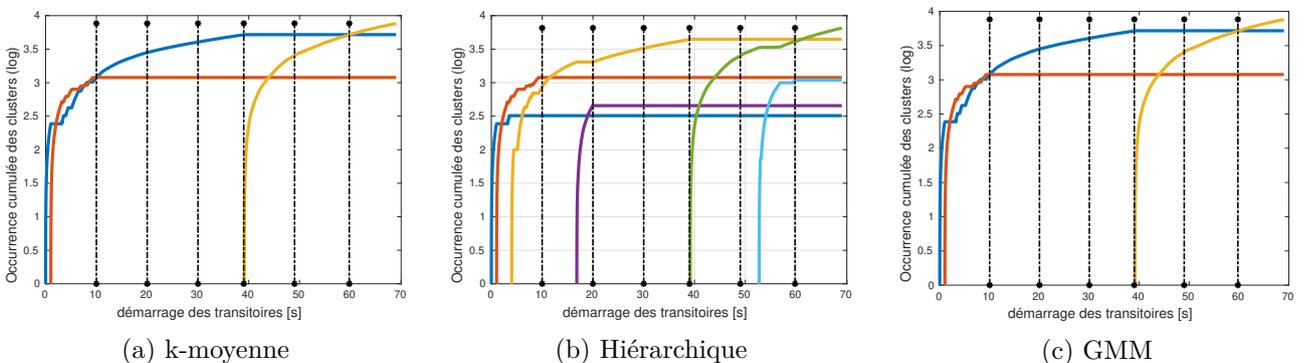


Figure 1.17 – **Résultats du partitionnement après validation par la méthode de vote proposée dans [3] pour l'essai B en utilisant les différents algorithmes** | Les k-moyennes à gauche, le modèle hiérarchique au milieu et les GMM à droite

L'interprétation des partitions sélectionnées pour l'essai B donne un avis assez mitigé sur la capacité de cette méthode d'analyse des EA, à trouver des partitions corrélées aux niveaux de serrage. Nous nous proposons d'observer l'ensemble des partitions pour avoir une vision globale. Pour cela, nous observons en premier l'ensemble des neuf partitions ( $k=2..10$ ) du meilleur sous-ensemble de descripteurs (Figure 1.18a); en second l'ensemble des partitions sélectionnées pour les différents sous-ensembles (Figure 1.18b). Dans les deux cas, le critère ARI sert de référence.

Nous avons illustré le vote pour la détermination du nombre de clusters uniquement avec les k-moyennes. Les critères de validation étant d'ordres de grandeur très différents, ils sont normalisés dans l'intervalle  $[0, 1]$  pour une meilleure visualisation.

Le vote se porte sur  $K = 3$ , à ce niveau les critères Davies-Bouldin (minimal) et le Silhouette (maximal) sont optimisés. Le critère ARI est faible, ce qui confirme une faible corrélation entre le sous-ensemble de descripteur « *optimal* » et les serrages (Figure 1.18a). Le critère Calinski-Harabasz est optimal à  $K=8$  clusters, cette valeur est la plus proche du nombre de serrages et la valeur optimale ( $K=9$ ) donnée par le critère ARI.

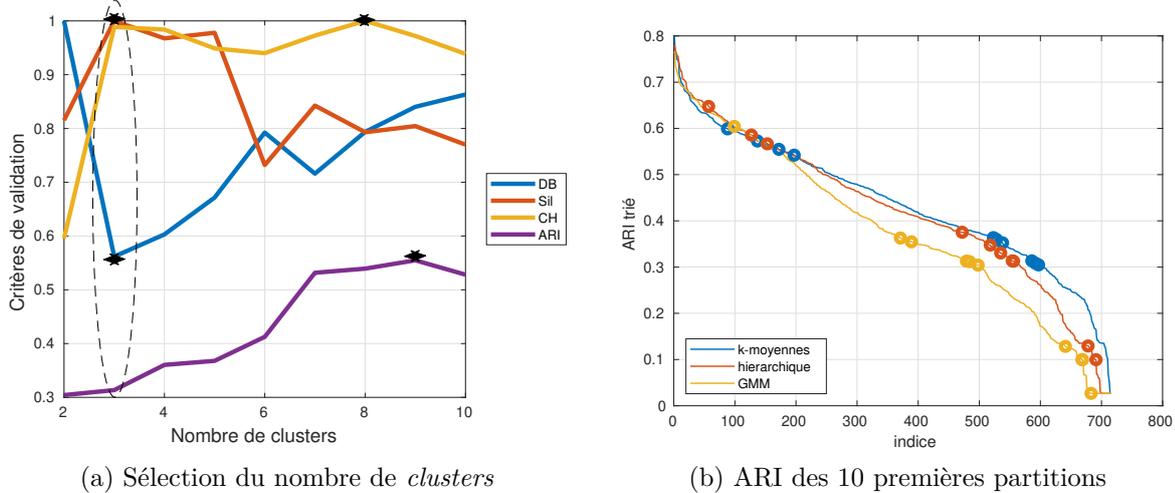


Figure 1.18 – Performances de la validation pour l'essai B. L'axe de la figure de droite correspond aux différents sous-ensembles possibles de descripteurs (sous-ensembles de cardinalité  $d$ ).

Les critères ARI des différentes partitions sélectionnées pour chaque sous-ensemble de descripteurs sont calculés, triés par ordre croissant et tracés. Les dix premières partitions du tri ont été représentées par des petits cercles sur la figure 1.18b. Quelques valeurs du ARI autour de 0.8 sont présentes, ce qui signifie que les algorithmes de partitionnement parviennent à discriminer les niveaux de serrage. Cependant les dix premiers choix du vote ont des valeurs d'ARI réparties le long du graphe, ce qui signifie que la méthode de validation peut aussi bien fournir des partitions en lien direct avec le serrage que des partitions sans lien apparent. La méthode de partitionnement est peut-être correcte mais il est nécessaire d'optimiser la validation.

## Essais C, D E et F

Pour le reste des essais, seules les occurrences cumulées pour chaque méthode ainsi que l'évolution du critère ARI seront illustrées.

Pour l'essai C, sept *clusters* sont obtenus avec les K-moyennes associés au vote (figure 1.19a). Le premier *cluster* (bleu) se déclenche au début de l'essai et reste actif durant tout le premier serrage. Il est inactif pendant les 2 serrages suivants, soit à 50 cNm et 40 cNm. Il s'active de nouveau au cinquième et septième serrage. Le second (rouge) est actif la moitié du temps au cours des premier et troisième serrages. Le troisième *cluster* correspond au second niveau de serrage. Le quatrième *cluster* correspond au troisième serrage, bien qu'il se déclenche précocement à 10s. Les trois derniers *clusters* réunis correspondent au cinquième niveau de serrage (10 cNm).

Le partitionnement hiérarchique (figure 1.19b) donne six *clusters* dont les quatre premiers sont similaires à ceux donnés par les K-moyennes. Les deux derniers *clusters* correspondent au cinquième serrage. Un même constat est observé pour les deux algorithmes de partitionnement : les serrages se trouvant dans le même *cluster* ne sont pas toujours successifs. Les GMM donnent six *clusters* (figure 1.19c). Le premier *cluster* est actif pendant le premier et le troisième serrage. Le second et troisième sont communs aux trois algorithmes. Le quatrième *cluster* regroupe les quatrième et sixième serrage. Le cinquième serrage est représenté par les deux derniers *clusters*.

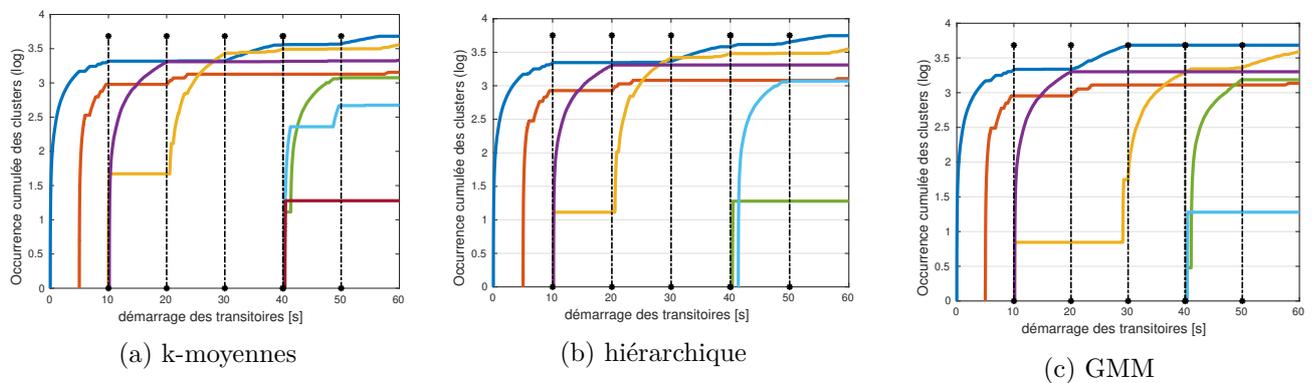


Figure 1.19 – Résultats du clustering après validation par la méthode de vote proposée dans [3] pour l'essai C en utilisant les différents algorithmes| Les k-moyennes à gauche, le modèle hiérarchique au milieu et les GMM à droite

Pour l'essai D (figure 1.20), la même partition à trois *clusters* est obtenue pour les trois algorithmes de partitionnement. Le troisième *cluster* correspond au troisième serrage. Le premier *cluster*, actif durant tout l'essai, regroupe l'essentiel des points des autres serrages. Le second « cluster » est constitué de quelques observations des différents serrages potentiellement aberrantes.

Les K-moyennes appliquées (figure 1.21a) à l'essai E fournissent 4 clusters liés aux serrages. Le premier *cluster* correspond au premier et sixième serrages. Le second *cluster* démarre à 10s et correspond au second et dernier serrages. Le troisième est constitué du troisième et quatrième niveaux de serrage.

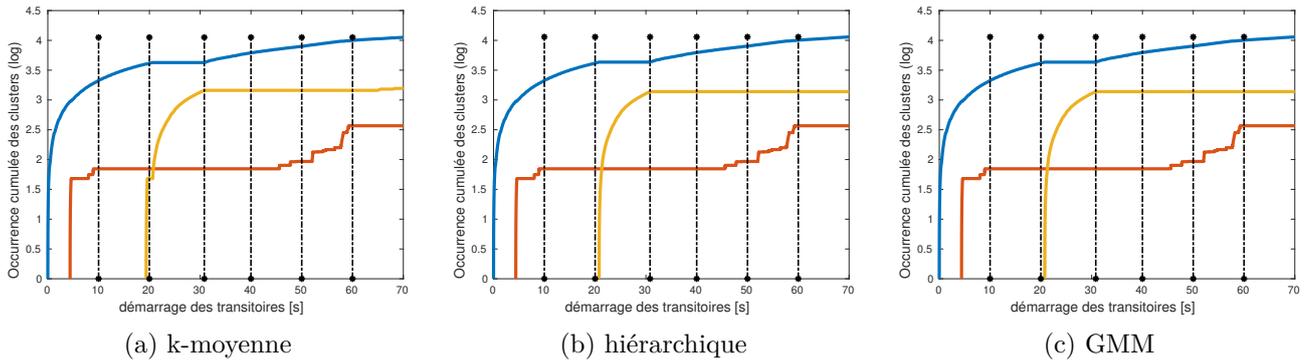


Figure 1.20 – Résultats du clustering après validation par la méthode de vote proposée dans [3] pour l'essai D en utilisant les différents algorithmes| Les k-moyennes à gauche, le modèle hiérarchique au milieu et les GMM à droite

Le quatrième *cluster* est actif au cours du cinquième serrage. Tous les instants de déclenchement de *clusters* correspondent à des changements de niveaux de serrage. Il existe très peu de mélange : les points d'un même serrage sont attribués au même *cluster*. L'essai E fournit des partitions identiques à six *clusters* pour le partitionnement hiérarchique (figure 1.21c) et les GMM (figure 1.21b), qui bien que débutant aux bons instants, n'ont pas de lien direct avec des niveaux de serrage.

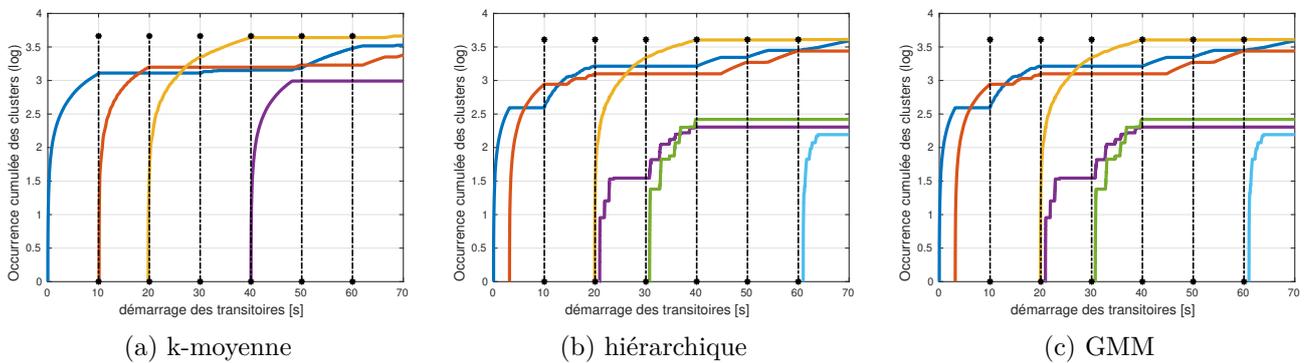


Figure 1.21 – Résultats du clustering après validation par la méthode de vote proposée dans [3] pour l'essai E en utilisant les différents algorithmes| Les k-moyennes à gauche, le modèle hiérarchique au milieu et les GMM à droite

L'essai F donne deux *clusters* qui démarrent tous au cours du premier serrage (figure 1.22). Le premier *cluster* est actif du début à la fin de l'essai (70s). Le second s'active après quelques secondes et devient inactif lorsqu'on passe du premier au second serrage jusqu'à la fin de l'essai. Il est difficile de lui donner un sens physique en particulier.

Dans le tableau 1.4, nous avons mis en lumière quelques caractéristiques de la « meilleure » partition pour chacune des méthodes de partitionnement sur chaque essai. Ces caractéristiques sont jugées importantes parce qu'elles permettent d'avoir une vision globale de la qualité de la partition autant dans l'espace des descripteurs que sa corrélation avec les niveaux de serrage. On constate que dans la

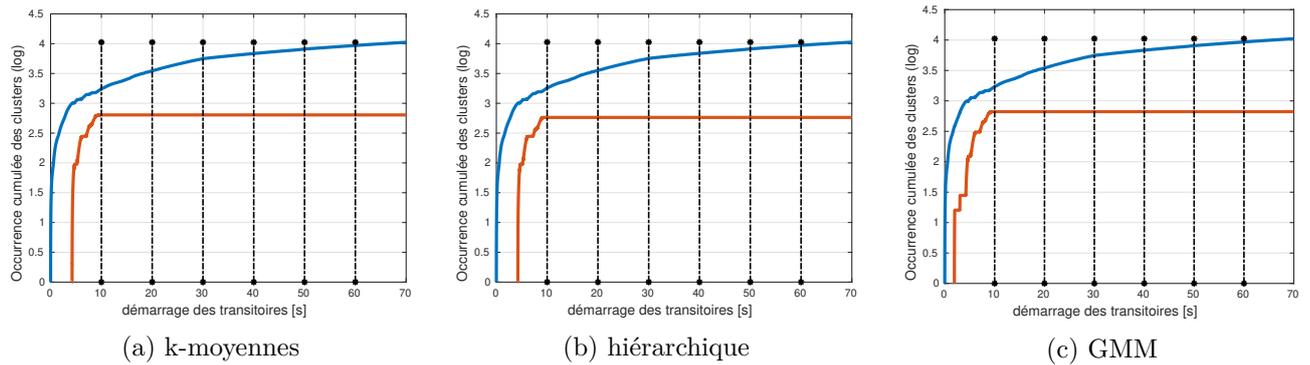


Figure 1.22 – Résultats du clustering après validation par la méthode de vote proposée dans [3] pour l'essai F en utilisant les différents algorithmes| Les k-moyennes à gauche, le modèle hiérarchique au milieu et les GMM à droite

plupart des cas le nombre de clusters est  $K = 3$ , ce qui est loin du nombre de niveaux de serrage. Les instants de changement de niveaux sont détectés correctement uniquement pour l'essai E. Le nombre de clusters étant faible, il y a regroupement de niveaux au sein des clusters.

Le critère ARI reste assez faible, exception faite pour l'essai E. Ce tableau met en évidence la non-relation de causalité entre la séparabilité spatiale et les niveaux de serrage.

Une autre remarque est que les sous-ensembles de descripteurs sélectionnés ont au moins deux de ces cinq descripteurs en commun : le nombre de pics (2), l'énergie du signal (3) ; la puissance partielle [0 200] (9) ; l'énergie maximale (10) et la fréquence centroïde (13). Cette observation est valide indépendamment de la qualité des partitions obtenues. Une hypothèse est que les critères de forme favorisent ces descripteurs parce qu'ils sont mieux étalés dans l'espace des descripteurs.

Cette fois nous nous intéressons aux performances globales de la validation. Les ARI sont triés par ordre décroissant, les 10 meilleures partitions suivant le vote sont représentés par des marqueurs.

La figure 1.23a conduit aussi à l'idée que plusieurs partitions sont liées aux niveaux de serrage pour l'essai C. Les valeurs d'ARI des partitions sélectionnées sont réparties sur tout l'intervalle. Certaines des partitions sont proches de 0.8, signe qu'elles correspondent à la vérité de terrain (niveaux de serrage). En contre partie, certaines partitions ont des ARI proches de 0.

Pour l'essai D, tenant compte des ARI les valeurs maximales restent autour de 0.6, le vote ne parvient pas à choisir des valeurs optimales (figure 1.23b). Les valeurs d'ARI sont proches de 0 pour les meilleures partitions du tri suivant le vote.

Le critère est supérieur à 0.5 pour certains des dix premiers sous-ensembles pour l'essai E, ce qui montre la corrélation entre données et signaux d'EA observés (figure 1.23c).

Pour l'essai F (figure 1.23d), l'interprétation est similaire à celle de l'essai D. Les valeurs basses du ARI découlent du nombre de clusters sélectionné très petit par rapport au nombre de niveaux de serrage et au mélange des *clusters*.

Tableau 1.4 – Résultats obtenus pour les essais « ORION-EA »

Essai B			
	k-moyennes	Hiérarchique	GMM
$K$	3	6	3
Instants trouvés / 6	1	1	1
Descripteurs	2; 3; 9; 13	2; 9; 10; 13	2; 3; 9; 13
ARI	0.315	0.335	0.315
Essai C			
	k-moyennes	Hiérarchique	GMM
$K$	7	6	6
Instants trouvés / 5	2	2	2
Descripteurs	4; 5; 9; 13	2; 3; 5; 9	6; 9; 10; 13
ARI	0.4553	0.2048	0.0002
Essai D			
	k-moyennes	Hiérarchique	GMM
$K$	3	3	3
Instants trouvés / 6	1	1	1
Descripteurs	2; 9; 12; 13	5; 11; 12; 13	5; 7; 10; 12
ARI	0.107	0.107	0.107
Essai E			
	k-moyennes	Hiérarchique	GMM
$K$	5	5	6
Instants trouvés / 6	4	4	4
Descripteurs	5; 9; 12; 13	5; 9; 12; 13	1; 9; 10; 13
ARI	0.5044	0.5044	0.320
Essai F			
	k-moyennes	Hiérarchique	GMM
$K$	2	2	2
Instants trouvés / 6	0	0	0
Descripteurs	4; 5; 10; 13	4; 5; 10; 13	4; 9; 10; 13
ARI	0	0	0

Dans l'ensemble les ARI sont meilleurs pour les k-moyennes que pour le partitionnement hiérarchique et les GMM. Quels que soient les algorithmes considérés, ceux-ci mettent en évidence des partitions pertinentes au regard de la « vérité terrain » avec des critères ARI de l'ordre de 0.7 voire supérieurs. En revanche, ces partitions les plus pertinentes ne sont pas (excepté pour l'essai C) sélectionnées par les critères internes usuels. Ces critères internes s'avèrent donc inadaptés dans ce contexte pour lequel la chronologie des évènements est primordiale. Une autre remarque est que les 10 meilleurs sous-ensembles de descripteurs varient peu d'un algorithme de partitionnement à l'autre.

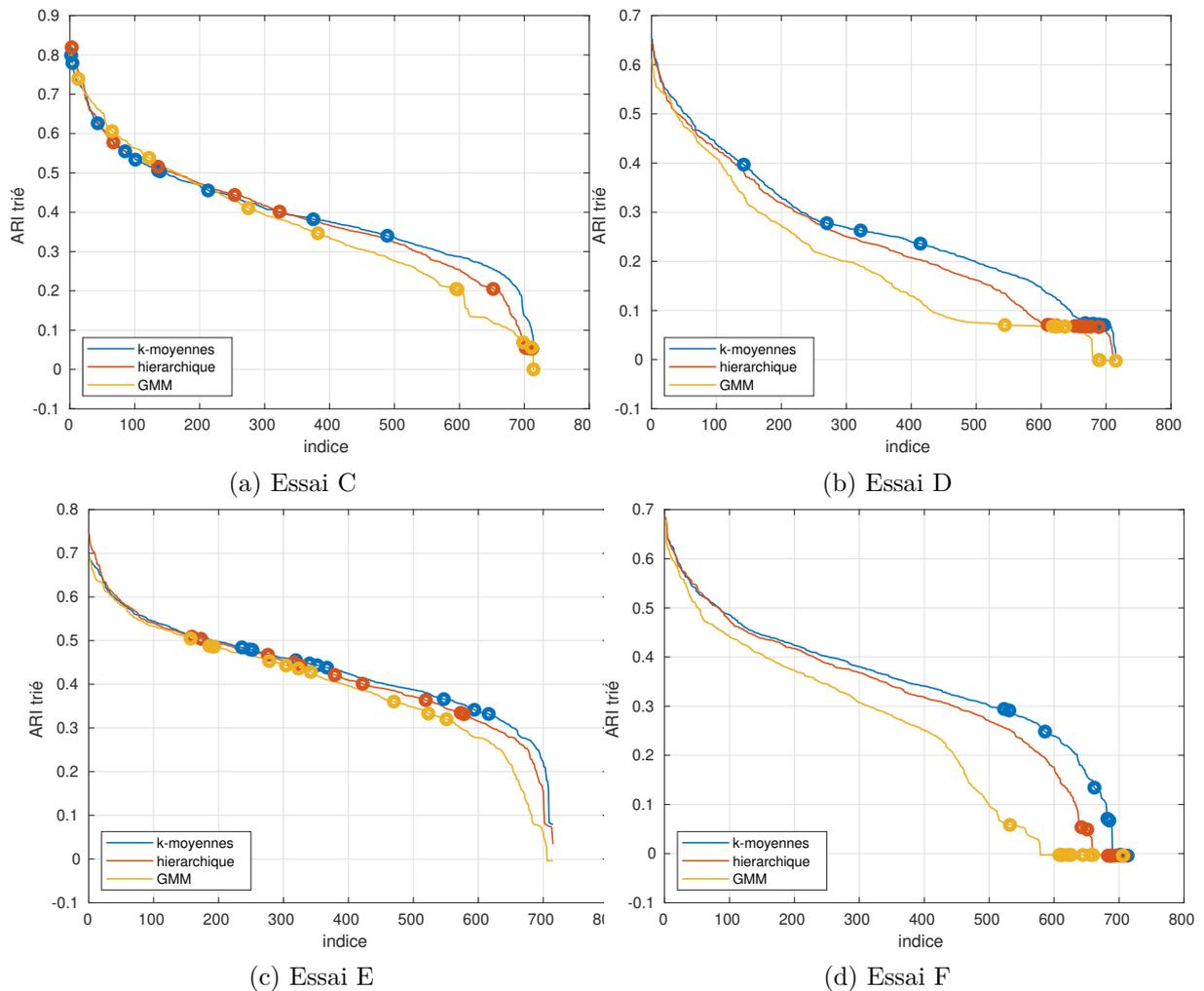


Figure 1.23 – Performances de la validation : ARI des 10 premières partitions

Globalement pour les cinq essais, il existe quelques partitions qui sont fortement corrélées aux niveaux de serrage. Cependant lors de la validation, ces partitions ne sont pas choisies ; ce qui montre que la méthode de validation basée uniquement sur des critères internes peut avoir des limites si on s'intéresse à l'aspect chronologique et à la détection du début des évènements de même nature (Figure 1.23).

En somme l'application du partitionnement associé à la validation par vote nous a permis de confirmer la relation d'indépendance entre la séparabilité spatiale et la séquentialité des évènements

(ici le niveau de serrage). Le nombre de « *clusters naturels* » n'est pas non plus une évidence. Elle permet de conclure que pour les données ORION-EA et les données séquentielles en général, une méthode de validation basée essentiellement sur des critères de distance n'est pas toujours pertinente. Il est donc nécessaire de faire appel à un ou plusieurs critères séquentiels.

### 1.3 Problématique et contributions

L'application sur les données ORION-EA illustre bien l'idée qu'un bon partitionnement suivant les critères de représentation basés sur la forme ne garantit pas nécessairement la visualisation des séquences au niveau temporel. Les méthodes de validation du partitionnement classiques ont déjà fait leurs preuves dans la caractérisation des matériaux et en *SHM*, cependant nous mettons en lumière leur insuffisance dans la détection des événements séquentiels. Nous avons aussi remarqué dans ces illustrations la variabilité du nombre de clusters en fonction du critère de validation, ce qui en fait un résultat incertain. Cette thèse vise à apporter une contribution à la résolution de ces problèmes évoqués.

Des approches de validation prenant en compte le temps et l'ordre d'apparition des clusters ont déjà été proposées [42, 69, 99, 100]. Il s'agit des critères basés sur les séquences, ils partent de l'intuition que la structure subit différents mécanismes d'endommagement qui ne se déclenchent pas spontanément au démarrage de l'essai mais repartis dans le temps. Cette approche est utilisée pour reconstruire des scénarios d'endommagement, la variable habituellement observée est l'occurrence cumulée des clusters au cours du temps [69]. La séparabilité des *clusters* dans l'espace des descripteurs ne garantit pas la séquentialité de ces derniers, car une partition peut aussi bien discriminer des événements microscopiques (qui se chevauchent) que les événements macroscopiques (souvent séquentiels). Dans les deux cas la technique d'EA peut être accompagnée d'une corrélation d'images ou une autre technique similaire permettant de confirmer le résultat obtenu.

Nous opterons donc pour une approche de validation basée sur un critère séquentiel dans les prochains chapitres et définirons ainsi un nouveau critère. Dans le deuxième chapitre, ce critère séquentiel sera utilisé uniquement pour le choix du meilleur sous-ensemble de descripteurs. Dans le troisième chapitre, un algorithme de partitionnement bayésien sera décrit pour déterminer le nombre de *clusters*.

La technique d'EA génère des données massives et relativement difficiles à analyser et interpréter. Ce chapitre a permis d'évaluer les méthodes de la littérature sur un jeu de données commun. Parmi les différentes étapes d'analyse de ces données, le partitionnement est d'une importance. Les méthodes de partitionnement que nous avons vues utilisent deux logiques : dure et floue, selon qu'un signal puisse appartenir à un seul ou à plusieurs *clusters*. Cependant dans les deux cas, les paramètres des modèles proposés sont constants. Dans le chapitre 3, nous aborderons une méthode de partitionnement

bayésien qui prend en compte cette variabilité du modèle en considérant les paramètres sous la forme de distributions, ce qui a pour conséquence la détermination automatique de  $K_{opt}$ , le nombre de clusters pour un sous-ensemble de descripteurs tout en évaluant l'incertitude sur chaque paramètre.



## Chapitre 2

# Validation du partitionnement par un critère séquentiel

### Sommaire

---

<b>2.1</b>	<b>Approches basées sur la séquence d'apparition des clusters</b>	<b>47</b>
2.1.1	Limites des critères de forme	49
2.1.2	Définition du critère séquentiel	51
<b>2.2</b>	<b>Validation séquentielle avec plusieurs partitions</b>	<b>56</b>
2.2.1	Principe de la validation sans estimation préalable de $K^*$	57
2.2.2	Algorithme proposé pour une validation séquentielle	58
2.2.3	Ajustement du critère pour l'estimation du nombre de clusters	60
<b>2.3</b>	<b>Illustration sur ORION-EA</b>	<b>61</b>
2.3.1	Lien entre les partitions et les serrages	62
2.3.2	Détection des niveaux de serrage	63
<b>2.4</b>	<b>Discussions et conclusions</b>	<b>67</b>

---

## Introduction

Comme présenté au chapitre précédent, les signaux d'émission acoustique (EA) sont des transitoires de tailles différentes, généralement représentés dans un espace commun de descripteurs. En variant les sous-ensembles de descripteurs et d'autres hyper-paramètres, plusieurs partitions sont obtenues. La validation consiste en la détermination du nombre de *clusters* pour chaque sous-ensemble de descripteurs et à la sélection du sous-ensemble optimal. Les deux doivent être choisis avec soin pour obtenir une représentativité et une interprétation pertinentes des *clusters*.

La validation du partitionnement consiste en la détermination de la partition optimale  $P^*$ , déduite de  $S^*$ , le meilleur sous ensemble de descripteurs ; caractérisée par  $K^*$ , le nombre de *clusters* et  $\theta^*$  (moyenne et dispersion), les paramètres des *clusters*. Dans la communauté de l'EA, cette estimation se fait à l'aide de critères, dits internes. Ces critères reposent sur la forme des *clusters* dans l'espace des descripteurs. Ils optimisent la notion de « *clusters naturels* ». Un *cluster* est dit naturel, s'il est compact et séparé des autres dans l'espace des descripteurs. Des critères tels que Davies-Bouldin, Silhouette et Calinski-Harabasz visent à optimiser cette notion. Ces critères ne prennent pas en compte un aspect possiblement séquentiel des événements à détecter. Or, il existe de nombreux contextes où la chronologie et plus particulièrement la séquentialité (figure 2.1) est une caractéristique discriminante. Pour des tâches comme le monitoring d'un procédé, la chronologie (*séquences d'apparition et séparabilité temporelle des clusters*) est essentielle, c'est le cas des essais tels qu'ORION-EA par exemple.

Nous proposons dans ce chapitre un nouveau critère, spécifique à l'évaluation du partitionnement des données d'EA. Il se focalise essentiellement sur la séquentialité des *clusters*, c'est-à-dire la manière dont ils se déclenchent dans le temps. Une partition sera pertinente si les premières occurrences de chacun des *clusters* sont distinctes et réparties dans le temps. Justifié à partir de la théorie de l'information, notre critère permet d'intégrer des *a priori* sur les déclenchements des *clusters* s'ils sont connus, par exemple, à partir de simulations numériques. Nous montrons expérimentalement que le critère proposé conduit à une meilleure interprétation des signaux du benchmark *ORION-EA*.

Un second point limitant concernant les méthodes de validation du chapitre 1, est la variabilité de la *partition optimale* en fonction du critère de validation, rendant l'interprétation subjective. Des méthodes multicritères telles que le vote ont été proposées pour la rendre plus objective. Cependant au terme de cette approche, une seule partition est toujours choisie. Expérimentalement nous avons vu que la validation avec une seule partition ne garantit pas la détection d'un maximum d'évènements (niveaux de serrage pour le cas *ORION-AE*). Nous proposons ainsi une méthode de validation prenant en compte les instants de déclenchements des *clusters* de plusieurs partitions. Cette méthode sera associée au critère de validation séquentiel.

Ce chapitre centré sur la validation du partitionnement des données d'émission acoustique répond ainsi à deux problématiques :

- Quel critère de validation utiliser pour tenir compte de la séquentialité des évènements acoustiques ?
- Comment valider le partitionnement des données d'émission acoustique en prenant en compte le maximum d'information présente dans les données ?

## 2.1 Approches basées sur la séquence d'apparition des clusters

La validation est utilisée pour sélectionner le meilleur **paramétrage** (le meilleur sous-ensemble de descripteurs et le nombre de *clusters* optimal) en fonction d'un ou plusieurs **critère(s)**. Elle repose sur deux types d'approches :

- **L'évaluation de la forme des clusters** : elle est basée sur des critères, comme l'indice Davies-Bouldin , l'indice silhouette... [101, 102] vus au chapitre 1. Ces critères nécessitent en entrées les descripteurs et la partition à évaluer, de ce fait ils sont dits "internes". De nombreuses méthodes dans la littérature entrent dans cette catégorie. On y trouve les méthodes dites "mono-critères", utilisant un unique critère pour sélectionner le nombre de *clusters* et le meilleur sous-ensemble de descripteurs [103]. Les méthodes dites "multi-critères", utilisent plusieurs critères de validation [104–106], à l'instar du vote présenté au chapitre 1.
- **L'évaluation de la chronologie des clusters** : l'occurrence cumulée de chaque *cluster* est tracée en fonction du temps et un critère est calculé en fonction de l'évolution des cumulées [42, 69, 99, 100].

Cette section présente l'enjeu autour du choix d'un critère de validation. Elle présente l'approche basée sur des critères internes, l'intuition qui la régit et les limites qu'elle peut présenter dans un contexte général. Par ailleurs, nous avons déjà vu au chapitre 1 que dans le cas particulier du benchmark ORION-EA les critères internes s'avèrent inefficaces à discriminer les partitions optimales en termes de déclenchements des clusters (c'est-à-dire dans la pratique en termes d'instant de desserrages).. Par la suite la nécessité d'un critère basé sur la séquentialité des *clusters* est davantage justifiée.

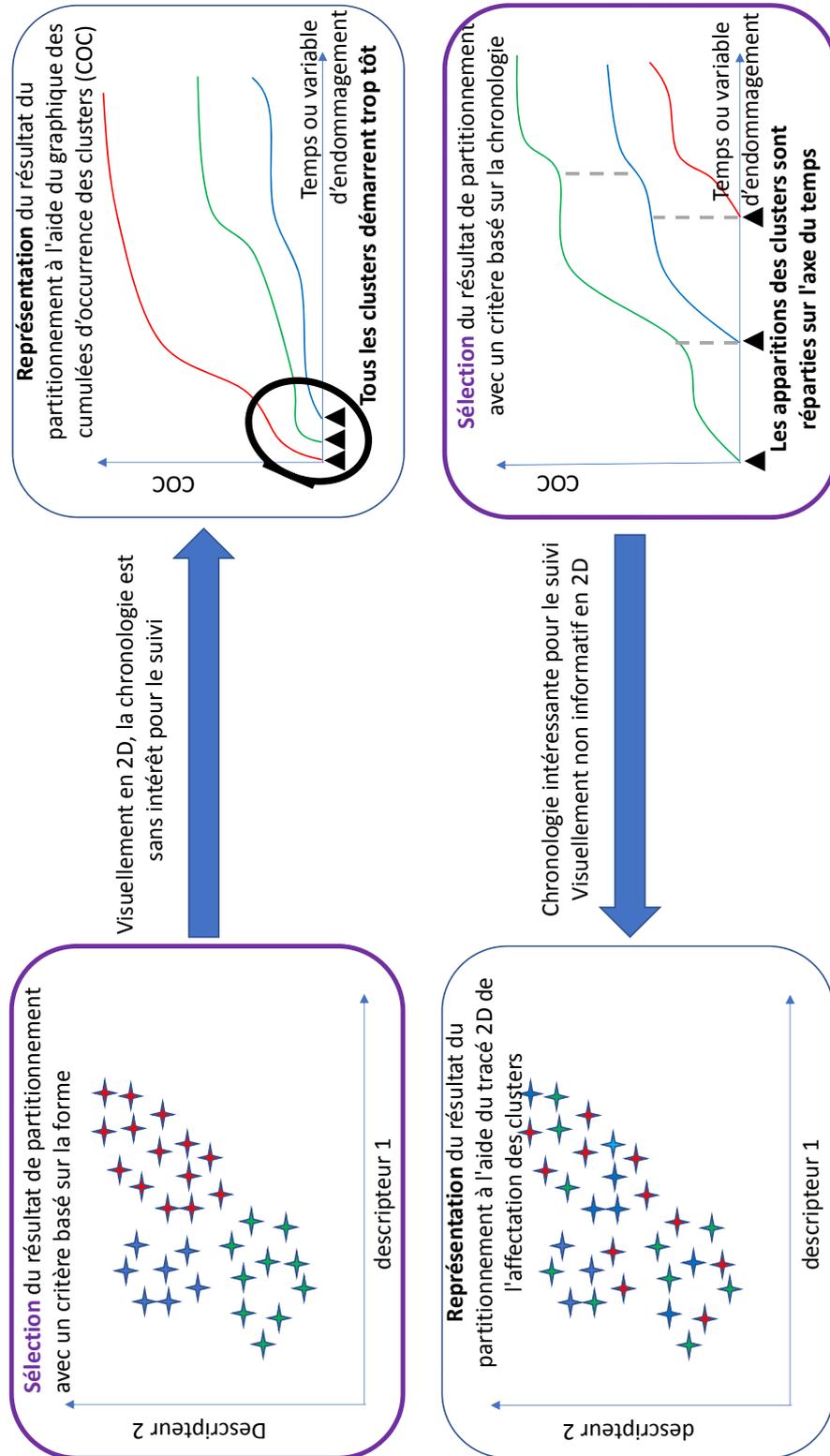


Figure 2.1 – **Paradigmes pour la qualité du partitionnement** | Ces représentations sont faites pour chaque sous-ensemble de descripteurs, pour chaque valeur de  $K$  lorsque des algorithmes standards (kmeans, GMM, hiérarchique) sont utilisés. Les cumulés d'occurrence des *clusters* ( $COC$ ) permettent d'illustrer la chronologie. Il s'agit de  $K$  compteurs  $C_k$  qui s'incrémentent de 1 chaque fois qu'une EA est attribuée au cluster  $k^{ième}$  cluster.

### 2.1.1 Limites des critères de forme

Les deux approches de validation sus-mentionnées conduisent généralement à des résultats différents. L'étude proposée dans [107] montre expérimentalement que les meilleures représentations dans l'espace des descripteurs ne garantissent pas une chronologie pertinente. Bien qu'attachée à une expérience particulière, cette étude pointe les limites des méthodes standards de validation basées sur les critères de forme.

La figure 2.1 illustre les deux points de vue. Le côté supérieur gauche est une représentation dans l'espace des descripteurs de la meilleure partition avec des critères qui favorisent des *clusters* compacts et distincts. Le côté supérieur droit montre que le graphique des occurrences cumulées des *clusters* ne présente pas nécessairement une chronologie pertinente à des fins de surveillance car les déclenchements des *clusters* sont trop proches et souvent détectés prématurément. Cela signifie que la plupart des signaux d'EA collectés sont affectés aux *clusters* apparaissant au début d'un test mécanique. Si, comme on le suppose généralement dans la littérature, un endommagement (ou condition expérimentale) est caractérisé par un *cluster*, cela signifie que tous les types d'endommagement commencent au début du test, ce qui est peut-être vrai pour certaines configurations mais faux en règle générale.

La partie inférieure droite de la figure 2.1 montre le problème abordé dans le sens inverse. Le meilleur paramétrage est déterminé en fonction de la répartition des *clusters* sur l'axe horizontal (temps). Les *clusters* apparaissent alors étendus et imbriqués dans l'espace des descripteurs, c'est-à-dire ne satisfaisant pas les critères de forme usuels de compacité et de séparabilité. À des fins de surveillance (diagnostic et pronostic), il nous apparaît plus pertinent de trouver des événements qui s'étalent chronologiquement. Les comportements similaires à ceux observés dans [107] qui est une configuration particulière mais tout de même complexe (essais quasi-statiques sur des anneaux composites), ont été remarqués sur de nombreux autres matériaux et configurations rencontrés au laboratoire Femto-st comme les bio-composites, le métal, les hybrides métal/composites avec des conceptions différentes, telles que des plaques, des anneaux, cylindres, et des profils de chargement comme la traction et la flexion en quasi-statique et en fatigue.

Généralement lorsqu'on fait varier le nombre de *clusters*, il est assez courant de considérer un petit nombre car il correspond souvent au nombre de familles d'endommagement attendu. Par exemple, dans le cas des matériaux composites, il est courant dans la littérature de considérer trois *clusters* liés à la fissuration de la matrice, au délaminage et aux ruptures de fibres. Il s'agit d'un a priori fort sur le lien direct entre les données d'EA et les endommagements. En pratique, les critères de formes courants, tels que Davies-Bouldin, fournissent souvent un petit nombre de *clusters*. Cependant dans quelques publications récentes, les auteurs ont analysé les résultats avec un nombre de *clusters* plus élevé que d'habitude. Par exemple, dans [108], les auteurs ont étudié la pertinence des résultats fournis par un modèle de mélange gaussien (GMM) lorsque le nombre de *clusters* augmente. Au lieu d'utiliser

les critères usuels, le nombre maximal de *clusters* distincts a été **qualitativement** étudié. L'évolution chronologique de chaque *cluster* a été observée. Une approche similaire est utilisée dans [42]. Une des idées derrière ces études est de considérer que le nombre de clusters ne correspond pas exactement au nombre de familles endommagements. Cette situation peut se produire par exemple lorsque certains des endommagements sont **multimodaux** suivant certains descripteurs. En effet, certains signaux détectés peuvent être incomplets ou correspondre à de fausses détections à cause des seuils de détection mal adaptés. L'apparition des défauts peut aussi être une source de non stationnarité, en particulier en fatigue. La signature des EA détectées lorsqu'elles arrivent en cascades peut varier des endommagements isolés (mono-modaux). Une rupture brusque peut potentiellement générer une signature différente d'un évènement lié à un phénomène d'usure.

Une interprétation possible est qu'un certain nombre des partitions obtenues par différents sous-ensembles de descripteurs ont du sens, mais ne sont pas représentatives des mêmes évènements. Une partition à deux *clusters* pourraient par exemple discriminer le bruit des endommagements. Certaines partitions ayant des *clusters* bien distincts dans l'espace des descripteurs pourrait correspondre à des évènements microscopiques ayant des signatures très différentes se déroulant simultanément. Certaines partitions peuvent correspondre aux évènements macroscopiques tels que le passage du domaine élastique au domaine plastique, la rupture ou un changement de phase dans un procédé industriel. Pour le benchmark ORION-EA, par exemple, certaines partitions peuvent correspondre aux niveaux de serrages et d'autres à des phénomènes de tribologie. Au chapitre 1, nous avons pu constater qu'il existe un lien entre certaines partitions obtenues et les niveaux de serrages, ce qui signifie qu'un *cluster* ne représente pas nécessairement un endommagement, mais un phénomène (ou évènement) observé.

Si on s'intéresse par exemple à un matériau composite, parmi les différentes partitions obtenues certaines discriminent les évènements microscopiques (rupture de fibres, fissure matricielle, délaminage et décohésion fibre/matrice) qui ne sont pas nécessairement distincts dans le temps en fonction du type d'essai. D'autres partitions discriminent des évènements macroscopiques (déformations élastique et plastique, rupture) qui sont séquentiels. L'utilisation des critères de formes ne peut pas garantir que le choix du sous ensemble  $S^*$  se portera sur les phénomènes macroscopiques qu'on désire observer. Les signaux détectés pouvant représenter des évènements de natures ou d'échelles différentes, l'EA peut s'accompagner d'autres méthodes (caméra thermique par exemple) pour caractériser les évènements qu'on veut observer ou **définir des critères adaptés** [107].

Le critère séquentiel a pour but d'améliorer la sélection des sous-ensembles produisant des *clusters* dont les déclenchements sont répartis dans le temps. Il peut servir à décrire les évènements de manière chronologique pour les cas où la répartition des déclenchements des *clusters* est plus importante que leur forme et leur distribution dans l'espace des descripteurs. L'accent mis sur la chronologie des clusters donne un aperçu de la façon dont les endommagements se produisent et se développent pro-

gressivement, ce qui ne peut pas être assuré avec des critères standards.

Comme les critères de la littérature, le critère que nous proposons n'intervient pas dans le partitionnement, il ne sert qu'à trier les partitions obtenues par des algorithmes de partitionnement (par exemple k-moyennes, GMM, hiérarchique...).

### 2.1.2 Définition du critère séquentiel

Reprenons les notations du chapitre 1,  $\mathbf{y} = \{y_1, \dots, y_N\}$  est l'ensemble des  $N$  signaux d'EA représentés dans l'espace des descripteurs. Chaque observation  $\mathbf{y}_n$  est constituée de  $D$  composantes  $[y_{n1}, y_{n2}, \dots, y_{nD}]$ ,  $\mathbf{y}$  est une matrice de dimension  $[N \times D]$ . Le nombre de *clusters*  $K \in [K_{min}, K_{max}]$  est fixé.

Les sous-ensembles de  $d$  descripteurs sont utilisés. Chacun des  $P = \binom{D}{d}$  sous-ensembles  $S_p$  de descripteurs génère  $(K_{max} - K_{min} + 1)$  partitions, soit un total de  $(K_{max} - K_{min} + 1) \binom{D}{d}$  partitions. Idéalement le critère proposé doit être capable de sélectionner  $K_p^* \in [K_{min}, K_{max}]$ , le nombre de *clusters* optimal pour un sous-ensemble  $S_p$  et également trouver le meilleur des sous-ensembles  $S^*$ . Une hypothèse est posée sur la chronologie des *clusters*.

**Hypothèse 2.1** *Les instants de déclenchement des clusters permettent de trier les partitions indépendamment de leur séparabilité et de leur compacité dans l'espace des descripteurs. Une partition est pertinente si elle est constituée de clusters dont les instants de déclenchement suivent une distribution a priori. Le critère se focalise essentiellement sur l'axe horizontal du graphique des cumulées d'occurrences des clusters. Ainsi, si plusieurs clusters sont caractérisés par des instants de déclenchement très proches, la partition est pénalisée [26, 69, 109].*

Cette hypothèse suppose un ordre dans la labellisation des *clusters*. Soit une partition  $\mathcal{P}_S$  donnée par un sous-ensemble de descripteurs  $S$  pour  $K$  fixé. Pour chacun des *clusters*  $C_{k=1\dots K}$ , nous définissons  $t_k$  comme l'instant de déclenchement. Les *clusters* sont numérotés selon l'ordre croissant des  $t_k$  ( $t_1 = \min(t_k), \dots, t_K = \max(t_k)$ ),  $C_1$  pour le premier *cluster* détecté,  $C_2$  pour le second, ainsi jusqu'à  $C_K$ . On définit une distribution  $\{\Delta(K, S_p)\}$  qui représente la répartition des déclenchements des *clusters*. On suppose que le premier déclenchement a eu lieu à  $t = 0$  en recalant les instants des signaux.

Normalisée, cette distribution s'écrit :

$$\Delta_k = \frac{t_{k+1} - t_k}{T}, \quad \text{avec } t_1 = 0, \text{ et } t_{K+1} = T \quad (2.1)$$

où  $t_k \in [0, T]$ ,  $T$  est l'instant d'apparition de la dernière EA détectée et  $t_{K+1} = T$ . La distribution  $\mathbf{p}_{est} = \{\Delta_k\}_{k=1}^K$ , estimée à partir du partitionnement, peut être interprétée comme une distribution de probabilité composée de  $K$  éléments puisque nous avons :

$$\sum_{k=1}^K \Delta_k = 1 \quad (2.2)$$

### a) Critère lorsque les instants de déclenchement sont connus

Supposons qu'on soit dans le cadre d'un essai maîtrisé ou que l'EA soit accompagnée d'une autre méthode de surveillance, ou encore que l'on dispose d'un modèle numérique ou physique. Dans ces cas, nous disposerions d'un *a priori* sur les instants de déclenchement des *clusters*.

Une distribution  $\mathbf{p}_{\text{ref}}$  peut alors être définie pour ces instants de référence, en suivant le même raisonnement que précédemment (également normalisée). La validation pourrait alors consister à comparer les distributions des différentes partitions  $\mathbf{p}_{\text{est}}$  estimées par rapport à la référence  $\mathbf{p}_{\text{ref}}$  afin de déterminer la partition la plus proche. Plusieurs mesures existent dans la littérature pour comparer les distributions par exemple, la norme  $L_1$ , la distance de Hellinger, la distance de Bhattacharyya, la distance de variation totale entre les probabilités, ou les divergences de Kullback-Leibler (KL) ou de Renyi [110–112].

La divergence KL représente la perte d'information introduite en utilisant  $\mathbf{p}_{\text{est}}$  au lieu de  $\mathbf{p}_{\text{ref}}$ . Aussi appelée **entropie relative**, elle sert de fonction objectif à minimiser en apprentissage automatique pour des modèles basés sur la théorie des probabilités et sur des réseaux de neurones [113]. La divergence KL est donnée par l'expression :

$$KL(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{ref}}) = \sum_{k=1}^K \mathbf{p}_{\text{est}}(k) \log_2 \left( \frac{\mathbf{p}_{\text{est}}(k)}{\mathbf{p}_{\text{ref}}(k)} \right) \quad (2.3)$$

En développant les termes de l'équation 2.3 on obtient

$$KL(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{ref}}) = \underbrace{- \sum_{k=1}^K \mathbf{p}_{\text{est}}(k) \log_2 (\mathbf{p}_{\text{ref}}(k))}_{\text{CE}} + \underbrace{\sum_{k=1}^K \mathbf{p}_{\text{est}}(k) \log_2 (\mathbf{p}_{\text{est}}(k))}_{-E} \quad (2.4)$$

Dans cette expression, on retrouve deux mesures d'incertitude très connues en apprentissage machine. Premièrement l'entropie croisée :

$$CE(\mathbf{p}_{\text{est}}||\mathbf{p}_{\text{ref}}) = - \sum_{k=1}^K \mathbf{p}_{\text{est}}(k) \log_2 (\mathbf{p}_{\text{ref}}(k)) \quad (2.5)$$

Elle représente l'espérance du logarithme de  $\mathbf{p}_{\text{ref}}$  en utilisant  $\mathbf{p}_{\text{est}}$  comme probabilité. Elle est minimale lorsque  $\mathbf{p}_{\text{est}} = \mathbf{p}_{\text{ref}}$  [113].

Le second terme est l'entropie de Shannon donnée par :

$$E(\mathbf{p}_{\text{est}}) = - \sum_{k=1}^K \mathbf{p}_{\text{est}}(k) \log_2 (\mathbf{p}_{\text{est}}(k)) \quad (2.6)$$

C'est une mesure de la quantité d'information contenue dans un code (distribution dans notre cas).

Celle-ci est maximale si la distribution  $\mathbf{p}_{est}$  est uniforme [114]. L'entropie croisée et celle de Shannon vérifient :

$$KL(\mathbf{p}_{est}||\mathbf{p}_{ref}) = CE(\mathbf{p}_{est}||\mathbf{p}_{ref}) - E(\mathbf{p}_{est}) \quad (2.7)$$

**Propriété 2.1** *Les propriétés suivantes sont propres à la divergence KL :*

— *La divergence KL est une grandeur positive ( $KL \geq 0$ ) et non symétrique.*

$$KL(\mathbf{p}_{est}||\mathbf{p}_{ref}) \neq KL(\mathbf{p}_{ref}||\mathbf{p}_{est}) \quad (2.8)$$

— *La divergence KL n'est pas une mesure de distance.*

—  *$KL(\mathbf{p}_{est}||\mathbf{p}_{ref}) = 0$  si  $\mathbf{p}_{est} = \mathbf{p}_{ref}$*

Ainsi, si  $\mathbf{p}_{ref}$  est connue (c'est-à-dire l'utilisateur peut fournir les instants de déclenchement des *clusters*) alors l'équation 2.7 peut être utilisée pour trier les partitions. Les valeurs les plus faibles de la *KL* donnent les partitions les plus pertinentes en terme de chronologie.

## b) Critère en cas d'incertitude totale

Généralement, les instants de déclenchement des *clusters* ne sont pas connus. De même, estimer ces instants par un modèle numérique ou physique s'avère généralement difficile (voire impossible) de par la complexité des structures (variabilité liée aux matériaux, l'assemblage, le procédé de fabrication, la dimension ...), les conditions d'usage et le couplage des phénomènes physiques mis en jeu.

**Hypothèse 2.2** *Sans a priori sur  $\mathbf{p}_{ref}$ , une solution pour estimer la divergence KL consiste à utiliser le principe de l'entropie maximale. Ce principe consiste à ne pas ajouter d'hypothèses arbitraires et à éviter d'introduire des biais autres que ceux déjà présents dans les données [115, 116]. Par conséquent, en l'absence d'a priori sur les instants de déclenchement, parmi toutes les distributions possibles de  $\mathbf{p}_{ref}$ , nous devrions sélectionner celle avec la plus grande entropie. Elle correspond à la **distribution uniforme**. Elle est définie par :*

$$\mathbf{p}_{ref}(k) = \{\text{Uniforme}\}_K = \frac{1}{K} \quad \forall k = 1 \dots K \quad (2.9)$$

Dans ce cas, l'équation 2.5 devient

$$CE(\mathbf{p}_{est}||\mathbf{p}_{ref}) = -\log_2 \frac{1}{K} = \log_2 K \quad (2.10)$$

La divergence *KL* est alors donnée par :

$$\mathbf{p}_{ref} = \{\text{Uniforme}\}_K \Rightarrow KL(\mathbf{p}_{est}||\mathbf{p}_{ref}) = \log_2 K - E(\mathbf{p}_{est}) \quad (2.11)$$

La recherche du sous-ensemble de descripteurs pertinents peut alors être posée comme :

$$(S^*, K^*, \theta^*) = \underset{S, \theta}{\operatorname{argmin}} [\log_2 K - E(\mathbf{p}_{\text{est}}(\cdot | S, K, \theta))] \quad (2.12)$$

où il est explicite que  $E$  dépend de  $K, S$  et  $\theta$ . En trouvant les paramètres et les hyper-paramètres tels que  $\mathbf{p}_{\text{est}}$  soit le plus proche possible de la distribution uniforme, on privilégie les partitions pour lesquelles les *déclenchements sont régulièrement distribués sur l'axe horizontal* (temps, ou grandeur physique). En conclusion, la KL tend vers 0 s'il existe un paramétrage de l'algorithme de partitionnement tel que la partition obtenue conduit à une distribution uniforme des instants de déclenchement.

Un critère de validation du partitionnement répond en général à deux exigences, choisir  $K^*$  pour un sous-ensemble de descripteurs et choisir le sous-ensemble descripteurs pertinents. La divergence KL proposée dans ce chapitre permet de répondre à l'exigence du choix des sous-ensembles pertinents lorsque le nombre de *clusters* est fixé. L'exemple 2.1 est proposé pour montrer la limite de ce critère pour choisir  $K^*$ .

**Exemple 2.1** *Considérons un jeu de données constitué de 4 clusters naturels suffisamment compacts dans l'espace des descripteurs pour qu'il n'y ait pas de mélange lors du partitionnement. On fait varier le nombre de clusters  $K \in [2, 5]$ . On suppose également que les déclenchements des 4 clusters sont uniformément répartis dans le temps. En général lorsqu'on fait varier le nombre de clusters, plusieurs cas de figures se présentent. Si  $K < K^*$ , il y a fusion entre les clusters proches. Si  $K > K^*$ , les clusters les plus dispersés se subdivisent. Dans certains cas les nouveaux clusters regroupent les points aux frontières lorsque celles-ci ne sont pas totalement disjointes. Nous nous focalisons uniquement sur les cas où de réels déclenchements sont détectés (il y a fusion des clusters pour  $K < 4$ ).  $S_{i=1,2,3}$  sont trois sous-ensembles de descripteurs utilisés pour le partitionnement.*

- $K = 2$ , le second instant détecté est à  $0.5T$  dans ce cas  $\Delta = \{0.5, 0.5\}$ . La seconde possibilité est que le second cluster soit détecté à  $0.25T$  ou à  $0.75T$ , dans ce cas  $\Delta = \{0.25, 0.75\}$  ou  $\Delta = \{0.75, 0.25\}$ . La même analogie peut être faite pour  $K = 3$ .
- $K = 4$ , tous les clusters sont détectés,  $\Delta = \{0.25, 0.25, 0.25, 0.25\}$
- $K = 5$  **ou plus**, tous les instants attendus sont trouvés, un cluster est subdivisé, créant ainsi un déclenchement supplémentaire.
- $K = 5$  **avec un sous-ensemble de descripteurs non pertinents** : nous avons choisi de représenter un ensemble où presque tous les clusters se déclenchent au début de l'essai.

Ces différents cas sont regroupés dans le tableau 2.1 dans le but d'évaluer le critère.

Si on se réfère au tableau 2.1, la divergence KL est bien minimale pour  $(S_2, K = 2)$  et  $(S_1, K = 4)$  qui correspondent aux distributions uniformes. La KL ne parvient pas à faire un choix entre ces deux cas uniformes, les deux donnant 0.

Pour  $(S_1; K = 2)$  et  $(S_1; K = 3)$ , on constate que la KL parvient à choisir la partition avec  $K = 3$  qui a le plus de valeurs identiques et proches de  $\frac{1}{K}$ .

Tableau 2.1 – Exemple d'un jeu à 4 classes

	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$	$\Delta_5$	CE	E	KL
$S_1, K=2$	0.25	0.75	/	/	/	1	0.8113	0.1887
$S_2, K=2$	0.5	0.5	/	/	/	1	<b>1</b>	<b>0</b>
$S_1, K=3$	0.25	0.25	0.5	/	/	1.5850	1.5000	0.0850
$S_1, K=4$	0.25	0.25	0.25	0.25	/	2	2	<b>0</b>
$S_1, K=5$	0.25	0.25	0.25	0.05	0.2	2.3219	<b>2.1805</b>	0.1414
$S_3, K=5$	0.1	0.1	0.1	0.05	0.65	2.3219	1.6166	0.7053

Pour les cas  $(S_1; K=3)$  et  $(S_1; K=5)$ , une discussion est possible sur la qualité des partitions. Pour  $K=5$ , quatre des cinq proportions sont très proches de  $1/5$ , ce qui la rend pertinente du point de vue séquentiel. La valeur de la divergence KL pour  $(S_1; K=5)$  est presque deux fois plus grande que celle obtenue pour  $K=3$  alors qu'elles sont de qualité similaire du point de vue séquentiel.

Cet exemple nous permet de montrer que la divergence KL, telle que formulée ne permet pas une comparaison fiable de partitions relatives à des nombres de clusters  $K$  différents. Cela signifie qu'il est nécessaire de normaliser la KL.

Pour une distribution uniforme,  $\log_2(K)$  est une borne supérieure de la KL, car l'entropie est toujours positive. La KL peut ainsi être normalisée en effectuant une division par le terme  $\log_2(K)$ .

La nouvelle fonction objectif à minimiser s'écrit alors :

$$(S^*, K^*, \theta^*) = \operatorname{argmin}_{S, K, \theta} \left[ 1 - \frac{E(\mathbf{p}_{\text{est}}(\cdot | S, K, \theta))}{\log_2 K} \right] \quad (2.13)$$

L'équation 2.13 équivaut à l'équation 2.14 :

$$Q_{\text{prop}}^{\text{Sequentiel}} : (S^*, K^*, \theta^*) = \operatorname{argmax}_{S, K, \theta} \left[ \frac{E(\mathbf{p}_{\text{est}}(\cdot | S, K, \theta))}{\log_2 K} \right] \quad (2.14)$$

Le critère  $Q_{\text{prop}}^{\text{Sequentiel}}$  proposé correspond à l'entropie de Shannon normalisée de la distribution  $\mathbf{p}_{\text{est}}$  [112]. L'objectif tel que défini dans l'équation 2.14 permet une meilleure comparaison des résultats. Cependant à performances égales pour deux valeurs de  $K$ , il est incapable de choisir le meilleur sous-ensemble (cas de  $(S_2, K=2)$  et  $(S_1, K=4)$ ).

Ainsi, lors d'une validation séquentielle du partitionnement sans *a priori* sur les instants de déclenchement des *clusters*, si le nombre de clusters  $K$  est fixé, minimiser la divergence KL revient à trouver le sous-ensemble  $S^*$  qui maximise l'entropie de Shannon  $E(\mathbf{p}_{\text{est}}(\cdot | S, K, \theta))$ .

Reprenons l'analyse faite à l'exemple 2.1, cette fois en estimant l'entropie normalisée.

**Exemple 2.2** Les résultats sont regroupés dans le tableau 2.2.

Tableau 2.2 – Exemple d’un jeu à 4 classes avec des critères normalisés

	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$	$\Delta_5$	$E_{norm}$
$S_1, K=2$	0.25	0.75	/	/	/	0.8113
$S_2, K=2$	0.5	0.5	/	/	/	<b>1</b>
$S_1, K=3$	0.25	0.25	0.5	/	/	0.9464
$S_1, K=4$	0.25	0.25	0.25	0.25	/	<b>1</b>
$S_1, K=5$	0.25	0.25	0.25	0.05	0.2	0.9391
$S_3, K=5$	0.1	0.1	0.1	0.05	0.65	0.6436

L’entropie normalisée, pour  $(S_1, K = 3)$  et  $(S_1, K = 5)$  donne des valeurs très proches, ce qui signifie que les deux partitions sont de qualité similaire. On peut conclure que la normalisation permet une meilleure comparaison des résultats, bien qu’elle reste indécise concernant le choix du nombre de clusters pour un sous-ensemble de descripteurs. Le critère reste toujours indécis pour les cas  $(S_2, K = 2)$  et  $(S_1, K = 4)$ .

Le critère proposé permet de faire un tri des partitions en fonction de leurs instants de déclenchement. La méthode de validation peut varier en fonction de l’objectif qu’on se fixe :

- Si  $K$  est fixé, la meilleure partition est celle qui maximise l’entropie de Shannon normalisée donnée par l’équation 2.14. Cette équation peut aussi être utilisée pour des partitions dont le nombre de clusters est assez proche. Elle ne permet cependant pas le choix du nombre de clusters si les performances sont similaires en terme de séquence.
- Si le but est de choisir le nombre de clusters à l’instar d’une analyse classique, une solution consiste à ajuster le critère avec un a priori sur  $K^*$ . Une autre approche serait de faire un partitionnement consensuel pour chacune des valeurs de  $K$  et choisir celle avec la plus petite enveloppe d’incertitude sur le graphe des cumulées d’occurrence des clusters. Cette seconde solution a l’avantage de conserver plus d’information sur les événements détectés.

La section suivante est consacrée à l’exploitation du critère séquentiel proposé (équation 2.14) pour une validation prenant en compte plusieurs partitions précédemment triées par ce critère.

## 2.2 Validation séquentielle avec plusieurs partitions

Conventionnellement un critère de validation répond à deux attentes :

- Pour un sous-ensemble de descripteurs, quel est le nombre de clusters optimal ?

- Étant donné un ensemble de descripteurs, quel est le sous-ensemble résumant l'information des données ?

Dans le premier chapitre nous avons présenté la validation par les méthodes mono/multicritères basées sur des critères de forme. Les limites qui avaient été mises en exergue sont la divergence des résultats suivant le critère et l'utilisation d'une seule partition pour représenter les résultats.

Le nombre de partitions à trier est généralement élevé (centaines ou milliers), les estimations des critères entre deux partitions successives du tri sont souvent très proches, donc pas assez discriminantes pour permettre une décision objective. Ainsi, il est légitime d'envisager des méthodes de validation prenant en compte plusieurs partitions.

### 2.2.1 Principe de la validation sans estimation préalable de $K^*$

L'objectif du partitionnement en SHM et PHM est la détection des événements. Si on considère un *cluster* comme une famille d'endommagement, les déclenchements de ces derniers sont alors considérés comme des événements. Utiliser une seule partition revient à affirmer que la validation est certaine. Or, comme illustré avec le benchmark ORION-EA au chapitre 1, et l'exemple 2.2, les critères de validation ne sont pas infaillibles. Il faut donc prendre en compte un certain nombre de partitions pour améliorer la qualité et la robustesse des résultats. Cependant le nombre de partitions ne doit pas être très grand pour éviter que les partitions non pertinentes ne détériorent les résultats. Des méthodologies basées sur le partitionnement consensuel ont été développées dans [26, 69, 109] pour répondre à cette problématique. La détermination de  $K^*$  est basée sur la minimisation des enveloppes de *clusters*.

Notre approche prend en compte les instants de déclenchement des *clusters* de plusieurs partitions pour différentes valeurs de  $K$ . La sélection de  $K^*$  est faite visuellement par l'utilisateur en se basant sur l'histogramme des instants de déclenchement des *clusters*, ou sur les enveloppes d'incertitude des cumulées d'occurrence des *clusters* des partitions consensuelles.

Les histogrammes des déclenchements de *clusters* sont tracés afin de localiser les instants pouvant correspondre aux changements d'étapes lors d'un essai (cas des niveaux de serrage pour le benchmark ORION-AE). L'algorithme part de l'hypothèse que le critère de validation n'est pas suffisant pour l'estimation de  $K^*$ . Ainsi, pour chaque valeur de  $K \in [K_{min}, K_{max}]$ , les instants de déclenchements des  $M$  (à définir) meilleures partitions sont relevés dans un vecteur  $vect_{[K_{min}, K_{max}]}$ . Une partition consensuelle  $\mathcal{P}_K$  peut également être construite en fusionnant les  $M$  meilleures partitions.

L'histogramme de  $vect_{[K_{min}, K_{max}]}$  est tracé, après une discrétisation de l'axe des temps. Les pics élevés correspondent aux instants de déclenchement les plus récurrents. L'utilisateur pourra juger visuellement les pics qu'il considère pertinents et choisir la partition consensuelle  $\mathcal{P}_K^*$  optimale, ayant les instants de déclenchements correspondant à ceux-ci, et ayant un graphe de cumulées d'occurrence de *clusters* avec une enveloppe d'incertitude fine, ce qui signifie que le partitionnement est robuste.

Cette approche associée au critère de validation séquentielle proposé (équation 2.14) a fait l'objet d'une communication [117].

## 2.2.2 Algorithme proposé pour une validation séquentielle

Une fois toutes les partitions calculées pour différents  $S$  et  $K$ , l'algorithme 1 proposé prend comme variables d'entrée le nombre de partitions  $M$  à considérer après tri;  $K_{max}$  et  $K_{min}$ , respectivement les valeurs maximale et minimale de  $K$ ; l'ensemble des partitions, soit  $(K_{max} - K_{min} + 1) \binom{D}{d}$ ;  $\{S\}$  l'ensemble constitué des  $\binom{D}{d}$  sous-ensembles de descripteurs;  $\mathbf{t}$ , les instants de démarrage des EA détectées; et  $T_e$  le pas de discrétisation de l'axe de temps. En sortie de l'algorithme nous avons un histogramme permettant de valider une partition. Les étapes sont données par l'algorithme 1.

Pour le tracé d'histogramme (ligne 28), un pas  $T_e$  de temps est fixé, celui-ci peut être défini comme la plus petite erreur d'estimation de temps tolérée pour la détection d'un évènement. L'axe temporel est alors discrétisé en  $N_e = \lceil \frac{T}{T_e} \rceil$  parties, où  $\lceil \cdot \rceil$  est la partie entière. Chaque déclenchement  $t_k \in [n_e \times T_e, (n_e + 1) \times T_e[$  est attribué à  $t_e = \frac{2n_e + 1}{2} T_e$ , avec  $n_e = 0, 1 \dots N_e - 1$ , un compteur enregistrant le nombre de détections. Plus ce nombre est grand, plus cet instant de déclenchement est pertinent.

L'histogramme des déclenchements de *clusters*, la partition consensuelle, et  $K^*$  sont ainsi obtenus. Cet algorithme a des *a priori* uniquement sur la distribution uniforme des instants de déclenchement et l'hypothèse que les enveloppes cumulées d'occurrence de *clusters* sont fines pour le bon nombre de *clusters*  $K^*$ .

### Remarques

Il est important de remarquer que :

- Prendre  $M = 1$  équivaut à ne pas faire de consensus.
- L'algorithme 1 estime les meilleurs sous-ensembles  $S_K^*$  pour  $K$  fixé. Dans cette version de l'algorithme, il est supposé impossible de discriminer deux partitions pertinentes (en terme de séquences) si elles ont un nombre de *clusters* différent.
- L'algorithme 1 peut être utilisé en remplaçant  $Q_{prop}^{Séquentiel}$  par un critère de forme, ou le score d'un vote. Pour  $K$  fixé, la validation par le vote se résume au tri des sous-ensembles de descripteurs. Elle consiste à trier les partitions suivant chacun des critères. Pour chaque tri, une note de 100 est attribuée à la meilleure partition, 99 à la seconde, 98 à la troisième et ainsi de suite jusqu'à la note de 1 à la centième partition du tri. Les  $(P - 100)$  partitions restantes ont une note de 0. Les partitions sont ensuite triées suivant leur score cumulé.

**Algorithme 1 : Validation séquentielle sans présélection du nombre de clusters par sous-ensemble de descripteurs** | La fusion des  $M$  meilleures partitions se fait pour différentes valeurs de  $K$ , en utilisant le partitionnement consensuel présenté au chapitre 1. La minimisation des enveloppes ou une observation de l'utilisateur permettent de choisir le meilleur modèle.

---

```

1 Entrée : Les instants  $\{\mathbf{t}\}$ , correspondant aux démarrages des EA, de dimension  $[N \times 1]$ ;
2 Entrée : La plage des valeurs de  $K$ , le nombre de clusters  $[K_{\min}, K_{\max}]$ ;
3 Entrée : Les sous-ensembles de descripteurs  $\{\mathcal{S}\}$ ;
4 Entrée : L'ensemble des  $(K_{\max} - K_{\min} + 1)$  matrices de partitions estimées  $[N \times \binom{D}{d}]$ ;
5 Entrée : La distribution  $P_{ref}$ , si elle est connue (ex. [69]);
6 Entrée : Le nombre de partitions  $M$ , à considérer pour les histogrammes;
7 Entrée : Le pas d'échantillonnage  $T_e$ , pour discrétiser les histogrammes;
8 Sortie : Les histogrammes des instants déclenchement;
9 Sortie : La partition consensuelle  $P_{\text{consensus}}^*$ , choisie et  $\{\theta^*\}$ , les paramètres de ses clusters;
10 pour chaque  $K \in [K_{\min}, K_{\max}]$  faire
11   pour chaque  $S \in \mathcal{S}$  faire
12     // Réordonner les clusters en fonction de leurs déclenchements;
13      $[P(S, K)] \leftarrow \text{Clusters\_labélisation}([P(S, K)]);$ 
14     // Calculer la distribution estimée  $\mathbf{p}_{est} = \{\Delta_k\}$ ;
15
16     
$$\Delta_k = \frac{t_{k+1} - t_k}{T}, \quad \text{avec } t_1 = 0, \text{ et } t_{K+1} = T \quad (2.1)$$

17     // Calculer le critère séquentiel de l'équation 2.3 si  $P_{ref}$  est connue et 2.14 sinon;
18      $Q_{\text{prop}}^{\text{Séquentiel}}(S, K) \leftarrow \text{Critère\_de\_Validation\_Séquentiel}(P, \mathbf{p}_{ref});$ 
19   fin
20   // Trier les partitions en fonction du critère  $\forall K$  et choisir les  $M$  premières;
21    $S_{\text{top}}(K) = \{S : \text{classés de 1 à } M \}$ ;
22   // Créer un vecteur  $\text{vect}_{[K_{\min}, K_{\max}]}$  qui regroupe tous les instants de déclenchement des
   clusters des  $M$  meilleures partitions  $\forall K$ . Elle rajoute à l'histogramme  $\text{vect}_{[K_{\min}, K]}$ 
   existant les instants de  $S_{\text{top}}(K)$ ;
23    $\text{vect}_{[K_{\min}, K_{\max}]} = \text{concaténation}(\text{vect}_{[K_{\min}, K]}; \text{vect}_K(S_{\text{top}}(K)));$ 
24   // Construire la partition consensuelle et calculer l'incertitude comme dans [69];
25    $F(K) \leftarrow \text{Partition\_consensuelle}(P(S_{\text{top}}(K), K));$ 
26 fin
27 // Résultats;
28 Calculer et tracer l'histogramme des déclenchements des clusters des  $M \times (K_{\max} - K_{\min} + 1)$ 
   meilleures partitions ;
29  $K^* \leftarrow$  Choisi visuellement à partir de l'histogramme;
30  $P_{\text{consensus}}^* \leftarrow F(K^*);$ 
31  $\theta^* \leftarrow \{\theta(s, K^*)\}_{s \in S_{\text{sub}}(K^*)}$  //les paramètres correspondant à la partition consensuelle (section
   1.2.2);

```

---

### 2.2.3 Ajustement du critère pour l'estimation du nombre de clusters

Empiriquement et sur les données d'EA nous avons vu que les critères classiques favorisent des faibles valeurs de  $K$ . Des études ont également considéré de grandes valeurs de  $K$  et discuté de leur pertinence vis à vis de l'EA [42, 108]. Un terme peut être ajouté pour favoriser les partitions avec un grand nombre de *clusters*, ce qui équivaut à une pénalisation pour des petites valeurs de  $K$ . Ce terme de pénalisation doit être assez faible pour ne pas être le seul facteur de décision. Des approches semblables ont déjà été faites avec des critères bayésiens tels-que l'AIC (*Akaike Information Criterion*) et le BIC (*Bayesian Information Criterion*) dont les termes de pénalisation de la log-vraisemblance sont respectivement  $K$  et  $\frac{K}{2} \log(N_p)$  (où  $N_p$  est le nombre de paramètres du modèle) [118–120]. Soit  $\lambda$  le terme de pénalisation à ajouter au critère. Afin de proposer de favoriser les valeurs de  $K$  élevées, nous proposons :

$$\lambda_1 = \log_2(1 + 1/K) \quad (2.15)$$

$$\lambda_2 = \frac{\log_2(1 + 1/K)}{K} \quad (2.16)$$

$$\lambda_3 = \frac{\log_2(1 + K)}{K} \quad (2.17)$$

Les trois termes de pénalisation sont illustrés dans la Figure 2.2 en fonction de  $K$ .

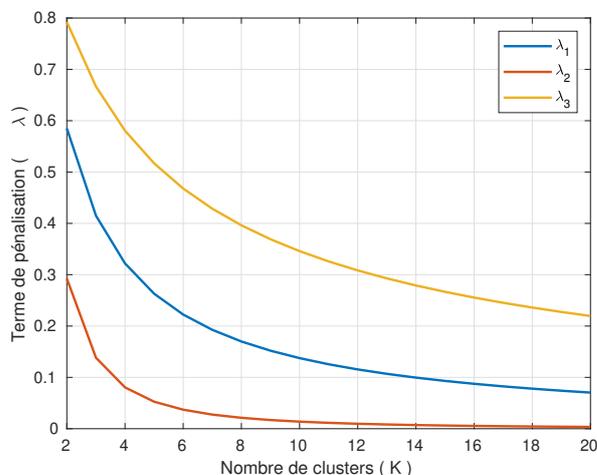


Figure 2.2 – Terme de pénalisation

En observant les trois termes proposés,  $\lambda_2$  est celui avec les plus petites valeurs, cependant sa décroissance est très rapide. Notre choix se porte donc sur  $\lambda_1$  qui a une pente plus progressive.

Le nouveau critère à maximiser est :

$$Q_{\text{pen}}^{\text{Séquentiel}} : (S^*, K^*, \theta^*) = \operatorname{argmax}_{S, K, \theta} \left[ \frac{E(\mathbf{p}_{\text{est}}(\cdot | S, K, \theta))}{\log_2 K} - \alpha \lambda_1 \right] \quad (2.18)$$

La multiplication de  $\lambda_1$  par une valeur  $\alpha \in [0, 1]$  est possible pour donner plus ou moins d'importance à la pénalisation.

On reprend notre précédent exemple 2.1, cette fois en estimant l'entropie normalisée munie de la pénalisation  $\alpha\lambda_1$ .

**Exemple 2.3** On choisit  $\alpha = 0.2$ , les résultats sont regroupés dans le tableau 2.3.

Tableau 2.3 – Exemple d'un jeu à 4 classes avec des critères normalisés munis d'une pénalité

	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$	$\Delta_5$	$E_{norm}$	$E_{norm}^{pen}$
$S_1, K=2$	0.25	0.75	/	/	/	0.8113	0.6943
$S_2, K=2$	0.5	0.5	/	/	/	<b>1</b>	<b>0.8830</b>
$S_1, K=3$	0.25	0.25	0.5	/	/	0.9464	0.8634
$S_1, K=4$	0.25	0.25	0.25	0.25	/	<b>1</b>	<b>0.9356</b>
$S_1, K=5$	0.25	0.25	0.25	0.05	0.2	0.9391	0.8865
$S_3, K=5$	0.1	0.1	0.1	0.05	0.65	0.6962	0.6436

Concernant les partitions  $(S_1, K=3)$  et  $(S_1, K=5)$ , le critère muni de la pénalité permet d'avoir une comparaison similaire à celle avec l'entropie de Shannon normalisée, la différence est que la partition avec 5 clusters est maintenant mieux notée. Le second point concerne les partitions dont l'entropie normalisée est 1, la pénalisation a permis de choisir  $K=4$ . On peut conclure sur la base de cet exemple que le critère de validation proposé à l'équation 2.18 peut servir au choix du nombre de clusters et le sous-ensemble de descripteurs pertinents pour cet exemple.

Les résultats obtenus par une validation conventionnelle (estimation de  $K^*$  et  $S^*$ ) avec le critère muni de la pénalisation seront illustrés à l'annexe A.1.

## 2.3 Illustration sur ORION-EA

La validation par vote (section 1.2.3, [17]) est faite à partir des trois critères de forme utilisés au chapitre 1 : *Silhouette*, *Calinski-Harabasz* et *Davies-Bouldin* déjà implémentés dans la *toolbox* [90] et sur Matlab. Nous utilisons 13 descripteurs, la configuration des données et l'ensemble des partitions sont identiques à ce qui a été fait à la section 1.2.4. Des sous-ensembles de 4 descripteurs sont utilisés pour le partitionnement, soit un total de 715 combinaisons. Le partitionnement se fait à partir de l'algorithme des GMM, pour un nombre de *clusters* variant entre  $K_{min} = 2$  et  $K_{max} = 10$  et  $M = 10$ .

L'évaluation repose sur deux aspects. Le premier est le lien entre les partitions sélectionnées et les niveaux de serrage. Pour cela les valeurs des ARI pour  $K \in [K_s - 1; K_s + 1]$  (avec  $K_s$  est le nombre de serrage) sont estimées de la même manière qu'au chapitre 1. Les résultats du vote sont comparés à ceux obtenus par le critère séquentiel à l'aide du critère ARI.

Le second aspect concerne les instants de déclenchement. Les histogrammes des instants de déclenchement des *clusters* sont ainsi tracés à partir de  $(K_{max} - K_{min} + 1) \times M$  partitions, soit un total de 90

(Sur les graphes, les valeurs des histogrammes sont divisées par 90). Plus un pic est haut, plus cet instant apparaît dans les partitions. Sur l’histogramme, un pas de 0.5s est fixé pour discrétiser l’axe des temps sur l’histogramme. Pour chacune des cinq campagnes (#B, #C, #D, #E, #F), l’histogramme obtenu par vote est comparé à celui donné par le critère séquentiel proposé.

### 2.3.1 Lien entre les partitions et les serrages

Les valeurs des ARI ont été calculées pour les 10 meilleures partitions suivant chacune des méthodes pour  $K \in [2; 10]$ . L’ARI médian est estimé pour les valeurs  $K=\{6,7,8\}$  pour les essais #B, #D, #E et #F; et  $\{5, 6, 7\}$  pour l’essai #C. Ces valeurs médianes, ainsi que les valeurs maximales sélectionnées ont été regroupées dans le tableau 2.4.

Tableau 2.4 – Lien entre partitions obtenues et les serrages

Essai #B		
	Séquentiel	Vote
ARI médian	<b>0.56</b>	0.53
ARI maximal	<b>0.7</b>	0.66
Essai #C		
	Séquentiel	Vote
ARI médian	0.45	<b>0.46</b>
ARI maximal	<b>0.74</b>	<b>0.74</b>
Essai #D		
	Séquentiel	Vote
ARI médian	<b>0.43</b>	0.27
ARI maximal	<b>0.61</b>	0.6
Essai #E		
	Séquentiel	Vote
ARI médian	<b>0.44</b>	<b>0.44</b>
ARI maximal	<b>0.56</b>	0.55
Essai #F		
	Séquentiel	Vote
ARI médian	<b>0.33</b>	0.26
ARI maximal	<b>0.6</b>	<b>0.6</b>

Les valeurs maximales sont très peu impactées par le critère choisi (écarts relatifs respectifs "au profit" du critère séquentiel de 5.9% - 0% - 1.7% - 1.8% - 0% pour les essais #B, #C, #D, #E et #F. Les valeurs médianes sont meilleures pour le critère séquentiel pour les essais #D et #F (écarts respectifs de (45.7 % et 23.7%) et très proches pour les essais #B, #C et #E (écarts respectifs de 5.5%, -2.2% et 0%).

En conclusion, s’intéresser aux instants de déclenchement (aspect séquentiel) n’est pas incompatible avec la pertinence des *clusters*. Cependant les valeurs des ARI que nous avons obtenues ne sont pas optimales. Une explication possible est que le critère n’est pas très spécifique aux données. Une contrainte sur la séparabilité temporelle des *clusters* peut par exemple être ajoutée, ou fixer un seuil

pour considérer un déclenchement de *clusters*. Cela permettrait d'éviter la sélection des partitions ayant des *clusters* presque vides et la sélection des partitions ayant des *clusters* détectés de manière précoce.

Par la suite, les partitions issues des critères de forme et séquentiel sont illustrés à l'aide des histogrammes des instants de déclenchements des *clusters* afin d'évaluer la détection des événements (ici niveaux de serrage). De manière générale, les classes sont inconnues dans l'analyse de l'EA, une analyse d'histogrammes peut servir à identifier les instants critiques de l'essai. L'utilisation des références (instants et classes) permet juste d'évaluer la pertinence de nos résultats. Aucune valeur de référence n'intervient dans l'estimation du critère, ni de la validation, elles sont utilisées pour une évaluation *a posteriori* des résultats, connaissant les niveaux de serrage, qui sont ici nos classes.

### 2.3.2 Détection des niveaux de serrage

On souhaite vérifier si les déclenchements de *clusters* coïncident avec les changements de niveaux de serrage. Pour cela, sur les histogrammes des droites verticales en traits interrompus ont été tracées pour marquer ces bornes. Ces traits verticaux sont donc la vérité de terrain des instants de déclenchements. Deux hyper-paramètres : le **pas de discrétisation** et un **seuil de pic** (fixé par l'utilisateur), sont très importants pour que l'analyse des histogrammes soit pertinente. Le pas d'échantillonnage représente l'erreur maximale qu'on peut tolérer pour la détection des événements, pour notre cas il est fixé à 0.5s. Le choix du pas conditionne la hauteur des pics. Plus le pas est grand, plus les pics seront élevés, en revanche les instants de déclenchements sont moins précis. Le second paramètre à considérer est un seuil pour la hauteur d'un pic, on considère qu'il y a un événement uniquement si les pics sont plus hauts que ce dernier. Le choix du pas peut conditionner celui du seuil, car la hauteur des pics en dépend. Nous illustrerons pour l'essai #B (uniquement), l'impact que peut avoir ce seuil pour la détermination du nombre de *clusters* optimal par exemple.

Pour l'essai #B, les instants à 10s, 30s et 60s ne sont pas mis en évidence (les pics sont en dessous de 0.2) avec le vote (Figure 2.3a). Pour ces trois changements de niveaux, on observe très peu de déclenchements de *clusters*. Plusieurs événements sont détectés par la validation par le vote (critères de formes) entre 0 et 10s (temps du premier serrage), ils sont cependant inexistantes (ou moins prononcés) sur l'histogramme de validation séquentielle (figure 2.3b). Tous les niveaux de serrage sont détectés par le critère séquentiel (avec les pics les plus élevés), à l'exception du dernier serrage. Les observations liées à ce serrage sont attribués au serrage précédent ou distribués entre les *clusters* liés aux autres serrages. Il reste cependant quelques pics ne correspondant à aucun changement de niveaux à 18s et 53s. Un seuil à 0.2 (équivalent à 20 % des pics), conduit à sept événements à 0s (début de l'essai), 1s, 2s, 2.5s, 11s, 20s et 40s pour le vote. Ce même paramétrage conduit à 8 événements localisés à 0s 10s, 18s, 20s, 30.5s, 39s, 49.5s et 52s.

En prenant un seuil plus faible, fixé à 0.1 par exemple, des événements supplémentaires sont détectés

à 10s (serrage à 40cNm) et et 49s (serrage à 10cNm) pour la validation par vote. Pour la validation séquentielle, deux évènements supplémentaires à 5s et 11s sont détectés.

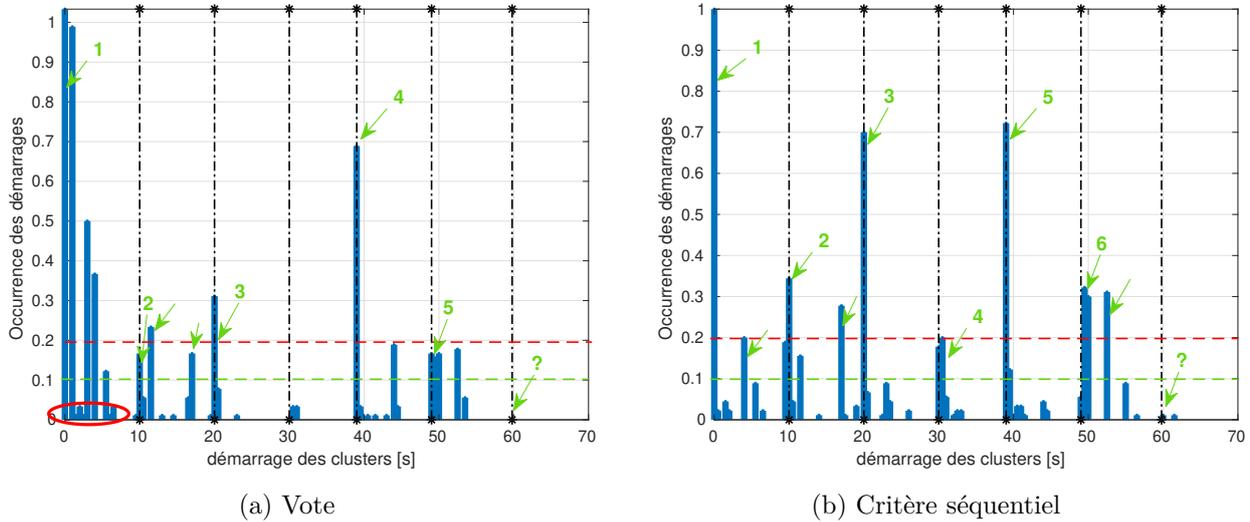


Figure 2.3 – Essai #B : Validation du partitionnement par les GMM en utilisant l’*algorithme 1*, en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de  $K \in [2, 10]$ .

Si on suppose que le nombre d’évènements détectés par le critère séquentiel correspond au nombre de *clusters* optimal, le seuil à 0.2 conduit à 8 *clusters*. On aura 10 *clusters* pour un seuil à 0.1. Les figures 2.4a et 2.4b montrant les cumulées d’occurrence des *clusters* pour  $K = 8$  et  $K = 10$  des meilleures partitions suivant le critère séquentiel expliquent certaines observations au niveau des histogrammes et des diagrammes des ARI. Pour  $K = 8$  et  $K = 10$  les changements de serrage (en dehors du dernier) sont correctement détectés. Les serrages sont constitués d’un ou plusieurs *clusters*. À  $K = 10$ , les *clusters* qui représentaient initialement le deuxième (entre 10 et 20s) et troisième (entre 20 et 30s) serrages sont segmentés, ce qui fera ainsi apparaître des pics à ces instants dans la représentation en histogrammes. Les mélanges de serrage dans les *clusters* (démarrant à 30s pour la figure 2.4a par exemple) réduisent les valeurs des ARI. Pour les essais restants, seuls les histogrammes seront présentés.

Le vote appliqué à l’essai #C (figure 2.5a), permet de détecter uniquement les changements de serrage à 10 et 40s pour un seuil de 0.2. Ceux à 20s, 30s sont détectés, mais pas suffisamment mis en évidence. En baissant ce seuil à 0.1, on a une détection supplémentaire à 30s. On note très peu d’évènements entre les changements de niveaux. À 10s le pic avoisine 1.7 (supérieur à 1), ce qui signifie que plusieurs *clusters* se déclenchent à cet instant pour une même partition.

La figure 2.5b présente les résultats avec le critère séquentiel où des déclenchements de *clusters* sont observés en cohérence avec la vérité de terrain. À 20s le changement de serrage n’est pas suffisamment mis en évidence. Le changement de serrage à 50s est détecté de manière précoce, ce qui signifie que les derniers points du serrage à 10 cNm sont assimilés à ceux du dernier serrage (5 cNm). En baissant

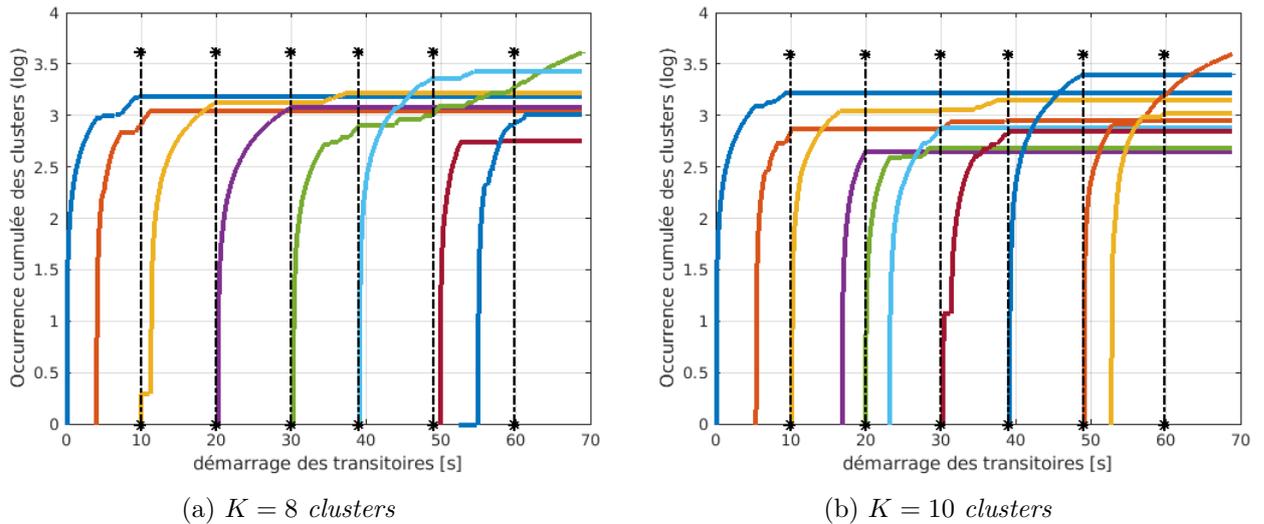


Figure 2.4 – Essai #B : Cumulées d’occurrence de la première partition du tri par le critère séquentiel pour  $K = 8$  et  $K = 10$  clusters.

le seuil à 0.1, il n’y a pas de serrage supplémentaire détecté.

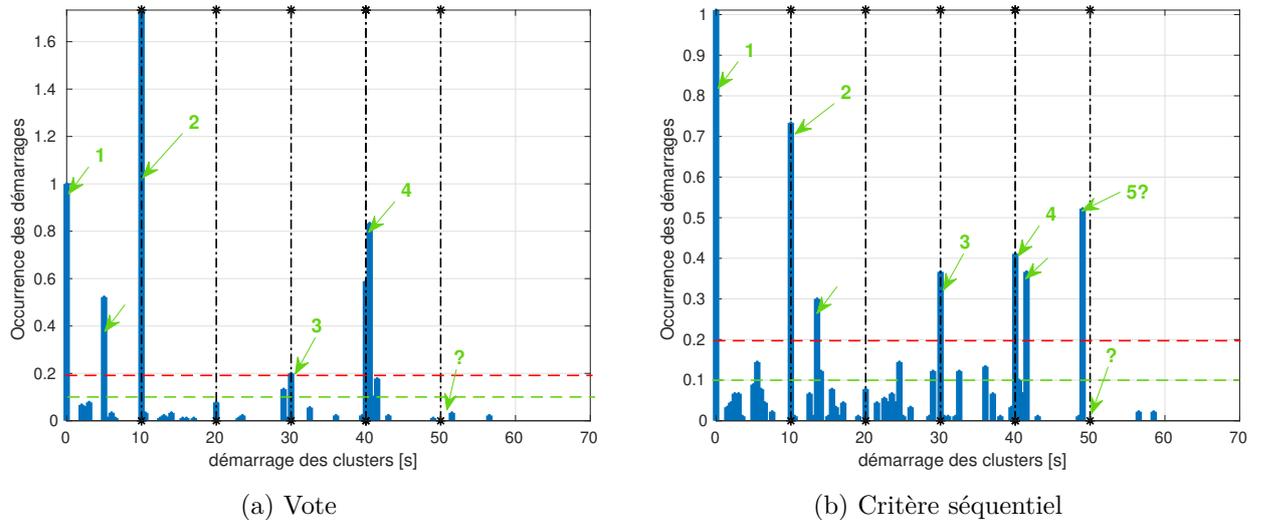


Figure 2.5 – Essai #C : Validation du partitionnement par les GMM en utilisant l’*algorithme 1*, en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de  $K \in [2, 10]$ .

La figure 2.6a correspond à l’histogramme des déclenchements de clusters obtenus par la validation par le vote pour l’essai #D. Seuls deux desserrages sont détectés à 20 et 30s pour les seuils 0.2 et 0.1. Quelques clusters se déclenchent au milieu d’un serrage.

Sur la figure 2.6b donnée par le critère séquentiel, pour un seuil de 0.1, des déclenchements de clusters sont observés à tous les instants de changements de serrage, excepté à 50s où le pic est légèrement en dessous. Cinq parmi six desserrages sont détectés par la validation séquentielle, contre deux pour les critères de forme. la hauteur des pics étant plus régulière qu’avec le vote. Il existe cependant quelques

pics ne correspondant à aucun changement de niveaux de serrage.

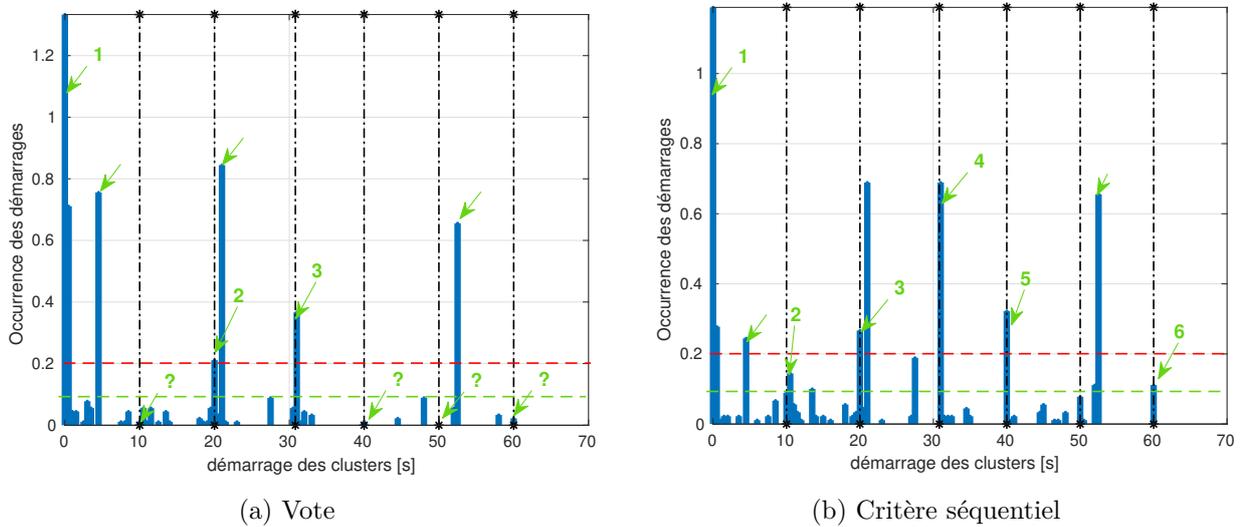


Figure 2.6 – Essai #D : Validation du partitionnement par les GMM en utilisant l’*algorithme 1*, en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de  $K \in [2, 10]$ .

Les figures 2.7a et 2.7b présentent les histogrammes obtenus par le vote et le critère séquentiel pour l’essai #E. Le vote ne permet pas de mettre en évidence les desserrages à 50s et 60s. Les évènements détectés sont les mêmes pour les seuils de 0.1 et 0.2.

Pour le critère séquentiel, tous les desserrages sont détectés. Le dernier est détecté avec un retard compris entre 0.5 et 1s. Les pics les plus élevés correspondent aux changements de serrage. Malgré les valeurs médianes et maximales très proches (tableau 2.4), les évènements sont mieux détectés avec le critère séquentiel proposé.

La figure 2.8a illustre la validation basée sur le vote pour l’essai #F. Plusieurs déclenchements de *clusters* sont détectés pendant le premier serrage pour les deux méthodes de validation. Ces pics sont cependant plus prononcés pour le vote. Avec un seuil à 0.2, seul le desserrage à 30s est détecté avec le vote. En le baissant à 0.1, on a des évènements à 40s et 50s, ce qui engendre également plusieurs détections durant le premier serrage.

Le résultat s’améliore nettement avec le critère séquentiel (Figure 2.8b). En effet les pics sont plus élevés et leur hauteur relativement régulière au niveau des bornes. Avec un seuil de 0.2, quatre desserrages sur six sont détectés, contre deux sur six pour le vote. En baissant ce seuil à 0.1, le desserrage à 20s est également mis en évidence. Le desserrage à 60s reste cependant non détecté.

L’absence de pics à certaines bornes, observée pour les différents essais est due à une possible défaillance de la méthode de partitionnement (GMM), comme illustré avec les figures 2.4a et 2.4b des cumulées d’occurrence des clusters pour l’essai #B. L’algorithme ne parvient pas toujours à discriminer certains serrages. Les critères de validation trient uniquement les partitions, ils n’interviennent

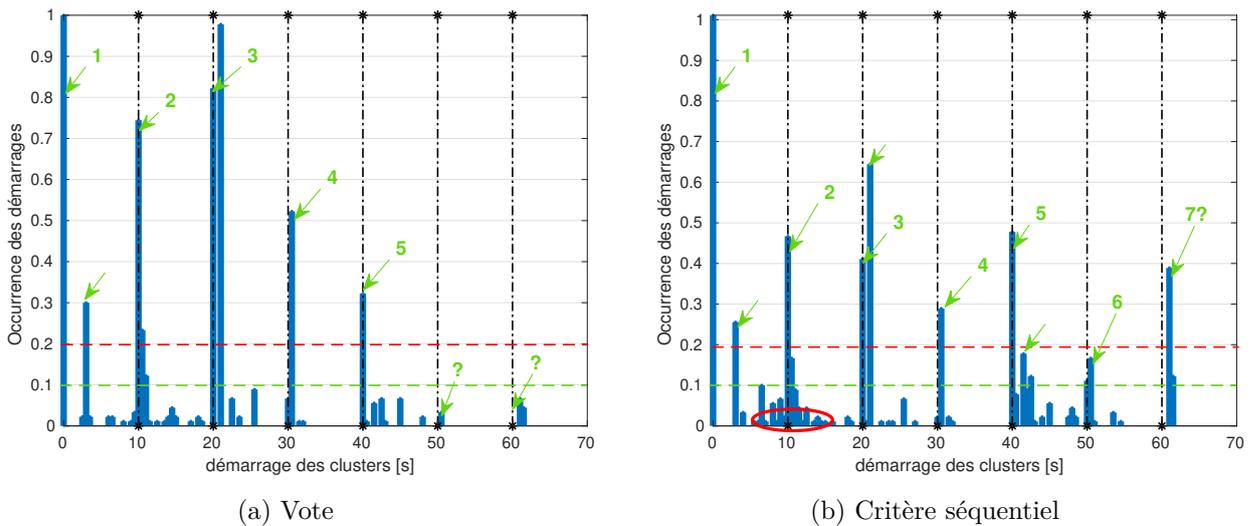


Figure 2.7 – Essai #E : Validation du partitionnement par les GMM en utilisant l’*algorithme 1*, en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de  $K \in [2, 10]$ .

pas dans la phase du partitionnement. Si aucun *cluster* ne se déclenche à ces instants, ou alors les partitions en question ne sont pas choisies par le critère séquentiel parmi les 10 meilleures ; rien ne sera observé sur l’histogramme.

Utiliser le critère séquentiel permet d’observer un plus grand nombre d’évènements (changement de niveaux). Le risque que peut présenter l’algorithme est qu’en prenant des valeurs très grandes de  $K$  comparées au vrai nombre de *clusters*, le nombre de fausses détections croît. Cependant ce qui nous intéresse ce sont les instants avec les pics les plus élevés. Pour les différents essais, nous avons pu constater qu’ils correspondent aux changements de niveaux de serrage lorsque le critère séquentiel est utilisé. Inversement si  $K_{max}$  est faible, certains évènements ne seront pas détectés. Cette partie expérimentale aura ainsi permis de montrer que l’utilisation du critère séquentiel proposé conduit à une meilleure analyse des évènements (changement de niveaux de serrage pour notre cas) que ce qui est possible avec les critères de forme.

## 2.4 Discussions et conclusions

Ce chapitre a permis de formaliser un critère capable de trouver des sous-ensembles de descripteurs conduisant à des instants de déclenchement de *clusters* pertinents du point de vue séquentiel. Il s’agit à notre connaissance du premier critère séquentiel dédié à l’EA. Si des *a priori* sont disponibles sur les instants de déclenchement des *clusters*, la divergence de Kullback-Leibler permet de trier les partitions. En l’absence de ces derniers, la distribution uniforme est privilégiée en se basant sur le principe de l’entropie maximale. Le critère prend alors la forme d’une entropie normalisée donnée par l’équation 2.14. Ce critère permet d’évaluer la qualité des partitions en terme de séquence, bien qu’il reste indécis sur le choix du nombre de clusters lorsque les performances sont proches. Une validation basée sur

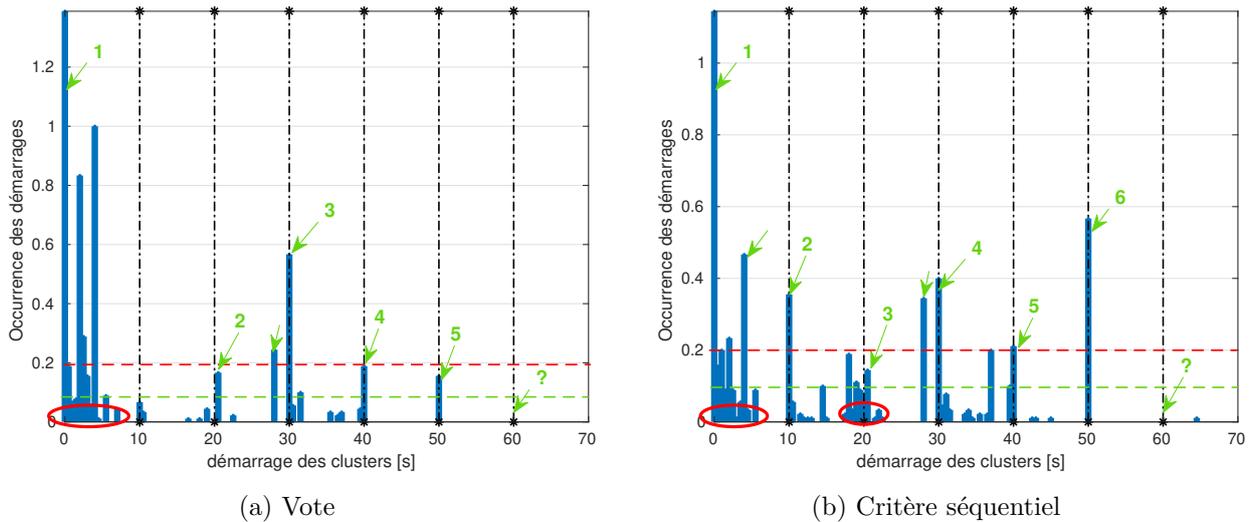


Figure 2.8 – Essai #F : Validation du partitionnement par les GMM en utilisant l’*algorithme 1*, en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de  $K \in [2, 10]$ .

les histogrammes des déclenchements des *clusters* a été proposée. Celle-ci fournit à l’utilisateur des instants d’évènements potentiellement majeurs, tout en lui laissant la responsabilité de les considérer ou non. Cette approche peut être très utile pour les tâches de SHM et PHM où la connaissance et la prédiction de ces instants de déclenchements sont nécessaires pour la prise de décision.

Cette approche a pour but de répondre aux limites relevées au chapitre 1 sur la validation mono / multicritère basée sur des critères internes et le choix d’une seule partition. Le premier point soulevé est l’absence de corrélation entre la compacité des *clusters* et leur pertinence du point de vue séquentiel. Le second concerne la subjectivité du choix d’une seule partition, la différence de qualité étant généralement très faible entre les premières partitions d’un tri, le choix variant aussi d’un critère à l’autre. Le critère séquentiel et la validation par des histogrammes sont des tentatives de réponses à ces deux préoccupations respectives.

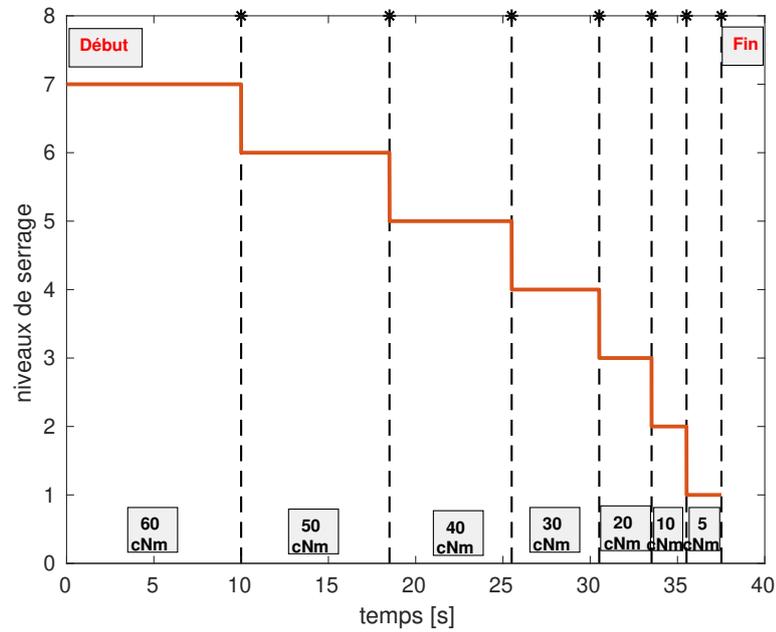
Expérimentalement, le benchmark ORION-EA a été utilisé. Il est montré que l’approche proposée avec le critère séquentiel conduit à une évaluation plus pertinente des séquences. Cette étude expérimentale montre la nécessité du développement d’outils spécifiques pour l’évaluation des EA dans le cadre du SHM et du monitoring des procédés. Une limite que peut cependant présenter notre approche est qu’en augmentant la valeur de  $K_{max}$ , on crée potentiellement des nouveaux *clusters* avec des instants de déclenchement non pertinents. Cependant c’est l’accumulation des instants de déclenchement des *clusters* sur des périodes courtes qui nous informe de la présence d’évènements particuliers.

L’interprétation non supervisée des EA est généralement définie comme une sélection d’un sous-ensemble de descripteurs et d’un nombre de *clusters* optimal, ce que ne permet pas notre critère. Le nombre de clusters est déterminé visuellement à partir des histogrammes. Nous avons également proposé un critère séquentiel (équation 2.18) pouvant déterminer  $K^*$  sous l’hypothèse de la pénalisa-

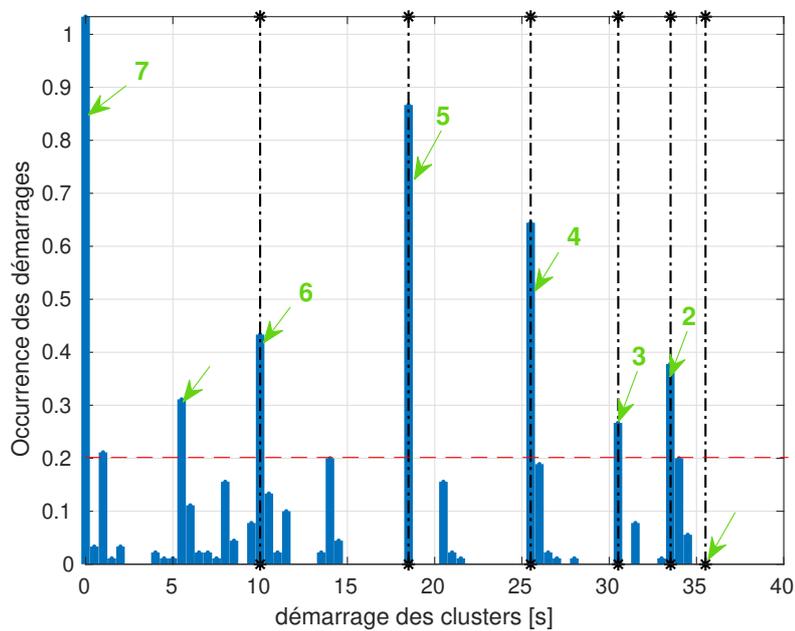
tion des nombres de *clusters* faibles. Pour l'essai #B avec une pénalisation de  $0.25 \times \log_2(1 + 1/K)$  (annexe A.1), les résultats montrent des partitions avec des ARI plus élevés avec ce critère que ceux obtenus avec la méthode de vote avec les critères de forme.

Une solution pour le choix du nombre de *clusters* optimal  $K_S^*$ , pour un sous-ensemble de descripteurs peut reposer sur le partitionnement bayésien. Comme décrit dans [78, Chap. 10], les méthodes de partitionnement bayésien estiment automatiquement  $K_S^*$ . Par conséquent, il suffit de définir une valeur élevée de  $K$  et les clusters inutiles disparaissent automatiquement lors de l'optimisation du modèle [78, Chap. 10]. Ce modèle a été étudié pour le partitionnement des données AE dans [117] et est présenté au chapitre 3. Il permet de contourner la limite sus-mentionnée pour l'approche proposée dans ce chapitre.

Enfin la méthode de validation proposée a été évaluée pour des événements ayant des durées qui ne sont pas uniformément réparties dans le temps mais avec des durées logarithmiques afin de simuler un desserrage qui s'approche plus de la réalité. L'idée est que le boulon met plus de temps à se desserrer lorsque le serrage est fort. Ainsi pour un couple de 60 cNm (niveau 7), la durée du serrage est de 10s, 8.5s pour 50cNm, 7s pour 40 cNm, 5s pour 30 cNm, 3s pour 20 cNm, 2s pour 10 cNm et 2s pour 5 cNm (figure 2.9a). Les résultats illustrés en annexe A.2 et la figure 2.9b, montrent que tous les niveaux de serrage sont détectés, à l'exception du niveau le plus faible. Ce résultat a été obtenu en supposant que les instants de déclenchement a priori n'étaient pas connus (la distribution de référence est donc uniforme). Les performances sont également meilleures en terme d'ARI et de séquence, comparés à ceux obtenus avec les critères de forme. Ces différents résultats valident la pertinence de la méthode proposée pour des tâches de surveillance, où la détection des événements est nécessaire.



(a) Niveaux de serrage



(b) Déclenchements des clusters

Figure 2.9 – Essai #B : Durée de serrage logarithmiques | Le boulon se desserre plus rapidement lorsque le niveau de serrage est faible.

## Chapitre 3

# Approche bayésienne pour l'interprétation des données d'EA

### Sommaire

---

<b>3.1</b>	<b>Concepts clés et outils nécessaires</b>	<b>73</b>
3.1.1	Entre non supervisé et supervisé	73
3.1.2	Labellisation douce et fonctions de masse	75
3.1.3	Combinaison des fonctions de croyance	77
<b>3.2</b>	<b>Méthode bayésienne variationnelle pour les mélanges gaussiens</b>	<b>79</b>
3.2.1	Lois de distribution des variables	79
3.2.2	Définition de la fonction objectif	81
3.2.3	Inférence approchée et algorithme VB-EM	82
3.2.4	Étape VB-E avec a priori sur les clusters	86
<b>3.3</b>	<b>Usage du VB-GMM-SOFT pour l'analyse de l'EA</b>	<b>88</b>
3.3.1	Détermination du nombre de clusters	88
3.3.2	Comment générer les <i>a priori</i>	89
<b>3.4</b>	<b>Résultats sur ORION-EA</b>	<b>90</b>
3.4.1	Influence de la probabilité d'erreur $\rho$	91
3.4.2	Influence de la corruption des étiquettes et de la taille des données	100
3.4.3	Influence des autres hyper-paramètres	100
<b>3.5</b>	<b>Discussions et conclusions</b>	<b>103</b>

---

## Introduction

Les algorithmes de partitionnement usuels (k-moyennes, hiérarchiques, mélanges gaussiens, partitionnement flou) vus au chapitre 1 utilisent un nombre de *clusters*  $K$ , fixé comme paramètre du modèle. Dans le cadre de l'EA, ce nombre est généralement inconnu pour un sous-ensemble de descripteurs, plusieurs partitions sont alors générées en le faisant varier dans un intervalle  $[K_{min}, K_{max}]$ . Une étape de validation permet de l'estimer en optimisant un critère de « qualité ». Dans ce même chapitre nous avons vu que ce nombre pouvait varier d'un critère à l'autre (voir figure 1.18a). Partant de ce constat, au chapitre 2, une approche de validation à base d'histogramme des instants de déclenchement des *clusters* de plusieurs partitions présélectionnées à partir d'un critère séquentiel a été proposée. Le critère séquentiel n'étant pas adapté à l'estimation de  $K$ , toutes les partitions à  $K \in [K_{min}, K_{max}]$  *clusters* sont prises en compte. Une limite qui avait été soulevée est qu'en utilisant un intervalle large  $[K_{min}, K_{max}]$ , on génère des instants de déclenchement des *clusters* potentiellement inexistantes, ce qui réduit la qualité du résultat tout en augmentant le coût du calcul.

Dans ce chapitre nous nous proposons d'explorer l'**algorithme bayésien variationnel (VB-GMM)**. Cet algorithme de partitionnement intègre la **détermination automatique du nombre de « clusters naturels »** pour un sous-ensemble de descripteurs. La validation se réduit dès lors à la sélection du meilleur sous-ensemble de descripteurs. Un second avantage applicatif qu'offre cet algorithme est la possibilité d'introduire des *a priori* sur les *clusters*.

Le second point abordé dans ce chapitre est la construction d'un modèle prenant en compte des *a priori* sur les *clusters*. En pratique, labelliser les données d'EA peut s'avérer laborieux pour deux contraintes principales. La première concerne la **taille des données**, les signaux d'EA détectés peuvent être très nombreux en fonction de l'émissivité du matériau et des contraintes subies par la structure. Pour les 5 campagnes ORION-EA par exemple, nous avons une détection moyenne d'environ 210 signaux par seconde. Pour des essais de longue durée, ce nombre peut devenir problématique en terme de coût si les données sont traitées manuellement par l'expert. Le second point concerne les **incertitudes sur les signaux**, liées aux seuils de détection et à la connaissance insuffisante du lien entre les signaux et les endommagements (chapitres 1 et 2). Ainsi, certaines étiquettes peuvent être erronées ou manquantes. Ces contraintes sont communes aux problèmes de diagnostic et pronostic, notamment dans les domaines de l'aéronautique et du génie civil. Cette connaissance, bien qu'incertaine peut être exploitée en y accordant **plus ou moins de crédit** de façon à ce qu'elle ne soit pas le seul facteur de décision.

La question a déjà été abordée dans [121] pour les modèles de mélanges gaussiens (GMM). L'extension de cette approche à un modèle bayésien (**VB-GMM avec *a priori***) que nous proposons, peut être intéressante pour deux principales raisons. La première est la possibilité de considérer conjointement la connaissance sous forme d'*a priori* et les signaux d'EA. La seconde raison concerne la sélection automatique du nombre de *clusters* en cas de doute sur sa valeur *a priori*.

Ce chapitre vise à proposer une méthode de partitionnement qui répond à ces deux problématiques :

- Comment améliorer le partitionnement en prenant en compte la connaissance sous forme d'*a priori* incertains sur les *clusters* ?
- Comment estimer de manière automatique le nombre de *clusters* pour un sous-ensemble de descripteurs **avec et sans *a priori*** sur les *clusters* ?

### 3.1 Concepts clés et outils nécessaires

En général, lorsque les *classes* sont connues, on migre vers des modèles supervisés. Cela permet de générer une fonction d'apprentissage, qui est par la suite utilisée pour la caractérisation de nouveaux essais. Les signaux d'EA étant complexes à analyser et difficilement reproductibles, on a généralement des ***a priori* incertains ou incomplets sur les *clusters***. L'apprentissage partiellement supervisé est une des solutions existantes pour prendre en compte ces *a priori* incertains sur les *clusters* dans un modèle, en faible quantité pour ne pas être en contradiction avec les données d'EA.

Au cours de ce chapitre, le modèle exploité et celui proposé sont des versions générales des mélanges gaussiens (GMM vus au chapitre 1.2.2). Nous reprenons les définitions des chapitres précédents :  $\mathbf{y} = \{y_1, \dots, y_N\}$ , représente l'ensemble des observations ;  $\mathbf{x} = \{x_1, \dots, x_N\}$ , les *clusters* les ayant générés ;  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Lambda})$ , l'ensemble des paramètres (moyenne et dispersion) des  $K$  *clusters* ; et  $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^K$ , les proportions du mélange.  $\mathbf{x}$  est un ensemble de valeurs binaires  $x_n = \{x_{nk}\}_{k=1}^K$ , représentant l'appartenance aux *clusters* :  $x_{nk} = 1$  si l'observation  $y_n$  appartient au  $k^{ième}$  cluster, et 0 sinon.

#### 3.1.1 Entre non supervisé et supervisé

Plusieurs approches font le pont entre les approches supervisée et non supervisée. Elles reposent sur l'exploitation des données lorsque, seule une partie des données est étiquetée ou qu'un doute existe sur les étiquettes disponibles. Il s'agit des méthodes semi-supervisées et partiellement supervisées. Nous présentons ici les principales approches qui permettent d'utiliser en même temps les étiquettes et les données non étiquetées. Il s'agit de :

- **la classification et partitionnement avec pré-étiquetage** : un apprentissage supervisé est préalablement fait à partir des données étiquetées. Les données non étiquetées sont par la suite classées à partir de la fonction apprise lors de l'étape précédente. Le modèle est réajusté en utilisant cette fois toutes les données pour apprendre. *L'algorithme des plus proches voisins* peut s'inscrire dans ce type de méthodes [122].
- **L'apprentissage avec post-étiquetage à la majorité** : un partitionnement est fait sur l'ensemble des données, qui sont par la suite étiquetées à la majorité. Cet étiquetage consiste à attribuer à une donnée non-étiquetée, l'étiquette ayant le plus d'occurrences au sein du *cluster*.

Si après cette phase certains *clusters* ne possèdent aucune donnée étiquetée, une mesure de similarité permet de leur attribuer l'étiquette du *cluster* le plus proche. Après cette phase un apprentissage peut être effectué sur l'ensemble des données [122].

- **Le partitionnement semi-supervisé** consiste à répartir les données en prenant en compte  $M$  données étiquetées dans un ensemble de  $N$  observations pour orienter le partitionnement à l'aide des étiquettes. Pour un modèle semi-supervisé l'ensemble des données d'apprentissage est de la forme :

$$(\mathbf{y}, \mathbf{l}) = \underbrace{\{(y_1, \mathbf{l}_1), (y_2, \mathbf{l}_2), \dots, (y_M, \mathbf{l}_M)\}}_{\text{données étiquetées}}, y_{M+1}, \dots, y_N \quad (3.1)$$

Une représentation binaire  $\mathbf{l} = \{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_N\}$ , permet de représenter les classes  $\omega_k$ .  $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ , est l'ensemble des classes possibles donné par l'expert. Pour  $K$  classes données,  $\mathbf{l}_n$  peut prendre les valeurs  $\{1, 2, \dots, K\}$ . Sous forme matricielle, on a  $l_{nk} = 1$  si  $y_n \in \omega_k$ , sinon  $l_{nk} = 0$ . Dans ces modèles, le problème est d'estimer  $E(x_{nk})$  pour  $n = M + 1 \dots N$ , l'espérance que l'observation  $y_n$  appartienne à la  $k^{ieme}$  classe. La résolution est faite par un algorithme de type EM [122].

Pour un GMM semi-supervisé, la fonction objectif qu'on cherche à maximiser est la log-vraisemblance donnée par :

$$\mathcal{L}(\mathbf{y}, \mathbf{x}, \mathbf{l}) = \underbrace{\sum_{n=1}^M \sum_{k=1}^K l_{nk} \log \{\pi_k \mathcal{N}_k(y_n | \theta_k)\}}_{\text{données étiquetées}} + \underbrace{\sum_{n=M+1}^N \log \left\{ \sum_{k=1}^K \pi_k \mathcal{N}_k(y_n | \theta_k) \right\}}_{\text{données non-étiquetées}} \quad (3.2)$$

Dans l'équation 3.2, le premier terme somme les probabilités des  $M$  données étiquetées. Le second terme équivaut à la log-vraisemblance des observations non-étiquetées [122, 123]. Seules les espérances  $E(x_{nk}), (\forall n > M)$  des données non étiquetées sont estimées.  $E(x_{nk})$  est égale à  $l_{nk}$  quand la donnée est étiquetée.  $E(x_{nk})$  garde la même expression que dans l'équation 1.15 pour les données non-étiquetées [124, 125] :

$$E(x_{nk}) = p(x_{nk} = 1 | \mathbf{l}, \mathbf{y}) = \begin{cases} l_{nk} & \text{si } n \leq M \\ \frac{\pi_k \mathcal{N}(y_n | \theta_k)}{\sum_{l=1}^K \pi_l \mathcal{N}(y_n | \theta_l)} & \text{si } n > M \end{cases} \quad (3.3)$$

- **Le partitionnement partiellement supervisé** est un cas plus général que l'approche semi-supervisée. Contrairement au cas supervisé et au semi-supervisé, le doute sur la classe est pris en compte, en attribuant à une observation les différentes étiquettes possibles. Une autre différence par rapport au semi-supervisé réside sur le fait que les *a priori* sur les classes sont complets.

L'ensemble d'apprentissage est sous la forme :

$$(\mathbf{y}, \mathbf{l}) = \{(y_1, \mathbf{l}_1), (y_2, \mathbf{l}_2), \dots, (y_M, \mathbf{l}_N)\} \quad (3.4)$$

où  $\mathbf{l}_n = \{l_{nk}\}_{k=1}^K$  est un vecteur contenant  $K$  valeurs binaires, représentant l'appartenance à une classe. La log-vraisemblance est donnée par :

$$\mathcal{L}(\mathbf{y}) = \sum_{n=1}^N \log \left\{ \sum_{k=1}^K l_{nk} \pi_k \mathcal{N}_k(y_n | \theta_k) \right\} \quad (3.5)$$

Une observation pouvant appartenir à une ou plusieurs classes, dans cette représentation :

- Si on a une donnée non étiquetée, alors  $l_{nk} = 1, \forall k$
- Si la donnée est étiquetée et appartient à la classe  $\omega_{k'}$ , alors  $l_{nk} = 1$  si  $k = k'$ , et 0 sinon.

On remarque qu'en remplaçant dans l'équation 3.5, les  $(N - M)$  dernières valeurs par  $l_{nk} = 1, \forall k$ , on obtient l'équation 3.2. Le modèle semi-supervisé peut donc être vu comme un cas particulier du partiellement supervisé [123–125].

Dans les approches partiellement supervisées, la connaissance des classes peut être mise sous forme d'*a priori* **binaire** ( $l_{nk} \in \{0, 1\}$ ) ou **non-binaire** sur les *clusters* [126, chap.4]. Pour les cas non-binaires, cette connaissance est représentée par des **fonctions de masse**  $\mathbf{m}_n(l_{nk}) \in [0; 1]$ , représentant le degré d'appartenance de l'observation  $y_n$  à différentes classes. Une des théories permettant l'estimation de ces masses et de leur exploitation dans les approches d'apprentissage est la **théorie des fonctions de croyance (TFC)** proposée par A.P. Dempster [127] et G. Shafer [128]. La TFC considère les classes et les sous-ensembles qui représentent le doute explicite sur les classes ( $\omega_k \cup \omega_{k'}$  par exemple) dans l'estimation des masses. Ceci permet dans le cadre de l'EA d'exploiter les *a priori incertains ou incomplets* fournis par un expert.

Le modèle avec *a priori sur les clusters* que nous proposons dans ce chapitre est basée sur cette théorie. Nous présenterons un aperçu de la méthode et en faisons une analogie avec les GMM avec *a priori* développée dans [123–125].

### 3.1.2 Labellisation douce et fonctions de masse

Dans l'idéal chaque endommagement a une signature caractéristique qui lui est propre. Les classes et *clusters* sont par conséquent certains, compacts et distincts dans l'espace des descripteurs. Cette description est rarement observée pour des données réelles, les frontières entre classes (ou *clusters*) sont souvent floues dans l'espace des descripteurs. La **labellisation douce** prend en compte cette réalité en considérant les données d'apprentissage comme un ensemble d'informations partielles dans les modèles de partitionnement. Une information est dite partielle si elle est incertaine, imprécise ou

incomplète. Ainsi pour une labellisation douce, une observation peut avoir une ou plusieurs étiquettes.

Pour chaque observation  $y_n$ , une fonction de masse  $\mathbf{m}_n$  est attribuée à tout sous-ensemble  $A$ , de *clusters* et définie comme suit [128] :

$$\begin{aligned} \mathbf{m}_n &: 2^\Omega \rightarrow [0, 1] \\ &A \mapsto \mathbf{m}_n(A) \\ \text{s.t.} \quad &\sum_{A \subseteq \Omega} \mathbf{m}_n(A) = 1, 0 \leq \mathbf{m}_n(\emptyset) < 1 \end{aligned} \tag{3.6}$$

où  $\Omega = \{\omega_1, \dots, \omega_k, \dots, \omega_K\}$  est le **cadre de discernement**, constitué de  $K$  classes (ou *clusters*). Une masse est encodée par un vecteur suivant une structure spécifique qui englobe tous les sous-ensembles de *clusters*. Par exemple, pour  $K = 2$  *clusters*, on a  $\mathbf{m}_n$  codé pour :  $\{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1 \cup \omega_2\}\}$ . Une fonction de masse  $\mathbf{m}_n$  est ainsi constituée de  $2^{|\Omega|}$  éléments où  $|\cdot|$  est la cardinalité, correspondant au nombre de *clusters* [129].

Considérons l'exemple illustré dans la Figure 3.1 qui montre deux *clusters* dans un espace 2D.

**Exemple 3.1** *Quatre points sont mis en évidence, pour ces derniers nous disposons des informations préalables suivantes encodées à l'aide des fonctions de masse :*

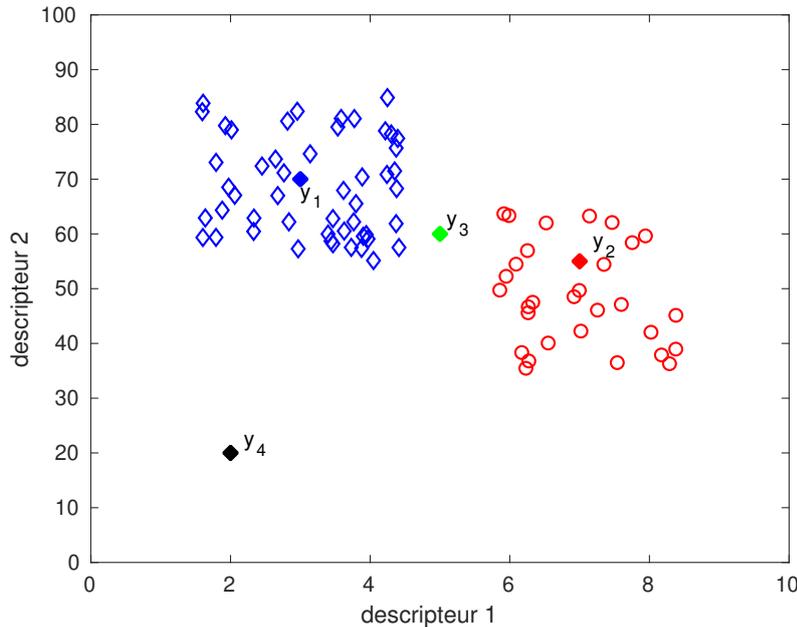


Figure 3.1 – Exemple des fonctions de masse avec observations  $\mathbf{y}_n$  pour 2 *clusters*.

- $\mathbf{m}_1 = [0, 1, 0, 0]$ , qui encode le fait que le point  $\mathbf{y}_1$  appartienne avec certitude au cluster 1.
- $\mathbf{m}_2 = [0, 0, 0.8, 0.2]$ , qui encode le fait que le point  $\mathbf{y}_2$  appartienne au cluster 2 avec une masse égale à 0.8, le reste de la masse étant affecté au cadre de discernement  $\Omega$ . Cette partie de la masse n'appartient ni à  $\omega_1$  ni à  $\omega_2$  mais seulement à l'union.
- $\mathbf{m}_3 = [0, 0, 0, 1]$ , qui encode le fait qu'on ne sait pas à quel cluster appartient  $\mathbf{y}_3$ , donc toute la masse est au cadre  $\Omega$ , sans privilégier un cluster en particulier.

- $\mathbf{m}_4 = [1, 0, 0, 0]$ , qui encode le fait que le point  $\mathbf{y}_4$  est une valeur aberrante et donc toute la masse est affectée à l'ensemble vide ( $\emptyset$ ).

L'exemple 3.1 et l'équation 3.6 traduisent un modèle où l'ensemble vide  $\emptyset$  et les  $\{\omega_j \cup \omega_k\}$  sont pris en compte.

### 3.1.3 Combinaison des fonctions de croyance

Pour les modèles d'apprentissage partiellement supervisé, l'information sur les *clusters* est portée par deux fonctions de masses : la connaissance  $\mathbf{m}_n$ , sous forme d'*a priori* sur les *clusters* fourni par l'expert et  $\pi$ , l'*a priori* sur les *clusters* estimé par le modèle, supposées indépendantes. Ces deux informations peuvent être combinées par la règle de Dempster.

La connaissance (étiquette douce) peut être fournie en termes de plausibilités sur les classes. Ceci permet de ne stocker que des valeurs de dimension  $[N \times |\Omega|]$  au lieu de  $[N \times 2^{|\Omega|}]$ , par rapport à l'utilisation des FC. L'incertitude totale, les valeurs aberrantes et l'absence d'étiquette sont traitées de la même façon en distribuant uniformément la plausibilité sur les  $K$  *clusters*. La règle de Dempster [127, 128] permettant la combinaison conjonctive de  $\mathbf{m}_n$  et  $\pi$  a été formulée par E. Côme et T. Denœux [123–125] de la manière suivante. Étant donné la FC *a priori*  $\mathbf{m}_n$  fournie par la connaissance, et la distribution des probabilités  $\pi$  sur les *clusters*, toutes deux de sources d'évidence indépendantes ; la combinaison de Dempster est donnée par

$$(\mathbf{m}_n \oplus \pi)(A) \propto \sum_{B, C \subseteq \Omega, s.t. B \cap C = A \neq \emptyset} m_n(B) \times \pi(C) = \begin{cases} pl_n(A) \times \pi(A) & \text{si } |A| = 1 \\ 0 & \text{sinon.} \end{cases} \quad (3.7)$$

Le symbole  $\propto$  dans la règle de Dempster signifie que la somme est normalisée puisque certaines des intersections donnent l'ensemble vide qui est exclu de la somme. L'équation 3.7 peut être reformulée ainsi :

$$(\mathbf{m}_n \oplus \pi)(\omega_k) \propto pl_{nk} \pi_k \quad (3.8)$$

Le résultat (équations 3.7 et 3.8) est un cas particulier de la combinaison conjonctive où l'une des deux FC est bayésienne, ici les proportions  $\pi$  [125].

**Proposition 3.1** *Étant donnée la FC a priori  $\mathbf{m}_n$  fournie par la connaissance experte, on définit la plausibilité  $\{pl_n(\omega_k)\}, \forall \omega_k \in \Omega$ , comme une fonction dérivée des fonctions de masse  $\mathbf{m}_n$ . C'est le degré de croyance susceptible que l'observation  $\mathbf{y}_n$  ait été générée par la classe  $\omega_k$  après prise en compte de toute l'information disponible :*

$$pl_n(\omega_k) = \sum_{A \subseteq \Omega, s.t. \omega_k \in A} m_n(A), \forall \omega_k \in \Omega \quad (3.9)$$

La plausibilité  $pl_n(\omega_k)$ , de chaque classe (ou *cluster*)  $\omega_k$ , somme toutes les masses des parties de  $\Omega$  contenant  $\omega_k$ , elle sera notée  $pl_{nk} = pl_n(\omega_k)$ .

**Exemple 3.2** Reconsidérons l'exemple 3.1 précédent avec :  $\mathbf{m}_1 = [0, 1, 0, 0]$ ,  $\mathbf{m}_2 = [0, 0, 0.8, 0.2]$ ,  $\mathbf{m}_3 = [0, 0, 0, 1]$  et  $\mathbf{m}_4 = [1, 0, 0, 0]$ . Les plausibilités sont ainsi données par  $\mathbf{pl}_1 = [0, 1, 0, 1]$ ,  $\mathbf{pl}_2 = [0, 0.2, 1, 1]$ ,  $\mathbf{pl}_3 = [0, 1, 1, 1]$ , et  $\mathbf{pl}_4 = [0, 0, 0, 0]$ . L'observation  $y_4$  sera donc traitée comme une donnée sans étiquette ou retirée de l'ensemble d'apprentissage. Le tableau 3.1 recense ces différents cas d'usage avec la plausibilité  $\mathbf{pl}$  normalisée.

Tableau 3.1 – Cas d'usage des étiquettes pour un partitionnement

	$\omega_1$	$\omega_2$	Commentaires
$p(x_{ik} l_k = 1) := pl_{ik}$	1	0	Cas avec certitude sur la classe $\omega_1$ ; $\mathbf{m}_1$
$p(x_{jk} l_k = 1) := pl_{jk}$	0.17	0.83	Cas avec un a priori majoritaire pour la classe $\omega_2$ ; $\mathbf{m}_2$
$p(x_{lk} l_k = 1) := pl_{lk}$	0.5	0.5	Cas de l'ignorance totale ; $\mathbf{m}_3$
$p(x_{nk} l_k = 1) := pl_{nk}$	0.5	0.5	Valeur aberrante ou absente ; $\mathbf{m}_4$

Reprenons les mélanges gaussiens (GMM), le résultat de l'équation 3.8 permet ainsi d'estimer la loi jointe sur  $(\mathbf{x}, \mathbf{l}, \mathbf{y})$  :

$$p(x_{nk} = 1, \mathbf{l}_n, y_n) = \pi_k pl_{nk} \mathcal{N}_k(\mathbf{y}_n | \theta_k) \quad (3.10)$$

où  $x_{nk} = 1$  si l'observation  $y_n \in \omega_k$ , et 0 sinon. Le résultat présenté au chapitre 1.2.2 (équation 1.15) est modifié afin de prendre en compte la connaissance sous forme d'*a priori*. La loi *a posteriori* sur les *clusters* formalisée dans [123–125] aboutit au résultat :

$$\mathbb{E}(x_{nk}) = p(x_{nk} = 1 | \mathbf{l}_n, y_n) = \frac{p(x_{nk}, \mathbf{l}_n, y_n)}{p(\mathbf{l}_n, y_n)} \quad (3.11a)$$

$$= \frac{\pi_k pl_{nk} \mathcal{N}_k(\mathbf{y}_n | \theta_k)}{\sum_{l=1}^K \pi_l pl_{nl} \mathcal{N}_k(\mathbf{y}_n | \theta_l)}. \quad (3.11b)$$

L'équation 3.11 correspond au GMM partiellement supervisé avec des *a priori* non-binaires sur les *clusters*, estimés sous forme de FC. Si les *a priori* sont binaires, les plausibilités  $pl_{nk}$  sont remplacées dans l'équation 3.11 par les étiquettes  $l_{nk}$ .

Un algorithme de type EM permet d'estimer ces valeurs de façon itérative. À l'étape E (Espérance), les espérances  $E(x_{nk})$  sont estimées à partir des paramètres  $\{\boldsymbol{\pi}, \boldsymbol{\theta}\}$  estimées à l'itération précédente. L'étape M (Maximisation) consiste en une nouvelle estimation de  $\{\boldsymbol{\pi}, \boldsymbol{\theta}\}$  à partir des  $E(x_{nk})$  estimés à l'étape E.

**Propriété 3.1** Si les plausibilités et les étiquettes ne sont pas disponibles, les plausibilités sont toutes fixées :  $pl_{nk} = 1$ . On retrouve l'équation 1.15 des GMM. Le GMM peut ainsi être considéré comme un cas particulier des GMM avec *a priori* [123–125].

## 3.2 Méthode bayésienne variationnelle pour les mélanges gaussiens

Dans le cas du GMM, les paramètres des *clusters*  $\theta = \{\pi_k, \mu_k, \Lambda_k\}_{k=1}^K$  sont des scalaires, considérés comme des valeurs certaines. Pour le modèle bayésien, ces paramètres sont considérés incertains,  $\{\pi_k, \mu_k, \Lambda_k\}$  sont alors représentés par des distributions. La fonction objectif à maximiser est la **log-vraisemblance des données complètes**, en introduisant des distributions  $\{p(\pi_k), p(\mu_k, \Lambda_k)\}$  qui sont des fonctions des paramètres estimés par la règle de Bayes.

Ce GMM avec des paramètres incertains doit sa désignation **VBGMM** (de l'anglais *Variational Bayesian Gaussian Mixture Model*) à son processus d'apprentissage, qui est basé sur une formulation bayésienne variationnelle [78, chap.10]. Ce processus d'apprentissage est expliqué dans cette section, et son intérêt pour l'analyse des données d'EA à la section 3.3.

Cette section fait appel à l'article de K. Murphy [130], au livre [78, chap.10] et aux travaux de thèse de M.J. Beal [131] formalisant une version bayésienne du GMM. En second, nous proposons un algorithme de partitionnement partiellement supervisé combinant l'approche bayésienne et la TFC. Ce modèle avec *a priori* sur les *clusters* permet une estimation automatique du nombre de *clusters* et la prise en compte des *a priori* fournis par un expert lors du partitionnement.

### 3.2.1 Lois de distribution des variables

Pour le VB-GMM, les paramètres du modèle  $\theta = (\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  sont des valeurs incertaines qui suivent des lois de distribution de forme prédéfinie.

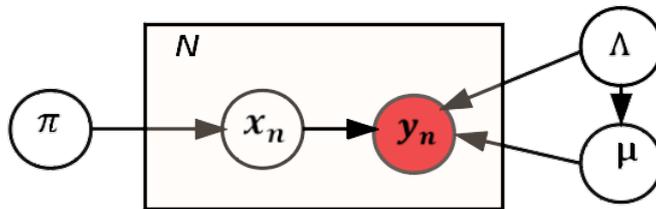


Figure 3.2 – Graphe acyclique dirigé représentant le VB-GMM | source : [78, Chap. 10]

Le graphe acyclique dirigé (Figure 3.2) permet de visualiser le sens des probabilités conditionnelles. Chaque observation  $y_n$  (avec  $n = \{1, 2, \dots, N\}$ ) suit une loi conditionnelle à la valeur prise par un état  $x_n$ , qui est un vecteur de  $K$  valeurs binaires, représentant les *clusters*. La variable  $\boldsymbol{\pi}$  représente les proportions incertaines dans chaque *cluster*. Les paramètres des *clusters*  $\boldsymbol{\theta} = \{\theta_k\} = \{(\pi_k, \mu_k, \boldsymbol{\Lambda}_k)\}, k = 1 \dots K$  sont représentés par des lois de distribution, impactant ainsi la probabilité des observations. La zone à l'intérieur du cadre dans la Figure 3.2 correspond à un GMM standard.

À partir de la Figure 3.2, on peut déduire la loi conjointe sur toutes les variables incertaines du modèle :

$$p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\mathbf{x} | \boldsymbol{\pi})p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\boldsymbol{\pi})p(\boldsymbol{\mu} | \boldsymbol{\Lambda})p(\boldsymbol{\Lambda}) \quad (3.12)$$

Les distributions  $p(\boldsymbol{\pi})$  et  $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  sont des lois *a posteriori* estimées par la loi de Bayes. Pour les distributions  $p(\mathbf{x} | \boldsymbol{\pi})$  et  $p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ , qui suivent une loi multinomiale et des lois gaussiennes respectivement, les probabilités des paramètres sont de la même forme que celles-ci. Les distributions  $p(\boldsymbol{\pi})$  et  $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  sont dites : **conjuguées**. Le tableau 3.2 regroupe les distributions  $p(\mathbf{x} | \boldsymbol{\pi})$  et  $p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ , ainsi que les distributions conjuguées  $p(\boldsymbol{\pi})$  et  $p(\boldsymbol{\mu} | \boldsymbol{\Lambda})p(\boldsymbol{\Lambda})$  [130].

Tableau 3.2 – Lois de distribution d'un modèle bayésien variationnel (VB-GMM)

Variabes	Distributions	Distributions conjuguées
Proportions : $\boldsymbol{\pi}$	Multinomial $\frac{N_k}{N}$	$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi}   \boldsymbol{\alpha}_0)$ est une distribution de Dirichlet
Composantes	Gaussiennes $\mathcal{N}_k(\mathbf{y}   \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$	$p(\boldsymbol{\mu}   \boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\mu}_k   \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1})$ , $p(\boldsymbol{\Lambda}) = \mathcal{W}(\boldsymbol{\Lambda}_k   \mathbf{W}_0, \nu_0)$ distribution Gaussienne-Wishart

Les proportions  $\boldsymbol{\pi}$  du mélange suivent une loi multinomiale et les *clusters* restent des gaussiennes  $\mathcal{N}_k(\mathbf{y} | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$  (voir Chapitre 1.2.2). Ces probabilités sont données par les équations 3.13 et 3.14, il s'agit d'une estimation sur l'ensemble des données :

$$p(\mathbf{x} | \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{k=1}^K \pi_k^{x_{nk}} \quad (3.13)$$

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{n=1}^N \prod_{k=1}^K \left[ \mathcal{N}_k(y_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}) \right]^{x_{nk}} \quad (3.14)$$

$x_{nk} = 1$  si l'observation  $y_n$  appartient au  $k^{\text{ième}}$  cluster, et 0 sinon. Si  $x_{nk} = 0$ , alors  $\left[ \mathcal{N}_k(y_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}) \right]^{x_{nk}} = 1$  et  $\pi_k^{x_{nk}} = 1$ , ainsi ces probabilités ne prennent en compte que les *clusters* actifs.

La distribution sur les paramètres est un *a priori* conjugué donné par la règle de Bayes. Par conséquent, l'incertitude sur les paramètres d'une variable multinomiale est exprimée par la distribution de Dirichlet :

$$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}_0) = C(\boldsymbol{\alpha}_0) \prod_{k=1}^K \pi_k^{\alpha_0 - 1} \quad (3.15)$$

où  $\alpha_0 > 0$  est la constante de Dirichlet. Elle représente le nombre de fois où chaque composante  $\pi_k$  a été observée. Plus elle est grande ( $\alpha_0 \rightarrow \text{inf}$ ), plus le modèle tendra vers une équiprobabilité d'observation des composantes ( $\mathbb{E}(\pi_k) = 1/K$ ). Plus cette valeur est faible, plus le modèle se concentrera uniquement sur les données observées.  $C(\boldsymbol{\alpha}_0)$  est une constante de normalisation.

De même, l'incertitude sur la moyenne et la précision d'une gaussienne est une distribution Gaussienne-Wishart :

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &= p(\boldsymbol{\mu} \mid \boldsymbol{\Lambda})p(\boldsymbol{\Lambda}) \\ &= \prod_{k=1}^K \mathcal{N}(\boldsymbol{\mu}_k \mid \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k \mid \mathbf{W}_0, \nu_0) \end{aligned} \quad (3.16)$$

Dans cette expression,  $\mathbf{W}_0$  est une matrice de mise à l'échelle définie positive  $[D \times D]$  généralement initialisée par une matrice dont tous les coefficients sont égaux à 1.  $\nu_0$  est un *a priori* sur le nombre de degrés de liberté de la distribution ( $\nu_0 > d - 1$ , par défaut  $\nu_0 = d$ ). Le paramètre  $\mathbf{m}_0$  est un vecteur de taille  $K$ , *a priori* sur l'emplacement des moyennes (généralement fixé à 0 partout, pour un souci de symétrie) et  $\beta_0$ , souvent égal à 1 est un facteur d'échelle des matrices de précision  $\boldsymbol{\Lambda}$ .

### 3.2.2 Définition de la fonction objectif

Le partitionnement a pour principal objectif, de trouver le paramétrage  $\theta^* = (\boldsymbol{\pi}^*, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$  qui maximise la vraisemblance  $p(\mathbf{y}, \theta)$  : la probabilité que les observations  $\mathbf{y}$  aient été générées par le modèle de paramètres  $\theta$ . Dans les modèles précédents, on a maximisé la loi marginale pour chaque observation  $p(y_n) = \sum_{\theta} p(\mathbf{y}_n, \theta_k)$  jusqu'à l'optimisation de  $p(\mathbf{y}) = \prod_n p(y_n)$ .

Pour le VB-GMM, les paramètres incertains  $\{\theta_k\}$  sont des distributions, ce qui nécessite le calcul d'une intégrale sur toutes les valeurs possibles de ces paramètres :

$$p(\mathbf{y}) = \prod_n p(y_n) = \prod_n \int_{\theta} d\theta p(y_n, \theta) \quad (3.17)$$

L'estimation de l'équation 3.17 et sa maximisation sur l'ensemble des données exigent des calculs complexes prenant en compte toutes les distributions intervenant dans la loi conjointe (équation 3.12). Une solution analytique dans le cas général serait coûteuse en temps de calcul et difficile à estimer.

Une solution consiste à approximer la loi *a posteriori*  $p(\mathbf{x}, \theta \mid \mathbf{y})$ . On introduit une variable latente  $\mathbf{z} = (\mathbf{x}, \theta)$  qui englobe  $\mathbf{x}$  (les *clusters*) et les paramètres  $\theta$ . La loi de probabilité  $p(\mathbf{y})$  peut être ré-écrite en fonction de la loi *a posteriori*  $p(\mathbf{z} \mid \mathbf{y})$  en utilisant la règle d'inversion de Bayes :

$$p(\mathbf{y}) = \frac{p(\mathbf{z}, \mathbf{y})}{p(\mathbf{z} \mid \mathbf{y})} \quad (3.18)$$

En introduisant le logarithme de part et d'autre de l'équation 3.18, on obtient :

$$\log p(\mathbf{y}) = \log \left[ \frac{p(\mathbf{y}, \mathbf{z})}{q(\mathbf{z})} \frac{q(\mathbf{z})}{p(\mathbf{z} \mid \mathbf{y})} \right] \quad (3.19a)$$

$$= \log \left[ \frac{p(\mathbf{y}, \mathbf{z})}{q(\mathbf{z})} \right] + \log \left[ \frac{q(\mathbf{z})}{p(\mathbf{z} \mid \mathbf{y})} \right] \quad (3.19b)$$

où  $q(\mathbf{z}) = q(\mathbf{x}, \theta)$  est une distribution vérifiant  $\int_{\mathbf{z}} q(\mathbf{z}) = 1$ .

L'espérance de  $\log p(\mathbf{y})$  est donnée par :

$$\mathbb{E}_{q(\mathbf{z})} [\log p(\mathbf{y})] = \log p(\mathbf{y}) = \underbrace{\int_{\mathbf{z}} q(\mathbf{z}) \log \left[ \frac{p(\mathbf{y}, \mathbf{z})}{q(\mathbf{z})} \right] d\mathbf{z}}_{\mathcal{L}(q)} + \underbrace{\int_{\mathbf{z}} q(\mathbf{z}) \log \left[ \frac{q(\mathbf{z})}{p(\mathbf{z}|\mathbf{y})} \right] d\mathbf{z}}_{KL(q||p)} \quad (3.20)$$

$\mathcal{L}(q)$  est appelée énergie libre et  $KL(q||p)$  représente la divergence de Kullback-Leibler, introduite au chapitre 2. Les deux termes sont des fonctions de  $q(\mathbf{z})$  [130]. Une propriété de la KL est qu'elle est toujours positive ( $KL(q||p) \geq 0$ ). Cette propriété fait de  $\mathcal{L}(q)$  un minorant de  $\log p(\mathbf{y})$ .

$$\log p(\mathbf{y}) \geq \mathcal{L}(q) = \int_{\mathbf{z}} q(\mathbf{z}) \log \left[ \frac{p(\mathbf{y}, \mathbf{z})}{q(\mathbf{z})} \right] d\mathbf{z} \quad (3.21)$$

Ainsi, si on maximise  $\mathcal{L}(q)$  (au lieu de  $\log p(\mathbf{y})$ , difficile à réaliser),  $\log p(\mathbf{y})$  ne décroîtra pas. Au mieux elle augmente, au pire elle stagne, mais la vraisemblance ne sera pas détériorée. Cette valeur optimale de  $\mathcal{L}(q)$  est atteinte lorsque  $KL(q||p) = 0$  (cas où  $q(\mathbf{z}) = p(\mathbf{z}|\mathbf{y})$ ) [78, chap.10]. En remplaçant  $q(\mathbf{z}) = p(\mathbf{z}|\mathbf{y})$  dans l'équation 3.20, on obtient  $\mathcal{L}(q) = \log p(\mathbf{y})$ .

### 3.2.3 Inférence approchée et algorithme VB-EM

Lorsque les paramètres sont incertains, estimer la loi *a posteriori*  $p(\mathbf{z}|\mathbf{y}) = p(\mathbf{x}, \theta|\mathbf{y})$  est difficile dans le cas général car elle dépend d'une intégrale sur  $\theta$ . La **théorie des champs moyens** permet de trouver une approximation  $q(\mathbf{z}) = q(\mathbf{x}, \theta)$  dans un espace de fonctions où la complexité des calculs diminue. L'hypothèse simplificatrice est **l'indépendance des lois de distribution des termes incorporés dans  $\mathbf{z}$** . Ainsi la loi *a posteriori* est décomposée en facteurs indépendants qui peuvent être estimés analytiquement. Dans le cas du VB-GMM qui nous intéresse, on aura  $q_{\mathbf{x}}(\mathbf{x})$ , le facteur relatif aux *clusters* et  $q_{\theta}(\theta)$ , le facteur relatif aux paramètres des *clusters*.

$$q(\mathbf{x}, \theta) = q_{\mathbf{x}}(\mathbf{x})q_{\theta}(\theta) , \quad \text{avec } q_{\theta}(\theta) = q_{\pi}(\boldsymbol{\pi})q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \quad (3.22)$$

**Propriété 3.2** Il a été prouvé dans [130, 131] et [78, chap.10] que maximiser l'espérance de la vraisemblance  $\mathcal{L}(q)$  revient à choisir  $q_{\mathbf{x}}$  et  $q_{\theta}$  tels que :

$$\log q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) = \mathbb{E}_{\theta} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \quad (3.23)$$

$$\log q_{\theta}^{(t+1)}(\theta) = \mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \quad (3.24)$$

où  $t + 1$  est l'itération suivante.  $\mathbb{E}_{\theta} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})]$  désigne l'espérance en utilisant la distribution  $q_{\mathbf{x}}^{(t)}(\mathbf{x})$ .  $\mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})]$  est l'espérance calculée en utilisant  $q_{\theta}^{(t)}(\theta)$  calculé à partir des paramètres estimés à l'étape  $E$  et  $c$  est une constante de normalisation. L'expression de  $p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  est donnée par l'équation 3.12.

La convergence de la loi *a posteriori*  $q_\theta(\theta)$  (équation 3.24), est évaluée à chaque itération de l'algorithme VB-EM (algorithme 2).

---

**Algorithme 2 : Algorithme VB-EM** | C'est un algorithme itératif en deux étapes postérieures à l'initiation de  $q^{(0)}$ . La première étape VB-E vise à estimer l'espérance des états  $\mathbf{x}$ . La seconde VB-M se sert des espérances estimées à l'étape VB-E pour réajuster le modèle.

---

```

1 Entrée :  $K_{max}$ , le nombre maximal de clusters;
2 Entrée : Les sous-ensembles de descripteurs  $\{\mathcal{S}\}$ ;
3 Sortie : La matrice  $R_{nk} = \mathbb{E}[x_{nk}]$  contenant les espérances des  $x_{nk}$  (initialisés
aléatoirement);
4 Sortie :  $\mathcal{L}(q)$ , la log-vraisemblance;
5 début
6    $t \leftarrow 0$  ;
7   //Initialiser les paramètres des distribution  $(\alpha_0, \beta_0, m_0, W_0, \nu_0)$ ;
8    $q^{(0)} =$  Initialisation;
9   tant que  $(t \leq t_{max})$  et  $(\log q^{(t+1)}(\theta) - \log q^{(t)}(\theta) > tol)$  faire
10     //Étape VB-E : on estime l'espérance des états  $\mathbf{x}$  ;
11     
$$\log q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) = \mathbb{E}_\theta [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \tag{3.23}$$

12     //Étape VB-M : on réajuste le modèle  $\theta = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}\}$  ;
13     
$$\log q_\theta^{(t+1)}(\theta) = \mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \tag{3.24}$$

14      $t \leftarrow t + 1$  ;
15   fin
16 // Estimer  $\mathcal{L}(q)$  par l'équation A.29;
17 fin

```

---

### a) L'étape VB-E sans a priori sur les clusters

L'étape VB-E (Variational bayesian Expectation) a pour but l'estimation des espérances  $\mathbb{E}[x_{nk}]$ , des *clusters*  $\mathbf{x}_n$  pour chaque observation  $y_n$ . Le produit de ces espérances sur l'ensemble des observations permet d'estimer  $q_{\mathbf{x}}(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$ . À partir des équations 3.23 et 3.12, en remplaçant  $\theta$  par  $(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ , on peut ré-écrire :

$$\log q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) = \mathbb{E}_{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \tag{3.25a}$$

$$= \mathbb{E}_{\boldsymbol{\pi}} [\log p(\mathbf{x} | \boldsymbol{\pi})] + \mathbb{E}_{\boldsymbol{\mu}, \boldsymbol{\Lambda}} [\log p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \tag{3.25b}$$

où  $c$  est une constante indépendante de  $\mathbf{x}$ . L'expression précédente peut être développée comme démontré dans [78, pages.476-478]. En remplaçant respectivement  $p(\mathbf{x}|\boldsymbol{\pi})$  et  $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  par leurs

expressions dans les équations 3.13 et 3.14. On obtient :

$$\log q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) = \sum_{n=1}^N \sum_{k=1}^K x_{nk} \log \rho_{nk} + c \quad (3.26)$$

où l'expression de  $\rho_{nk}$  est donnée par :

$$\begin{aligned} \log \rho_{nk} = & \mathbb{E} [\log \pi_k] + \frac{1}{2} \mathbb{E} [\log |\mathbf{\Lambda}_k|] - \frac{D}{2} \log(2\pi) \\ & - \frac{1}{2} \mathbb{E}_{\boldsymbol{\mu}_k, \mathbf{\Lambda}_k} [(\mathbf{y}_n - \boldsymbol{\mu}_k)^\top \mathbf{\Lambda}_k (\mathbf{y}_n - \boldsymbol{\mu}_k)] \end{aligned} \quad (3.27)$$

Dans l'équation 3.27,  $D$  est la dimension des données,  $(\boldsymbol{\mu}_k, \mathbf{\Lambda}_k)$ , les moyennes et dispersions des *clusters*,  $\forall k \in [1, K]$ . Ces paramètres sont mis à jour à l'étape de maximisation (VB-M). Les espérances de ces différents paramètres sont déduites des équations 3.28 :

$$\mathbb{E} [\log \pi_k] = \log \tilde{\pi}_k = \psi(\alpha_k) - \psi\left(\sum_{k=1}^K \alpha_k\right) \quad (3.28a)$$

$$\mathbb{E} [\log |\mathbf{\Lambda}_k|] = \log |\tilde{\mathbf{\Lambda}}_k| = \sum_{i=1}^D \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D \log 2 + \log |\mathbf{W}_k| \quad (3.28b)$$

$$\mathbb{E}_{\boldsymbol{\mu}_k, \mathbf{\Lambda}_k} [(\mathbf{y}_n - \boldsymbol{\mu}_k)^\top \mathbf{\Lambda}_k (\mathbf{y}_n - \boldsymbol{\mu}_k)] = D\beta_k^{-1} + \nu_k (\mathbf{y}_n - \mathbf{m}_k)^\top \mathbf{W}_k (\mathbf{y}_n - \mathbf{m}_k) \quad (3.28c)$$

où  $\psi(a) = \frac{d}{da} \Gamma(a)$  est la fonction Gamma. Les valeurs  $(\alpha_k, \beta_k, \mathbf{W}_k, \mathbf{m}_k)$  sont les paramètres des distributions de Dirichlet et Gaussienne-Wishart décrivant la distribution des paramètres du modèle. Estimés à l'étape VB-M, ils sont fonction des valeurs *a priori*  $(\alpha_0, \beta_0, \mathbf{W}_0, \mathbf{m}_0)$  des équations 3.15 et 3.16; et de  $q_{\mathbf{x}}^{(t+1)}(\mathbf{x})$  estimé à l'itération précédente.

En remplaçant les termes de l'équation 3.27 par leurs expressions dans 3.28, puis en mettant en exponentielle les deux termes de l'équation, on déduit :

$$\mathbb{E} [x_{nk}] \propto \rho_{nk} = \tilde{\pi}_k |\tilde{\mathbf{\Lambda}}_k|^{1/2} \exp\left[-\frac{D}{2\beta_k} - \frac{\nu_k}{2} (\mathbf{y}_n - \mathbf{m}_k)^\top \mathbf{W}_k (\mathbf{y}_n - \mathbf{m}_k)\right] \quad (3.29)$$

De même  $q_{\mathbf{x}}^{(t+1)}(\mathbf{x})$  est déduit à partir de l'équation 3.26 :

$$q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) \propto \prod_{n=1}^N \prod_{k=1}^K \rho_{nk}^{x_{nk}} \quad (3.30)$$

Le symbole  $\propto$  signifie que les termes de part et d'autre sont équivalents à une constante près. En

normalisant la distribution  $\rho_{nk}$ , on aboutit à :

$$q_{\mathbf{x}}^{(t+1)}(\mathbf{x}) = \prod_{n=1}^N \prod_{k=1}^K r_{nk}^{x_{nk}} \quad (3.31a)$$

$$\text{avec } r_{nk} = \frac{\rho_{nk}}{\sum_{i=1}^K \rho_{ni}} \quad (3.31b)$$

$r_{nk} = \mathbb{E}[x_{nk}]$  est le degré d'appartenance de l'observation  $y_n$  au *cluster*  $C_k$ . La normalisation permet de présenter le résultat sous forme de probabilité :  $0 \leq r_{nk} \leq 1$  et  $\sum_{k=1}^K r_{nk} = 1$ . Ainsi  $x_{nk'} = 1$  si et seulement si  $k' = \underset{k}{\operatorname{argmax}}(r_{nk})$ .

## b) L'étape VB-M

Le but de l'étape de maximisation bayésienne variationnelle (VB-M) est la ré-estimation des paramètres  $\theta = \{\theta_k\} = \{\pi_k, \mu_k, \Lambda_k\}$  et de la loi *a posteriori* du modèle  $q_{\theta}^{(t+1)}(\theta)$  à partir de  $q_{\mathbf{x}}^{(t)}(\mathbf{x})$  estimé à l'étape VB-E. Cette nouvelle estimation met à jours les paramètres en tenant compte des  $r_{nk}$ , les nouvelles valeurs des *clusters*.

À partir des équations 3.22 et 3.24, on obtient :

$$\log q_{\theta}^{(t+1)}(\pi, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + c \quad (3.32a)$$

$$= \log q_{\pi}^{(t+1)}(\pi) + \log q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}^{(t+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \quad (3.32b)$$

La loi *a posteriori* est décomposée en termes liés aux proportions des mélanges  $\boldsymbol{\pi}$  et des caractéristiques  $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  des *clusters*.

$$\log q_{\pi}^{(t+1)}(\pi) = \log p(\boldsymbol{\pi}) + \mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}|\boldsymbol{\pi})] \quad (3.33a)$$

$$= (\alpha_0 - 1) \sum_{k=1}^K \log \pi_k + \sum_{n=1}^N \sum_{k=1}^K r_{nk} \log \pi_k + c \quad (3.33b)$$

L'équation 3.33 correspond à une distribution de Dirichlet,  $q_{\pi}^{(t+1)}(\pi) = \operatorname{Dir}(\pi \mid \alpha_1, \dots, \alpha_K)$  avec des poids des composants  $\alpha_k$  donnés par :

$$\alpha_k = \alpha_0 + N_k \quad (3.34)$$

Le nombre d'observations du  $k^{\text{ième}}$  *cluster*  $N_k = \sum_{n=1}^N r_{nk}$ , et les proportions des mélanges  $\pi_k = \frac{N_k}{N}$ , gardent les mêmes expressions que dans le GMM.

L'expression de  $q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}^{(t+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  est déduite des équations 3.14 et 3.16 :

$$\log q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}^{(t+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \sum_{k=1}^K \log p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) + \sum_{n=1}^N \sum_{k=1}^K r_{nk} \log \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}) \quad (3.35)$$

Cette probabilité *a posteriori*  $q_{\mu, \Lambda}^{(t+1)}(\mu, \Lambda)$ , correspond à une distribution gaussienne-Wishart avec les paramètres suivants [130], [78, pages.481-482] :

$$\beta_k = \beta_0 + N_k \quad (3.36a)$$

$$\mathbf{m}_k = \frac{1}{\beta_k} (\beta_0 \mathbf{m}_0 + N_k \bar{\mathbf{x}}_k) \quad (3.36b)$$

$$\mathbf{W}_k^{-1} = \mathbf{W}_0^{-1} + N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{\mathbf{x}}_k - \mathbf{m}_0) (\bar{\mathbf{x}}_k - \mathbf{m}_0)^T \quad (3.36c)$$

$$\nu_k = \nu_0 + N_k \quad (3.36d)$$

Les paramètres ainsi calculés sont utilisés à l'itération suivante dans l'étape VB-E. Les valeurs des espérances des *clusters*  $r_{nk}$  ne sont pas affectées pendant cette étape.

### 3.2.4 Étape VB-E avec *a priori* sur les *clusters*

Cette fois nous nous plaçons dans le cadre partiellement supervisé. On cherche à introduire l'avis d'un expert sous forme d'*a priori* sur les *clusters* lors du partitionnement. Comme vu à la section précédente l'*a priori* sur les *clusters* et les proportions estimées sont considérés comme deux masses de croyance. L'intérêt d'un tel modèle est la possibilité d'introduire des *a priori* sur les *clusters* tout en gardant le principal avantage du VB-GMM qui est l'estimation automatique du nombre de *clusters*. En fixant  $K$ , le nombre de *clusters* à une valeur supérieure au nombre de classes, on donne une possibilité au modèle de trouver des nouveaux *clusters*. Un tel modèle est dit ouvert.

Comme pour le GMM-SOFT de [124], l'apprentissage a été modifié pour incorporer un *a priori* sur les *clusters*, sous forme de fonctions de croyance  $\mathbf{m}_n$ . Les masses sont représentées par leurs plausibilités  $pl_{nk} = p(\mathbf{l}|\mathbf{x})$ , lors de l'estimation des combinaisons conjointes. Le VB-GMM avec *a priori* incertains sur les *clusters* sera désigné par **VB-GMM-SOFT** (*VB-GMM with soft-labels*).

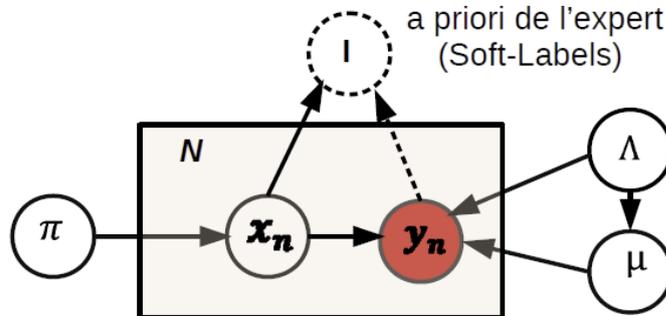


Figure 3.3 – Graphe acyclique dirigé représentant le VB-GMM avec des *a priori*

Le nouveau graphe acyclique est présenté par la Figure 3.3. On note la présence de l'*a priori* des

étiquettes  $\mathbf{l}$  donné par une expertise. Les *a priori*  $\mathbf{l}_n$  sur les *clusters* dépendent des observations et des *clusters* les ayant générés. La flèche en trait continu entre  $\mathbf{l}$  et  $\mathbf{x}$  montre qu'il existe une relation entre les deux qui se traduit par une loi conditionnelle  $p(\mathbf{l}|\mathbf{x})$ . Une nouvelle loi conjointe est ainsi définie par :

$$p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{l}) = p(\mathbf{x} | \boldsymbol{\pi})p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\boldsymbol{\pi})p(\boldsymbol{\mu} | \boldsymbol{\Lambda})p(\boldsymbol{\Lambda})p(\mathbf{l}|\mathbf{x}) \quad (3.37)$$

Dans l'équation 3.37, les distributions  $p(\mathbf{x} | \boldsymbol{\pi})$ ,  $p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ ,  $p(\boldsymbol{\pi})$ ,  $p(\boldsymbol{\mu} | \boldsymbol{\Lambda})$  et  $p(\boldsymbol{\Lambda})$  gardent le même sens et les mêmes expressions analytiques (équations 3.12). L'*a priori* sur les *clusters* de l'ensemble des observations est donné par :

$$p(\mathbf{l}|\mathbf{x}) = \prod_{n=1}^N \prod_{k=1}^K [pl_{nk}]^{x_{nk}} \quad (3.38)$$

Dans cette expression, les étiquettes sont supposées complètes, car le modèle est partiellement supervisé. Les lois *a posteriori* sur les *clusters*  $q_{\mathbf{x}}^{t+1}(\mathbf{x})$  et sur les paramètres  $q_{\theta}^{(t+1)}(\theta)$  sont estimées par la formulation variationnelle bayésienne.  $q_{\mathbf{x}}^{t+1}(\mathbf{x})$  est estimé à partir de l'espérance de  $p(\mathbf{x}|\boldsymbol{\pi})$ ,  $p(\mathbf{l}|\mathbf{x})$  et  $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  en utilisant  $q_{\theta}^{(t)}(\theta)$ . L'expression obtenue à partir de la loi conjointe 3.37 est :

$$\log q_{\mathbf{x}}^{t+1}(\mathbf{x}) = \mathbb{E}_{\boldsymbol{\pi}} [\log p(\mathbf{x} | \boldsymbol{\pi})] + \mathbb{E}_{\boldsymbol{\mu}, \boldsymbol{\Lambda}} [\log p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + \log p(\mathbf{l} | \mathbf{x}) + c \quad (3.39a)$$

$$= \sum_{n=1}^N \sum_{k=1}^K x_{nk} \log \rho_{nk} + \sum_{n=1}^N \sum_{k=1}^K x_{nk} \log pl_{nk} + c \quad (3.39b)$$

On obtient finalement :

$$\log q_{\mathbf{x}}^{t+1}(\mathbf{x}) = \sum_{n=1}^N \sum_{k=1}^K x_{nk} [\log \rho_{nk} + \log pl_{nk}] + c \quad (3.40)$$

Dans cette expression, on note la sommation des logarithmes des deux *a priori*  $\boldsymbol{\pi}$  et  $\mathbf{pl}$  donnés par le modèle et un expert. Ce qui signifie que :

$$q_{\mathbf{x}}^{t+1}(\mathbf{x}) \propto \rho_{nk} pl_{nk} \quad (3.41)$$

En utilisant les mêmes règles de normalisation l'équation 3.29 du VB-GMM se transforme en 3.42 pour le VB-GMM-Soft.

$$\tilde{r}_{nk} \propto pl_{nk} \tilde{\pi}_k |\tilde{\Lambda}_k|^{1/2} \exp \left[ -\frac{D}{2\beta_k} - \frac{\nu_k}{2} (\mathbf{y}_n - \mathbf{m}_k)^T \mathbf{W}_k (\mathbf{y}_n - \mathbf{m}_k) \right] \quad (3.42)$$

L'estimation des lois *a posteriori* des paramètres des *clusters* sont des fonctions de  $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ ,  $p(\boldsymbol{\pi})$  et  $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  en prenant  $q_{\mathbf{x}}$  comme distribution. Il est évident que ces termes sont indépendants des *a priori*  $pl_{nk}$  sur les *clusters*. Ainsi l'introduction des *a priori* sur les *clusters* modifie uniquement les expressions des termes calculés dans l'étape VB-E-SOFT. Les espérances estimées sont ensuite utilisées dans l'étape VBM. Les équations de l'étape VB-M-SOFT sont identiques à celles du VB-GMM

non supervisé :

$$\log q_{\theta}^{(t+1)}(\theta) = \mathbb{E}_{\mathbf{x}} [\log p(\mathbf{x}, \mathbf{y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) | \mathbf{y}, \mathbf{m}_n] + c.$$

**Proposition 3.2** *Étant donnée la FC a priori  $\mathbf{m}_n$  sur les clusters fournie par une expertise, et les probabilités sur les clusters  $\boldsymbol{\pi}$ , l'équation 3.26 devient :*

$$\log \tilde{\rho}_{nk} = \log \rho_{nk} + \log pl_{nk} \quad (3.43)$$

**Propriété 3.3** *Cas particuliers :*

- Si la plausibilité est constante  $pl_n(\omega_k) = 1, \forall n, \forall k$ , le modèle est non supervisé donc équivalent au VB-GMM.
- Lorsque la plausibilité est constante ( $pl_n(\omega_k) = 1, \forall n, \forall k$ ) et les paramètres certains, le modèle équivaut au GMM.
- Si  $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$  sont certains, l'équation 3.43 devient 3.11 qui correspond au GMM-SOFT de [124].

Ce modèle a les avantages du GMM-SOFT et du VB-GMM. Les propriétés 3.3 énumérées montrent également que les modèles GMM, GMM-SOFT et VB-GMM en sont des cas particuliers.

### 3.3 Usage du VB-GMM-SOFT pour l'analyse de l'EA

L'intérêt de l'algorithme proposé pour l'EA repose essentiellement sur la détermination du nombre de clusters et la possibilité d'ajouter des *a priori* venant d'un expert.

#### 3.3.1 Détermination du nombre de clusters

A chaque itération  $t$ , de l'algorithme VB-EM, les paramètres de Dirichlet  $\{\alpha_k\}$  pour différents *clusters* et les probabilités  $r_{nk} = \mathbb{E}[x_{nk}]$  sont estimés. Le nombre d'observations du  $k^{ieme}$  *cluster*  $N_k^{(t)}$  et l'espérance d'une proportion de mélange  $\mathbb{E}^{(t)}[\boldsymbol{\pi}_k]$  sont déduits pour chaque *cluster* à partir des équations :

$$\mathbb{E}^{(t)}[\boldsymbol{\pi}_k] = \frac{\alpha_k + N_k^{(t)}}{K\alpha_0 + N} \quad (3.44)$$

$$N_k^{(t)} = \sum_{n=1}^N r_{nk}^{(t)} \quad (3.45)$$

Lorsque la taille des données n'est pas importante, la valeur du paramètre de la distribution de Dirichlet  $\alpha_0$  a un impact significatif sur le résultat. Si  $\alpha_0 \gg 1$ , les *clusters* vides ( $N_k^{(t)} \simeq 0$ ) ont une espérance non nulle ( $\mathbb{E}^{(t)}[\boldsymbol{\pi}_k] \rightarrow \frac{\alpha_0}{K\alpha_0 + N}$ ). Ainsi toutes les composantes jouent un rôle dans le modèle. Utiliser  $\alpha_0 < 1$  favorise les solutions où  $N_k = 0$  pour plusieurs *clusters* [78, Chap. 10]. Il est suggéré d'utiliser un *a priori* non informatif ( $\alpha_0 \leq 1$ ). Selon les articles [132, 133], les valeurs couramment

utilisées dans la littérature sont :  $\alpha_0 = \frac{1}{K}$  (Perks, [134]),  $\alpha_0 = \frac{1}{2}$  (Jeffreys, [135]),  $\alpha_0 = 1$  (Bayes, [136]).

Dans ce manuscrit  $\alpha_0 = \frac{1}{K}$ , l'*a priori* de Perks est utilisé. Cette valeur est suffisamment petite pour limiter l'impact des *clusters* vides et suffisamment grande pour ne pas supprimer les *clusters* ayant peu d'observations, c'est celle proposée dans [132, 133]. De plus, les données utilisées sont volumineuses ( $N \gg K$ ), les résultats obtenus en utilisant les valeurs de  $\alpha_0$  proposées par Perks, Jeffreys et Bayes convergent vers des solutions proches [132].

Si  $\alpha_0$  est bien choisi, l'équation 3.44 montre que : si une composante n'a pas beaucoup d'éléments ( $N_k \simeq 0$ ), son espérance sera presque nulle  $\mathbb{E}[\boldsymbol{\pi}_k] \rightarrow 0$ . Les valeurs  $r_{nk}$  étant proportionnelles à  $\boldsymbol{\pi}_k$ , donc la composante n'aura aucun impact sur le modèle. Ainsi pour l'estimation VB-EM, une valeur élevée  $K = K_{max}$  correspondant au nombre maximal de *clusters* possibles, peut être attribuée au nombre de composantes. Le nombre de composantes avec des coefficients de mélange  $\pi_k \neq 0$  diminue au cours des itérations, jusqu'à ce que la log-vraisemblance converge. Le nombre final de *clusters* correspond au nombre de composantes du modèle ayant des proportions non nulles ( $\mathbb{E}[\boldsymbol{\pi}_k] \neq 0$ ) après convergence de l'algorithme.

Cet algorithme est intéressant pour l'interprétation des données d'EA dont le nombre de *clusters* est incertain. Cela permet de trouver le nombre de *clusters*  $K$  et les "*clusters* naturels" pour un jeu de données sans intervention d'un critère de validation.

### 3.3.2 Comment générer les *a priori*

Pour inclure des *a priori* sur les *clusters* lors de l'apprentissage bayésien variationnel, une matrice de plausibilité doit être fournie. Comme indiqué précédemment, sans *a priori*, chaque élément de cette matrice est égal à 1. Dans les cas où l'*a priori* sur les *clusters* est incertain ou imprécis, nous avons également vu comment exploiter les fonctions de croyance pour représenter et incorporer l'*a priori* dans le VB-GMM-SOFT.

Dans [123, 124], une méthode est proposée pour ajouter de l'incertitude sur les étiquettes en générant quelques fois des étiquettes aléatoires. Ceci est particulièrement utile lors de la validation expérimentale afin d'évaluer la robustesse d'un algorithme d'apprentissage semi-supervisé pour des étiquettes manquantes ou incertaines. La prise en compte des *a priori* bruités est un sujet étudié dans la communauté de l'apprentissage automatique pour des applications pratiques où l'étiquetage des données est difficile, conduisant à des étiquettes erronées, corrompues ou manquantes [137]. Avec l'intérêt récent de coupler l'apprentissage automatique avec la science des matériaux pour une meilleure interprétation des données d'émission acoustique, de telles étiquettes pourraient être obtenues.

La génération des plausibilités  $pl$  bruitées à partir d'étiquettes données est basée sur une méthode développée dans [124]. On définit  $\rho$  le niveau de doute moyen accordé à la connaissance *a priori*.

Lorsque  $\rho = 1$ , nous sommes dans un cas d'incertitude totale, ce qui équivaut à du non supervisé. Si  $\rho \rightarrow 0$ , l'information est très fiable, on tend vers une classification supervisée. On définit également  $v$ , la variance des informations *a priori* de l'expert.

Les plausibilités sont définies à partir d'une distribution bêta  $\beta(a, b)$  avec les paramètres  $a$  et  $b$  tels que

$$\begin{cases} a = \rho^2 \frac{(1-\rho)}{v} - \rho \\ b = \rho \frac{(1-\rho)}{v} - 1 - a \end{cases} \quad (3.46)$$

On définit  $Perr$  un vecteur de nombres aléatoires choisis parmi les distributions bêta avec les paramètres  $a$  et  $b$  pour chaque échantillon ( $n = 1 \dots N$ ) :

$$Perr = \beta(a, b) \quad (3.47)$$

La plausibilité est alors donnée par :

$$\begin{cases} pl_{nk} = \frac{Perr(n)}{K}; & \text{si } k \neq label(n) \\ pl_{nk} = 1 - Perr(n) + \frac{Perr(n)}{K}; & \text{sinon} \end{cases} \quad (3.48)$$

Cette estimation de la plausibilité  $pl$  permet d'avoir une équiprobabilité *états* lorsque la probabilité d'erreur est forte ( $\rho \rightarrow 1$ ). Une forte probabilité est attribuée à l'étiquette correspondante lorsque  $\rho \rightarrow 0$  (*a priori* quasi certain). On peut donc ajuster l'impact de la connaissance sur le modèle en ajustant la valeur de l'espérance de la probabilité d'erreur  $\rho$ .

La corruption des étiquettes a été retirée dans un premier temps. Ce cas correspond à un cas réel avec les étiquettes complètes, mais une incertitude sur la fiabilité de l'expertise.

Par la suite les étiquettes corrompues sont utilisées dans le but de couvrir les cas où certaines observations sont mal classées. On attribue alors des étiquettes aléatoires à certaines observations.

### 3.4 Résultats sur ORION-EA

Cette section regroupe les illustrations sur les données ORION-EA déjà présentées dans les chapitres 1 et 2. Les données ont été sous-échantillonnées : cela consiste à prendre une observation sur 10 sur les données. Ce sous-échantillonnage a un but : l'évaluation du maximum d'hyper-paramètres possible grâce au gain en temps de calcul. L'objectif est d'évaluer en premier la capacité du partitionnement bayésien à produire des résultats avec et sans *a priori incertains* sur les *clusters*. Par la suite l'influence des autres hyper-paramètres  $K_{max}, N, \alpha, d$  respectivement le nombre de clusters initial, le nombre d'observations, le paramètre de Dirichlet et le nombre de descripteurs par sous-ensemble sont

évalués.

Le nombre de *clusters* a été initialisé à  $K_{max} = 10$ , un nombre suffisamment grand par rapport au nombre de *clusters* potentiels. L'algorithme de partitionnement se charge de l'estimation du nombre de *clusters* optimal pour chaque sous-ensemble de descripteurs. Les *a priori* ont été générés en suivant la méthode présentée à la section précédente. Pour chacune des configurations, l'évaluation se fait à base des histogrammes des instants de déclenchement et le graphe des ARI triés avec des marqueurs pour les 10 meilleures partitions suivant le critère séquentiel défini au chapitre 2.

L'algorithme VB-GMM fournit une partition par sous-ensemble, ce qui fait un total de 715 partitions à trier. L'entropie de Shannon normalisée (équation 2.13) permet de trier les meilleures partitions indépendamment du nombre de *clusters* trouvé. Après le tri, seule la première moitié des partitions a été conservée. Pour cette première moitié, les 10 meilleures partitions à  $K \in [2, K_{max}]$  *clusters*, si elles existent sont sélectionnées pour le tracé des histogrammes des instants de déclenchement des *clusters*. Ce qui fait un maximum de  $((K_{max} - 1) \times 10) = 90$  partitions par graphe d'histogrammes. Les résultats montrent qu'il y a en général la convergence du nombre de *clusters* vers un nombre de *clusters* inférieur à  $K_{max}$ .

### 3.4.1 Influence de la probabilité d'erreur $\rho$

Pour chacun des essais nous avons choisi de faire varier la probabilité d'erreur  $\rho = \{1; 0.9; 0.6; 0.3\}$ . Lorsque  $\rho = 1$  l'information portée par les étiquettes est totalement incertaine, le modèle équivaut à un VB-GMM. Le choix de ces valeurs permet d'établir une évaluation en présence d'*a priori* absents, faibles, moyens et forts. Les Figures 3.4, 3.6, 3.7, 3.8 et 3.9 représentent respectivement les résultats obtenus pour les essais #B, #C, #D, #E et #F. Ces résultats sont similaires en terme d'histogrammes des instants de déclenchement des *clusters* et en terme d'ARI (qualité du partitionnement).

**Essais #B :** Sur le graphe 3.4a de l'essai #B **sans *a priori***, des pics élevés sont observés à 0s, 10s, 20s, 40s, 50s. Ces instants correspondent aux changements de niveaux de serrage. Le VB-GMM sans *a priori* sur les *clusters*, parvient à détecter ces évènements en mettant en évidence tous les niveaux de serrage. Des évènements supplémentaires sont détectés à 1s, 5s, 12s, 17s, 55s, comme au chapitre 2 certains serrages sont constitués de deux ou plusieurs *clusters*. On constate aussi qu'à 60s le pic est assez bas, ce qui signifie que les sous-ensembles de descripteurs sélectionnés ne permettent pas de discriminer le serrage le plus faible. Un second constat, seules 47 partitions ont été sélectionnées en faisant varier le nombre de *clusters*, ce qui signifie qu'il y a moins de 10 partitions à  $K$  *clusters* présélectionnées pour certaines valeurs de  $K \in [K_{min}, K_{max}]$ . Cette observation est due à deux raisons dont la première est la convergence de l'algorithme VB-GMM. En effet le nombre de *clusters* estimé est inférieur à 10 pour bon nombre de sous-ensembles de descripteurs. La seconde raison est que le tri a

mis en évidence les partitions avec un nombre de clusters se rapprochant de 7 qui est la valeur attendue.

En ajoutant peu d'*a priori* (Figure 3.4b,  $\rho = 0.9$ ), la plupart des évènements détectés entre les niveaux de serrage disparaissent. En effet les clusters supplémentaires (pour la plupart avec très peu d'échantillons) sont majoritairement supprimés. Les *a priori* agissent également en subdivisant les *clusters* qui correspondent à 2 ou plusieurs niveaux de serrage. La quasi-totalité des partitions de la première moitié compte entre 6 et 8 clusters. En sélectionnant les 10 premières partitions lorsqu'elles existent, seules 20 partitions sont utilisées pour le graphe des histogrammes. Les mêmes observations sont faites plus on augmente l'influence des *a priori* (Figure 3.4c). Dans ces 2 configurations ( $\rho = \{0.9; 0.6\}$ ), le nombre de clusters varie entre 6 et 8. Il est aussi à noter que les partitions à plus de 7 *clusters* contiennent des *clusters* presque vides (1 à 3 observations). Sur la Figure 3.4d (*a priori fort*), quelques pics sont observés à des instants ne correspondant pas aux changements de niveaux de serrage (pour  $\rho = 0.6$ ), ce qui montre que même en présence d'un fort *a priori*, l'algorithme VB-GMM-SOFT parvient à maintenir un résultat différent si les  $pl_{nk}$  sont en totale contradiction avec les données. Quand  $\rho = 0.3$ , toute la première moitié du tri est constitué des partitions à sept *clusters*, ce qui conduit à 10 partitions avec les instants de déclenchement correspondant aux serrages.

Le cinquième graphe (Figure 3.4e) représente les valeurs triées fournies par l'ARI entre la vérité terrain et le partitionnement (une valeur de 1 signifie un partitionnement parfait) selon le sous-ensemble de descripteurs (il y a 715 partitions). Les courbes correspondent aux quatre configurations  $\rho = \{1, 0.9, 0.6, 0.3\}$ . L'augmentation de la quantité d'*a priori* (diminution de la probabilité d'erreur  $\rho$ ) conduit logiquement à une amélioration du partitionnement. Selon la quantité d'*a priori*, l'évolution des valeurs de l'ARI diffère sensiblement. Sans *a priori* (courbe bleue), l'ARI diminue de manière importante, alors que pour 0.6 et 0.3 (jaune et violet) l'ARI reste assez stable. Cela signifie que les *a priori* rendent les résultats robustes par rapport au choix des sous-ensembles de descripteurs. Par ailleurs, on peut observer que, sans *a priori* (courbe bleue), les performances du VB-GMM sont assez hautes, avec un ARI supérieur à 0.6 pour  $\frac{1}{4}$  des partitions. Sur ces courbes on constate également que les cercles correspondant aux 10 premières partitions du tri sont bien du côté gauche de la figure (ARI élevés). Globalement, on voit que la méthode de partitionnement est pertinente car les caractéristiques sélectionnées conduisent à des ARI élevés, pour différentes valeurs de  $\rho$ .

Les occurrences cumulées de la première partition du tri ont été tracées dans les figures 3.5, pour différentes valeurs de  $\rho$  en complément aux histogrammes. Lorsqu'on n'a aucun *a priori* (figure 3.5a), la partition illustrée présente 6 *clusters* estimés par le VB-GMM. Le premier cluster correspond au premier serrage. Le second est détecté de manière précoce, une centaine de points du premier et du second serrages lui sont attribués, mais il correspond au quatrième serrage débutant à 30s. Aucune détection n'est faite à 30s, ceci explique la présence d'un pic à 5s sur l'histogramme 3.4a et un pic plus bas à 30s comparé aux autres bornes détectées. Les troisième, quatrième et cinquième serrage

correspondent chacun à un serrage. Le sixième *cluster* combine les deux derniers serrages. En ajoutant un peu d'*a priori* (figure 3.5b), la centaine de points mal attribués initialement au quatrième serrage retrouvent leurs serrages respectifs, ce qui a pour conséquence l'amélioration des ARI (figure 3.4e). L'influence des *a priori* n'est cependant pas encore suffisant pour discriminer les deux derniers serrages. Pour  $\rho = 0.6$ , l'*a priori* devient suffisamment fort pour séparer ces serrages. Cette observation n'est valable que pour quelques sous-ensembles, pour la plupart cet *a priori* n'est pas suffisant, ce qui justifie le pic assez bas à 60s (figure 3.4d). Ce n'est qu'en présence d'*a priori* très fort ( $\rho = 0.3$ ) que les *clusters* correspondent aux serrages, ce qui se traduit par une asymptote quasi-horizontale sur le graphe des ARI (figure 3.4e).

**Essai #C :** Sans *a priori* (Figure 3.6a), seul le dernier serrage à 50s n'est pas mis en évidence. Peu d'*a priori* améliore la détection des changements de niveaux, tout en mettant moins en évidence les détections entre les bornes. Le graphe des ARI montre que les résultats sont globalement bons, la moitié des partitions rend compte de valeurs supérieures à 0.5.

**Essai #D :** La figure 3.7 montre des résultats corrects concernant les histogrammes sans *a priori* ( $\rho = 1$ ). Les pics correspondant aux niveaux de serrage sont bien mis en évidence. En ajoutant peu d'*a priori*, les pics à 10 et 50s sont mieux mis en évidence. Plus l'*a priori* augmente, plus la hauteur des pics s'uniformise, ce qui signifie que les partitions sélectionnées détectent les mêmes instants de déclenchements. Il y a de moins en moins des déclenchements à des instants localisés entre deux niveaux de serrage en augmentant l'*a priori*. Le seul point à souligner est que même en mettant des *a priori* forts ( $\rho = 0.3$ ) il n'y a pas de détection à 60s. Une explication est que certaines observations du dernier serrage sont toujours mal attribuées malgré les *a priori* forts, ce qui explique aussi la présence de quelques déclenchements entre les serrages. Globalement en observant les ARI, on constate que le partitionnement avec le VB-GMM produit déjà des résultats corrects, les *a priori* améliorent les performances.

**Essai #E :** Illustré à la figure 3.8, l'essai E produit des résultats similaires à ceux précédemment présentés. Tous les serrages sont mis en évidence. Deux nuances sont à noter, d'une part, plusieurs évènements sont détectés, et d'autre part, les données sont un peu moins sensibles aux *a priori*, ce qui crée des pics hauts entre les bornes, ce même en présence des *a priori* forts. On obtient exactement 20 partitions de 6 à 8 *clusters* pour  $\rho = 0.3$ .

**Essai #F :** Cet essai produit des histogrammes mettant en évidence tous les changements de niveaux de serrage. Il y a cependant plusieurs autres évènements détectés pendant le premier serrage, ceux-ci disparaissent quand on ajoute des *a priori*. Avec  $\rho = 0.3$ , on obtient 22 partitions dont la quasi-totalité des instants de déclenchements des *clusters* correspondent aux niveaux de serrages attendus. Plus l'*a priori* augmente, plus élevées sont les valeurs de ARI correspondant aux meilleures partitions.

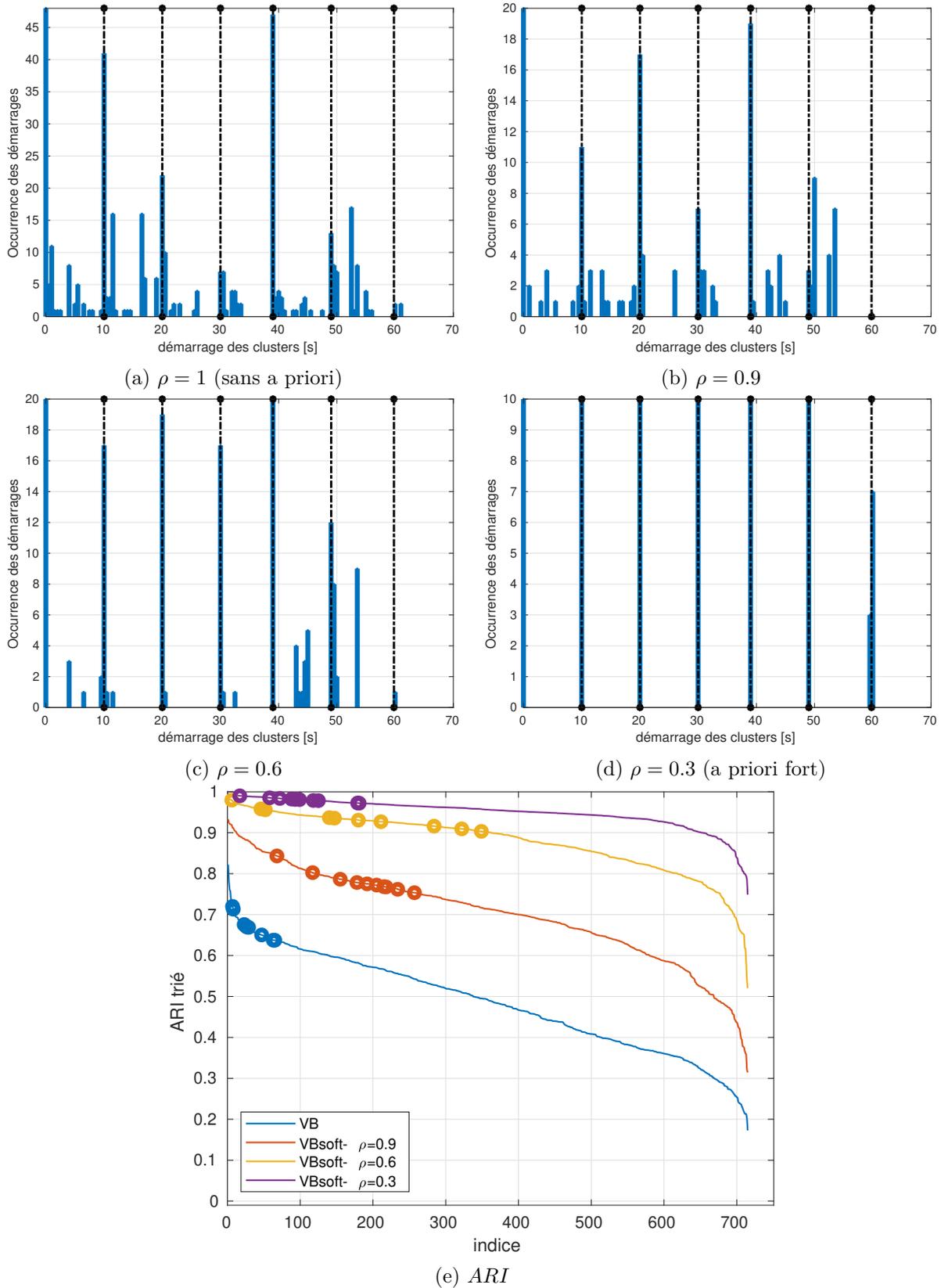


Figure 3.4 – Essai #B : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$ . Les histogrammes montrent que sans a priori les performances sont très satisfaisantes (pics correspondant aux changements de niveaux de serrage), et que l'ajout des a priori permet de normaliser la hauteur des pics.

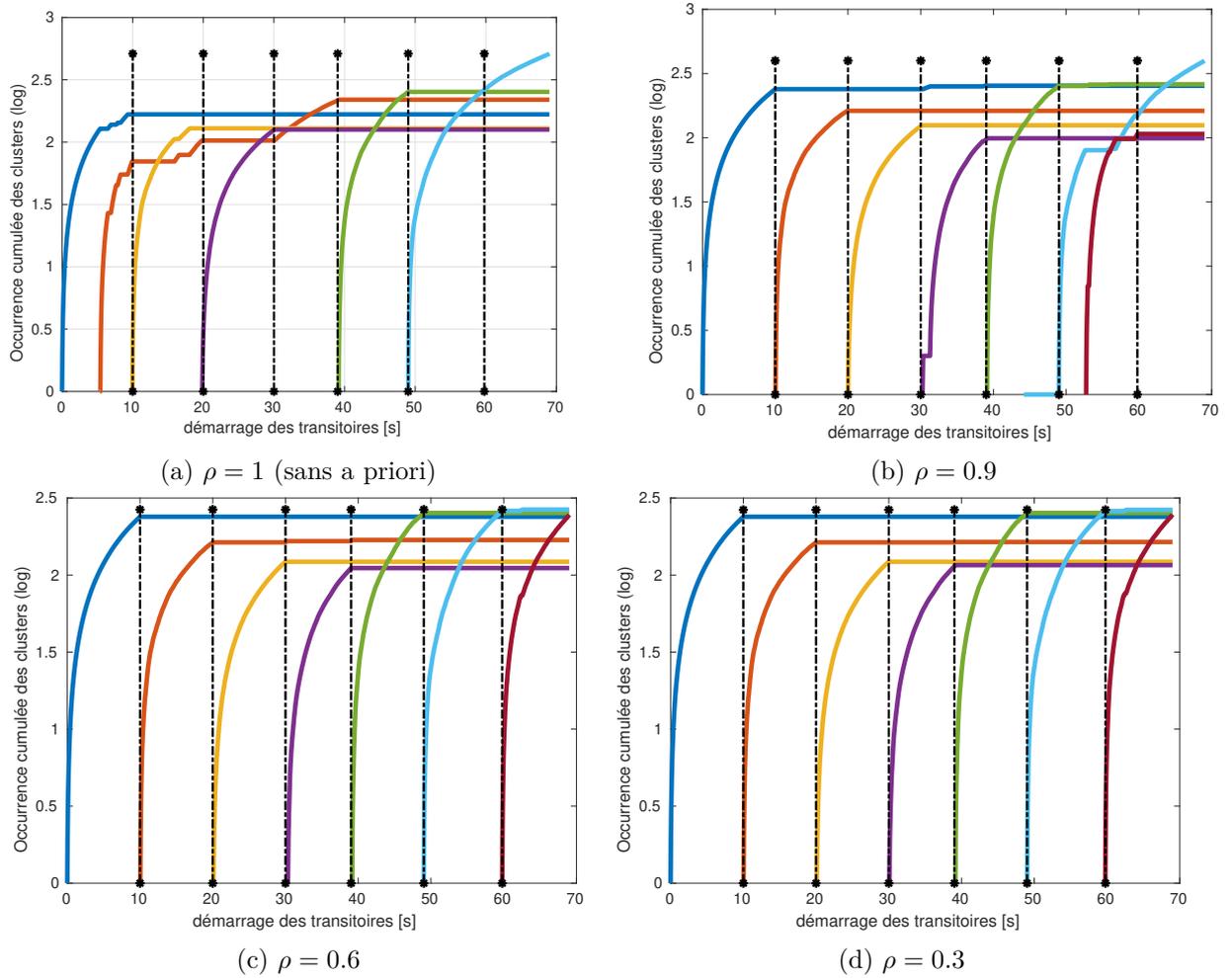


Figure 3.5 – **Essai #B** : Cumulée des occurrences des *clusters* de la première partition du tri pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$

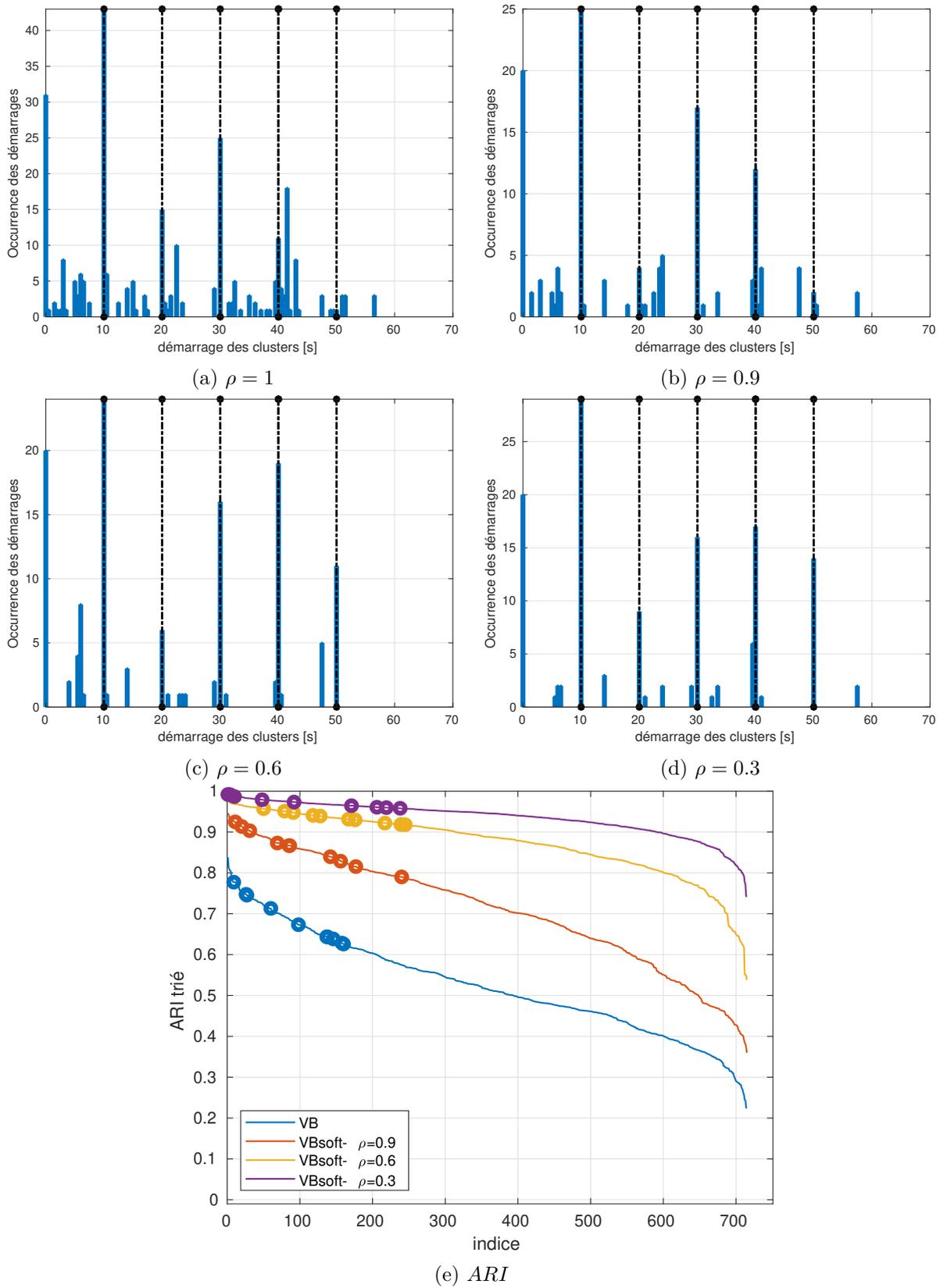


Figure 3.6 – Essai #C : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$

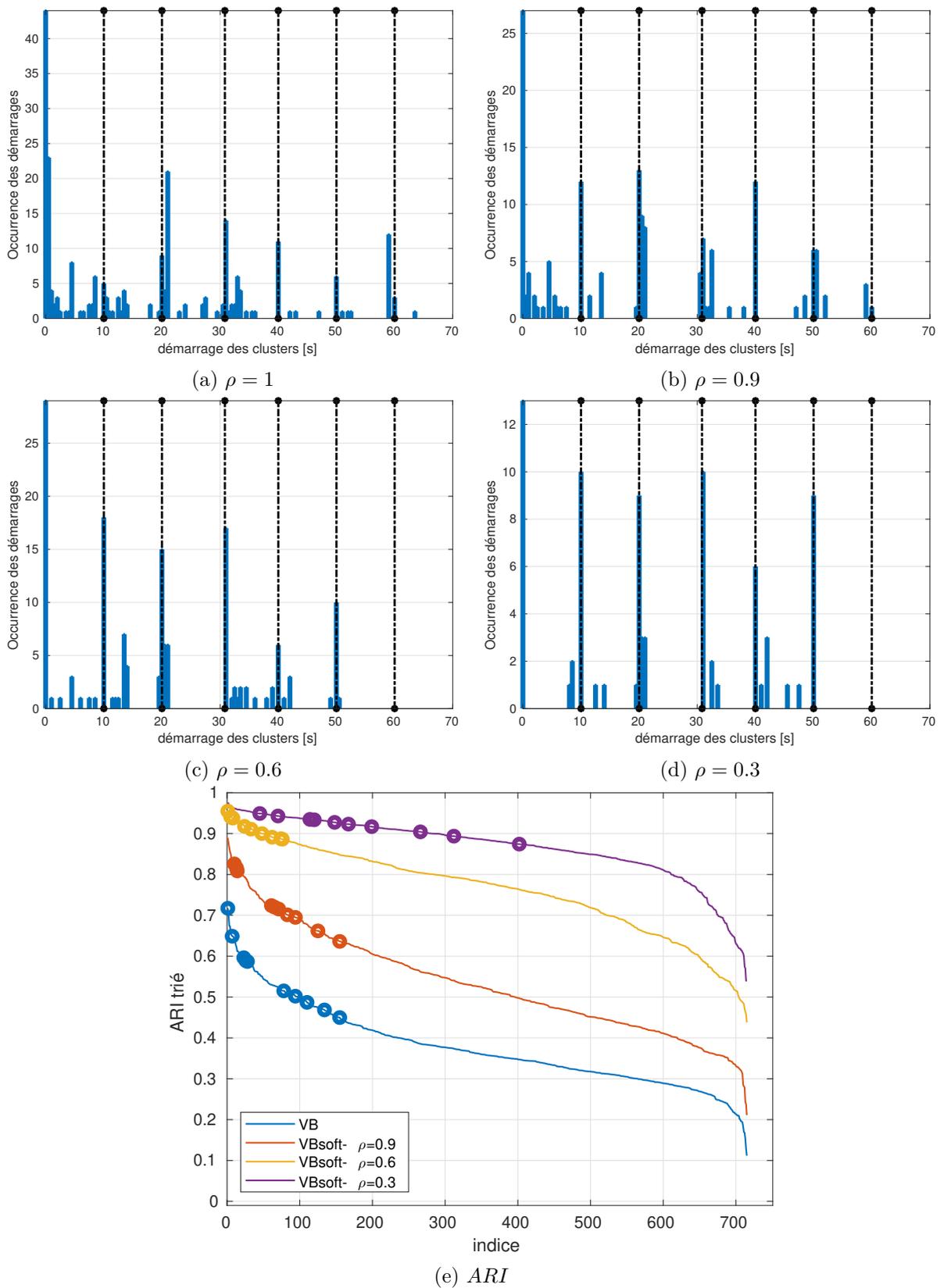


Figure 3.7 – Essai #D : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$

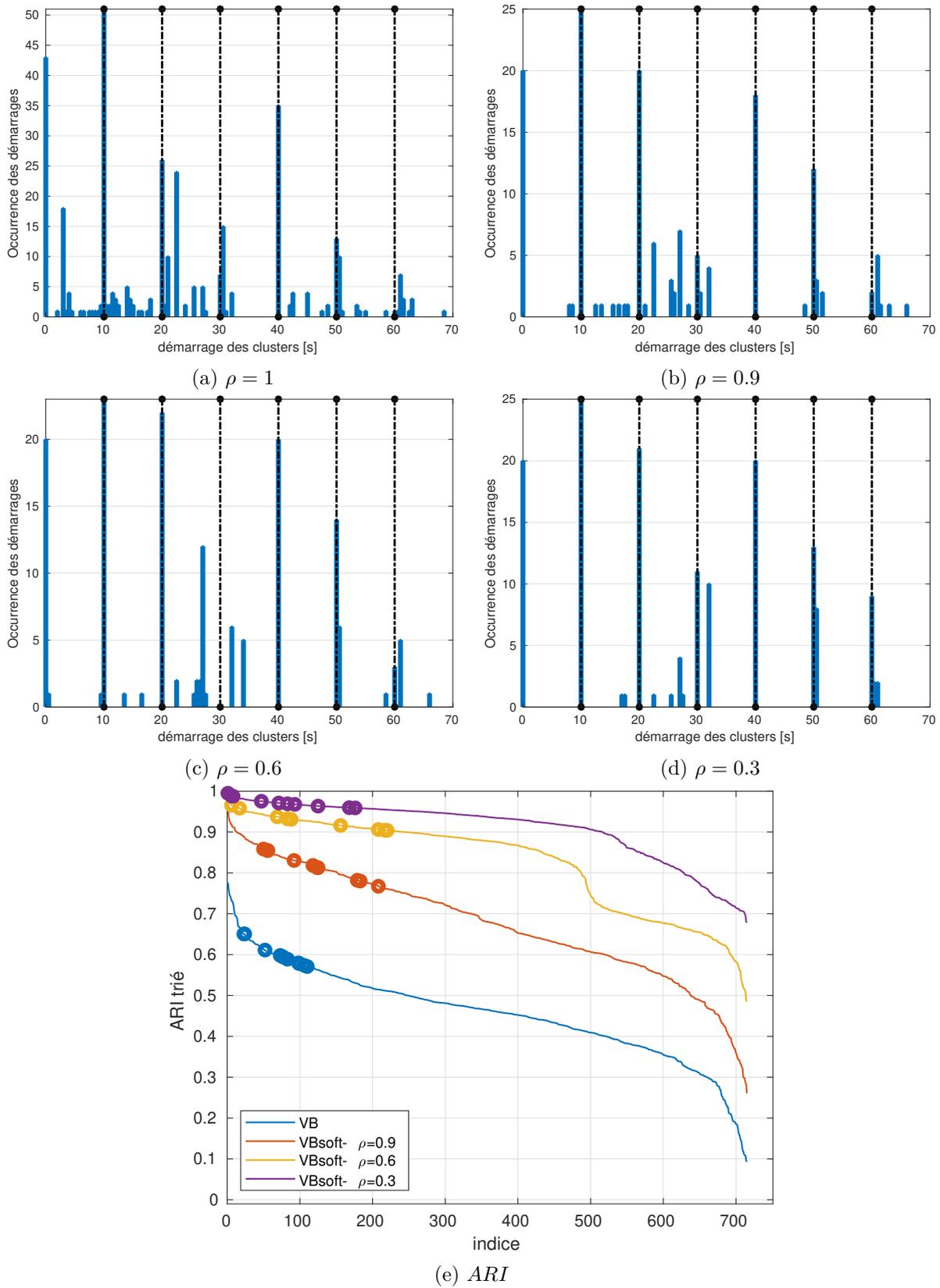


Figure 3.8 – Essai #E : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$

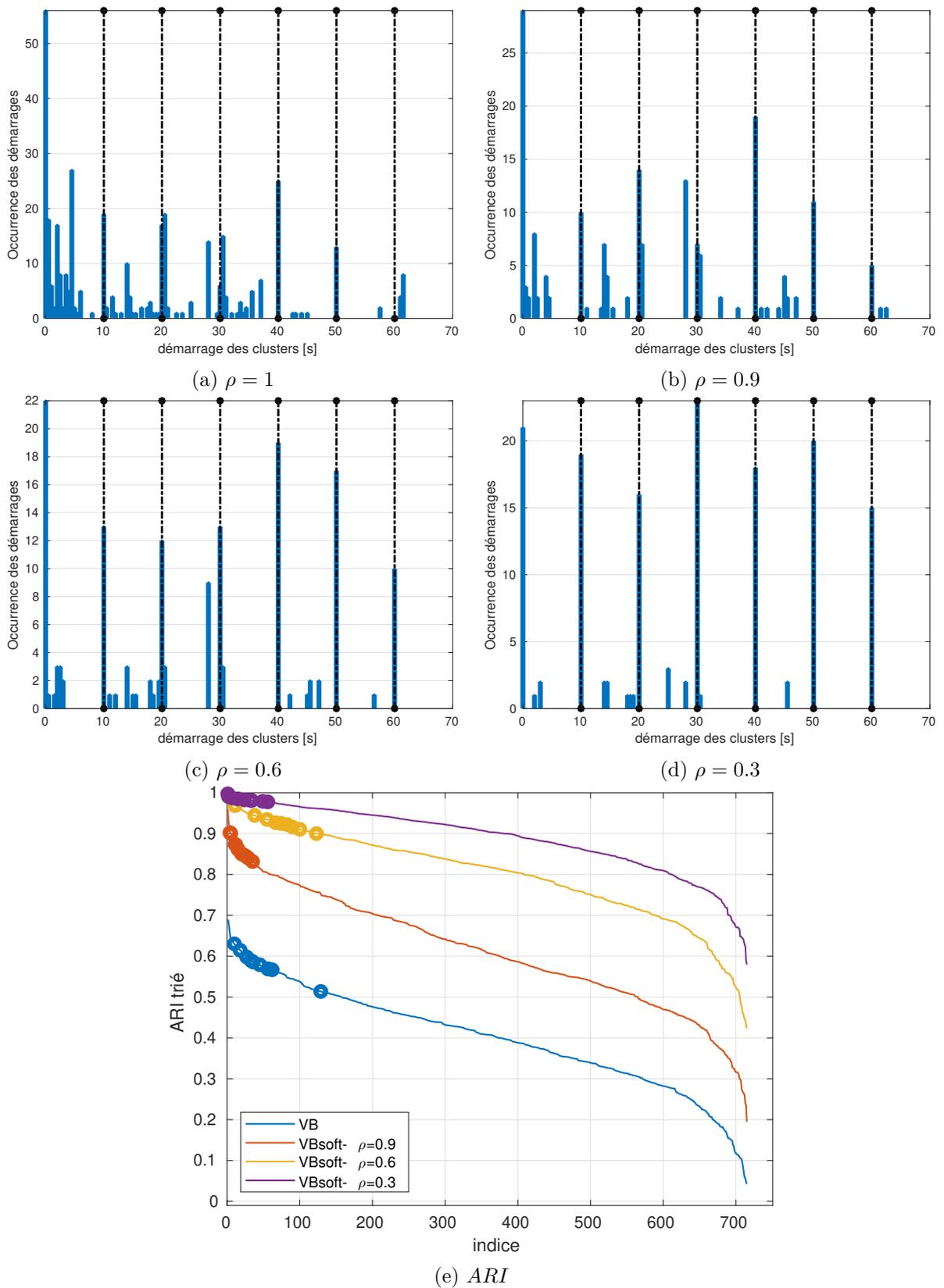


Figure 3.9 – Essai #F : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$

**Remarques** Les observations faites avec des étiquettes correctes permettent d'affirmer que :

1. L'algorithme VB-GMM donne des *clusters* corrects corrélés aux niveaux de serrage.
2. L'ajout d'*a priori incertains* permet d'avoir des résultats qui sont inversement corrélés à la probabilité d'erreur  $\rho$ .

### 3.4.2 Influence de la corruption des étiquettes et de la taille des données

**Influence de la corruption des étiquettes :** Pour les cinq essais B-F nous avons repris les estimations en conservant les hyper-paramètres : le nombre de *clusters* initial  $K_{max} = 10$  et l'a priori de Dirichlet  $\alpha = \frac{1}{K_{max}}$ . On fait de nouveau varier l'influence de l'a priori pour les valeurs de  $\rho = \{1; 0.9; 0.6; 0.3\}$ . La seule différence est la corruption de quelques étiquettes afin de simuler un cas où les étiquettes sont incomplètes ou mal étiquetées. La corruption des étiquettes est faite aléatoirement. Pour chaque valeur  $n$ , on tire aléatoirement un nombre réel dans l'intervalle  $[0, 1]$ . Si le nombre tiré est supérieur à la valeur maximale de  $pl_{nk}$ , l'étiquette est distribuée aléatoirement pour ce point. C'est la même corruption des données qui est utilisée dans [121, 125]. La nouvelle ré-estimation avec la valeur de  $\rho = 1$  a uniquement pour but d'évaluer la robustesse du VB-GMM avec des initialisations aléatoires des clusters. On remarque que le VB-GMM parvient à des résultats presque identiques avec des initialisations aléatoires différentes (Figure 3.10a). Les résultats étant très similaires pour les cinq essais, seul l'essai #B est illustré.

En ajoutant peu d'*a priori* (Figure 3.10b), on ne note pas un grand changement au niveau des histogrammes. La hauteur des pics diminue assez lentement sans s'uniformiser. Cette observation est contraire à celle avec les labels corrects (Figure 3.4). Cependant en observant les graphes des ARI (Figure 3.10e), l'amélioration du partitionnement est constatée. Les labels, même corrompus parviennent à améliorer le partitionnement. L'utilisation des étiquettes bruitées permet d'améliorer le partitionnement, mais pas nécessairement le graphe des histogrammes des instants de déclenchement des *clusters*. Aucun déclenchement n'est détecté à 60s, même en présence d'*a priori* forts. On remarque également la diminution des hauteurs de pics entre les bornes.

**Influence de la taille des données :** Dans les résultats présentés jusqu'ici nous avons effectué un sous-échantillonnage de 1/10. Cette fois, nous utilisons l'ensemble des observations. Le nombre de partitions sélectionnées varie très peu (autour de 20-30, figures 3.11). Plus les *a priori* augmentent, plus les pics des histogrammes tendent à s'uniformiser aux instants des changements de niveaux de serrage, exception faite pour le dernier serrage démarrant à 60s qui n'est pas détecté.

### 3.4.3 Influence des autres hyper-paramètres

Nous nous proposons d'analyser l'influence de  $\alpha_0$ , l'a priori de Dirichlet,  $K_{max}$  le nombre de clusters initial et  $d$  le nombre de descripteurs utilisés dans une combinaison pour le partitionnement. Pour chacune des évaluations, seul le paramètre évalué varie.

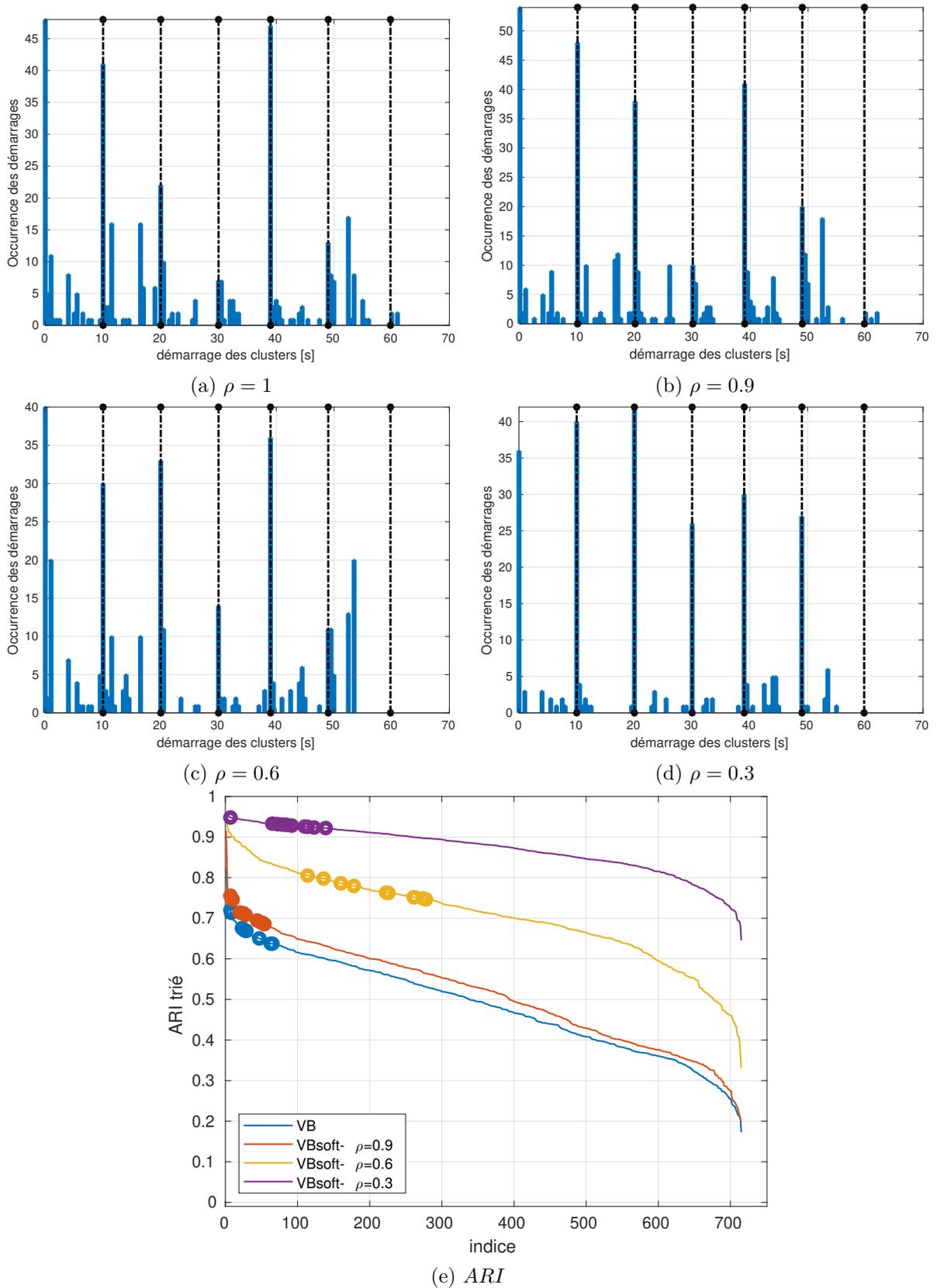


Figure 3.10 – Essai #B : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$ . Les classes ont été corrompues, 10% des étiquettes introduites comme *a priori* ne correspondent pas au niveau de serrage correct.

**Influence de l'*a priori* de Dirichlet :** Nous avons fait varier  $\alpha_0$ , l'*a priori* de Dirichlet pour trois valeurs  $\{\frac{1}{10}; 1; 10\}$  (voir figure 3.12). Lorsque  $\alpha_0 = 10$  plusieurs sous-ensembles de descripteurs ont tendance à donner un nombre de clusters assez grand. Ainsi le nombre de partitions sélectionnées pour le tracé des histogrammes diminue. Dans la majeure partie il s'agit d'un *cluster* qui est divisé en 2 plus petits. Au niveau des graphes des histogrammes et ARI on a une petite baisse de qualité. En prenant  $\alpha_0 = 1$ , on a les résultats très proches à ceux obtenus pour  $\alpha_0 = 0.1$ . Quelquefois les clusters presque vides sont regroupés avec les plus proches. Pour  $K = 10$  l'impact de  $\alpha_0$  n'est pas très visible, cependant si  $K_{max}$  est assez grand (20 ou 40 par exemple), utiliser une valeur de  $\alpha_0 \geq 1$  peut s'avérer problématique.

**Influence du nombre de clusters initial  $K_{max}$  :** En fixant le nombre de *clusters* initial à 20 par exemple, la majorité des sous-ensembles générant un nombre de *clusters* égal à 10 lorsque l'initialisation est faite à 10 donne plus de 10 *clusters* sans *a priori*. Au niveau des ARI, on a des résultats très proches pour les 2 cas. Pour les différents essais le nombre de clusters converge toujours vers une valeur inférieure à 20. Les histogrammes sont plus denses vu qu'il y a de nombreux instants de déclenchement. Cependant dès qu'on utilise les *a priori*, les résultats pour  $K_{max} = 20$  et  $K_{max} = 10$  sont identiques en termes de nombre de *clusters* trouvé et de partitions obtenues. En initialisant  $N_{max} = 40$ , les résultats sont identiques à ceux illustrés pour  $N_{max} = 20$ . Nous avons choisi d'illustrer l'essai #B (résultats performants) et l'essai #D (le moins performant). Les figures 3.13 représentent les histogrammes et ARI sans *a priori* sur les *clusters* et les ARI pour différentes valeurs de  $\rho$ . Pour l'essai #B (figure 3.13a), le nombre de *clusters* estimé est toujours inférieur ou égal à 13, ce qui permet d'avoir 103 partitions sélectionnées au lieu de 190 (cas de non convergence). Les histogrammes sont similaires à ceux obtenus avec  $K = 10$  en termes de proportion. Toutes les bornes sont détectées et bien mises en évidence à l'exception du dernier serrage. Les histogrammes sont plus chargés qu'avant avec plusieurs pics entre les bornes. Fixer un seuil permet de décider des instants à considérer. Pour l'essai #D, le changement de serrage à 50s est détecté mais avec un pic assez faible. L'histogramme est assez chargé. Les résultats sont légèrement dégradés par rapport à ceux obtenus avec  $K_{max} = 10$ . En terme d'ARI les résultats sont du même ordre que ceux obtenus avec l'initialisation avec 10 *clusters*.

**Influence du nombre de descripteurs par sous-ensemble :** Les résultats avec 4 descripteurs ont déjà été présentés précédemment. Dans cette partie, des sous-ensembles à 2,3 et 10 descripteurs sont étudiés pour évaluer l'influence du nombre de descripteurs. En utilisant 2 descripteurs, on obtient 78 partitions. Celles-ci sont moins corrélées aux niveaux de serrage, ce qui se traduit par des valeurs d'ARI plus faibles. Pour l'essai #B, les changements de niveaux de serrage à 30s et 60s ont des pics assez bas (Figure 3.14a). Enfin l'influence des *a priori* est également moins prononcée (figure 3.14c). Une explication peut être qu'avec 2 descripteurs nous avons moins d'informations qu'avec 3 ou 10. Les résultats permettent quand même d'avoir la même interprétation que nous avons faite avec 3 descripteurs. En utilisant 10 descripteurs nous avons 286 partitions. Les valeurs d'ARI sont

plus élevées, ce qui signifie qu'augmenter le nombre de descripteurs améliore la caractérisation des niveaux de serrage. Au niveau des histogrammes, les pics correspondant aux bornes sont assez hauts et presque uniformes, ce qui signifie que le nombre de *clusters* et que les différentes partitions détectent les mêmes évènements. On note cependant l'absence d'un pic à 60s, comme illustré avec 4 descripteurs. En ajoutant très peu d'*a priori* les résultats sont largement améliorés, les valeurs des ARI avoisinent un pour presque toutes les partitions. À partir de  $\rho = 0.6$ , les partitions sont identiques aux classes, les ARI tendent tous vers 1. Utiliser un nombre élevé de descripteurs a donc un impact considérable sur la caractérisation des *clusters*. Il est cependant nécessaire de s'assurer que les descripteurs portent tous suffisamment d'information. L'algorithme proposé peut ainsi être combiné à des outils tels que l'analyse en composantes principales afin de l'optimiser.

## 3.5 Discussions et conclusions

### Lien avec les modèles de la littérature

Le modèle VB-GMM-SOFT a des propriétés qui permettent :

1. L'insertion des *a priori* sur  $\pi$ ,  $K$ ,  $\mu$  et  $\Lambda$  et leurs espérances. Dans l'étape de la maximisation de l'espérance (VB-M), l'incertitude est quantifiée non seulement sur les *clusters*, de la même manière que l'EM standard, mais aussi pour les paramètres du modèle. Cela permet de résoudre les problèmes où la quantité de données n'est pas suffisante pour couvrir l'espace des états et d'éviter aussi des problèmes numériques. De plus, comme l'un des avantages les plus importants, le modèle VB-EM permet de trouver un modèle qui décrit les *clusters*  $\mathbf{x}$  ayant généré les observations  $\mathbf{y}$  en estimant les *clusters* mais sans *sur-ajustement*. Un modèle sur-ajusté est un modèle dont la complexité est supérieure à la configuration réelle des données. Pour la régression il peut s'agir de l'utilisation d'un polynôme lorsqu'une droite est amplement suffisante pour représenter le nuage de points. Ceci a pour conséquence de mieux ajuster le modèle aux données. La problématique du partitionnement concerne le nombre de *clusters*. Le sur-ajustement est l'une des faiblesses de l'EM. En effet le maximum de vraisemblance a tendance à croître avec le nombre de *clusters*, les modèles surajustés peuvent donc être validés. Ceci peut être problématique pour les données réelles à grande dimension, l'EM est donc moins approprié que le VB-EM pour optimiser la structure d'un modèle [138].
2. La détermination de  $K$ , le nombre de *clusters* sans recours aux critères classiques tels que Silhouette et Davies-Bouldin.
3. La possibilité d'évaluer la propagation de l'inférence liée à un des paramètres du modèle, car l'espérance de la log-vraisemblance est composée des espérances des différentes distributions [78, Chap.10].

4. La prise en compte des *a priori* sur les *clusters* en joignant la labellisation douce à une approche bayésienne à l'aide de la théorie des fonctions de croyance.

Les possibles limites du modèle sont les mêmes que celles des approches paramétriques en général :

1. La supposition d'une distribution spécifique pour un paramètre, ce qui peut le rendre inapproprié pour certaines applications [78, Chap.2]
2. La complexité de l'algorithme qui le rend difficile à paramétrer et initialiser.

### Sur l'application en émission acoustique

Le VB-GMM permet de déterminer le nombre de *clusters* pour une combinaison de  $d$  descripteurs. L'astuce consiste à fixer un grand nombre de clusters. À la convergence de la log-vraisemblance, le nombre de *clusters* correspond au nombre de composantes de taille non nulle. Cela réduit le problème du partitionnement au choix du sous-ensemble de descripteurs pertinents. Il peut donc être utilisé dans un système de validation classique en omettant la boucle du choix de  $K_S$  le nombre de *clusters* par sous-ensemble. Il permet également de prendre des incertitudes autour des paramètres du modèle.

Nous avons également ajouté à cet algorithme l'insertion des *a priori* sur les *clusters*. L'ajout des *a priori* permet d'améliorer le partitionnement lorsqu'on dispose d'un avis de l'expert ou une règle fiable d'attribution d'étiquettes.

La validation a reposé sur deux types de graphes : les histogrammes des instants de déclenchement des *clusters* et l'ARI des *clusters* obtenus. Les histogrammes montrent globalement une bonne performance de l'algorithme sans *a priori* pour la détection des déclenchements. Cela valide l'utilisation d'un algorithme de partitionnement bayésien qui n'a pas été utilisé jusqu'à maintenant dans la littérature. Le couplage de cet algorithme avec notre critère de sélection des partitions (chapitre 2) permet de résoudre le problème de détection des instants de déclenchement dans ORION-AE.

Ensuite nous avons évalué la pertinence des clusters au travers de l'ARI. On a pu observer qu'il n'y avait pas toujours de lien direct entre l'ARI et les instants de déclenchement. Mais comme attendu, l'ajout d'*a priori* même en faible quantité ou corrompu permet d'améliorer à la fois les instants de déclenchement et l'ARI.

La validation a aussi pris en compte l'influence de l'hyper-paramètre  $\alpha_0$ , le nombre de points et le nombre de descripteurs. Les résultats montrent que ces paramètres jouent relativement peu sur les performances.

La principale limite de la méthode proposée est commune aux méthodes d'apprentissage visant à choisir des sous-ensembles de descripteurs pertinents. Lorsqu'on applique ces méthodes sur les données d'EA avec  $D$  le nombre total de descripteurs, il devient difficile d'utiliser les différentes combinaisons de descripteurs. Par exemple pour  $d = 5$  et  $D = 25$  descripteurs, nous avons 53130 sous-ensembles. La méthode devient très vite coûteuse en temps de calcul. La méthode la plus courante utilisée est la réduction de dimension à l'aide de l'analyse en composantes principales. Elle permet d'avoir un aperçu rapide de la structure des données. La réduction ainsi que les résultats obtenus avec le VB-GMM et le VB-GMM-SOFT ont été mis en annexe [A.3](#).

En conclusion, il s'agit de la première application d'un partitionnement bayésien sur les données d'EA. Elle démontre la pertinence de ce type d'algorithme pour l'interprétation des données d'EA. La contribution scientifique principale faite sur l'ajout d'*a priori* sur les *clusters* a montré sa pertinence sur les données ORION-AE.

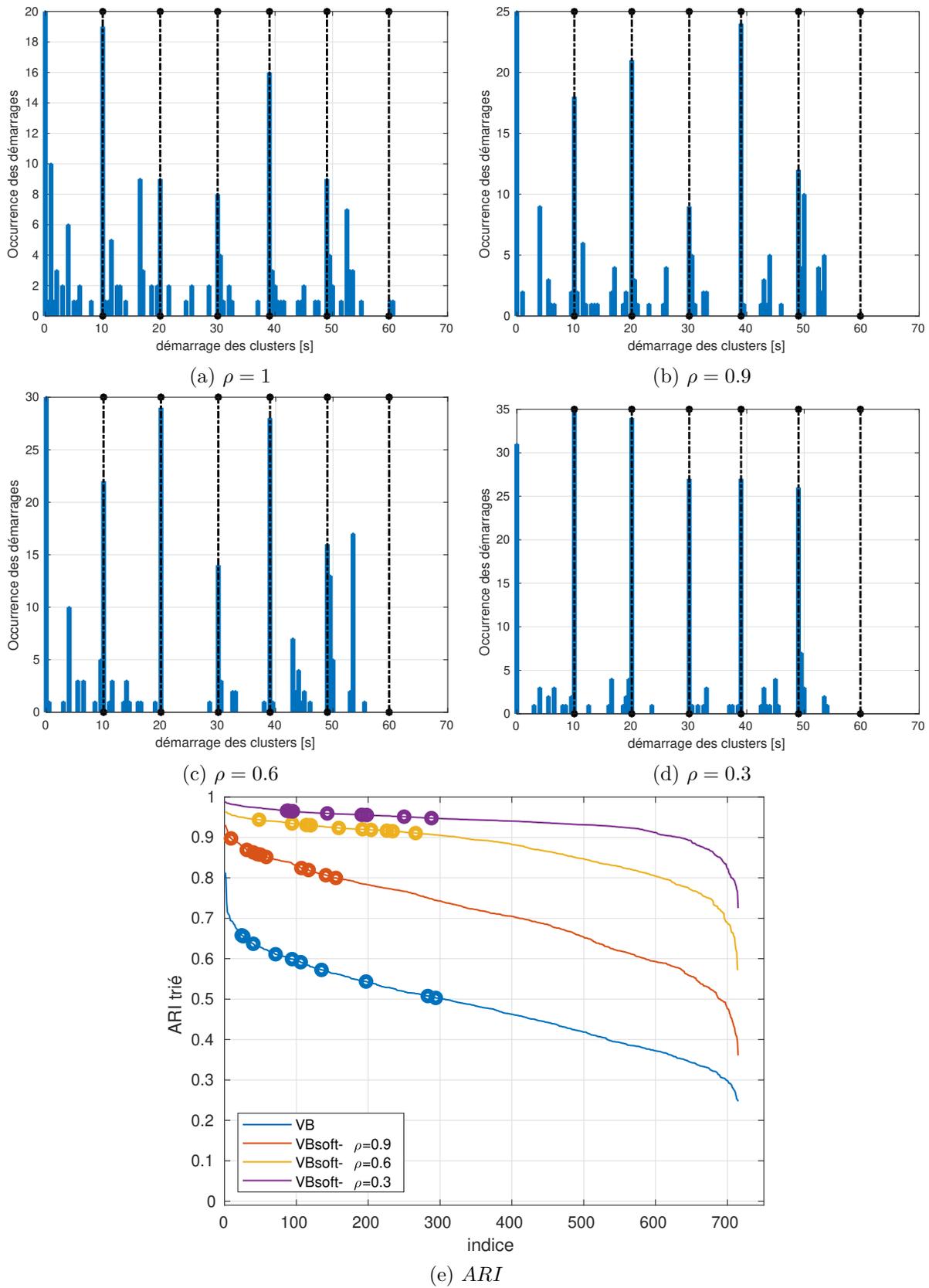


Figure 3.11 – Essai #B : Partitionnement pour les valeurs  $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour  $K_S^* \in [2, K_{max}]$ . Les données ont été prises en entier sans sous-échantillonnage ( $N = 14089$ ).

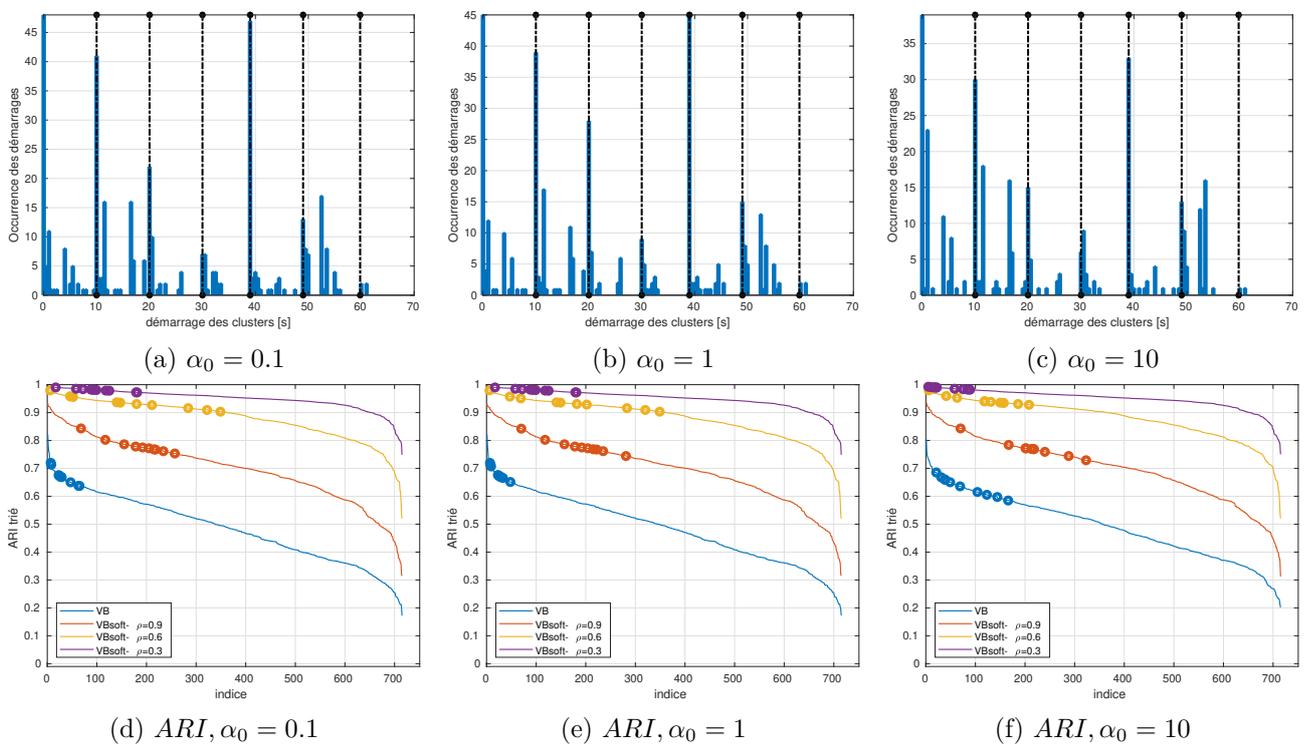


Figure 3.12 – Essai #B : Évaluation de l'influence l'a priori de Dirichlet  $\alpha_0 = \{0.1, 1, 10\}$ . Les histogrammes sont tracés sans *a priori*

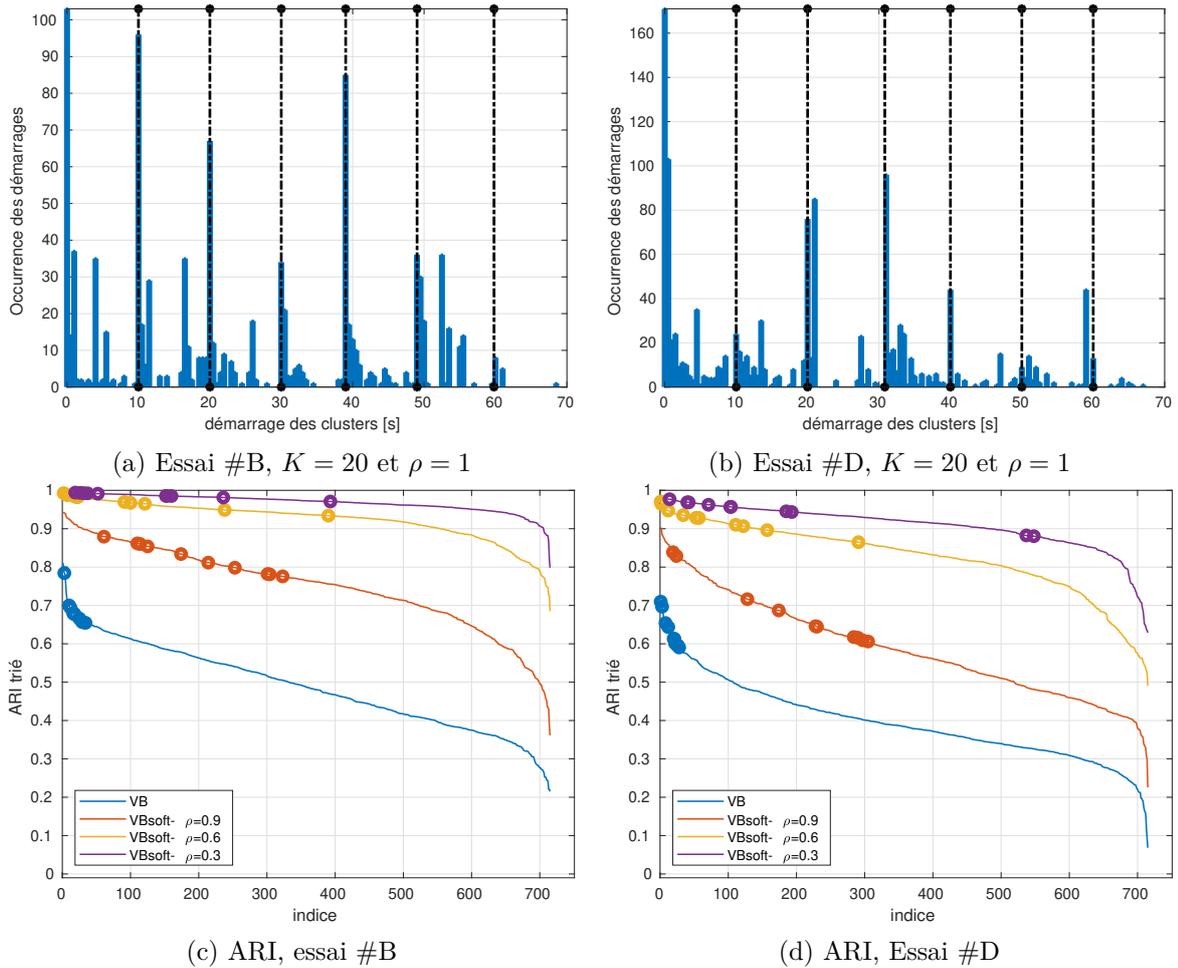


Figure 3.13 – Essais #B et #D : Illustration avec le nombre de *clusters* initialisé à 20. Les histogrammes sont réalisés sur les déclenchements de clusters avec un partitionnement sans *a priori*

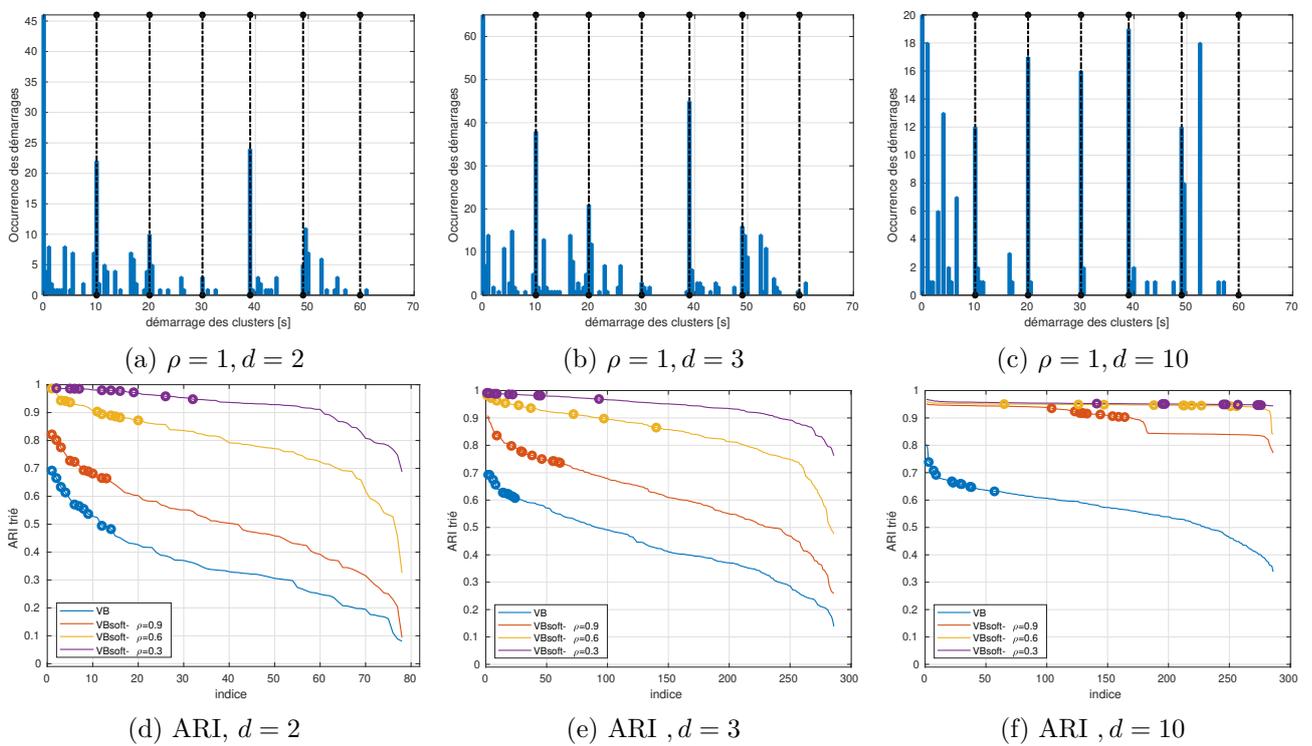


Figure 3.14 – Essai #B : Évaluation de l'influence du nombre de descripteurs choisis pour le partitionnement  $d = \{2, 3, 10\}$ .



# Conclusion

Au cours de cette thèse, nous nous sommes intéressés à l'évolution temporelle et aux séquences de déclenchement des *clusters* issus du partitionnement des données d'émission acoustique (EA). La chronologie des *clusters* est très importante pour l'interprétation des données d'EA, elle peut permettre à un utilisateur de faire le lien entre les *clusters* et les phénomènes observés. La plupart des critères de validation étant basés sur la forme des clusters dans l'espace des descripteurs, cet aspect temporel n'est pas pris en compte dans la validation. Pour certaines applications comme le monitoring des structures, se servant de la technique d'EA, l'aspect temporel est important. Nos travaux ont été inspirés par cette problématique relevée lors de l'étude bibliographique.

## Synthèse des travaux

### Sur la validation du partitionnement et les représentations

Au chapitre 1, trois méthodes de partitionnement (k-moyennes, hiérarchique et GMM) ont été illustrées sur le benchmark ORION-EA. La validation faite par le vote [3] est basée sur trois critères de « forme ». En s'intéressant à la partition obtenue, les différents algorithmes donnent des résultats proches avec un nombre de clusters souvent faible (autour de 3). Cependant l'analyse de l'ensemble des partitions a permis de constater que plusieurs partitions non-sélectionnées ont des clusters qui correspondent aux niveaux des serrages. Cette comparaison sur le benchmark ORION-EA permet de mettre en évidence qu'un « bon » *cluster* (basé sur la forme) ne conduit pas nécessairement à une chronologie pertinente.

Au chapitre 2, la contribution principale vise à améliorer la détection des événements d'un point de vue chronologique. Nous avons défini pour cela un critère séquentiel qui favorise la sélection des partitions dont les *clusters* ont des instants de déclenchement répartis dans le temps. Ce critère permet d'avoir des partitions plus pertinentes au vu de leur lien avec les serrages pour le benchmark ORION-EA. Le critère défini ne permet cependant pas de sélectionner le nombre de *clusters*. Un ajustement que nous avons proposé consiste à mettre en place un terme de pénalité pour favoriser un certain type de partitions lorsque les performances sont égales. Il s'agit à notre connaissance du premier critère séquentiel dédié à la validation du partitionnement en EA. Le travail fait l'objet d'un article en cours de rédaction.

Une représentation des résultats sous forme d'histogrammes des instants de déclenchement des *clusters* a été proposée pour contourner la difficulté liée à l'estimation du nombre de *clusters*. En faisant varier le nombre de clusters, les instants de déclenchement des clusters des meilleures partitions sont relevés et présentés sous forme d'histogrammes. La hauteur des pics correspond au nombre de fois qu'un cluster démarre à cet instant. Un avantage de cette validation est la prise en compte de plusieurs partitions, ce qui rend la méthode robuste. Un seuil de pic peut être fixé, celui-ci permet de déterminer le nombre d'évènements. Pour le benchmark ORION-EA, des pics élevés sont observés à la plupart des instants de changement de serrage validant l'approche proposée.

## Sur le partitionnement

Concernant le partitionnement, le premier apport est applicatif. Les modèles bayésiens [130, 131], [78, chap.10] permettent de déterminer automatiquement le nombre de *clusters* pour un sous-ensemble de descripteurs. Un tel modèle permet de s'affranchir d'une des principales difficultés de l'interprétation des données d'émission acoustique. L'illustration avec le benchmark ORION-EA permet de confirmer son applicabilité dans le domaine de l'EA et la caractérisation où la détermination du nombre de *clusters* est clé pour l'interprétation des données. Une estimation préalable du nombre de clusters permet aussi de ne garder sur l'histogramme que les déclenchements des « clusters naturels ».

La labellisation douce vise à répondre à la difficulté d'avoir des données annotées dans certains domaines, notamment l'EA. L'algorithme VB-GMM-SOFT proposé au chapitre 3 permet de répondre à cette problématique. En utilisant des informations partielles, même en faible quantité, l'algorithme parvient à améliorer les résultats du partitionnement. Une particularité de cet algorithme est qu'il laisse la possibilité de création de *clusters* supplémentaires qui ne sont pas des classes renseignées par l'expert. La seconde particularité est la possibilité que l'algorithme fournisse un résultat différent si l'avis de l'expert n'est pas en accord avec les données. L'algorithme peut donc être utilisé en accordant un certain crédit à l'expert, puis en initialisant un grand nombre de *clusters*, l'algorithme converge vers une solution qui établit un compromis entre mesures et expertise. Les histogrammes des instants de déclenchement obtenus sur le benchmark ORION-EA avec le VB-GMM donnent les pics les plus élevés aux instants correspondant aux changements des niveaux de serrage. Ce qui signifie que même sans *a priori* la méthode fournit des résultats pertinents. Nous avons montré qu'avec les modifications proposées dans cet algorithme, une faible quantité d'*a priori* permet de grandement améliorer les résultats et ce, à la fois en termes d'ARI et d'instant de déclenchement en couplant cet algorithme au critère séquentiel proposé.

## Perspectives

Certains travaux initiés dans cette thèse méritent d’être approfondis. Ils concernent les applications en EA et l’aspect algorithmique.

### À court terme

À court terme, l’idée est de (1) mieux définir la notion d’évènements et (2) décrire les caractéristiques spécifiques d’une bonne partition en fonction de la chronologie et du type de phénomènes qu’on veut discriminer. Le premier point vise une meilleure représentation des histogrammes des instants de déclenchement des *clusters*. Pour certaines partitions, il arrive qu’un cluster soit détecté de manière précoce. Graphiquement, à partir des occurrences cumulées, il est possible d’observer les « instants réels » de déclenchement. Une astuce serait alors de fixer un seuil d’occurrence pour considérer le déclenchement d’un *cluster*. Nous avons remarqué que tracer non seulement les histogrammes des instants de déclenchement mais aussi les instants d’arrêt des clusters pouvait nous informer sur les niveaux de serrage. Redéfinir la notion d’évènement en tenant compte des instants de fin permettra une meilleure interprétation des données.

En fonction du type d’essais et de ce qu’on souhaite discriminer (étapes d’un procédé par exemple), notre critère basé sur la chronologie des *clusters* peut être décliné de différentes manières pour s’adapter à l’application.

### À long terme sur ORION-EA

La disponibilité des serrages, des descripteurs et des signaux transitoires pour les différents essais ORION-EA permet d’envisager une orientation vers des méthodes supervisées pour analyser ces données. Nous avons cependant constaté que les valeurs des descripteurs d’un niveau de serrage changent d’un essai à l’autre. Utiliser des réseaux de neurones profonds pourrait permettre d’extraire des caractéristiques latentes communes à chaque niveau de serrage. Ces réseaux pourraient être entraînés en utilisant les trois capteurs disponibles dans le benchmark (fusion des capteurs).

Pour le benchmark ORION-EA, les desserrages sont faits manuellement. Il est envisagé d’effectuer des essais avec un desserrage progressif (naturel). Pour cela l’utilisation de boulons instrumentés peut permettre de suivre le niveau de serrage. Les données d’EA peuvent être collectées à des intervalles de temps réguliers pour limiter la taille des données. Une approche de ce type permettrait d’évaluer les algorithmes proposés en conditions réelles. Cela permettra également de les adapter plus facilement aux problématiques de surveillance en temps réel telles que rencontrées en SHM.

De manière générale, l’utilisation de l’algorithme VB-GMM peut se heurter à l’absence de connaissance des classes. Ces dernières peuvent être fournies soit par une autre méthode de diagnostic per-

mettant de connaître l'état de la structure, soit par un modèle physique ou numérique. Dans [139] par exemple, un modèle physique basé sur les éléments finis est appliqué sur une plaque afin d'évaluer l'impact du milieu de propagation. Après la localisation, des étiquettes sont attribuées aux « transitoires reconstruits », à partir d'un modèle supervisé préalablement fait. La difficulté de la mise en place d'un tel modèle réside en la connaissance du milieu de propagation et la construction d'un benchmark pour la caractérisation des différentes classes. Un tel modèle pourrait également servir à générer les *a priori* pour le critère séquentiel.

Dans le cadre non supervisé, nous avons constaté (annexe A.3) que la réduction de dimension avec les auto-encodeurs [140–144] permet de trouver des clusters pertinents. Les auto-encodeurs sont utilisés pour la réduction non linéaire des données. En utilisant des encodeurs avec plusieurs couches cachées la complexité des descripteurs extraits augmente. Un tel modèle peut être construit pour les données brutes en deux blocs, dont un extrait des descripteurs à partir des transitoires et le second associé à un apprentissage supervisé ou non supervisé. Le manque de théorie permettant de configurer le modèle peut être un facteur limitant. L'utilisation du benchmark fiable peut donc servir à l'optimisation de ce modèle en minimisant l'erreur de prédiction.

# Publications

## 3.6 Articles de conférence

- **Joint Optimization of the Number of Clusters and Their Parameters in Acoustic Emission Clustering** [117]

**Abstract :** Acoustic Emission (AE) is a non-destructive technique suitable for monitoring structural damages in complex systems. Interpretation of AE data is generally made by an unsupervised learning approach called clustering. A clustering method divides a dataset into different groups called clusters, which are expected to have a physical interpretation in terms of damages. The set of groups, called partition, is then evaluated through an external criterion. Since the number of clusters is a priori unknown, the standard approach for AE clustering consists in running a clustering method for different number of clusters and then to select the partition with the best value of the external criterion. The external criterion can vary across publications and aims at evidencing “natural” clusters. According to the criterion, the selected partition and its interpretation can be subjected to variability. In this publication, we explore a clustering method in which the parameters of clusters are optimised jointly with the number of clusters through an objective function and an optimisation procedure designed for that purpose. To our knowledge, it is the first method of this class in AE literature. Experimental results concern real data from tightening test.

- **A variational Bayesian clustering approach to acoustic emission interpretation including soft labels** [145]

**Abstract :** We investigate Gaussian Mixture Models (GMM) with uncertain parameters to evaluate whether this model can help in interpreting acoustic emission data used in non-destructive testing. This model, called VBGMM (variational Bayesian GMM) allows the end-user to automatically determine the number of clusters which makes it relevant for this type of application where clusters are related to damages. In this work, we modify the training procedure to include prior knowledge about clusters. Experiments are made on a recently published benchmark, ORION-AE, that aims at estimating the tightening levels in a bolted structure under vibrations. Preliminary results of the VBGMM with soft priors (VBGMM-SOFT) show good improvement over the standard VBGMM.

### 3.7 Articles de revue

- **A new external validity criterion for evaluating acoustic emission clustering methodologies based on the distribution of clusters onsets** (En cours d'écriture)
- **Variational Bayesian Acoustic Emission Data Clustering** (En cours d'écriture)

# Bibliographie

- [1] R. Liptai, D. Harris, B. Engle, and C. Tatro. Acoustic emission techniques in material research. 07 1970.
- [2] T. Le Gall, N. Godin, T. Monnier, and C. Fusco. Modélisation de l'Emission Acoustique : aide à l'identification de paramètres pertinents pour décrire l'endommagement des matériaux composites. In *Journées Nationales sur les Composites 2017*, 77455 Champs-sur-Marne, France, June 2017. École des Ponts ParisTech (ENPC).
- [3] M. Sause. *In Situ Monitoring of Fiber-Reinforced Composites : Theory, Basic Concepts, Methods, and Applications*. 2016.
- [4] G. Chevallier, E. Ramasso, and P. Butaud. Detection and analysis of loosening in jointed structures using acoustic emission sensors and smart bolts. In *37th Conference and Exposition on Structural Dynamics (IMAC)*, Orlando, USA, 2019.
- [5] G. Zwingelstein. Les principaux mécanismes de défaillance pour le diagnostic en maintenance. (ref. article : mt9132), 2020.
- [6] Takayuki Shiraiwa, Kazuki Ishikawa, Manabu Enoki, Ippei Shinozaki, and Shingo Kanazawa. Acoustic emission analysis using bayesian model selection for damage characterization in ceramic matrix composites. *Journal of the European Ceramic Society*, 40(8) :2791 – 2800, 2020.
- [7] K. Ono. Acoustic emission in materials research – a review. *Journal of acoustic emission*, 29, 2011.
- [8] P. Krawczak and S. Chaki. Diagnostiquer les composites, de la peau au cœur. (ref. article : mt9132), 2020.
- [9] W. P. Mason, H. J. Mckimin, and W. Shockley. Ultrasonic observation of twinning in tin. *Phys. Rev.*, 73 :1213–1214, May 1948.
- [10] H. Dunegan and D. Harris. Acoustic emission-a new nondestructive testing tool. *Ultrasonics*, 7(3) :160 – 166, 1969.
- [11] Xinglin Lei, Takahiro Funatsu, Shengli Ma, and Liqiang Liu. A laboratory acoustic emission experiment and numerical simulation of rock fracture driven by a high-pressure fluid source. *Journal of Rock Mechanics and Geotechnical Engineering*, 8(1) :27 – 34, 2016.
- [12] S. Invernizzi, F. Montagnoli, and A. Carpinteri. Fatigue assessment of the collapsed xxth century cable-stayed polcevera bridge in genoa. *Procedia Structural Integrity*, 18 :237–244, 01 2019.

- [13] Alain PROUST. Exemples d'applications shm. In *4ème Journée nationale Contrôle Santé et Monitoring des Structures, dernières avancées et solutions*, France, 07 2021.
- [14] K. Ono. Application of acoustic emission for structure diagnosis. *Diagnostyka*, 2 :3–18, 01 2011.
- [15] Coraline Simon. *Accelerated testing on ceramic matrix composites aimed at aeronautical applications*. PhD thesis, Université de Bordeaux, October 2017.
- [16] Tam Nguyen-Tat. *Application de la technique d'émission acoustique pour évaluer l'endommagement mécanique dans le béton*. PhD thesis, Toulouse, 03 2018. 2018TOU30119.
- [17] M.G.R. Sause, A. Gribov, A.R. Unwin, and S. Horn. Pattern recognition approach to identify natural clusters of acoustic emission signals. *Pat. Rec. Letters*, 33 :17–23, 2012.
- [18] K. Goswami and G.L Samuel. Monitoring of material-removal mechanism in micro-electrical discharge machining by pulse classification and acoustic emission signals. *Proceedings of the Institution of Mechanical Engineers, Part B : Journal of Engineering Manufacture*, 06 2020.
- [19] Michail Kozochkin, Sergey Grigoriev, Anna Okunkova, and Artur Porvatov. Monitoring of electric discharge machining by means of acoustic emission. *Russian Engineering Research*, 36 :244–248, 03 2016.
- [20] S. Momon, N. Godin, P. Reynaud, M. R'Mili, and G. Fantozzi. Unsupervised and supervised classification of ae data collected during fatigue test on cmc at high temperature. *Composites Part A : Applied Science and Manufacturing*, 43(2) :254–260, 2012.
- [21] T. Le Gall. *Simulation de l'émission acoustique : Aide à l'identification de la signature acoustique des mécanismes d'endommagement*. Thèse, 2016. url : <http://www.theses.fr/2016LYSEI005>.
- [22] Abida Satour, Silvio Montrésor, Mourad Bentahar, and Boubenider Fouad. Wavelet based clustering of acoustic emission hits to characterize damage mechanisms in composites. *Journal of Nondestructive Evaluation*, 39, 04 2020.
- [23] Anca Dobrovat. *Modélisation multi-échelle de l'endommagement et de l'émission acoustique dans les roches*. Thèse, Université de Grenoble, May 2011.
- [24] Nicolas Boulay. *Modélisation des capteurs d'émission acoustique en vue de la simulation d'un contrôle*. Thèse, 2017. url : <http://www.theses.fr/2017SACLS593/document>.
- [25] M. Kharrat, E. Ramasso, V. Placet, and M.L. Boubakar. 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, 2016.
- [26] E. Ramasso, P. Butaud, T. Jeannin, F. Sarasini, V. Placet, N. Godin, J. Tirillò, and X. Gabrion. Learning the representation of raw acoustic emission signals by direct generative modelling and its use in chronology-based clusters identification. *Engineering Applications of Artificial Intelligence*, 90, 2020.
- [27] M. Riethmuller. L'émission acoustique, applications aux équipements industriels. 2008. url : <https://www.ndt.net/article/cofrend2008/papers/079.pdf>.

- [28] Hamoudi Mazouz. *Frottement, usure et lubrification de couples de biomatériaux : Cas du PEHD / zircone et cas de l'acier inoxydable austénitique 316L / zircone*. thèse, 2003. 2003COMP1464.
- [29] M.H. Badi, G.G. Yaralioglu, Sanli Ergun, F. Degertekin, Ching-Hsiang Cheng, and Pierre Khuri-Yakub. A first experimental verification of micromachined capacitive lamb wave transducers. volume 1, pages 311 – 314 vol.1, 11 2000.
- [30] I. Md Mazharul and H. Huang. Pwas-based wireless acoustic emission sensor. Number 8345, 03 2012.
- [31] V. Giurgiutiu. *Piezoelectric Wafer Active Sensors – PWAS Transducers*, pages 357–394. 12 2014. doi : 10.1016/B978-0-12-418691-0.00007-1.
- [32] H. Mei, M. Faisal Haider, R. Joseph, A. Migot, and V. Giurgiutiu. Recent advances in piezoelectric wafer active sensors for structural health monitoring applications. *Sensors*, 19(2), 2019.
- [33] Emmanuel Ramasso, Thierry Dencœux, and Gaël Chevallier. Clustering acoustic emission data streams with sequentially appearing clusters using mixture models. *Mechanical Systems and Signal Processing*, 181 :109504, 2022.
- [34] H. Festjens, G. Chevallier, and J. Dion. A numerical tool for the design of assembled structures under dynamic loads. *Int. J. of Mechanical Sciences*, 75 :170–177, 2013.
- [35] C. Migeon. Émission acoustique et analyse vibratoire pour l'étude des défauts de roulements pour différents régimes moteurs. Mémoire master, 06 2011.
- [36] Rapee Ujjin and M Eng. *Prediction of Bearing Failure in Pin-loaded Laminates*. PhD thesis, 04 2021.
- [37] R. Unnporsson. *Hit Detection and Determination in AE Bursts*. IntechOpen, Rijeka, 2013.
- [38] Zhensheng Yang, Wei Yan, Li Jin, Feng Li, and Ziting Hou. A novel feature representation method based on original waveforms for acoustic emission signals. *Mechanical Systems and Signal Processing*, 135 :106365, 2020.
- [39] N. Godin, S. Huguet, R. Gaertner, and L. Salmon. Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers. *NDT & E International*, 37(4) :253–264, 2004.
- [40] D.M. Egan and J.H. Williams-Jr. Acoustic-emission spectral analysis of fiber composite failure mechanisms. Technical Report 2938, NASA Contractor Report, MIT, Cambridge, Massachusetts, USA, 1978.
- [41] Karen M. Holford, Mark J. Eaton, James J. Hensman, Rhys Pullin, Sam L. Evans, Nikolaos Dervilis, and Keith Worden. 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, 2017.

- [42] I.A. Rastegaev, D.L. Merson, A.V. Danyuk, M.A. Afanasyev, and A. Vinogradov. Using acoustic emission signal categorization for reconstruction of wear development timeline in tribosystems : Case studies and application examples. *Wear*, 410 :83–92, 2018.
- [43] Mistras Group. Pci-2 based ae system user’s manual. april 2007. <https://material.karlov.mff.cuni.cz/people/mathis/Navody/PCI2.pdf>.
- [44] S. Kattis. Noesis : Advanced data analysis, pattern recognition & neural networks software for acoustic emission applications. In *Kolloquium Schallemission, Statusberichte zur Entwicklung und Anwendung der Schallemissionsanalyse*, Fulda, 9-10 March 2017.
- [45] N. Godin, S. Huguet, R. Gaertner, and L. Salmon. Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers. *NDT & E International*, 37(4) :253–264, 2004.
- [46] E. Ramasso. *Méthodologies d’analyse des séries temporelles sous incertitudes aléatoires et épistémiques pour le suivi et le pronostic de l’état de systèmes et structures*. Habilitation à diriger des recherches de l’université de franche-comté, Besançon, France, Novembre 2016.
- [47] Xiaoran Wang, Xiandong Liu, Tian He, Junfei Tai, and Yingchun Shan. A novel joint localization method for acoustic emission source based on time difference of arrival and beamforming. *Applied Sciences*, 10(22), 2020.
- [48] L. Dong, Q. Tao, Q. Hu, S. Deng, Y. Chen, Q. Luo, and X. Zhang. Acoustic emission source location method and experimental verification for structures containing unknown empty areas. *International Journal of Mining Science and Technology*, 2022.
- [49] Li Yang and Feiyun Xu. A novel acoustic emission sources localization and identification method in metallic plates based on stacked denoising autoencoders. *IEEE Access*, 8 :141123–141142, 2020.
- [50] K. Komai, K. Minoshima, and T. Shibutani. Investigations of the fracture mechanism of carbon/epoxy composites by ae signal analyses. *JSME international journal. Ser. 1, Solid mechanics, strength of materials*, 34(3) :381–388, 1991.
- [51] S. Barré and M. Benzeggagh. On the use of acoustic emission to investigate damage mechanisms in glass-fibre-reinforced polypropylene. *Composites Science and Technology*, 52(3) :369–376, 1994.
- [52] P. De Groot, P. Wijnen, and R. Janssen. Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites. *Composites Science and Technology*, 55 :405–412, 1995.
- [53] Mitsuharu S., O. Chen, T. Kishi, S. Carpenter, S. Mitsuno, H. Ichikawa, Y. Tae Lee, S. Tea Kim, and T. Soon Lee. Fracture mechanisms in unnotched and notched sic/sic composites studied by acoustic emission analysis. *Materials Transactions, JIM*, 36(4) :511–517, 1995.

- [54] D. Short and J. Summerscales. Amplitude distribution acoustic emission signatures of unidirectional fibre composite hybrid materials. *Composites*, 15(3) :200–206, 1984.
- [55] E. Pomponi and A. Vinogradov. A real-time approach to acoustic emission clustering. *Mechanical Systems and Signal Processing*, 40(2) :791–804, 2013.
- [56] P. Nimdum and J. Renard. Use of acoustic emission to discriminate damage modes in carbon fibre reinforced epoxy laminate during tensile and buckling loading. *ECCM 2012 - Composites at Venice, Proceedings of the 15th European Conference on Composite Materials*, 06 2012.
- [57] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. A review of vibration-based damage detection in civil structures : From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, 147 :107077, 2021.
- [58] Karen M. Holford, Mark J. Eaton, James J. Hensman, Rhys Pullin, Sam L. Evans, Nikolaos Dervilis, and Keith Worden. 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, 2017.
- [59] A. Anastassopoulos and T. Philippidis. Clustering methodology for the evaluation of acoustic emission from composites. *Journal of acoustic emission*, 13 :11–22, 1995.
- [60] M. Steinbach, L. Ertöz, and V. Kumar. The challenges of clustering high dimensional data. *Univ. Minnesota Supercomp. Inst. Res. Rep.*, 213, 01 2003.
- [61] N. Tomašev and M. Radovanovic. *Clustering Evaluation in High-Dimensional Data*, pages 71–107. 04 2016.
- [62] M. Eaton, R. Pullin, K. Hensman, J. and Holford, K. Worden, and S. Evans. Principal component analysis of acoustic emission signals from landing gear components : an aid to fatigue fracture detection. *Strain*, 47 :e588–e594, 2011.
- [63] O.I. Traore, P. Cristini, N. Favretto-Cristini, L. Pantera, P. Vieu, and S. Viguier-Pla. Clustering acoustic emission signals by mixing two stages dimension reduction and nonparametric approaches. *Computational Statistics*, 01 2019.
- [64] S. Borucki, J. Łuczak, and M. Lorenc. The application of selected hierarchical clustering methods for classification the acoustic emission signals generated by partial discharges. *Archives of Acoustics*, vol. 46(No 3) :409–417, 2021.
- [65] H. Hamdan and G. Govaert. Mixture model approach for acoustic emission control of cylindrical pressure equipment. *Journal of Pressure Vessel Technology-transactions of The Asme - J Pressure Vessel Technol*, 128, 08 2006.
- [66] A. Farhidzadeh, E. Dehghan-Niri, and S. Salamone. Gaussian mixture modeling of acoustic emissions for structural health monitoring of reinforced concrete structures. *Proceedings of SPIE - The International Society for Optical Engineering*, 8692, 03 2013.

- [67] S. Omkar, S. Suresh, and V. Raghavendra, T.and Mani. Acoustic emission signal classification using fuzzy c-means clustering. volume 4, pages 1827 – 1831 vol.4, 12 2002.
- [68] H. Mohamed. Acoustic emission data clustering for the analysis of damage mechanisms using fuzzy c-means method. 12 2018. <https://www.researchgate.net/publication>.
- [69] E. Ramasso, V. Placet, and M. L. Boubakar. 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, Dec 2015.
- [70] R. Gray and D. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6) :2325–2383, 1998.
- [71] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- [72] Michael Steinbach, George Karypis, and Vipin Kumar. A comparison of document clustering techniques. *Proceedings of the International KDD Workshop on Text Mining*, 06 2000.
- [73] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301) :236–244, 1963.
- [74] L. Candillier. *Contextualisation, Visualisation et Evaluation en Apprentissage Non Supervisé*. Theses, Université Charles de Gaulle - Lille III, September 2006.
- [75] Diab Abuaiadah. Using bisect k-means clustering technique in the analysis of arabic documents. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 15 :1–13, 01 2016.
- [76] J. Verbeek and B. Vlassis, N.and Krose. Efficient greedy learning of gaussian mixture models. *Neural computation*, 15 :469–85, 03 2003.
- [77] Geoffrey McLachlan and David Peel. *Finite mixture models : McLachlan/finite mixture models*. 2005.
- [78] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.
- [79] S. Amidi and A. Amedi. Cs 229 - machine learning. <https://stanford.edu/~shervine/l/fr/teaching/cs-229/pense-bete-apprentissage-non-supervise>.
- [80] D. E. Gustafson and W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pages 761–766, 1978.
- [81] R. Babuka, P. J. van der Veen, and U. Kaymak. Improved covariance estimation for gustafson-kessel clustering. In *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291)*, volume 2, pages 1081–1085 vol.2, 2002.

- [82] S. Monti, C. Tamayo, J. Mesirov, and J. Golub. Consensus clustering : A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, (52) :91–118, 2003.
- [83] Andrey Goder and Vladimir Filkov. Consensus clustering algorithms : Comparison and refinement. In *Proceedings of the Meeting on Algorithm Engineering and Experiments*, page 109–117, USA, 2008. Society for Industrial and Applied Mathematics.
- [84] Davy Dalmas, Salim Benmedakhene, Caroline Richard, Abdelouahed Laksimi, Gérard Béranger, and Thierry Grégoire. Caractérisation par émission acoustique de l’adhérence et de l’endommagement d’un revêtement : cas d’un revêtement wc-co sur acier. *Comptes Rendus de l’Académie des Sciences - Series IIC - Chemistry*, 4(5) :345 – 350, 2001.
- [85] A. Monti, A. El Mahi, L. Guillamat, and Z. Jendli. Analyse de l’endommagement d’un composite à matrice thermoplastique renforcée de fibres de lin par émission acoustique. In *22th French Congress of Mechanics*, Lyon, France, 2017.
- [86] M. Aslan. Investigation of damage mechanism of flax fibre lpet commingled composites by acoustic emission. *Composites : Part B*, 54(1) :289–297, 2013.
- [87] A. Bravo, L. Toubal, D. Koffi, and F. Erchiqui. Characterization of tensile damage for a short birch fiber-reinforced polyethylene composite with acoustic emission. *International Journal of Material Science*, 3(13) :79–89, 2010.
- [88] Ghazi Mufti, Patrice Bertrand, and L Moubarki. Determining the number of groups from measures of cluster validity. *Proceedings of ASMDA 2005*, 01 2005.
- [89] Bernard Desgraupes. Clustering indices. 11 2017. <https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf>.
- [90] K. Wang, B. Wang, and L. Peng. Cvap : Validation for cluster analyses. *Data Science Journal*, 8 :88–93, 05 2009.
- [91] P. Rousseeuw. Silhouettes : A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20 :53–65, 1987.
- [92] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1) :1–27, 1974.
- [93] D. Davies and D. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2) :224–227, 1979.
- [94] L. Calabrese, G. Campanella, and E. Proverbio. *Use of Acoustic Emission Data Clustering to Identify Damage Mode in Concrete Structures*, volume 6, pages 329–334. 01 2013.
- [95] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336) :846–850, 1971.

- [96] G. Forestier, C. Wemmert, and P. Gancarski. Background knowledge integration in clustering using purity indexes. 10 2018. <https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf>.
- [97] A. Fred and A. Jain. Combining multiple clusterings using evidence accumulation. *IEEE transactions on pattern analysis and machine intelligence*, 27 :835–50, 07 2005.
- [98] P. Gancarski, A. Cornuéjols, and Y. Bennani. Clustering collaboratif : Principes et mise en oeuvre. Nancy, France, 11 2017. BDA'17.
- [99] Alexander Hohl, Adam D Griffith, Martha Cary Eppes, and Eric Delmelle. Computationally enabled 4d visualizations facilitate the detection of rock fracture patterns from acoustic emissions. *Rock Mechanics and Rock Engineering*, 51(9) :2733–2746, 2018.
- [100] I. Rastegaev, D. Merson, I. Rastegaeva, and A. Vinogradov. A time-frequency based approach for acoustic emission assessment of sliding wear. *Lubricants*, 52(7) :1–24, 2020.
- [101] M.K. Pakhira, S. Bandyopadhyay, and U. Maulik. Validity index for crisp and fuzzy clusters. *Pattern Recognition*, 37(3) :487 – 501, 2004.
- [102] J.C. Bezdek and N.R. Pal. Some new indexes of cluster validity. *IEEE Trans. on Syst., Man, and Cybernetics, Part B (Cybernetics)*, 28(3) :301–315, 1998.
- [103] Xiaoyang Yu, Mourad Bentahara, Charfeddine Mechrib, and Silvio Montrésor. Passive monitoring of nonlinear relaxation of cracked polymer concrete samples using acoustic emission. *The Journal of the Acoustical Society of America, Express Letters*, 146 :323–328, 2019.
- [104] J.B. Machado, W.C. Amaral, and R.J.G.B. Campello. Design of obf-ts fuzzy models based on multiple clustering validity criteria. In *19th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI 2007)*, volume 2, pages 336–339. IEEE, 2007.
- [105] K. Kryszczuk and P. Hurley. Estimation of the number of clusters using multiple clustering validity indices. In *Int. Workshop on Multiple Classifier Systems*, pages 114–123. Springer, 2010.
- [106] M. Sause, A. Gribov, A. Unwin, and S. Horn. Pattern recognition approach to identify natural clusters of acoustic emission signals. *Pattern Reco. Lett.*, 33 :17–23, 2012.
- [107] E. Ramasso, X. Gabrion, DD. Doan, M. Kharrat, V. Placet, and L. Boubakar. Reconnaissance des sources acoustiques dans les composites à matrice organique : quel(s) critère(s) utiliser pour une classification non-supervisée des signaux? In *Congrès Français d'Acoustique*, pages 1611–1617. 2014.
- [108] H. Sawan, M. E Walter, and B. Marquette. Unsupervised learning for classification of acoustic emission events from tensile and bending experiments with open-hole carbon fiber composite samples. *Composites Science and Technology*, 107 :89–97, 2015.

- [109] Neha Chandarana. *Combining passive and active methods for damage mode diagnosis in tubular composites*. PhD thesis, Manchester University, Faculty of Science and Engineering, Departement of Materials, 18th November 2019.
- [110] C. Cachin. Entropy measures and unconditional security in cryptography. 1997. Reprint as vol. 1 of ETH Series in Information Security and Cryptography, ISBN 3-89649-185-7, Hartung-Gorre Verlag, Konstanz, 1997.
- [111] F. Nielsen and S. Boltz. The burbea-rao and bhattacharyya centroids. *CoRR*, abs/1004.5049, 2010. <http://arxiv.org/abs/1004.5049>.
- [112] J. Smith. Some observations on the concepts of information-theoretic entropy and randomness. *Entropy*, 3(1) :1–11, 2001.
- [113] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1) :79 – 86, 1951.
- [114] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3) :379–423, 1948.
- [115] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4) :620, 1957.
- [116] A. De Martino and D. De Martino. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, 4(4) :e00596, 2018.
- [117] M. Mbarga Nkogo, E. Ramasso, P. Le Moal, G. Bourbon, B. Verdin, and G. Chevallier. Joint optimization of the number of clusters and their parameters in acoustic emission clustering. In Piervincenzo Rizzo and Alberto Milazzo, editors, *European Workshop on Structural Health Monitoring*, pages 109–116, Cham, 2021. Springer International Publishing.
- [118] J. Galbraith and V. Zinde-Walsh. Evaluation de critère d’information pour les modèles de séries chronologiques. *Actualité économique*, 2-3 2004.
- [119] S. Tufféry. *Data Mining et Statistique décisionnelle : La science des données*. 2017. <https://books.google.fr/books?id=ijOMtAEACAAJ>.
- [120] L. Ljung. *System Identification : Theory for the User*. 1998. Pearson Education.
- [121] E. Côme, L. Oukhellou, T. Denoeux, and P. Akinin. Learning from partially supervised data using mixture models and belief functions. *Pattern recognition*, 42(3) :334–348, 2009.
- [122] Wemmert Cédric and Forestier Germain. Slemc : Apprentissage semi-supervisé enrichi par de multiples clusterings. *Revue des Nouvelles Technologies de l’Information*, Fouille de données complexes. Complexité liée aux données multiples, RNTI-E-21 :147–170, 2011.
- [123] E. Côme. *Apprentissage de modèles génératifs pour le diagnostic de systèmes complexes avec labellisation douce et contraintes spatiales*. Compiègne, France, 2009. <http://www.theses.fr/2009COMP1796>.

- [124] E. Côme, L. Oukhellou, T. Denoeux, and P. Akinin. Mixture model estimation with soft labels. In *Fourth International Workshop on Soft Methods in Probabilities and Statistics*, pages 165–174, Toulouse, France, 09 2008.
- [125] T. Denoeux. Maximum likelihood estimation from uncertain data in the belief function framework. *Knowledge and Data Engineering, IEEE Trans. on*, 25(1) :119–130, 2013.
- [126] G. Forestier. *Connaissances et clustering collaboratif d’objets complexes multisources*. PhD thesis, Strasbourg, France, 2010. <http://www.theses.fr/2010STRA6075/document>.
- [127] A. P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(2) :205–247, 1968.
- [128] G. Shafer. *A mathematical theory of Evidence*. Princeton University Press, 1976.
- [129] Philippe Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(1-2) :1–30, 2002.
- [130] K. Murphy. Conjugate bayesian analysis of the gaussian distribution. 11 2007. <https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf>.
- [131] Matthew J. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.
- [132] D. Alvares, C. Armero, A. Forte, and L. Rubio. Dirichlet-multinomial model : The impact of prior distributions. In *Conference : 11th International Workshop on Objective Bayes Methodology*, Valencia, Spain, 2015.
- [133] J. Berger, J. Bernardo, and D. Sun. Overall objective priors. *Bayesian Analysis*, 10, 04 2015.
- [134] W. Perks. Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries*, 72, 01 1947.
- [135] H. Jeffreys. The theory of probability. *Journal of the American Statistical Association*, 57, 01 1948.
- [136] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53 :370–418, 01 1900.
- [137] Brendan Van Rooyen and Robert C Williamson. A theory of learning with corrupted labels. *J. Mach. Learn. Res.*, 18(1) :8501–8550, 2017.
- [138] H. Attias. A variational bayesian framework for graphical models. In *NIPS*, volume 12, 1999.
- [139] Joshua J. Radice, Victor Farm-Guoo Tseng, Trevon Drummond, Daniel Diamond, Natalie Schieuer, and Sarah Bedair. Multiphysics modeling of ultrasonic elastic wave attenuation for wireless power transfer including viscoelasticity and acoustic emission. *Mechanics Research Communications*, 115 :103600, 2021.
- [140] Tarek Berghout, Leïla-Hayet Mouss, Ouahab Kadri, Lotfi Saïdi, and Mohamed Benbouzid. Aircraft engines remaining useful life prediction with an adaptive denoising online sequential extreme learning machine. *Engineering Applications of Artificial Intelligence*, 96 :103936, 2020.

- [141] Yuan Hao W. and Yi Qing N. Variational autoencoder-based approach for rail defect identification. In Fu-Kuo Chang, Alfredo Guemes, and Fotis Kopsaftopoulos, editors, *Structural Health Monitoring 2019*, pages 2818–2824. DEStech Publications Inc., January 2019.
- [142] Xirui Ma, Lin Yizhou, Zhenhua Nie, and Hongwei Ma. Structural damage identification based on unsupervised feature-extraction via variational auto-encoder. *Measurement*, 160 :107811, 04 2020.
- [143] L. Wen, L. Gao, and X. Li. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 49(1) :136–144, 2019.
- [144] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. <http://www.deeplearningbook.org>.
- [145] Martin Mbarga Nkogo, Emmanuel Ramasso, Patrice Le Moal, and Gilles Bourbon. A variational bayesian clustering approach to acoustic emission interpretation including soft labels. In *International Conference on Belief Functions*, 2022. Accepted.
- [146] Yoshua Bengio. Learning deep architectures for ai. *Foundations and Trends® in Machine Learning*, 2(1) :1–127, 2009.
- [147] Lilian Weng. From autoencoder to beta-vae. *lilianweng.github.io*, 2018. <https://lilianweng.github.io/posts/2018-08-12-vae>.
- [148] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set : A strategy employed. *Vision Research*, 37(23) :3311–3325, 1997.



# Annexes

## A.1 Validation conventionnelle à l'aide du critère séquentielle avec pénalité

Les résultats obtenus avec la validation à partir du critère séquentiel muni d'une pénalité de  $0.5\lambda_1$  ( avec  $\lambda_1 = \log_2(1 + 1/K)$  ) pour l'essai #B sont regroupés dans la figure A.1. Avec cette validation les cinq *clusters* obtenus ont un lien direct avec les serrages (voir Figure A.1a). Le premier cluster correspond au premier serrage, le second au deuxième et quatrième serrage. Le troisième et le quatrième *clusters* représentent les troisième et cinquième niveaux de serrage. Le dernier *cluster* est constitué des deux derniers serrages.

Les performances globales sont évaluées avec les diagrammes en bloc des dix premières partitions du tri (figure A.1b). La formulation d'un troisième critère a été faite en ajoutant une contrainte sur la séparabilité des clusters dans le temps. La courbe A.1a est obtenue en deux boucles. La première en sélectionnant  $K$  qui maximise le critère avec pénalité pour le nombre de *clusters*. Puis en maximisant ce même critère pour le choix du sous-ensemble de descripteurs. Le graphe A.1b donne les valeurs moyennes d'ARI obtenues avec différentes validations. Les valeurs médianes sont données par des points au milieu des diagrammes en boîtes. Les valeurs les moins bonnes sont données par le vote basé sur des critères de forme. En utilisant le critère séquentiel avec pénalité (ESp) les valeurs performances sont nettement améliorées. En dernier une nouvelle spécificité est ajoutée au critère, la séparabilité temporelle des *clusters*. Cette contrainte est traduite par la construction d'une partition de référence  $P_{ref}$  dont chaque cluster est constitué de tous les points après son déclenchement jusqu'à l'apparition du cluster suivant.  $P$  est la partition estimée.  $EspAri = Esp + 0.5 \times ari(P, P_{ref})$ . Avec cette nouvelle spécificité, les performances sont améliorées pour le GMM et la méthode hiérarchique.

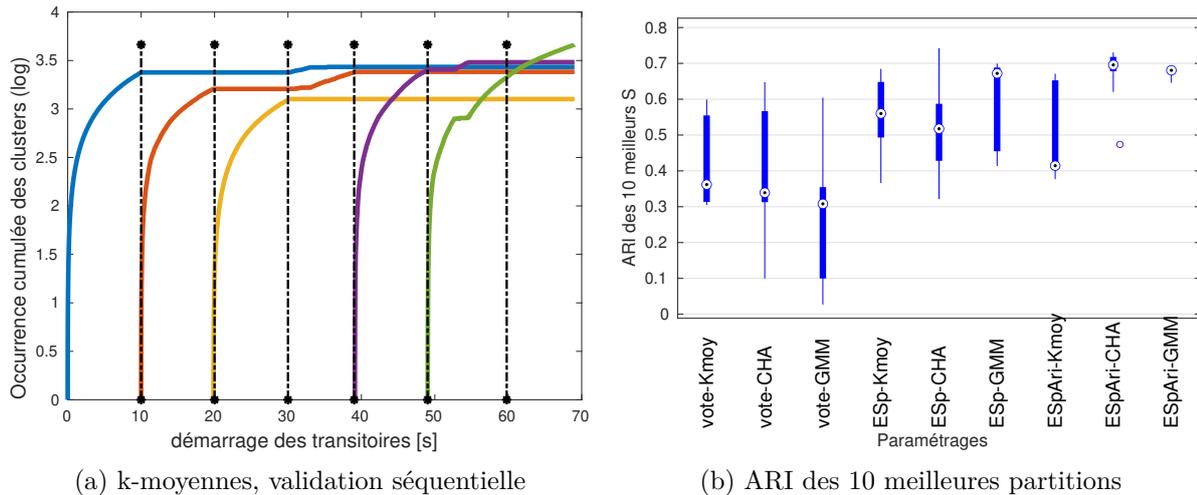


Figure A.1 – Essai #B : Validation conventionnelle avec le critère séquentiel muni d'une pénalité  $0.5\lambda_1$ .

## A.2 Validation consensuelle avec des instants de démarrage séquentiel logarithmiques

Les résultats obtenus pour l'essai #B avec des temps logarithmiques sont regroupés à la figure A.2. De manière globale, les meilleurs résultats sont obtenus avec le critère séquentiel par les deux méthodes d'évaluation (ARI et histogrammes).

## A.3 Réduction de dimension

La réduction de dimension consiste à réduire le nombre de variables (descripteurs) d'apprentissage, ce qui permet minimiser les coûts de calcul et de stockage des données, et de gérer les problèmes liés au « fléau de dimension » [60,61]. Le nouvel espace réduit permet une meilleure visualisation des données, améliorant ainsi les analyses et les prises de décision.

Reprenons les notations du chapitre 1,  $\mathbf{y}[N \times D]$  est un ensemble de  $N$  observations dans  $\mathbb{R}^D$  générées par des états  $\mathbf{x}$ . Soit une fonction bijective  $f : \mathbb{R}^D \rightarrow \mathbb{R}^d$  qui associe à chaque observation  $y_n$  une représentation  $z_n$  (avec  $n = 1 \dots N$  et  $d < D$ ). La représentation  $\mathbf{z}[N \times d]$  obtenue est appelée *code*. Elle doit permettre de reconstruire l'entrée lors d'une phase de *décodage*. L'encodage et le décodage mis ensemble doivent minimiser l'erreur quadratique moyenne entre les données et les données reconstruites. Indépendamment de l'approche utilisée, la réduction de dimension se résume à trouver un code optimal  $\mathbf{z}^*$  (respectivement une fonction  $f$ ) tel que

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{n=1}^N \|\tilde{y}_n - y_n\|^2 \quad (\text{A.1})$$

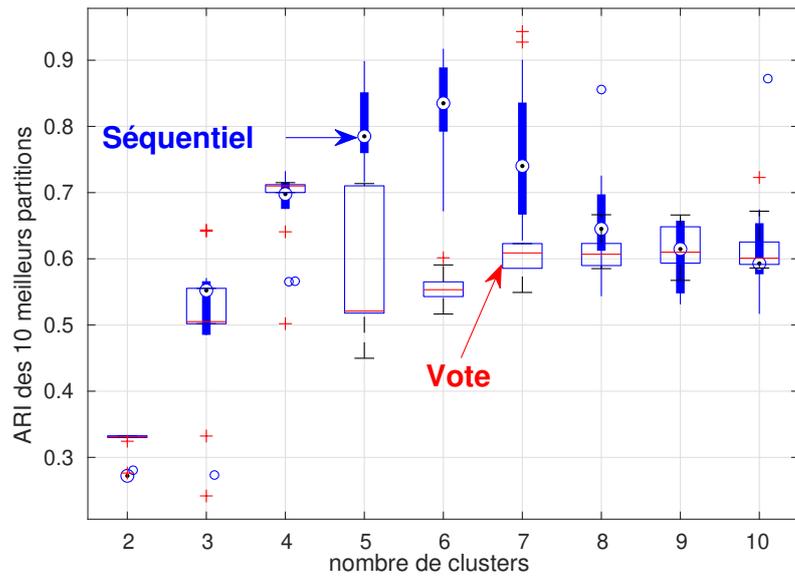
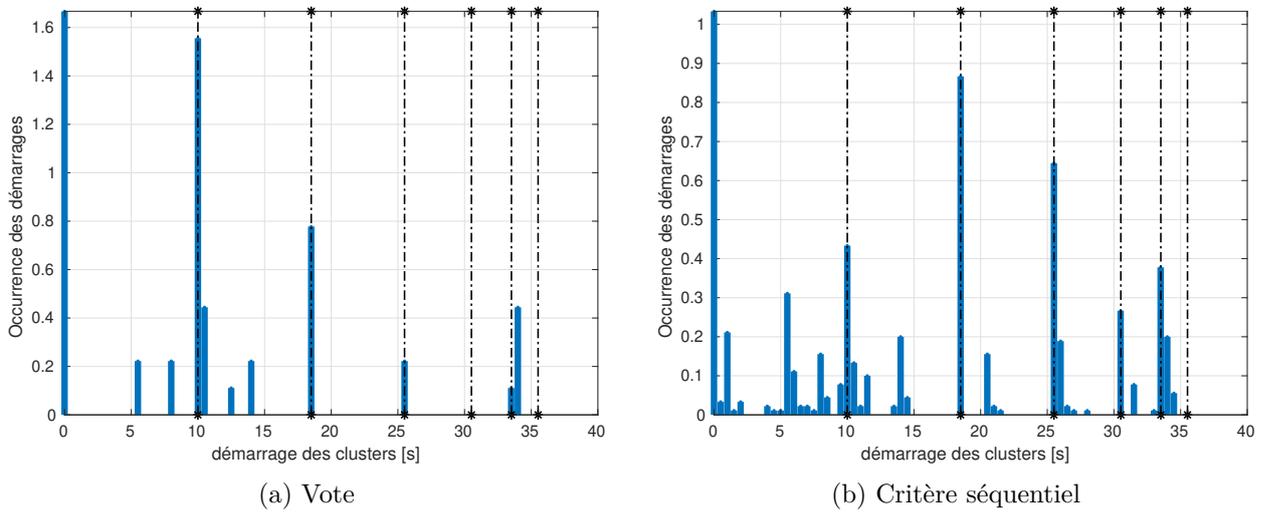


Figure A.2 – Essai #B : Validation consensuelle avec des durées de serrage logarithmiques. L'idée est de simuler un desserrage naturel. Le boulon se desserre plus rapidement lorsque le niveau de serrage est faible. Les boîtes bleues compactes A.2c correspondent au critère séquentiel et celles larges au vote

où  $\tilde{\mathbf{y}} = g(\mathbf{z})$  est l'estimation de l'entrée  $\mathbf{y}$  et  $g$  la fonction de décodage. Les approches de réduction de dimension sont dites non-supervisées car elles n'apprennent pas une sortie spécifique en fonction des entrées.

Dans ce manuscrit nous nous intéressons à l'analyse en composantes principales (ACP) et aux auto-encodeurs, deux réponses à la problématique de réduction de dimension. L'ACP est l'approche la plus utilisée pour l'analyse exploratoire des données d'EA. Elle produit un code dont les variables sont orthogonales et linéairement indépendantes entre elles. L'ACP permet de réduire la redondance au niveau des variables.

Les auto-encodeurs sont des nouvelles approches basées sur les *réseaux de neurones artificiels* (RNA) pouvant trouver dans certaines conditions des corrélations non linéaires complexes dans les données. Ces corrélations sont parfois difficiles (voire impossibles) à trouver à partir d'une ACP. Cette section est un résumé des chapitres 2 et 14 de [144] portant sur la PCA et les auto-encodeurs, respectivement.

### A.3.1 L'analyse en composantes principales

L'ACP est une méthode statistique permettant de transformer des observations corrélées en nouvelles variables linéairement indépendantes. La dimension  $d$  du nouvel espace de représentation est inférieure ou égale à celle des observations  $D$ . La particularité d'une ACP est que les nouvelles variables appelées "*composantes principales*" sont orthogonales (indépendantes) les unes des autres. L'approche présente un intérêt géométrique grâce à la visualisation des données dans un meilleur espace de représentation suivant les directions qui maximisent l'inertie. Statiquement l'intérêt réside dans la possibilité d'expliquer la variabilité des données en quantifiant l'impact de chaque composante principale sur la variance totale des données. Chacune des composantes principales  $z_i = f_i(\mathbf{y})$  (avec  $i = 1 \dots D$ ) est fonction des observations. L'évaluation de l'impact d'une composante est faite en estimant  $f_i^{-1}(z_i)$  la fonction inverse de  $f_i$ . Conventionnellement les premières composantes sont celles qui expliquent le mieux la variance. Pour la réduction de dimension, seules les  $d < D$  premières composantes sont prises en compte en concédant une perte d'information (variance / inertie). Les premiers travaux sur l'ACP remontent au début du vingtième siècle. Dès lors l'approche a été utilisée dans divers domaines : traitement d'image, compression des données, analyse vocale, l'apprentissage automatique (*machine learning*). L'ACP est la méthode systématique pour la réduction de dimension dans la communauté de l'EA et du SHM [20, 62].

#### Estimation des paramètres

La fonction d'encodage  $f$  à  $\mathbf{y}$  associe un code  $\mathbf{z} = f(\mathbf{y})$ ; et  $g(\mathbf{z})$  la fonction de décodage reconstruit les entrées  $\tilde{\mathbf{y}}$  avec une erreur  $\|\tilde{\mathbf{y}} - \mathbf{y}\|^2$ . L'ACP propose une solution qui minimise cette erreur. Les composantes principales étant des fonctions linéaires,  $f$  et  $g$  sont définies ainsi

$$\begin{cases} \mathbf{z} = f(\mathbf{x}) = C^T \mathbf{x}, \text{ l'encodage} \\ \tilde{\mathbf{y}} = g(\mathbf{z}) = Q\mathbf{z}, \text{ le décodage} \end{cases} \quad (\text{A.2})$$

$C$  et  $Q$  sont des matrices réelles de dimension  $[d \times D]$  et  $.^T$  l'opérateur transposée.

L'ACP a pour objectif de trouver la paire  $(C, Q)$  qui minimise  $\|\tilde{\mathbf{y}} - \mathbf{y}\|^2$  comme décrit dans l'équation A.1. Les matrices  $C$  et  $Q$  étant dépendantes, cette optimisation peut se réduire à une optimisation de  $Q$ . Dans la méthode proposée dans [144, chap.2], l'optimisation se concentre sur l'optimisation du décodeur  $Q$ .

Plusieurs solutions sont possibles. L'ACP ajoute deux spécificités à la solution

- Les colonnes de la matrice  $Q$  sont orthogonales entre elles.
- La matrice  $Q$  est de taille  $[D \times D]$  et chacune de ses colonnes est un vecteur de norme 1. En effet chacun des vecteurs proportionnels à une des colonnes est proportionnel aux autres. La normalisation conduit à une solution unique.

le choix de la norme qu'on cherche à maximiser se porte sur la norme  $L2$ . En remplaçant  $\tilde{\mathbf{y}}$  par  $g(\mathbf{z})$  l'équation A.1 à optimiser devient

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{n=1}^N \|y_n - g(z_n)\|_2^2 \quad (\text{A.3})$$

On se propose d'utiliser une notation simplifiée sans l'indication des sommes.

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} (\mathbf{y} - g(\mathbf{z}))^2 \quad (\text{A.4})$$

La fonction objectif peut s'écrire :

$$(\mathbf{y} - g(\mathbf{z}))^2 = (\mathbf{y} - g(\mathbf{z}))^T (\mathbf{y} - g(\mathbf{z})) \quad (\text{A.5})$$

Après développement et simplification on obtient

$$(\mathbf{y} - g(\mathbf{z}))^2 = \underbrace{\mathbf{y}^T \mathbf{y}}_{\text{indépendant de } \mathbf{z}} - \underbrace{2\mathbf{y}^T g(\mathbf{z}) + g(\mathbf{z})^T g(\mathbf{z})}_{\text{nouvelle fonction objectif}} \quad (\text{A.6})$$

Le premier terme de l'expression est une constante qui dépend uniquement des observations. Il faut donc optimiser la seconde partie de l'équation.

Le nouvel objectif peut s'écrire

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} -2\mathbf{y}^T g(\mathbf{z}) + g(\mathbf{z})^T g(\mathbf{z}) \quad (\text{A.7})$$

En remplaçant  $g(\mathbf{z})$  par  $Q\mathbf{z}$  dans l'équation précédente, on obtient :

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} -2\mathbf{y}^T Q\mathbf{z} + \mathbf{z}^T Q^T Q\mathbf{z} \quad (\text{A.8})$$

Les contraintes liées à la normalisation et l'orthogonalité de  $Q$  font que  $Q^T Q = I_d$ , la matrice identité.

Après simplification, on obtient

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} -2\mathbf{y}^T Q\mathbf{z} + \mathbf{z}^T \mathbf{z} \quad (\text{A.9})$$

La solution est optimale lorsque la dérivée en fonction de  $\mathbf{z}$  est nulle.

$$\nabla_{\mathbf{z}}(-2\mathbf{y}^T Q\mathbf{z} + \mathbf{z}^T \mathbf{z}) = 0 \quad (\text{A.10})$$

On obtient :

$$-2\mathbf{y}^T Q + 2\mathbf{z} = 0 \quad (\text{A.11})$$

Le code optimal est obtenu lorsque :

$$\mathbf{z} = Q^T \mathbf{y} \quad (\text{A.12})$$

Le système d'équations A.2 de l'encodage et du décodage devient

$$\begin{cases} \mathbf{z} = f(\mathbf{x}) = Q^T \mathbf{x} , \text{ l'encodage} \\ \tilde{\mathbf{y}} = g(\mathbf{z}) = Q\mathbf{z} , \text{ le décodage} \end{cases} \quad (\text{A.13})$$

Cette résolution conduit à une nouvelle estimation de la fonction d'estimation (encodage + décodage)

$$\tilde{\mathbf{y}} = g \circ f(\mathbf{y}) = QQ^T \mathbf{y} \approx \mathbf{y} \quad (\text{A.14})$$

La seconde étape consiste à trouver la fonction d'encodage optimale  $f$  (c'est-à-dire la matrice  $Q$ ) pour la minimisation de l'erreur. Une autre façon d'aborder le problème est de maximiser la variance (qui est maximale lorsque l'erreur est faible). On cherche à projeter les observations  $\mathbf{y}[N \times D]$  dans un espace réduit. Il faut donc trouver les directions qui maximisent la conservation de la variété (variance des données). On se propose alors de trouver cette matrice  $Q$  en maximisant la matrice de covariance dans l'espace de projection défini par

$$\begin{cases} \mathbf{S}_p = \frac{1}{N} \sum_{n=1}^N (Q^T y_n - Q^T \bar{y})^2 \\ = Q^T \mathbf{S} Q , \text{ la matrice de covariance} \\ \mathbf{S} = \frac{1}{N} \sum_{n=1}^N (y_n - \bar{y})(y_n - \bar{y})^T , \text{ la matrice de covariance} \end{cases} \quad (\text{A.15})$$

$\bar{y}$  représente la moyenne des observations, on rappelle aussi que  $Q^T Q = I_D$ . En introduisant un coefficient de Lagrange  $\lambda$ , la valeur à maximiser devient

$$\mathbf{S}_p = Q^T \mathbf{S} Q + \lambda(I_D - Q^T Q) \quad (\text{A.16})$$

La valeur recherchée est

$$Q^* = \underset{Q}{\operatorname{argmin}} Q^T \mathbf{S} Q + \lambda(I_D - Q^T Q) \quad (\text{A.17})$$

Elle est atteinte lorsque la dérivée suivant le paramètre  $Q$  est nulle.

$$\nabla_Q(Q^T \mathbf{S} Q + \lambda(I_D - Q^T Q)) = 0 \quad (\text{A.18})$$

On obtient l'expression

$$\begin{cases} \mathbf{S} Q = \lambda \\ Q^T \mathbf{S} Q = \lambda, \text{ en multipliant par } Q^T \end{cases} \quad (\text{A.19})$$

L'équation A.19 correspond à la définition de *valeur propre* pour  $\lambda$  et de la matrice des *vecteurs propres* pour  $Q$ . Il existe  $D$  vecteurs propres. Une valeur propre peut correspondre à un ou plusieurs vecteurs propres.

Il est démontré que les composantes  $Q_{i=1\dots D}$  expliquant le mieux la variance correspondent aux valeurs propres les plus grandes. Une analyse en composantes principales classique consiste donc en trois phases dont la première est l'estimation des valeurs propres de  $\mathbf{S}$  la matrice de covariance et à fixer  $d$  la taille de l'espace de projection qu'on souhaite. La seconde phase consiste à estimer les vecteurs propres correspondant à ces valeurs et à trier les  $d$  premiers suivant les valeurs croissantes de  $\lambda$ . La dernière étape est l'estimation du code  $\mathbf{z}$ .

$$\mathbf{z} = Q_{\{d\}}^T \mathbf{y} \quad (\text{A.20})$$

$Q_{\{d\}}$  est la matrice constituée des premiers vecteurs propres (*composantes principales*).

Il est à noter que la complexité d'un tel calcul est  $O(D^3)$ . Le calcul peut donc s'avérer très difficile, voire impossible pour des observations du type images par exemple où les dimensions sont trop grandes. Il existe une version de l'algorithme permettant de contourner le problème sans utiliser la méthode analytique présentée dans ce chapitre.

### Application à l' émission acoustique

L'ACP est utilisée lors de l'analyse des données d'EA principalement pour la réduction de dimension de l'espace des descripteurs, l'extraction des descripteurs (rarement) et la visualisation en 2D des données. Cette étape précède l'apprentissage (supervisé ou non supervisé) [20, 62].

Le choix des composantes principales en EA suit une approche distincte de celle présentée à la précédente section. Après avoir déterminé toutes les valeurs propres et trié les vecteurs propres suivant leur impact sur la variance, on estime cet impact en pourcentage pour chacune des composantes. La

somme des impacts des  $D$  composantes explique entièrement la variance.

$$\sum_{i=1}^D \text{Impact}(Q_i) = 100\% \quad (\text{A.21})$$

Une perte acceptable est fixée, souvent entre 1% et 20%. Le nombre de composantes  $d$  correspond aux premiers vecteurs propres dont la somme cumulée des impacts est supérieure ou égale à la perte tolérée

$$\sum_{i=1}^d \text{Impact}(Q_i) = 100\% - \text{perte} \quad (\text{A.22})$$

Si l'ACP est utilisée pour la visualisation, on se limite à deux à trois composantes. Des graphes en 2D et 3D permettent de visualiser les nuages de points.

L'utilisation de l'ACP lorsqu'elle précède la partition, élimine l'étape de sélection des descripteurs pertinents. En faisant varier le nombre de *clusters*, plusieurs sont faites chacune générant une partition. Le choix de la « meilleure partition » se fait par une validation mono/multicritère [20, 62].

### A.3.2 Illustration sur ORION-EA

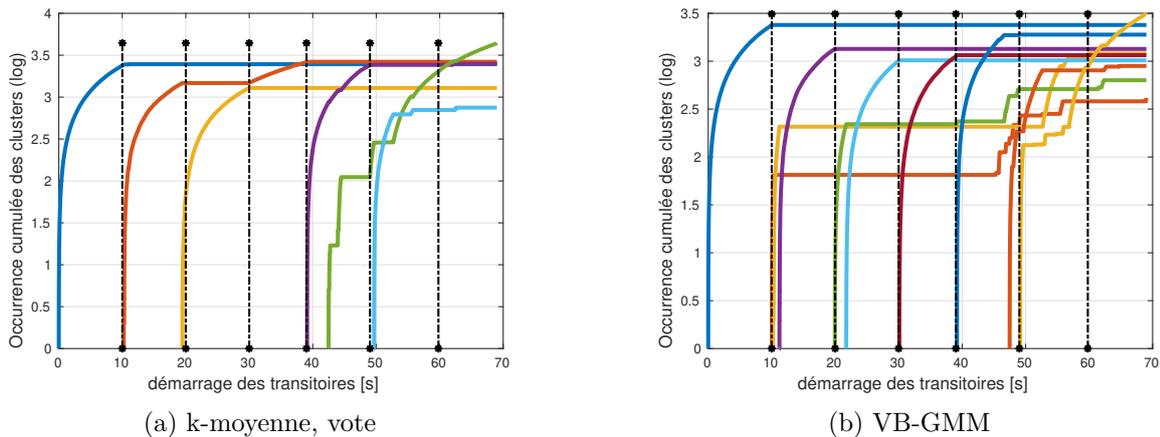
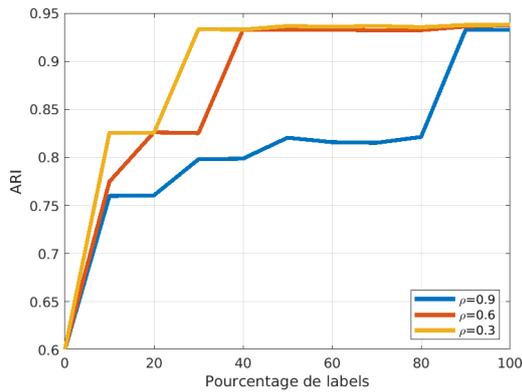


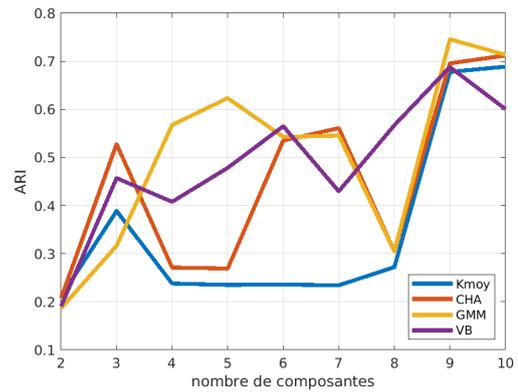
Figure A.3 – Essai #B : Résultats obtenus par le partitionnement après réduction de dimension par une analyse en composantes principales avec une conservation de la variance à 95 %, ce qui fait 10 composantes principales. Les performances et graphes sont quasi-identiques pour les algorithmes k-moyennes, le partitionnement hiérarchique et les GMM. Seuls les k-moyennes et le VB-GMM sans a priori sont illustrés.

### A.3.3 L'auto-encodeur

Aussi appelé **auto-associeur**, "un auto-encodeur (AE) est un réseau de neurones artificiels multicouches ayant pour but l'apprentissage d'une représentation d'un ensemble de données [144, 146]. Cet apprentissage consiste en une étape d'**encodage** qui crée une représentation réduite  $\mathbf{z}[N \times d]$  des données d'entrée  $\mathbf{y}[N \times D]$  (avec  $d < D$ ). La seconde phase : le **décodage** fait une estimation  $\tilde{\mathbf{y}}[N \times D]$  des entrées en minimisant l'erreur quadratique moyenne  $\frac{1}{N} \sum_{n=1}^N (\tilde{y}_n - y_n)^2$  entre l'entrée et la sortie.



(a) ARI=f(pourcentage d'étiquettes)



(b) ARI =f(composantes principales)

Figure A.4 – Essai #B : Résultats globaux en fonction du nombre de composantes principales et de la quantité de données annotées. Plus les a priori sont forts, moins il faut des données pour améliorer le partitionnement (A.4a). Pour les méthodes classiques, les valeurs des ARI ont des variations importantes. Le VB-GMM, bien qu'il ne soit pas celui avec les meilleurs ARI est l'algorithme qui garantit un minimum de résultats. Les ARI ne chutent pas A.4b.

Cette application du machine learning est principalement utilisée pour le traitement d'images et la reconnaissance vocale. N'apprenant pas une réponse spécifique à l'entrée (régression ou classification), l'auto-encodeur est considéré comme un modèle non supervisé. Un tel modèle présente deux atouts spécifiques : la réduction non linéaire de la dimension des données et la mise en place de deux unités de traitement, une qui extrait des « descripteurs cachés » et la seconde centrée sur l'apprentissage (supervisé ou non supervisé). L'encodage permet aussi d'éliminer l'étape de sélection des descripteurs pertinents si on choisit une faible dimension  $d$  du code. Un auto-encodeur est constitué d'au moins trois couches :

- Une entrée constituée de  $D$  nœuds, correspondant à la dimension d'une observation  $y_n$ . Les entrées sont représentées dans un même espace.
- Une couche cachée constituée de  $d < D$  nœuds. Elle comporte le code  $\mathbf{z}$  qui est la représentation compressée des entrées.
- Une couche de sortie de même dimension que l'entrée, elle donne l'estimation  $\tilde{\mathbf{y}}$  de  $\mathbf{y}$ .

Chaque couche de neurones a comme entrée les nœuds de la couche précédente (figure A.5).

## Principe

Schématiquement l'AE englobe deux phases :

- La phase d'encodage, dite « *convolutive* » au cours de laquelle la dimension des couches diminue progressivement jusqu'à l'obtention du code  $\mathbf{z}$ . Si les fonctions d'activation sont non-linéaires,  $\mathbf{z}$  extrait des informations dont la complexité dépend du nombre de couches de l'AE.
- La seconde phase « *déconvolutive* » s'accompagne de l'augmentation de la dimension des couches jusqu'à ce qu'elle soit identique à celle d'une observation.

Pour le cas le plus simple avec trois couches, les équations des différentes couches sont données par :

$$\begin{cases} z_j = f(\sum \omega_j y_j + b_j) \text{ l'encodage} \\ \tilde{y}_j = g(\sum \omega'_j z_j + b'_j) \text{ le décodage} \end{cases} \quad (\text{A.23})$$

$f$  et  $g$  sont les fonctions d'activation de chacune des couches, elles peuvent être identiques ou différentes. L'apprentissage est similaire à celui d'un réseau de neurones artificiels avec une rétro-propagation de l'erreur.

**Apprentissage** | L'apprentissage consiste en une suite d'opérations itératives visant à estimer les matrices de poids de chaque couche. À chaque itération l'erreur est calculée et rétro-propagée dans l'estimation des poids à l'aide d'un système de récompense. Les neurones se voient récompensés si le résultat estimé est correct et pénalisés dans le cas contraire.

Plusieurs théories permettent l'estimation de ces poids. La fonction coût à minimiser étant l'erreur quadratique moyenne, les paramètres à estimer sont trouvés à l'aide de la *méthode de gradient stochastique* en respectant la règle de Hebb [144]. Associée à la règle des dérivées chaînées, la règle de

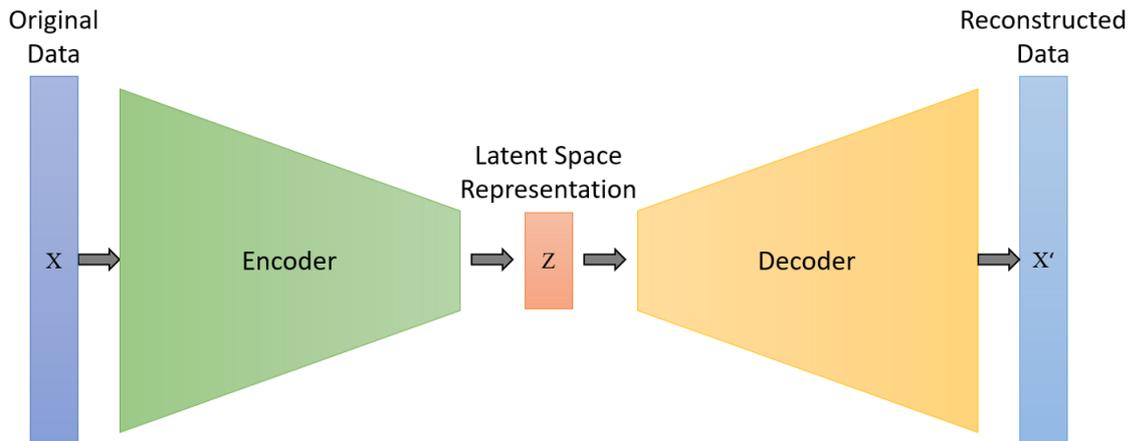


Figure A.5 – Principe d'un auto-encodeur [147]

Hebb donne ces résultats :

$$\begin{cases} \Delta\omega_{jk} = \varepsilon\delta_k a_j \text{ avec } \varepsilon \text{ le taux d'apprentissage et} \\ \delta_k = (d_k - a_k) f'(\sum a_j \omega_{jk}) \text{ l'erreur propagée sur la couche } k \end{cases} \quad (\text{A.24})$$

Pour le cas du perceptron mono-couche, en utilisant la même règle nous retrouvons la fonction d'apprentissage suivante :

$$\Delta\omega = \varepsilon(d_k - o_k)x_n \quad (\text{A.25})$$

avec  $o_k$  le résultat trouvé par l'algorithme du perceptron.

## Variantes d'auto-encodeur et application à l'émission acoustique

Il existe plusieurs variantes de l'auto-encodeur suivant la tâche que l'on souhaite réaliser :

1. L'AE débruiteur (DAE : *denoising autoencoder*), il apprend à reconstruire l'observation réelle à partir d'une entrée bruitée. [140, 144]
2. L'AE épars (SAE : *sparse autoencoder*), il désactive certains poids aléatoirement afin d'apprendre une représentation généralisable des entrées [143, 144, 148].
3. L'AE variationnel (VAE : *variational autoencoder*), il utilise une approche bayésienne pour estimer le code  $\mathbf{z}$  [141, 142, 144].

La principale difficulté lors de l'usage d'un auto-encodeur est la structure du modèle (nombre de couche, la taille de chaque couches, le type d'activation...). Cette difficulté est partagée avec tous les modèles utilisant des neurones artificiels.

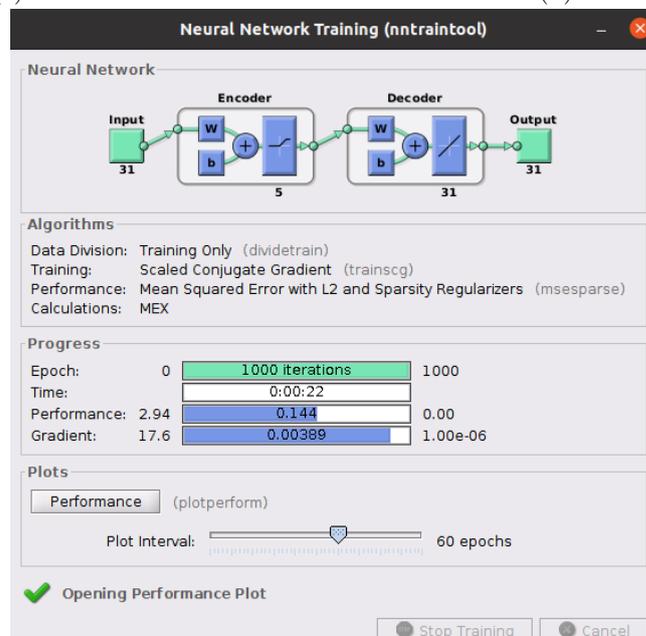
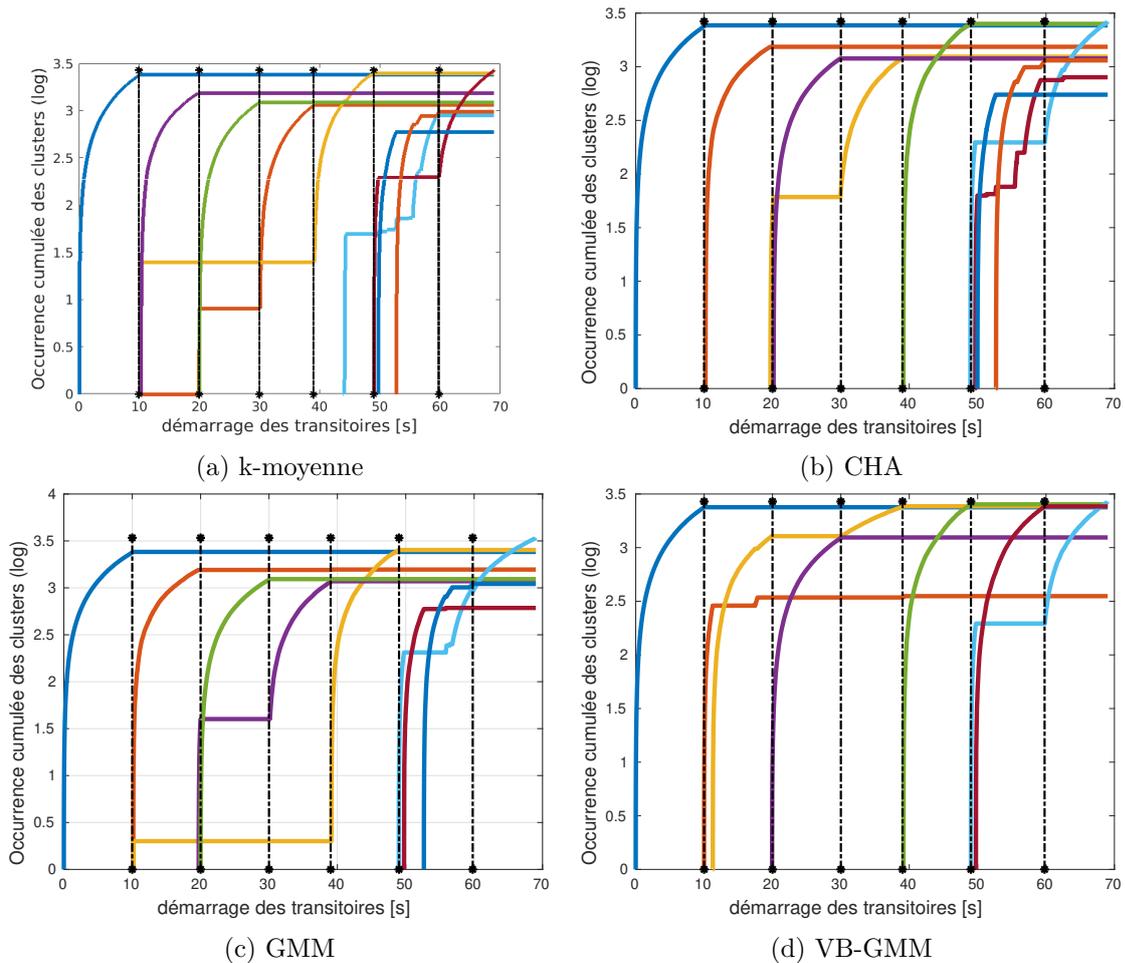
**Applications en émission acoustique et diagnostic** | Dans le cadre de l'émission acoustique, les auto-encodeurs sont utilisés comme la PCA pour réduire le nombre de descripteurs et l'extraction des descripteurs [141, 148]. Cette réduction de dimension permet d'avoir de nouveaux descripteurs orthogonaux et plus discriminants. Les auto-encodeurs sont aussi utilisés sur des signaux bruts afin d'extraire des descripteurs.

Dans l'article [143] par exemple, les auteurs utilisent des auto-encodeurs épars (SAE) pour extraire des descripteurs. Le modèle généré est appliqué sur d'autres données, on parle alors de transfert d'apprentissage (*Transfert Learning*).

Dans l'article [49], les auteurs utilisent les auto-encodeurs variationnels convolutifs pour la caractérisation des endommagements et la localisation des sources d'EA. Ceux-ci ont la particularité d'analyser le signal par morceaux. Un premier bloc constitué d'un auto-encodeur permet d'extraire des descripteurs, qui sont par la suite utilisés par deux blocs dont un sert à la localisation et l'autre à la caractérisation. Dans les articles [140–142], les mêmes méthodes avec des auto-encodeurs sont utilisées pour la caractérisation et l'identification des endommagements pour des turboréacteurs d'avion commerciaux, des rails ferroviaires et des bâtiments respectivement.

L'association de ces différents articles suscités nous a permis d'appréhender deux méthodes d'analyse de données en utilisant dans un premier temps les descripteurs et dans un second nous travaillons avec des salves brutes.

## Illustration avec ORION-EA



(e) Configuration

Figure A.6 – Essai #B : Partitionnement obtenu avec les différents algorithmes en utilisant un auto-encodeur épars avec une couche cachée de 5 neurones et une entrée à 31 descripteurs initiaux. Les serrages sont tous détectés à l'exception du dernier. Les résultats sont très prometteurs, sachant que la configuration n'a pas été optimisée.

## A.4 Estimation de la log-vraisemblance pour le VB-GMM

On cherche  $p(y)$  maximale, on peut suivre un procédé qui consiste à écrire la loi jointe sur  $(y, z)$  :

$$p(\mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{z})}{p(\mathbf{z}|\mathbf{y})} \quad (\text{A.26})$$

On a :

$$\log p(\mathbf{y}) = \mathcal{L}(q) + KL(p(\mathbf{z}|\mathbf{y})) \quad (\text{A.27})$$

L'expression  $q(\mathbf{x}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  de l'équation 3.22 est insérée dans le calcul de la log-vraisemblance des données complètes. On obtient alors :

$$\mathcal{L}(q) = \sum_{\mathbf{x}} \int \int \int q(\mathbf{x}, \pi, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \log \left\{ \frac{p(\mathbf{y}, \mathbf{x}, \pi, \boldsymbol{\mu}, \boldsymbol{\Lambda})}{q(\mathbf{x}, \pi, \boldsymbol{\mu}, \boldsymbol{\Lambda})} \right\} d\pi d\boldsymbol{\mu} d\boldsymbol{\Lambda} \quad (\text{A.28a})$$

$$= \mathbb{E} [\log p(\mathbf{y}, \mathbf{x}, \pi, \boldsymbol{\mu}, \boldsymbol{\Lambda})] - \mathbb{E} [\log q(\mathbf{x}, \pi, \boldsymbol{\mu}, \boldsymbol{\Lambda})] \quad (\text{A.28b})$$

où  $\mathbb{E}$  représente l'espérance. En décomposant chacune des probabilités on obtient :

$$\begin{aligned} \mathcal{L}(q) = & \mathbb{E} [\log p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + \mathbb{E} [\log p(\mathbf{x} | \boldsymbol{\pi})] + \mathbb{E} [\log p(\boldsymbol{\pi})] + \mathbb{E} [\log p(\boldsymbol{\mu}, \boldsymbol{\Lambda})] \\ & - \mathbb{E} [\log q(\mathbf{x})] - \mathbb{E} [\log q(\boldsymbol{\pi})] - \mathbb{E} [\log q(\boldsymbol{\mu}, \boldsymbol{\Lambda})] \end{aligned} \quad (\text{A.29})$$

Le résultat obtenu par l'équation A.29 représente la fonction objectif que l'on cherche à maximiser. L'idée centrale de cet algorithme repose sur le calcul des distributions optimales  $q^{t+1}$  sur une variable (état ou paramètres incertains) en prenant l'espérance du logarithme de la distribution conjointe (3.12) sur les autres variables. La série d'équations suivantes permet d'avoir une estimation de chacun des termes utilisés dans l'équation A.29 [78, chap.10].

L'espérance que les observations  $\mathbf{y}$  soient un mélange des gaussiennes :

$$\begin{aligned} \mathbb{E} [\log p(\mathbf{y} | \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] = & \frac{1}{2} \sum_{k=1}^K N_k \left\{ \log \tilde{\Lambda}_k - D\beta_k^{-1} - D \log (2\pi) \right. \\ & \left. - \nu_k \text{Tr}(\mathbf{S}_k \mathbf{W}_k) - \nu_k (\bar{\mathbf{x}}_k - \mathbf{m}_k)^T \mathbf{W}_k (\bar{\mathbf{x}}_k - \mathbf{m}_k) \right\} \end{aligned} \quad (\text{A.30})$$

L'espérance que les *clusters* soient répartis suivant les proportions  $\boldsymbol{\pi}$  :

$$\mathbb{E} [\log p(\mathbf{x} | \boldsymbol{\pi})] = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \log \tilde{\pi}_k \quad (\text{A.31})$$

L'espérance sur les proportions  $\boldsymbol{\pi}$  (avec  $p(\boldsymbol{\pi})$  une loi de Dirichlet) :

$$\mathbb{E} [\log p(\boldsymbol{\pi})] = \log C(\alpha_0) + (\alpha_0 - 1) \sum_{k=1}^K \log \tilde{\pi}_k \quad (\text{A.32})$$

L'espérance des paramètres ( $\boldsymbol{\mu}$  et  $\boldsymbol{\Lambda}$ ) des gaussiennes,  $p(\boldsymbol{\mu})$  est une gaussienne et  $p(\boldsymbol{\Lambda})$  une distribution

de Wishart :

$$\begin{aligned} \mathbb{E} [\log p(\boldsymbol{\mu}, \boldsymbol{\Lambda})] &= \frac{1}{2} \sum_{k=1}^K \left\{ \log \tilde{\Lambda}_k - D \log \left( \frac{\beta_0}{2\pi} \right) - \frac{D\beta_0}{\beta_k} \right. \\ &\quad \left. - \beta_0 \nu_k (\mathbf{m}_k - \mathbf{m}_0)^\top \mathbf{W}_k (\mathbf{m}_k - \mathbf{m}_0) \right\} + K \log B(\mathbf{W}_0, \nu_0) \\ &\quad + \frac{\nu_0 - D - 1}{2} \sum_{k=1}^K \log \tilde{\Lambda}_k - \frac{1}{2} \sum_{k=1}^K \nu_k \text{Tr}(\mathbf{W}_0^{-1} \mathbf{W}_k) \end{aligned} \quad (\text{A.33})$$

L'espérance sur  $q(\mathbf{x})$  la distribution estimée pour approximer  $p(\mathbf{x})$

$$\mathbb{E} [\log q(\mathbf{x})] = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \log r_{nk} \quad (\text{A.34})$$

L'espérance sur  $q(\boldsymbol{\pi})$  la distribution estimée pour approximer  $p(\boldsymbol{\pi})$

$$\mathbb{E} [\log q(\boldsymbol{\pi})] = \sum_{k=1}^K (\alpha_k - 1) \log \tilde{\pi}_k + \log C(\boldsymbol{\alpha}) \quad (\text{A.35})$$

L'espérance sur  $q(\boldsymbol{\mu}, \boldsymbol{\Lambda})$  la distribution estimée pour approximer  $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$

$$\mathbb{E} [\log q(\boldsymbol{\mu}, \boldsymbol{\Lambda})] = \sum_{k=1}^K \left\{ \frac{1}{2} \log \tilde{\Lambda}_k + \frac{D}{2} \log \left( \frac{\beta_k}{2\pi} \right) - \frac{D}{2} - \mathbb{H}[q(\Lambda_k)] \right\} \quad (\text{A.36})$$

avec  $\mathbb{H}$  l'entropie de la distribution de Wishart :

$$\mathbb{H} = -\log B(\mathbf{W}, \nu) - \frac{(\nu - D - 1)}{2} \mathbb{E} [\log |\boldsymbol{\Lambda}|] + \frac{\nu D}{2} \quad (\text{A.37})$$

et  $B(\mathbf{W}_0, \nu_0)$  est défini par :

$$B(\mathbf{W}_0, \nu_0) = |\mathbf{W}|^{-\nu/2} \left( 2^{\nu D/2} \pi^{D(D-1)/4} \prod_{i=1}^D \Gamma \left( \frac{\nu + 1 - i}{2} \right) \right)^{-1} \quad (\text{A.38})$$

et  $C(\boldsymbol{\alpha})$  est définie par :

$$\begin{aligned} C(\boldsymbol{\alpha}) &= \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \\ \hat{\alpha} &= \sum_{k=1}^K \alpha_k \end{aligned} \quad (\text{A.39})$$

L'algorithme *variationnel bayésien EM* propose un modèle génératif optimisant l'espérance de  $p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$ .

Les espérances des différents paramètres sont données par :

$$\mathbb{E} [\log \pi_k] = \psi(\alpha_k) - \psi \left( \sum_{k=1}^K \alpha_k \right) \quad (\text{A.40a})$$

$$\mathbb{E} [\log |\boldsymbol{\Lambda}_k|] = \sum_{i=1}^D \psi \left( \frac{\nu_k + 1 - i}{2} \right) + D \log 2 + \log |\mathbf{W}_k| \quad (\text{A.40b})$$

où  $\psi(a) = \frac{d}{da}\Gamma(a)$  est la fonction Gamma, et l'espérance utilisée dans l'équation (3.27) est donnée par :

$$\mathbb{E}[(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k)] = D\beta_k^{-1} + \nu_k (\mathbf{x}_n - \mathbf{m}_k)^\top \mathbf{W}_k (\mathbf{x}_n - \mathbf{m}_k) \quad (\text{A.41})$$

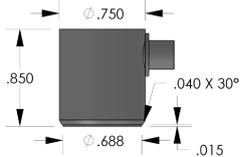
## A.5 Fiche technique du capteur F50 $\alpha$



### PRODUCT DATA SHEET

#### F50 $\alpha$ Sensor

##### General Purpose Sensor

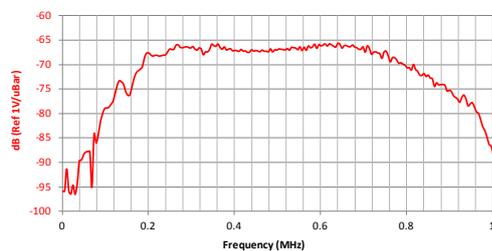


#### DESCRIPTION AND FEATURES

The F-series sensors are designed to incorporate the latest advances in materials and transducer design principles. These sensors unlike the resonant sensors (R-series) have a very high sensitivity and flat response in the bandwidth of operation. The F50A is a wideband sensor with a flat frequency response over the range of 200 – 800 kHz. This sensor is an ideal candidate for applications requiring high bandwidth for frequency analysis of the AE signals for noise discrimination and source identification. This sensor features a rugged steel construction and an SMA connector exiting from the side of the sensor.

#### APPLICATIONS

This sensor is well suited for structural health monitoring of large structures like storage tanks, pipelines etc. Flat response of these sensors help eliminate the sharp rise and fall in the sensitivity with frequency commonly seen in resonant sensors, providing additional information for engineers to analyze. Wideband sensors are well suited for research applications where a high fidelity AE response is required. It can be easily mounted using epoxy.



#### OPERATING SPECIFICATIONS

##### Dynamic

Peak Sensitivity, Ref V/(m/s)	65 dB
Peak Sensitivity, Ref V/ $\mu$ bar	-65 dB
Operating Frequency Range	200-800 kHz
Resonant Frequency, Ref V/(m/s)	100 kHz
Directionality	+/-1.5 dB

##### Environmental

Temperature Range	-65 to 177°C
Shock Limit	500 g
Completely enclosed crystal for RFI/EMI immunity	

##### Physical

Dimensions	0.75" OD X 0.85" H 19 mm OD X 21.6 mm H
Weight	34 grams
Case Material	Stainless steel
Face Material	Ceramic
Connector	SMA
Connector Locations	Side

#### ORDERING INFORMATION AND ACCESSORIES

F50 $\alpha$	F50 $\alpha$ or F50a
Cable (specify length in 'm')	1232-X-SMA
Magnetic Hold-Down	MHR15A
Pre-amplifier	0/2/4, 2/4/6

##### Sensors include

NIST Calibration Certificate & Warranty



**WORLDWIDE HEADQUARTERS:**  
195 Clarksville Rd •  
Princeton Jct, NJ 08550 • USA  
T: +1.609.716.4000 • F: +1.609.716.0706  
E-MAIL: sales.systems@mistrasgroup.com

**CANADA** T: +1.403.556.1350  
**CHINA** T: +86.10.5877.3672  
**FRANCE** T: +331.498.26040  
**GERMANY** T: +49.040.2000.4025  
**GREECE** T: +30.210.2846.801

**HOLLAND** T: +31.010.245.0325  
**INDIA** T: +91.22.2586.2444  
**JAPAN** T: +81.33.498.3570  
**MIDDLE EAST** T: +44(0)1954.231.612  
**RUSSIA** T: +7495.789.4549

**SCANDINAVIA** T: +46(0)31.252040  
**S. AMERICA** T: +55.11.3082.5111  
**UK** T: +44(0)1954.231.612

# Table des figures

1.1	Illustration de la technique de l'émission acoustique en <i>SHM</i> . . . . .	3
1.2	Endommagements dans un matériau composite . . . . .	5
1.3	Exemples de capteurs pour des applications standards . . . . .	6
1.4	Chaîne de reconnaissance de formes . . . . .	8
1.5	Structure étudiée . . . . .	9
1.6	Niveaux de serrage, émission acoustique et signal du vibromètre . . . . .	11
1.7	Débruitage des signaux   zoom autour de 0.3s pour l'essai #B pour un couple de serrage de 60 cNm. La détection du seuil est faite à partir du transitoire filtré (b) et l'extraction des descripteurs à partir du signal brut (a). . . . .	12
1.8	Émission continue et discrète . . . . .	13
1.9	Image d'un transitoire . . . . .	14
1.10	Illustration des modèles supervisé et non supervisé . . . . .	17
1.11	Localisation d'un signal d'EA sur une poutre isotrope . . . . .	17
1.12	Illustration de l'algorithme hiérarchique . . . . .	23
1.13	Illustration 2D de l'algorithme Espérance Maximisation . . . . .	26
1.14	Illustration du partitionnement consensuel . . . . .	28
1.15	Partition obtenue sur l'essai B par les k-moyennes et la validation par vote . . . . .	34
1.16	Partition référente pour l'essai B . . . . .	35
1.17	Cumul d'occurrence de la partition obtenue sur l'essai B d'après la validation par vote . . . . .	35
1.18	Performances de la validation pour l'essai B. L'axe de la figure de droite correspond aux différents sous-ensembles possibles de descripteurs (sous-ensembles de cardinalité $d$ ). . . . .	36
1.19	Cumul d'occurrence de la partition obtenue sur l'essai C d'après la validation par vote . . . . .	37
1.20	Cumulée d'occurrence de la partition obtenue sur l'essai D d'après la validation par vote . . . . .	38
1.21	Cumul d'occurrence de la partition obtenue sur l'essai E d'après la validation par vote . . . . .	38
1.22	Cumul d'occurrence de la partition obtenue sur l'essai F d'après la validation par vote . . . . .	39
1.23	Performances de la validation : ARI des 10 premières partitions . . . . .	41
2.1	Paradigmes pour la qualité du partitionnement . . . . .	48
2.2	Terme de pénalisation . . . . .	60

2.3	Essai #B : Validation du partitionnement par les GMM en utilisant l' <i>algorithme 1</i> , en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de $K \in [2, 10]$ . . .	64
2.4	Essai #B : Cumulées d'occurrence de la première partition du tri par le critère séquentiel pour $K = 8$ et $K = 10$ <i>clusters</i> . . . . .	65
2.5	Essai #C : Validation du partitionnement par les GMM en utilisant l' <i>algorithme 1</i> , en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de $K \in [2, 10]$ . . .	65
2.6	Essai #D : Validation du partitionnement par les GMM en utilisant l' <i>algorithme 1</i> , en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de $K \in [2, 10]$ . . .	66
2.7	Essai #E : Validation du partitionnement par les GMM en utilisant l' <i>algorithme 1</i> , en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de $K \in [2, 10]$ . . .	67
2.8	Essai #F : Validation du partitionnement par les GMM en utilisant l' <i>algorithme 1</i> , en sélectionnant les 10 meilleurs sous-ensembles pour chaque valeur de $K \in [2, 10]$ . . .	68
2.9	Durées de serrage logarithmiques . . . . .	70
3.1	Exemple des fonctions de masse avec observations $\mathbf{y}_n$ pour 2 <i>clusters</i> . . . . .	76
3.2	<b>Graphe acyclique dirigé représentant le VB-GMM</b>   source : [78, Chap. 10] . . .	79
3.3	<b>Graphe acyclique dirigé représentant le VB-GMM avec des a priori</b> . . . . .	86
3.4	Essai #B : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . Les histogrammes montrent que sans a priori les performances sont très satisfaisantes (pics correspondant aux changements de niveaux de serrage), et que l'ajout des a priori permet de normaliser la hauteur des pics. . . . .	94
3.5	<b>Essai #B</b> : Cumulée des occurrences des <i>clusters</i> de la première partition du tri pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . . . . .	95
3.6	Essai #C : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . .	96
3.7	Essai #D : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . .	97
3.8	Essai #E : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . .	98
3.9	Essai #F : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . .	99
3.10	Essai #B : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . Les classes ont été corrompues, 10% des étiquettes introduites comme <i>a priori</i> ne correspondent pas au niveau de serrage correct. . . . .	101

3.11	Essai #B : Partitionnement pour les valeurs $\rho = \{1, 0.9, 0.6, 0.3\}$ . Histogrammes des premières occurrences de clusters des 10 meilleures partitions pour $K_S^* \in [2, K_{max}]$ . Les données ont été prises en entier sans sous-échantillonnage ( $N = 14089$ ). . . . .	106
3.12	Essai #B : Évaluation de l'influence l'a priori de Dirichlet $\alpha_0 = \{0.1, 1, 10\}$ . Les histogrammes sont tracés sans <i>a priori</i> . . . . .	107
3.13	Essais #B et #D : Illustration avec le nombre de <i>clusters</i> initialisé à 20. Les histogrammes sont réalisés sur les déclenchements de clusters avec un partitionnement sans <i>a priori</i> . . . . .	108
3.14	Essai #B : Évaluation de l'influence du nombre de descripteurs choisis pour le partitionnement $d = \{2, 3, 10\}$ . . . . .	109
A.1	Validation conventionnelle avec le critère séquentiel . . . . .	130
A.2	Essai #B : Validation consensuelle avec des durées de serrage logarithmiques. L'idée est de simuler un desserrage naturel. Le boulon se desserre plus rapidement lorsque le niveau de serrage est faible. Les boîtes bleues compactes A.2c correspondent au critère séquentiel et celles larges au vote . . . . .	131
A.3	Essai #B : Résultats obtenus par le partitionnement après réduction de dimension par une analyse en composantes principales avec une conservation de la variance à 95 %, ce qui fait 10 composantes principales. Les performances et graphes sont quasi-identiques pour les algorithmes k-moyennes, le partitionnement hiérarchique et les GMM. Seuls les k-moyennes et le VB-GMM sans <i>a priori</i> sont illustrés. . . . .	136
A.4	Essai #B : Résultats globaux en fonction du nombre de composantes principales et de la quantité de données annotées. Plus les <i>a priori</i> sont forts, moins il faut des données pour améliorer le partitionnement (A.4a). Pour les méthodes classiques, les valeurs des ARI ont des variations importantes. Le VB-GMM, bien qu'il ne soit pas celui avec les meilleurs ARI est l'algorithme qui garantit un minimum de résultats. Les ARI ne chutent pas A.4b. . . . .	137
A.5	Principe d'un auto-encodeur [147] . . . . .	138
A.6	Essai #B : Partitionnement obtenu avec les différents algorithmes en utilisant un auto-encodeur épars avec une couche cachée de 5 neurones et une entrée à 31 descripteurs initiaux. Les serrages sont tous détectés à l'exception du dernier. Les résultats sont très prometteurs, sachant que la configuration n'a pas été optimisée. . . . .	140



# Liste des tableaux

1.1	Récapitulatif des conditions expérimentales des essais « ORION-EA » . . . . .	10
1.2	Descripteurs obtenus pour les essais "ORION-EA" . . . . .	15
1.3	Récapitulatif des méthodes de partitionnement classiques . . . . .	28
1.4	Résultats obtenus pour les essais « ORION-EA » . . . . .	40
2.1	Exemple d'un jeu à 4 classes . . . . .	55
2.2	Exemple d'un jeu à 4 classes avec des critères normalisés . . . . .	56
2.3	Exemple d'un jeu à 4 classes avec des critères normalisés munis d'une pénalité . . . . .	61
2.4	Lien entre partitions obtenues et les serrages . . . . .	62
3.1	Cas d'usage des étiquettes pour un partitionnement . . . . .	78
3.2	Lois de distribution d'un modèle bayésien variationnel (VB-GMM) . . . . .	80

**Titre :** Aspects temporels dans l'interprétation de données d'émission acoustique : importance de critères de validation adaptés et pertinence de méthodes variationnelles bayésiennes avec et sans a priori.

**Mots-clés :** Validation du partitionnement, émission acoustique, incertitude, comparaison d'algorithmes, monitoring.

La technique de l'émission acoustique est une méthode passive et non destructive d'évaluation de l'état de santé des structures. Sous contrainte, les structures libèrent de l'énergie sous forme d'ondes transitoires appelées émissions acoustiques, portant les caractéristiques de leurs sources. Ces ondes sont représentées dans un espace commun de descripteurs. Le nombre de sources et les descripteurs pertinents étant inconnus, pour différents sous-ensembles de descripteurs, un partitionnement est fait en faisant varier le nombre de *clusters*. La validation consiste à déterminer le nombre de *clusters* et le sous-ensemble de descripteurs pertinents. Cette validation a longtemps

reposé sur des critères basés sur la forme des *clusters*, sans tenir compte de la chronologie de ces derniers. Dans cette thèse nous avons défini des critères de validation séquentiels dédiés à l'émission acoustique. Une méthode d'évaluation du partitionnement basée sur les histogrammes de déclenchement des clusters a également été proposée pour un suivi chronologique des événements. Enfin un modèle de partitionnement partiellement supervisé a été proposé pour orienter le partitionnement avec des informations partielles. Ce modèle permet aussi de déterminer le nombre de *clusters naturels* de façon automatique, sans l'intervention d'un critère.

**Title :** Time-based interpretation of acoustic emission data : importance of suitable validation criteria and relevance of Bayesian variational methods with and without a priori.

**Keywords :** Clustering validation, acoustic emission, uncertainty, benchmarking, monitoring.

The acoustic emission technique is a passive and non-destructive method for structural health monitoring and material characterization. Under stress, structures release energy in the form of waves transients called acoustic emissions, which carry the characteristics of their sources. These waves are represented in a common feature space. The number of sources and the relevant features being unknown, so for different subsets of features, a clustering is done by varying the number of clusters. Validation consists of determining the number of clusters and the subset of relevant features. This validation has long been made with criteria

based on the shape of the clusters, without taking into account clusters chronology. In this thesis we have defined sequential validation criteria dedicated to acoustic emission. A clustering validation method based on cluster trigger histograms has also been proposed for chronological monitoring of events. Finally, a partially supervised learning model has been proposed to help the clustering with partial information. This model also makes it possible to determine the number of natural clusters automatically, without the intervention of a criterion.